MUR
Ministero dell'Università e della Ricerca

Piano Nazionale
di Ripresa e Resilienza

ICSC
Centro Nazionale HPC, Big Data e Quantum Computing

INFN
Istituto Nazionale di Fisica Nucleare

A. Scientific quality

This document describing the **National Centre on HPC, Big Data and Quantum Computing** project is organized as follows: Section A presents the scientific aspects of the proposed National Centre (CN in the following), highlighting the objectives and motivations of the research program, planned research activities as well as the methodology to implement them. Section B deals with the characteristics, feasibility, and control, giving the details of the Centre's structure and of the implementation plan of the research activities. Section C illustrates the program impact both in the national and international context.

In particular, Section A is divided in the following four paragraphs:

A.1 The scenario and motivation

A.2 The research activities

A.3 The methodology

A.4 The data management

A.1 The scenario and motivation

In the next few years, an unprecedented amount of data will be produced by scientific, industrial, and institutional actors, so we will have to face the challenge of extracting social and economic value from this data explosion. In this context, supercomputing, numerical simulation, Artificial Intelligence, high-performance data analytics and Big Data management will be essential and strategic for understanding and responding to grand societal challenges and in stimulating a people-centered process of sustainable growth and human development, allowing academia, industry and institutions to develop services and discoveries.

Since many years US, China and Japan are making great strides in these frontiers; Europe started slightly later and in the last years has defined a clear strategy. Among the main pillars are: EuroHPC Joint Undertaking[1], the European Open Science Cloud (EOSC)[2], the European Processor Initiative (EPI)[3] and the Quantum Flagship[4], according to the European Data Strategy and the European Approach to AI. EU Member States have made significant investments in European Petascale and pre-exascale infrastructures. EuroHPC has put exascale supercomputers on the roadmap, as well as Quantum Computing. Within this framework Italy is investing in HPC co-funding with EuroHPC Leonardo, a pre-exascale world-class Tier-0 system, which will offer highly competitive HPC services to Italian and European public and private researchers. The ongoing project of Polo Strategico Nazionale (PSN) will collect Italian Public Administration's data assets in a national cloud infrastructure. The new "PA data hub" should be seamlessly linked with the CN HPC, Big Data, Quantum Computing infrastructure, leveraging data and cloud services, enabling new business and research opportunities for Research Centers, SMEs and big Corporates.

To follow up on these actions and to design the Italian National strategy for HPC and Big Data it will be necessary:

- build a world-class supercomputing cloud infrastructure to store, manage and analyse all the data produced, in combination with hardware general purpose cloud infrastructure
- set up centres of excellence with high level teams of experts to develop domain applications

---

[1] EuroHPC: https://eurohpc-ju.europa.eu
[2] EOSC: https://eosc.eu
[3] EPI: https://www.european-processor-initiative.eu/
[4] QT : https://qt.eu

- set up strong links between the scientific community and the industrial system, (addressing both corporates that need HPC and SMEs that need easy, fast and scalable services in a non-HPC DevOps infrastructure)
- train young scientists as well as managers to become experts in these fields
- implement organic and structural measures for innovation and for increasing the Technology Readiness Level (TRL), also for SMEs without specific HPC skills and resources.
- build capacity by sharing expertise within the broader scientific community and strengthening the areas where Italy is lagging behind
- assess and maximize the social and economic impact of the CN at local, national and EU level with special attention to ethical implications
- set-up a group of experts with the specific task of analyzing and monitoring the ethical implications of big-data use and management

The CN aims to create the national digital infrastructure for research and innovation, starting from the existing HPC, HTC and Big Data infrastructures evolving towards a cloud datalake model accessible by the scientific and industrial communities through flexible and uniform cloud web interfaces, relying on a high-level support team, form a globally attractive ecosystem based on strategic public-private partnerships to fully exploit top level digital infrastructure for scientific and technical computing and promote the development of new computing technologies. The key elements to pursue these objectives are summarized in Table 1.

Table 1 The key elements of National Centre on HPC, Big Data and Quantum Computing

World-class Infrastructure	Living Labs	Centers of Excellence	Unique Ecosystem	Leadership
1	2	3	4	5
Unlocking the research and innovation potential	Co-designing future supercomputing and big data technologies	Creating value from data and maximizing socio-economic impact	Empowering and training people, attracting and retaining international talent, inspiring young entrepreneurs	Strengthening Italian competitiveness and leading Europe to be a role model for the data-driven society

In this scenario, the CN provides a pivotal opportunity for the national scientific, industrial and socio-economic system to address current and upcoming scientific and societal challenges, strengthening and expanding existing competences and infrastructural resources. The CN will be structured according to the *hub and spoke* model: the hub is responsible for the validation and management of the research program, whose activities are elaborated and implemented by the spokes and their affiliate institutions, as well as through open calls. The hub implements all the activities on education and training, entrepreneurship, knowledge transfer, policy and outreach, and coordinates a transversal research group on Societal Implication and Impact. Hub and spokes consist of Universities, Research Institutions as well as private and public operators. The proposed CN includes one cross spoke, Spoke 0 "Supercomputing Cloud Infrastructure", and 10 thematic Spokes, as shown in Fig. 1 and described in the following.

The CN has two main goals:

i) to create a national computing infrastructure, Datalake-like, by grouping together the existing High-Performance Computing (HPC), High Throughput Computing (HTC), Big Data and network infrastructures and new targeted resources procured by means of the CN funding, and by providing to the scientific and industrial communities a flexible and uniform Cloud interface to serve the full spectrum of applications, that span from HPC computing (needed by computational-intensive application, typically created by Fundamental Research and large/specialized Corporates) to general purpose cloud infrastructure (addressing computational needs of Science Applications, other Corporates and SMEs). This infrastructure should also allow the proper service level needed by each customer, in terms of cost, speed, accessibility, scalability and resiliency, particularly important for businesses with applications in production.

ii) to create around the infrastructure a globally attractive ecosystem which supports the academia and the industrial system and fosters the exploitation of computing resources and technologies with the goal of encouraging processes of sustainable economic growth and human development.

The Spoke 0 is built upon a supercomputing cloud infrastructure that includes resources from the major national players. The 10 thematic spokes involve all the relevant institutions and national experts in their respective fields and they will provide an ecosystem to guide and support the scientific communities and the industrial system, including SMEs and Public Administration. The Spokes from 2 to 9 develop advanced applications based on the infrastructures made available and maintained by Spoke 0. These Spokes represent the scientific domains coherent with the PNRR guidelines and chosen for the relevance with the respect to the objectives of the CN. Spokes 1 and 10 are focused on the development of high potential future technologies: advanced codesign of high-performance and high-throughput hardware and software systems, and development of expertise on hardware and software quantum computing, respectively.

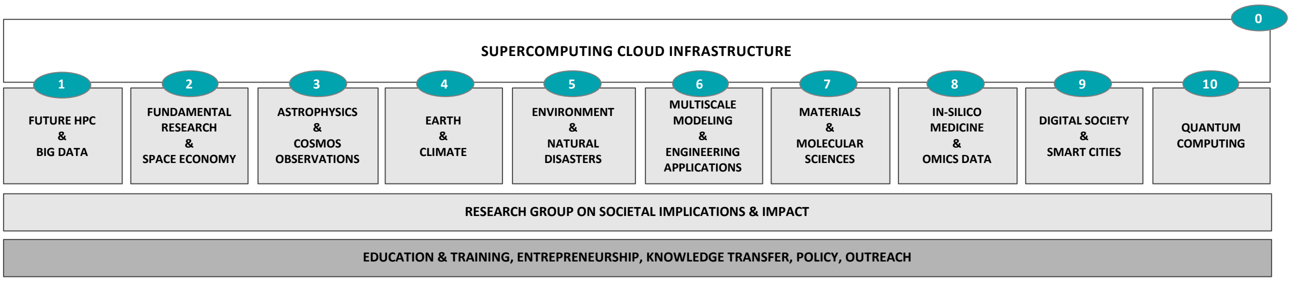


Figure 1 The organization chart of the National Centre

In addition to the Universities and Research Institutions, the CN foresees the involvement of a number of Italian companies aiming to establish a synergy between the scientific community and the industrial system to the advantage of national research and productive systems. The CN will contribute to make Italy competitive in the economic challenges of the coming years, supporting the production system in making the most of the data collected along the specific processes, and master the hardware and software technologies connected to them, providing the right combination between computing power and cost effective, quick and easy accessibility to data and application services. Dedicated targeted activities to increase the TRL to market levels will be developed. Finally, a basic element of the CN programme is to form a new cohort of computing savvy researchers and Ph.D. laureates, able to steer the transfer methods and technological solutions from academia to the productive sectors, mastering the most up-to-date DevOps, DataOps and MLOps framework. More generally, the programme is aimed at building capacity in order to strengthen Italian competitiveness and guaranteeing long-term sustainability.

A.2 The research activities

The research activities plan is elaborated by the Spokes and their Affiliates under the coordination of the Spokes' Coordination Board, that is in charge to stimulate and promote synergies, common initiatives and proper sharing among Spokes' activities, as well as to harmonize the proposals from each Spoke and to support Spokes for transversal actions by creating cross-spokes working groups, when necessary. The Hub will validate the proposed research activities plan based on the Governance model of the CN as described in the specific Section. Then the plan will be executed by the Spokes and their Affiliates as well as through open calls, under the coordination of the Spokes' Coordination Board and the support offered by Hub's Committees and Boards (see the specific Section below) on common issues and transversal themes. The Hub will monitor continuously the Research Programme advancement and will promptly react to manage emerging risks or delays. The plan provides for the inclusion of private companies and industrial partners. For those activities for which it is relevant, targeted programs to increase the TRL to market levels are developed.

In order to manage and supervise all activities a proper Management structure will be implemented as described in the specific Section.

Common goals of all spokes are listed in the following and summarized in Figure 2:

- bring together world-class knowledge and expertise in applying established mechanisms, user-driven development, performance tools, programming models and co-design activities for real systems based on leading edge technologies;
- promote the use of advanced (extreme performance) computing capabilities and scale up of existing codes and technologies;
- address the computational skill gap in the targeted domain for increased adoption of advanced computing in Academia, Industry and Public Administration;
- support Education, Training, Dissemination and Innovation initiatives coordinated by the Hub;
- develop software solutions for:
    - data analysis and numerical simulations able to effectively exploit HPC resources in the perspective of the Exascale era, so to consolidate and promote some Italian worldwide excellencies and flagship projects;
    - cloud application to leverage both private corporate data and Public Administration data (thanks to PSN), to enable business opportunities for the entire economic system
- adopt innovative approaches (e.g., Machine Learning (ML), Deep Learning (DL) and Artificial Intelligence) for the analysis of large and complex data volumes taking into account and actively promoting a lower energy impact and address convergence among the HPC, Big Data and Artificial Intelligence domains;
- develop Key Enabling Technologies.

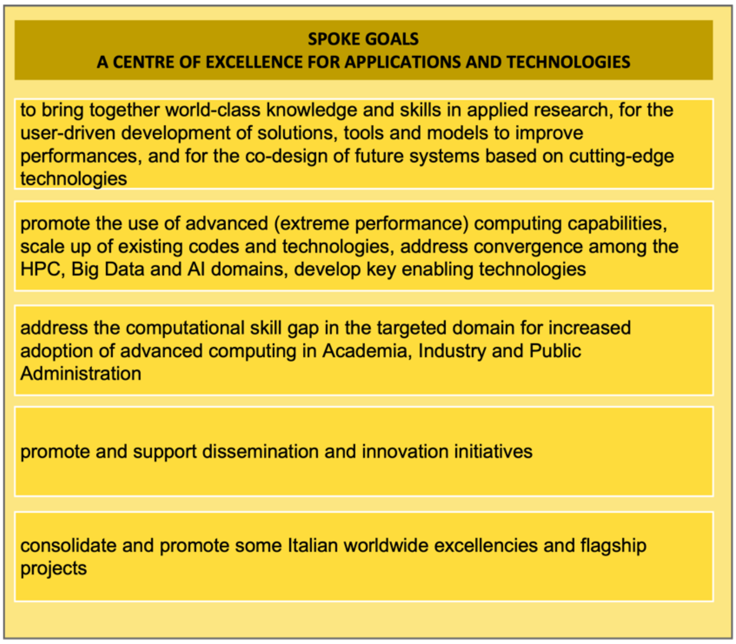


Figure 2 Common goals of the spokes

Figure 3 shows the common instruments of the Spokes.

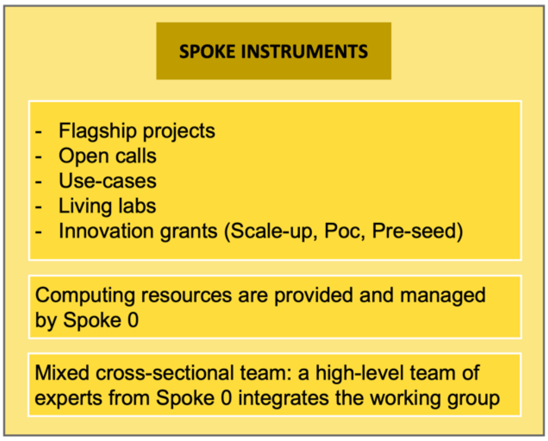


Figure 3 Common instruments of the spokes

In the following, the main research activities and objectives of each spoke of the CN are described.

Spoke 0 – Supercomputing Cloud Infrastructure

Supercomputing, numerical simulation, Artificial Intelligence, high-performance data analytics and Big Data management are essential and strategic for understanding and responding to grand societal challenges and in stimulating economic growth, allowing academia and industry to develop services and discoveries. Because of such an impact, the deployment of current HPC and Big Data technologies and the associated human expertise have also to be linked towards future exascale and post-exascale supercomputing technologies, including the emerging domain of quantum computing.
Applications of computing underpin all aspects of our lives. It is a cornerstone of academia and industry where everything, from fundamental and climate research to social and economic science, depends on algorithms to deliver solutions that are accurate, fast, safe, and affordable. Moreover, scientists rely heavily on computing to process and analyze data and then translate information into knowledge and technological innovations.

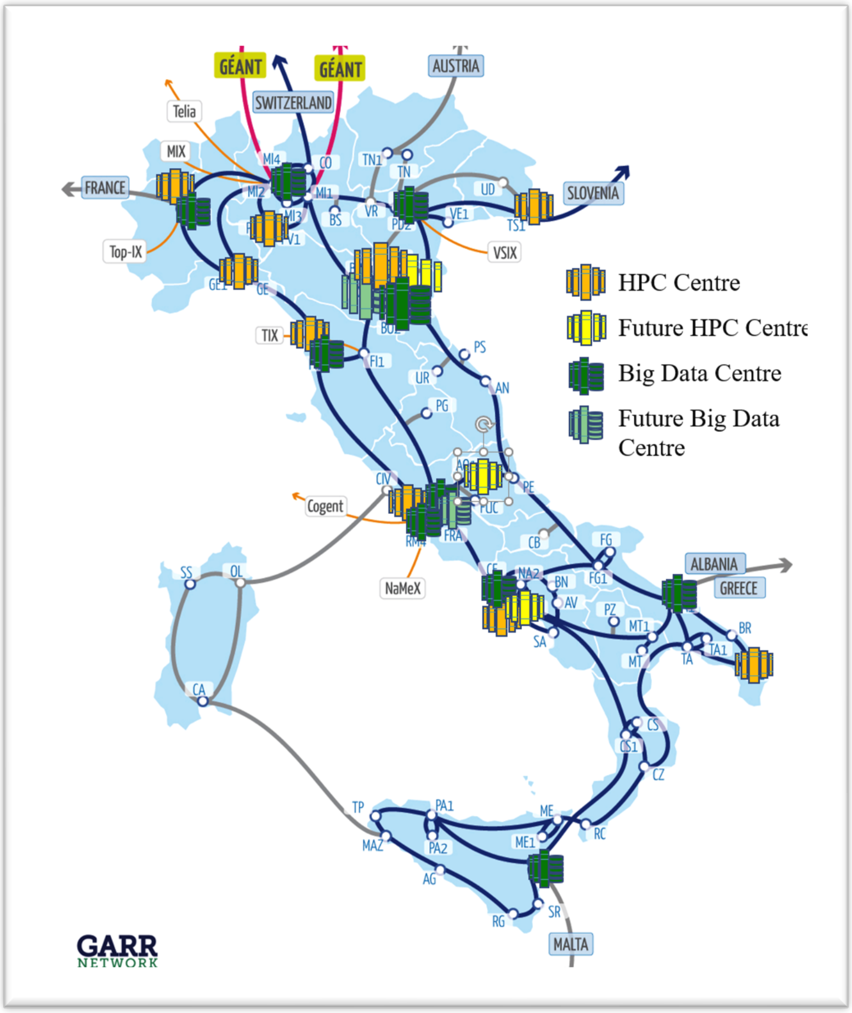


Figure 4 Distribution of relevant sites of the National Centre

The US, China and Japan are making great strides in these frontiers. In order to respond to the urgent need for an expanded European advanced computing infrastructure, recently the EC, EuroHPC Joint Undertaking (JU) and EU Member States have made significant investments in European petascale and pre-exascale infrastructure. EuroHPC has put exascale supercomputers on the roadmap, as well as Quantum Computing. Italy decided to invest in HPC co-funding with EuroHPC Leonardo, the world-class Tier-0 system that projects Italy towards the exascale class of HPC for research and innovation and will allow offering highly competitive HPC services to Italian, and European also, public and private researchers.

Besides, at national level, a valuable infrastructure of Tier-1 and Tier-2 HPC and Big Data systems are distributed across the country (Fig. 4) and will be object of a large and persistent upgrading process. This network of infrastructures will also be integrated in a federated way and managed in a coordinated system in order to: progress in the services provided and allow the sharing of data and IT resources; persistently scalable to meet the growing needs of users, accessible in cloud mode to allow the implementation of complex workflows, secure and regulations compliant for data privacy and confidentiality, with certified quality of service delivery.

Goals and Objectives

The infrastructural Spoke 0 is focused on the objectives of consolidating the presence in Italy of a federation of supercomputing centres (from a Tier0 of exascale class to Tier1 and Tier2) and data intensive facilities (currently providing storage capacity of ~200 PB and computing power of ~100,000 CPU cores) as well as to

ensure appropriate upgrade of the national research network (GARR) by bringing the connectivity of the national backbone and of the last mile of the main centres to the scale of the terabit per second.
Valuable is also the opportunity of sharing the experiences acquired in the various spokes, with particular reference to support for the emerging sectors of the wide disciplinary spectrum, those of significant potential industrial impact, such as the sectors of the space economy, materials science and multiscale engineering, those of expected social impact, such as the health and life sciences sectors, and finally those linked to sustainability, such as sectors relating to climate change and the digital twins of complex urban systems.
A mandatory element to be taken into account is the mitigation and overcoming of territorial differences relating to the regions of south Italy.
A starting point is the analysis of the usage of the existing computing and data centres.
On the one hand, CINECA, the main Italian supercomputing centre, in 2021 provided 450 million node hours, with a substantial increase respect to previous years, to communities in several disciplinary fields (mainly computational chemistry, biochemistry, and physics of condensed matter, but also particle physics and astrophysics) able to exploit accelerated computing systems such as MARCONI100.
On the other hand, the network of INFN data centres offers resources (HTC and storage) to several scientific collaborations in the field of particle, nuclear and astroparticle physics; INFN has twenty years of experience in distributed computing, having its data centres integrated into the worldwide computing grid since 2003.
From the analysis of the present situation, some guidelines can be derived to make the action of the Spoke 0 and of the entire National Centre more effective.
Line 1: transfer of competence related to computational methods and Big Data processing from disciplinary areas related to materials science, fundamental physics and astrophysics to other disciplinary areas, with particular reference to areas related to bioinformatics and life sciences and medicine in silico, digital twins of urban systems, resilience to extreme natural events.
Line 2: particular attention will be considered in the allocation of work time and the enabling specialist support provided in the context of the Spoke 0 in accordance with the knowledge transfer action, as this staff is also committed to supporting emerging scientific communities.
Line 3: relating to some critical domains, such as the nuclear fusion sector, pertaining to the field of applied research in a fundamental area in the process of ecological and energy transition, and the earth sciences filed, with particular reference to the relative aspects of understanding of the phenomena relating to climate change, also in relation to the evolution of the national system which has recently seen the creation of the Italia Meteo agency for the coordination of national activities in this disciplinary context, domains that constitute disciplinary areas where the demand for HPC infrastructure and Big Data by European researchers is particularly significant, but which sees a drastic reduction in demand from the Italian community, a specific action will be carried out, also in synergy with large European-level actions such as Destination Earth, Copernicus and Eurofusion. Furthermore, for the infrastructure aspects, the HPC point of presence of the CMCC located in Lecce will be strengthened, also with a territorial impact perspective.
Line 4: with particular reference to the aspects relating to the overcoming of the geographical digital divide, especially in the regions of southern Italy, the priority for the creation of a CINECA point of presence in Naples in collaboration with the CNR and the academic system of the Campania region, with a Tier1 system of petascale class and an adequate specialist support provided by CINECA personnel.
Line 5: the access policy to the Tier0 class infrastructure (Leonardo supercomputer also integrated with a quantum computing system), through open calls for HPC and Big Data Processing projects managed by the national ISCRA action will have an appropriate access quota reserved for researchers working in the southern regions, while keeping the rigor of the assignment based on scientific merit.
Line 6: in coordination and synergy with the vertical spokes' specific open calls and grants for innovation will be activated to support those areas where the Italian community has less capacity.
In particular this Spoke has the responsibility to coordinate and implement the activities for providing the most competitive federated computing (both HPC and HTC) and Big data processing infrastructure. It will rely on secure, production-level, state-of-the-art hardware and open-source software components, which will allow flexible service composition and a comprehensive set of access policies. This will facilitate access to high performance resources and solutions for the scientists working in the thematic spokes of this and other National Centres, across all research fields. The architectural and implementation flexibility that characterizes this

Spoke will also allow to promote activities related to distributed computing and storage requirements also beyond the scientific domains covered by the National Centres, including private sectors of strategic interest to the country, such as Industry 4.0, Health or Manufacturing.
The actions of the spoke will focus on following pillars and can be summarized as follows.
Network - Upgrade of the national high speed research network infrastructure towards terabits, for interconnecting the national data centres and data repositories to the European system of Data Centres for research through GÉANT and its global international connectivity.
Data Centre Facilities - Upgrade of the facilities of data centres directly managed by the Spoke 0 and in coordination with the other spokes, with particular focus on the use of energy-efficient technologies.
HTC, HPC and Big Data resources - Increase of state-of-the-art computational and data storing resources towards exascale to improve research competitiveness and serve the needs of scientists across all domains and enabling new discovery and technology innovation. To this purpose, it is foreseen to procure and operate diversified and heterogeneous computing resources, including disruptive technologies such as Quantum Computing systems.
Middleware and application software stack – Deployment and operation of open-source middleware supporting composable federation services compatible with EuroHPC, EOSC and GAIA-X principles and architecture to address the complexity of today's workloads. It will provide easy and transparent access to the distributed resources of the HTC, HPC and Big Data Infrastructures. Integration and development of existing middleware solutions will be used as a benchmark for the infrastructures coordinated by the spoke. The architectural middleware will be able to provide batch or interactive cloud access to a self-configuring virtual environment, which will be able to provide secure virtual partitions, with possible encrypted file systems and specific application stacks for the development of services, for example in the field of AI & ML and processing of genomic data and IoT, also through cloud edge interfaces.
High Level Support Teams - Leadership in HTC, HPC and Big Data use and skills is central as it is essential to extend the use of supercomputing to a wider range of scientific and industrial users, for instance by helping Industries and SMEs develop innovative business cases using supercomputers and Big Data techniques and providing them with training opportunities and critical skills they need. Moreover, leadership in computing and Big Data use and skills can then bridge the gaps between infrastructure needs, federation and services, technologies, and applications in order to enhance National competitiveness. As part of the spoke, an appropriately staffed team of high-level specialist support will be set up with the aim High level support and cross-sectional activities for software portability and optimization, co-design, software deployment, code development, data visualization and access interfaces.
Two key common aspects will characterize the technological pillars mentioned above.
Energy efficiency: through the acquisition of green equipment and power saving facilities and through middleware-level operational optimizations towards energy to solution as primary priority.
Cybersecurity: through compliance to International and National standards and regulations, such as ISO-27001, 27017 and 27018 certifications, ENISA best practices, and adherence to national initiatives such as *Perimetro di sicurezza cibernetica nazionale.* This action will be combined with a progressive extension of certifications also to the HPC infrastructure and with the use of extended two-factor authentication practices, encrypted file systems and with bastion hosts and firewall perimeters.

A preliminary list of existing HPC and Big Data infrastructures is shown in Figure 4. Further updates, with possible allocation of resources to entities not participating to the National Centre via dedicated open calls, will be prepared during the project execution.

Spoke 1 - Future HPC & Big Data

According to the EU vision, High-Performance Computing rests on five pillars: skills, applications, infrastructure, technology, and federation of resources. All the pillars are represented in the Future HPC spoke. The main focus is on the technology pillar across all the layers of standard (i.e., non-Quantum) computing systems: Circuits, Architecture, Programming Model, and Execution Model up to just before the last tier, i.e., Applications, which are addressed in other spokes of the centre. In this focused strategy, the FutureHPC & BD spoke is tightly linked with the work program of the technological pillar of the EuroHPC Joint Undertaking, which is expected to contribute to the sustainability of the FutureHPC & BD spoke. The technological pillar is crucial for EU digital sovereignty and is paramount for engaging industry in achieving (by 2030) European leadership and autonomy in HPC infrastructure, data, and services and fueling innovation across the computing continuum.

1) To create new laboratories as an integral part of a world-class national federated centre of competence for the advanced codesign of high-performance and high-throughput hardware and software systems, to strengthen Italy's leadership in the EuroHPC Joint Undertaking and the data infrastructure ecosystem.
2) To explore and roadmap key enabling technologies (hardware and software) for the next generation of high-performance/high-throughput computing systems: the goal is to design and create proofs-of-concept, prototypes, and demonstrators, raising the readiness level of the most promising technologies and facilitate industrial take-up and development.
3) To collaborate with the industry for an innovation strategy not limited to supercomputers but with a strong impact on large volume markets, such as edge servers, IoT gateways, autonomous vehicles, HPC-enabled cloud, and data analysis services that critically need reliable, energy-efficient, high-performance computing hardware and software. The goal is to widen Italy's footprint in the Key Digital Technologies (KDT) Joint Undertaking, consolidate Italy's leadership in the EuroHPC Joint Undertaking, and ultimately strengthen Italy's competitiveness in advanced computing with an open innovation approach.
4) To build up a large pool of talents with strong expertise in state-of-the-art hardware and software technologies and tools for advanced computing. The goal is to reduce the severe skill gap in this area and increase the number of highly trained professionals essential for Italy's industrial growth and competitiveness in the digital economy.

Research Topics and Lines

Hardware technologies and Systems - Computer architecture: Von Neumann, neuromorphic, in-memory, reconfigurable computing. Design energy-efficient and reliable parallel processors, accelerators, memory, storage hierarchy, and interconnects. Open instruction sets (RISC-V), open architectures, and open hardware for advanced computing. Integration of classical and disruptive computing technologies: new devices, three-dimensional integration, hybrid quantum/optical/classical computing systems for post-exascale HPC.
Co-design for added-value HPC and HTC systems: mini-applications design, workload profiling and analysis, hardware optimization and tradeoff analysis, system-level performance, power and reliability simulation, and design space exploration. Energy-aware hardware-dependent software: energy-efficiency extensions for programming libraries and languages, run-times; energy and power-aware scheduling and resource allocation. Systems for holistic monitoring and management of HPC resources; Data collection systems for power and energy monitoring and accounting.
Software Technologies and Tools - Programming models for modern HPC applications (shared-memory, message-passing, with-accelerators (e.g., GPU and FPGA); technologies and tools HPC-enabled digital twins; workflow management systems; high-performance I/O, ad-hoc file systems and high-performance streaming; parallel algorithms and libraries for scientific computing; high-performance compilers and run-time support systems; domain-specific languages and tools, benchmarking and software development methods and optimization for HPC-powered innovative applications; middleware for scalable BigData and AI/DL and their convergence with HPC systems; performance modeling, analysis, and simulation for complex parallel systems; heterogeneous computing and resource scheduling; integration of quantum computing kernels into traditional software pipelines; tools and libraries for distributed and Federated Machine Learning.
Cloud-HPC integration and convergence: data centre modeling and management; reliability assessment and predictive maintenance; data-driven autonomics, operator-less configuration and management, anomaly prediction, and classification; workload managers (e.g., locality-preserving, reservation-oriented, batch+interactive); multi-tenancy and management of critical data in HPC systems.

Spoke 2 - Fundamental Research & Space Economy

Science, and in particular science at the frontier of knowledge, is becoming more and more a computing intensive discipline. Current and next-generation experiments show processing and data needs comparable with the top global players and need a stack of solutions which are not typical of the curriculum of scientists. The trend has indeed started more than 15 years ago, with the development of solutions needed to satisfy the science of Collider Physics; since then, similar needs have been documented in other scientific domains, with Astroparticle physics showing by the end of the 2020s similar if not larger resource deployments. The activities in Spoke 2 “Fundamental Research and Space Economy" focus on boosting the science capabilities of current and future science initiatives, using the opportunities that PNRR in general and the National Centre for Big Data, HPC and Quantum Computing (CN) in particular offer in the next three years.

GOALS AND OBJECTIVES

The activities in the “Fundamental Research & Space Economy” Spoke will be cast within the context of state-of-the-art research in basic science, and, in particular, of the domains of theoretical and experimental physics with accelerators and with space- and ground-based detectors for astroparticle physics and gravitational wave investigations. Within different time scales, all of these areas have or will have to face problems regarding the scaling and efficiency of computing infrastructures. These areas demand a scaling of the computing infrastructures and an improvement in efficiency. The Spoke intends to address these needs designing, developing and testing solutions apt to the current and next-generation experiments, and fitting the contingent situation in Italy arising from the opportunities provided by the PNRR in general and the National Centre (CN) “Big Data, HPC and Quantum Computing”.

In addition, the Spoke activities aim at demonstrating that the same set of solutions are of value in domains outside the bounds of basic science and of the specific research use cases. We plan to address at least two specific situations:

- the handling of data and processing in the context of the Space Economy Italian Strategy, including the handling and processing of data from the Mirror Copernicus program, by fostering the conditions to enable radically innovative services;
- the seeding of similar solutions in the productive context, with industry-research shared testbeds and proofs of concept; this will cover the productive domains in which data and processing resources need to be geographically dispersed, need a secure and granular solution for the data access, and experience computational problems with an unsustainable predicted future scaling (due to cost, efficiency, or performance).

The crucial aspects of its mission concern the creation and/or optimization of algorithms and, in general, computing solutions capable of maximizing the potential physics output from experimental data and theoretical and phenomenological simulations, by using the tools made available by the Centre: e.g., heterogeneous and high-performance computing (via standard programming and AI-based solutions) and the ability to process large quantities of data beyond the capabilities of traditional methods.

The relevant scientific areas for the Spoke are fundamental, theoretical and experimental physics, in applications such as high-energy or high-intensity collisions (in experiments at CERN, KEK, JLAB, BNL, FNAL, IHEP, …), theoretical studies of fundamental interactions (lattice field theories, nuclear and particle physics, ...) and gravitational waves (VIRGO and Einstein Telescope), astroparticle physics and cosmology (including space-borne experiments, terrestrial telescopes such as the Cherenkov Telescope Array (CTA), and those hosted underground, as in the INFN Gran Sasso laboratories), physics of complex systems.

In these areas, the Spoke aims at providing innovative instruments driven by the quest for a deeper understanding of the fundamental laws of Nature, through tools and solutions based on the Centre’s infrastructure; a common denominator will be the utilization of more efficient strategies, reducing the computational costs and their power consumption footprint.

The development of algorithms and computational solutions designed will be shared with all the scientific domains in the Centre; indeed, it is expected that their application to further domains will add more value to the scientific advancement per se. In particular, we aim to foster a profitable exchange of technology and experience with the productive sector, by showing that solutions from research can be reused in the most data and computing intensive activities from the CN private partners.

RESEARCH TOPICS AND LINES

The Spoke activities will align on 6 different topics, which are identified as Work Packages in the following of this document. They are:

1) Theoretical Physics:
    a. Development of algorithms, codes and computational strategies for the simulation of physical theories and models, towards pre-Exascale and Exascale architectures.
    b. Theoretical research projects in domains already using HPC solutions, such as:
        i. lattice field theory (flavour physics, QCD phase diagrams, hadronic physics, interactions beyond the Standard Model, machine learning in quantum field theories, electromagnetic effects in hadronic processes);
        ii. collider physics phenomenology;
        iii. gravitational waves, cosmology and astroparticle physics (neutron-star physics, primordial universe, dark matter and energy, neutrino physics);
        iv. nuclear physics;
        v. physics of complex systems (fluid dynamics, disordered systems, quantitative biology);
        vi. condensed matter in low dimensional systems;
        vii. quantum systems (entanglement, quantum simulations, quantum information).
2) Experimental High Energy Physics: selection, data reduction, simulation and reconstruction algorithms (either via explicit programming or large-scale Machine Learning solutions) for HEP experiments (LHC, Future Colliders, KEK, IHEP, neutrino experiments...), with applications ranging from innovative triggers to distributed analysis techniques.
3) Experimental Astro-Particle Physics: data reduction, reconstruction and time cross-correlation algorithms, data selection and simulations of astroparticle and gravitational waves experiments, tools for cross-correlations and pattern recognition in multi-messenger physics, including novel implementations using techniques like Machine Learning.
4) Boosting the computational performance of Theoretical and Experimental Physics algorithms: porting of applications to GPUs and heterogeneous architectures (e.g., scalability of scientific codes and applications on GPU/CPU many-cores clusters, local and remote offloading, mission-critical algorithms on FPGAs, ...). The solutions and tools implemented during the project will be easily extendable to other scientific domains of the Centre and to the industrial partners in the Spoke; moreover, the personnel trained within the Centre will help to spread and boost the application of HPC methodologies to Italian academic and industrial fields, for a comprehensive advancement of the Italian system.
5) Architectural Support for Theoretical and Experimental Physics Data Management on the Distributed CN infrastructure: support for the adaptation of existing applications on the data-lake distributed infrastructure, and via innovative computational models (for example sharing of gauge configurations in lattice field theories, long-term data preservation, streaming access to data, tiered storage solutions, …). The solutions implemented will be tailored to the needs of the scientific fields, easily extendible not only to the nearby scientific domains in the Centre, but also to all academic and industrial realities where needs to access distributed computing and large amounts of data exist. In particular, the industrial partners in the Spoke have expressed interest in using the same technologies for their specific use cases.

6) Cross–domain Initiatives: optimization and adaptation of widely used software packages on the national Centre infrastructure, like Geant4 or FLUKA or generic high-performance techniques for data access/analysis; statistical and AI-based tools; data-interpretations tools. In the context of the Space Economy Italian Strategy, develop and deploy techniques to access, analyze and process the data from the Mirror Copernicus program, creating the conditions to enable radically innovative services. In particular, enable thorough and continuous observation programs for global and local processes, allowing external partners to operate a large variety of services, including the planning for emergencies, risks and resources.

Spoke 3 - Astrophysics & Cosmos Observations

High Performance Computing (HPC)-based and Big Data technologies are outstanding instruments to model the complex dynamic systems studied in Astrophysics and Cosmology today. Their use is needed by the majority of today's activities related to astrophysics: from the reduction and analysis of astronomical data up to their interpretation and comparison to theoretical predictions, including simulations and theoretical modeling. Furthermore, advanced numerical methods have become mandatory during both the preparatory and the operational phases of new scientific experiments, efficiently guiding and shaping the design and construction of instruments and observatories. They are also important to fully capture and analyse the torrent of complex observational data that the new generation of observatories will produce [e.g. the Square Kilometer Array Observatory (SKAO), the new MeerKAT+ and LOFAR2.0 observatories, the Cherenkov Telescope Array (CTA), the European Extremely Large Telescope (E-ELT), the Vera Rubin Observatory, the low-frequency space-born Laser Interferometer Space Antenna (LISA), the Euclid and WFIRST satellite missions, the balloon-borne LSPE experiment, the LiteBIRD satellite, DAMPE and HERD], which will be game-changing for our understanding of astronomical phenomena, the formation and evolution of the universe, and the fundamental laws of physics. The new generation of telescopes and instruments will produce exponentially more data than their predecessors and will require outstanding resources for the data to be post-processed, analyzed and stored. The design and deployment of those new scientific infrastructures is requiring a large worldwide financial and scientific investment (hundreds of millions of Euros) and, as consequence, a corresponding HPC and BigData investment should be pursued in order to maximize their outcomes.
This will lead to new IT (Information Technology) laboratories (codes/algorithms/tools/services) enabling the investigation of the physical processes originating the observed phenomena with unprecedented quality, resolution and reliability, allowing their interpretation and paving the way for new scientific discoveries. The efficient and effective exploitation of exascale (and beyond) computing capabilities will be fundamental for these laboratories: the combination of high-performance data processing with interactive real-time analysis and data visualization will be needed in the exploitation of the expected extremely large data volumes. The comparison of observations with large scale simulations will reach a new level of development thanks to the high-performance archiving and data access capabilities that are necessary to make the next generation of astrophysical data stores easily accessible to the whole astronomical community.
This effort requires designing, implementing and promoting a full ecosystem capable of: *(i)* delivering complex simulations capable of high predictive accuracy to address the complexity of the Universe; *(ii)* exploiting and/or driving the evolution of current and future computing architectures and algorithms; *(iii)* exploiting the wealth of data produced by computations and observations; *(iv)* effectively engaging with the Astronomy, Astrophysics, and Astroparticle physics (AAA) community in the HPC environment of codes and resources; *(v)* adapting and implementing existing and new algorithms for the new challenging (exascale and post exascale) HPC infrastructures; *(vi)* providing innovative data storage and archiving systems to face the big data challenges.
The realization of such an ecosystem is the objective of the Astrophysics and Cosmos Observations SPOKE (hereafter ACO-S).
The main objectives of ACO-S are the exploitation of cutting-edge solutions in HPC and Big Data processing and analysis for problems of interest in the following research area: Cosmology; Stars and Galaxies; Space physics (Earth, Solar and Planetary); Radio Astronomy; Observational Astrophysics and Time Domain; High

Energy Astrophysics, Cosmic Microwave Background; Large Scale Structure, Clusters and Galaxies; Multimessenger Astrophysics; Numerical Simulations and Modeling.
The ACO-S will also undertake a specific outreach and training program, aimed at creating a broad and skilled talent pool in Italy, and ultimately will boost the use of high performance and high throughput, cloud solutions in research centres, academia and private companies.
It will pursue a user-driven approach in order to tightly couple to the community and it will adopt a co-design methodology for the development of the selected applications, combining the requirements and the expertise of the scientists and the community code developers to the innovative software and hardware solutions and services envisioned by HPC and Cloud stakeholders, including green computing approaches, addressing the synergic and coordinated development of applications and technology.
A sustainable software development strategy will be adopted to cope with technological evolution, so that exascale ready software solutions can be easily adopted by the different codes, thus realizing a substantial *economy of scale*, shortening the time needed to adapt to a rapidly changing technology and to the new challenges.
The ACO-S main objectives are:

- **Objective 1.** High and Extreme performance computing, via developing software solutions for data analysis and numerical simulations able to effectively exploit HPC resources in the perspective of the Exascale era. (WP1, WP2, WP5)
    - Action 1.1: To optimize current codes on the present generation of HPC HW.
    - Action 1.2: To establish a co-design code development methodology by combining the expertise of scientists, community code developers, HPC experts, hardware manufacturers and software developers.
    - Action 1.3: To support accelerators, alternative architectures (e.g., GPUs and ARM architectures respectively) and experiment innovative supercomputing solutions.
    - Action 1.4: To develop high performance I/O modules compliant with the selected standards and interoperability protocols.
    - Action 1.5: To address energy efficiency and green computing.
- **Objective 2.** Big data processing and visualization, via adopting innovative approaches (e.g. Artificial Intelligence, inference via Bayesian statistics) for the analysis of large and complex data volumes and for their exploration (e.g. in-situ visualization), capable of efficiently exploiting HPC solutions. (WP3, WP4, WP5)
    - Action 2.1: To address the convergence (heterogeneous computing) of HPC, HTC, HPDA and Cloud by adopting and applying a common set of standards and interoperability protocols for data and applications.
    - Action 2.2: To develop new methods and/or optimize existing prototype Machine Learning applications for the automated processing of large and complex data, produced on the Exascale systems by the ACO-S community.
    - Action 2.3: To develop and optimize existing solutions for high performance visualization, addressing on-site and remote visualization for Exascale and post-Exascale systems.
    - Action 2.4: To develop integrated solutions starting from the outcome of the previous actions, in order to provide a unique and optimized eco-system for Exascale platforms for big and complex data sets.
- **Objective 3.** High Performance storage, Big Data management, and archiving applying the Open Science principles and implementing them in the Big Data Archives. Experimenting and adopting novel technologies and computational approaches for fast and scalable I/O (WP4, WP5)
    - Action 3.1: To develop and implement an automatic and efficient system to manage big data from storage to archive for structured and not structured data.
    - Action 3.2: To develop automatic and semi-automatic processes to improve data curation using community experience.

    - Action 3.3: To develop efficient user interfaces to access storage and archive following the Virtual Observatory and FAIR (Findable, Accessible Interoperable, Reusable) principles[5].
    - Action 3.4: To provide secure access to data and systems.
    - Action 3.5: To develop and implement efficient access following the Open Science initiative standard and protocol.
- **Objective 4.** Training and dissemination. Creating a community of scientists and code developers prepared to adopt and exploit innovative computational solutions. Fostering the engagement of the scientific community in Astronomy, Astrophysics and Astroparticle Physics.
    - Action 4.1: To organize new specific master courses and/or promote the access to existing ones so as to increase the researchers in HPC and BigData area. Promote PhD positions, courses and schools.
    - Action 4.2: To organize international conferences, workshops, schools and meetings.
    - Action 4.3: To set up International Network and Communities to extend the collaboration of scientists and to increase the excellence of the Italian community in ACO-S context (network of excellences).

RESEARCH TOPICS AND LINES

ACO-S is organized to address the main following topics (research lines):
HPC Codes Enabling and Optimization: it selects a number of codes that require large computational resources to face the next generation of scientific challenges and performs their redesign, reimplementation and optimisation in order to effectively exploit state-of-the-art HPC solutions, including accelerators and alternative architecture (e.g., GPUs and ARM architectures respectively).
Design of innovative Algorithms, Methodologies, Codes toward Exascale and beyond: it identifies innovative algorithms and methodologies upgrading their capability to exploit, and scale on, the exascale and post exascale architectures, merging the resulting improved features in codes, workflows and pipelines.
Big Data Analysis, Machine Learning and Visualization: it addresses the application of advanced Artificial Intelligence and Visualization solutions to large astrophysical data volumes, facilitating the convergence (heterogeneous computing) of HPC, HTC, HPDA and Cloud tools and exploiting Exascale platforms.
Big Data Management, Storage and Archiving: it starts from already implemented best practices and frameworks to manage data and software with FAIR and Open Science principles, to develop innovative frameworks able to satisfy the Big Data Challenge.

Spoke 4 - Earth & Climate

Climate change and its impact on countless sectors of society has enormously increased the demand for comprehensive, robust, timely and reliable climate data for decision making to inform adaptation and mitigation policies. In this context, the continued development and implementation of Earth System Models (ESMs) are key to address the complex challenges the society is facing in a changing climate. ESMs allow to simulate the cycling of energy, water and major geochemical elements (e.g., carbon, macronutrients) across different components of the Earth System (atmosphere, hydrosphere, cryosphere, biosphere), taking into account their major feedback mechanisms and related non-linear processes. Therefore, these numerical tools are of fundamental importance in order to simulate and predict the evolution of the climate system, investigate the effects that anthropogenic forcings may have on it and provide climate change projections for future decades.
Simulations performed with ESMs cover a wide range of timescales and computation requirements set by time constraints for the solution, integrations length, spatial and temporal resolutions and increased complexity of

[5] https://www.go-fair.org/fair-principles/

the represented processes and interactions/feedbacks within the climate system. These challenges can be met only with the advanced hybrid computational systems at pre- and full exascale.
In order, then, to provide a more appropriate set of climate predictions and climate change projections, it is extremely important to also be able to provide a reliable estimate of the uncertainties associated with the simulations produced. To this aim, it is useful to have a range of different ESMs, formed of different components, by means of which it is possible to investigate the role and effects that the different assumptions underlying the development of the models can have on the characteristics of the simulated climate. Therefore, it is very important that the modeling community has tools that allow easy configuration of different EMSs by means of a composition of different model components. Such approach, in fact, will allow to: i) ease the possibility to address and quantify uncertainties and errors in model parameterization that are inherent and impossible to fix in single-component approaches, and ii) improve the predictive capabilities of models, also in relationship to parameters of specific components.

GOALS AND OBJECTIVES

- Raise to the challenge by setting up an interdisciplinary modeling framework that integrates state-of-the-art Earth system modeling components, providing a flexible, reliable, and powerful tool to the scientific and operational community.
- Implement a digital infrastructure integrated in the HPC facilities made available through the ICSC National Centre that will: 1) assess and store robust data on the quality and health status of terrestrial and hydrosphere biodiversity and ecosystems, 2) facilitate the development and sharing of ESM components (e.g., models of the atmosphere, oceans, biogeochemistry, sea-ice, land-surface, vegetation, etc.); 3) facilitate the production and management of numerical simulations; 4) be a national asset available to the entire Italian community engaged in research, education and operational activities in the field of climate predictions and climate change, positioning our country at the forefront of climate research.
- Establish technologies that will make it possible to fulfil the climate objectives identified by the PNRR for the digital and green transition.

RESEARCH TOPICS AND LINES

The spoke activities will be aimed at the development of a shared interdisciplinary framework for advanced Earth system modelling and numerical experimentation. The framework will be based on digital infrastructures and smart workflows to streamline the production, facilitate the training, accelerate the understanding, and improve the quality of climate simulations and predictions. The implementation of the activities will contribute to the long-term sustainability approach promoted by the ICSC National Centre, also exploiting the industrial partnerships. In coherence and synergy with the European strategic agenda of Horizon Europe, Destination Earth and "A European Strategy for Data", activities will include, but are not going to be limited to:
- develop, optimize and share components and parametrizations for a new generation of high-resolution ESMs enabled by the large computing infrastructure;
- exploitation of large data sets of observations with digital technologies through the convergence of novel high-performance computing, big data, and artificial intelligence methodologies;
- development of end-to-end capabilities including increase in observations, assimilation, prediction, post-processing, and data handling needed to address near-term and long-term impacts together with evaluating the likelihood of high-impact-low-probability events and tipping points.

Spoke 5 - Environment & Natural Disasters

Nowadays, many different research communities generate and collect massive amounts of data with different granularities (from in-situ and remote from physical observations, building inventory to monitoring of critical infrastructures, advanced metering infrastructures, urban presence and mobility). Such communities are very diverse and include specialists in data analysis and network monitoring, in urban and city planningers, geophysics, geology, geotechnical and structural engineering, and experts in Earth Observation (EO) and Earth processes, as well as planners, and experts in model development focusing on a variety of risks related studies: from geomagnetic and geophysical hazards, landslides, or earthquakes, volcanoes, floods and extreme climate events, to pollution, space weather and more. Combining research efforts in different disciplinary contexts with data spanning different time and space scales is therefore the *primary aim* of the Spoke. The availability of easily accessible, up-to-date, understandable, science-based, non-sensitive risk information, together with the implementation of an open exchange and dissemination of disaggregated data are the fundamental requirements for developing a rational multi-hazard approach and of an inclusive risk-informed decision-making protocols. Indeed, one of the most critical tasks for a resilient society is in fact the ability of producing, processing, integrating, and managing and processing high quality and reliable data.

The main challenge of the problem at hand is to manage the multiple interacting factors that generate disasters. The overall behaviour of a system cannot be simply inferred from the response of a single phenomenon but must be derived from the studying of the relationships between interacting events (i.e., multi-hazard assessment) as typical of dynamical complex systems. Understanding risks involves the concept of "Digital Twins" as digital replicas of physical entities on which performing probabilistic simulations. The challenge is then to design and exploit Digital Twins at different, interacting, time and spatial scales (regional, urban, structures or infrastructures) and integrating real-time observations and monitoring, thanks to high-performance computing. To be validated, Digital Twins require the availability and management of large amounts of physical characteristic parameters, together with numerous additional interacting heterogeneous information collected from different and distinct sources.

In addition to the previously mentioned multi-scale and multi-disciplinary aspects, the specificities of the Italian territory add a layer of complexity to the problem. Italy is subjected to yearly frequent events with a local impact, such as landslides, floods and lava flows are regularly observed or as well as events with a high regional impact such as explosive eruptions, earthquakes, fires and in addition to risks on a temporal scale that cannot be precisely quantified, such as tsunamis. The Italian territory, and in particular Southern Italy, is prone to multiple hazards. Therefore, a multi-risk strategy is essential for territorial and emergency planning. Once again, these frameworks require huge computational resources.

Goals and Objectives

The ‘Environment & Natural Disasters’ Spoke has two main dependent goals: 1) To enhance both the research potential and the efficiency of the scientific community currently engaged in the modelling, simulation and management of natural and anthropic disasters as well as of their effects on the entire ecosystem, from higher organisms to microbial communities; 2) To support society and stakeholders in the definition of Disaster Risk Reduction (DRR) policies.

The designed strategies will consist in building reliable scenarios and simulations, physics-based hazard assessment, big data and machine learning analysis, real time forecasting and early warning, resilience quantification, preparedness and emergency response planning, participatory geographic information systems, virtual reality and EO data processing tools, economic and social computational models.

The Spoke not only will efficiently integrate different expertise and heterogeneous multi-source data, but it also will be a cardinal instrument to create novel scientific frameworks and technical approaches to interpret and manage the complexity of DRR and to provide civil society and stakeholders with effective, innovative tools, thus helping civil protection authorities and decision makers in adopting suitable mitigation measures.

To achieve these goals, following the Sendai Framework[6] principles, the Spoke research activities will encompass the implementation of HPC frameworks and tools that perform the following tasks : i) High resolution dynamical hazard maps of natural disasters threatening the Italian territory and ecosystem, ii) Gathering of multi-source Big Data, enrichment, image processing and machine learning for the vulnerability inventory of elements at risk; iii) Advanced site-scale modelling of disaster-inducing processes, such as

[6] https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030

landslides, earthquakes, etc., interacting with structures and infrastructures for deterministic risk assessment; iv) Advanced modelling of environmental disasters (e.g. pollution spills) effect on ecosystems and possible strategies for improved resilience; v) Typological-mechanical, Meta-modelling and multi-hazard probabilistic estimation of engineering parameters and losses (physical, economic, societal); vi) Elaboration of large scale scenarios for the analysis of citizen preparedness and ecosystems resilience to disasters; vii) Deep learning schemes for the deconvolution of multiparametric sensors' data from the monitoring networks that assist geophysical and environmental risks.

As a final milestone, the protocols and tools will be demonstrated and applied to relevant cases of natural and environmental disasters threatening the Italian territory, as specified in the following paragraph, and addressing to the main disasters, each of which represents a work package of the Spoke-5.

Application perspectives:

- Physical environment at regional and urban scale;
- Resilient infrastructures, lifelines and strategic facilities;
- Built environment in metropolitan and coastal regions;
- Built environment in inner areas;
- Architectural and historical heritage;
- Critical Infrastructure Protection;
- Industrial and production systems;
- Reconstruction planning, economic growth and societal progress;
- Physical environment and human health interactions.
- Safety of ecosystems in the "One Health" approach

Objectives:

- Big Data gathering/analysis: data management and real-time/massive data processing solutions compliant with Open Data/Open Science FAIR principles, including the initiatives for regulation and standardizing;
- HPC/QC simulations/applications/workflows: development and optimization of HPC and Quantum computing simulations, multi-scale and multi-physic models, ensemble simulations, inverse problems and stochastic modeling for geophysical, geotechnical phenomena as extreme weathers, earthquakes, tsunamis, volcanic eruptions, hydro-geological disasters, wildfires, space weather, dynamics of complex systems and chemical contamination;
- Advanced multi-scale modelling of materials and structures with a specific emphasis on damage, mechanical and dynamic nonlinearity; history-dependence; soil- structure dynamic interaction;
- Accessing and querying data and maps via WebGIS-based interoperable platforms
- Supporting political and civil protection decisions, informed by data analysis and scientific modelling;
- Supporting the Distribution System Operators in case of natural disaster to restore distribution grids and to evaluate cascading effects on interdependent systems;
- Supporting prevention planning and its relationship with ordinary planning;
- Promoting participatory reconstruction planning, economic growth and sustainable development;
- Interoperability with other CN HPC spokes and PNRR National Centres.

RESEARCH TOPICS AND LINES

'Environment and Natural Disasters' Spoke (Spoke-5) activities are organized in 8 Research lines (Work Packages, WPs) plus two complementary WPs for the management and the dissemination of the achieved results.

- WP0 Project Management

    - Leader: UNIBA; Co-leader: UNIAQ
- WP1 Dynamical hazard scenarios assessment
    - Leader: CNR;Co-leader: UNIAQ
- WP2 Vulnerability inventory of elements at risk
    - Leader: UNIFI; Co-leader: UNIBA
- WP3 Modelling of disaster-inducing processes
    - Leader: INGV; Co-leader: POLIBA
- WP4 Modelling of environmental disasters
    - Leader: ENEA; Co-leader: CNR
- WP5 Multi-hazard modelling and analysis estimation of engineering and geophysical parameters of damages and losses;
    - Leader: POLIBA; Co-leader: SAPIENZA
- WP6 Analysis of citizen preparedness and ecosystems resilience to disasters;
    - Leader: UNIAQ; Co-leader UNIFI
- WP7 Machine learning, Quantum Computing and AI platform to design and exploit Digital Twins
    - Leader: UNIBA; Co-leader: INGV
- WP8 Impact assessment and modelling of alternative risk mitigation strategies
    - Leader: SAPIENZA; Co-leader: ENEA
- WP9 Dissemination
    - Leader: UNIBA; Co-leader: UNIAQ

Spoke 6 - Multiscale Modelling & Engineering Applications

The spoke "Multiscale modeling & Engineering applications" deals on the one hand with computational aspects related to "Multiscale modeling" and on the other hand with the use of HPC and Big Data tools for applications in the various fields of Science and Engineering.
The Spoke includes expertise in the development of digital twins and model-based design methodologies, and use of simulation techniques that make extensive use of advanced numerical models implemented at major research centres, and via the intensive use of HPC and Big Data facilities.
The characteristic of this Spoke is the ability to globally address problems that span several disciplines, involving multiple scales (from the sub-nanoscopic scale of electrons and atoms, to mesoscale and macroscale), heterogeneous and large datasets, and interdisciplinary and multi-physics approaches, combined with the unique ability to translate numerical and analytical technologies into applications of direct interest and use for different scientific and engineering fields at national and international level.
Bridging the gap between the scales and developing algorithms and codes able to extract the computational power of modern computer architectures (e.g. massively parallel and with GPUs, FPGAs or future quantum computing accelerators) is a key for the development of enabling technology for a huge variety of Engineering applications in the view of fostering the development of a knowledge-based society.

GOALS AND OBJECTIVES

Scientific Goals

The characterization of the dynamics and statics in complex cyber/physical systems takes place on multiple scales of time and space, and represents a great scientific challenge, whose understanding requires a synergy between theory, advanced computational tools and digital twin methodologies.

In recent years, the possibility of determining the parameters to be used in meso- and macroscopic modeling, through microscopic approaches from first principles, and the availability of strategies based on big data and machine learning have revolutionized the field of Data Science and HPC and promise to change the paradigm with which future scientific analysis will be approached.

The spoke will aim at creating close interactions with other European initiatives in the HPC, Big Data and Quantum fields like EuroHPC, European Chips Act, GAIA-X, European Processor Initiative, EDIH, to name a few.

Training Goals

This spoke encompasses a wide range of expertise spanning from fundamental research on multiscale computational tools for engineering to their technology transfer to the industry and society at large. In this context a question arises on the availability of human resources with proper preparation updated to the most advanced technological developments.

One of the CN key functions is to finalize the preparation of young PhDs/graduates in the field of HPC to the different productive sectors and, specifically for what is of interest to the industry and society at large. These young resources could be allocated on projects of industrial interest and can spend a significant period at the industries themselves. At the end of the period covered by the NRRP, these resources, depending on availability, will be able to choose whether to remain in research or enter the industrial sector (existing or start-ups), thus giving greater sustainability to the initiative.

This spoke foresees the set-up of a new national PhD program in “Multiscale Modeling and HPC Engineering Applications” involving as founding members the Universities and research centres that are part of the spoke plus the companies (like TAS-I, Leonardo, ENI, Fincantieri, Terna) that are co-founders of the CN and expressed specific interest in the spoke.

Societal Goals

Coherently with the societal goals of the PNRR, this Spoke addresses the issues of favoring women's empowerment, enhancing young people’s skills, abilities and employment prospects, and increasing the regional rebalancing and the development of Southern Italy, also thanks to a close collaboration with the Universities, research centre and industrial centres operating in Central and Northern Italy.

Dissemination Goals

With the aim of bringing the general public closer to the scientific themes of the WP, we plan to organize several disseminations events, e.g., *café scientifique*, interacting expositions and labs. The goal is to make people aware of the many aspects of our life where HPC and multiscale modeling play key roles.

RESEARCH TOPICS AND LINES

Research Topics: Multiscale modeling

The Spoke activities address challenges requiring complex multiscale modeling.

Mesoscale modeling with quantum accuracy: development of machine learning and high-performance data analytics methods to extend high-precision quantum simulations (ab-initio) to mesoscopic temporal and spatial scales, and to path the way towards the most promising directions in the device/system parameter space.

Development of multiscale numerical methods (from the atom to the device) to model charge/heat/mass transport properties, multiphase interactions and phase transitions, while allowing the simulation of a wide range of devices/systems like integrated electronics, electromagnetics, photonics, thermal and thermoelectric, and piezoelectric devices, heat dissipation management, energy storage and phase change devices.

Development of multiscale numerical methods for the investigation and the design of multifunctional materials with innovative/coupled mechanical, electrical, electromagnetic, optical, and thermal properties (composite materials, nano-structured materials, meta-materials, two-dimensional materials etc.) and nonlinear constitutive behavior (plasticity, damage, fracture, etc.), to be exploited (at higher level of abstraction) for engineering applications.

Multiscale applications for synthetic biology aimed at developing functional materials for use in devices, sensors, and drug-delivery for applications in industry and society at large.

Multiscale modeling of complex systems: from disordered spin-glass materials to glassy materials with extremely slow dynamics, from neuronal networks to multi-agent interaction systems.

Research Topics: Engineering applications - Challenges

The Spoke focuses on the development of advanced numerical methods and applications, open software tools and workflows, to integrate the calculation, simulation, collection and analysis of data of interest for engineering applications, also through distributed computing and HPC-as-a-service approaches with a digital continuum going from cloud to edge and IoT.

In particular, the spoke will focus on the following challenges:

- modeling and simulation of complex/large systems, either linear or nonlinear, in a static and/or dynamic (periodic, transient, resonant regime...), including multi-physics problems, based on effective constitutive parameters derived from multiscale modeling approaches.
- digital twin modeling for physics-based and data driven simulations.
- advanced and innovative numerically efficient algorithms for complex/large dimension forward and inverse problems in engineering and industrial applications.

Research Topics: Engineering applications - Fields

The developed methodologies will be applied to all different fields of engineering applications such as (by way of non-limiting examples and possibly coupling the different domains):

- multi-scale modeling, from the atomic up to the circuit levels, while providing the guidelines for the realization of high-performance and reliable devices, circuits and systems for the semiconductor industry, also aiming at a link between the spoke and the European chips act initiative.
- computational fluid dynamics of complex and multiphase flows for modeling, design, optimization and control of systems, devices and tools for: energy and power, including utilization of green energy carriers, turbomachinery, efficient design and green propulsion for aerospace and automotive; intensification of chemical processes and green chemistry, rheological complex flows; scale-up of (bio)chemical operations; pharmaceutical processes and drug delivery; lab-on-chip and biosensing for biomedical applications; (bio)nanotechnologies; assessment of environmental impact and safety risk of industrial processes; control strategies in waste-to-energy plants; renewable energy harvesting and structural resilience with respect to environmental actions.
- computational mechanics of complex systems and devices for different applications ranging from modeling, design and optimization of machines and industrial production plants; supply chains; structural components, systems and infrastructures; vulnerable cultural heritage structure.
- computational electromagnetics of complex electrical components and systems for different applications at both high and low frequency scales, ranging from modeling, design and optimization of: antenna array systems, high-frequency multichip modules and integrated electronic-photonic circuits and systems, satellites, air/ground/marine vehicles, efficient electric machines and sustainable and renewable power systems;
- thermodynamics simulation of complex devices and systems for industry4.0 and green energy, power, aerospace, automotive applications.
- earth observation space-time big data processing using machine-learning advanced techniques for multi-source multispectral data fusion, non-linear multiscale field nowcasting and multiphysics numerical forward and inverse modeling.
- vector computing hardware acceleration for scientific computing simulations, even using new numerical data formats, integrated in the software ecosystem of the free and open RISC-V instruction set architecture, leveraging a full software toolchain.
- massive data analytics for engineering applications such as surveillance and anomaly/intrusion detection for predictive maintenance and security, homomorphic encryption for data privacy and protection in multi-user HPC, verification for functional safety and reliability in cyber-physical systems.

- massive data analytics to reduce environmental impact in air, maritime and land mobility of people and goods (logistics)

Spoke 7 - Materials & Molecular Sciences

From the stone to the silicon age, human civilization has revolved around the discovery of new materials and of new functions for existing ones. The energy, environmental, and climate emergencies set the stage to a paradigm shift whereby the serendipitous discovery of natural materials and their subsequent extraction is being replaced by the deliberate design of artificial ones with customised properties for specific applications and of innovative processes to synthesise them. The digital revolution is the engine propelling this shift, with computers capable of processing billions of billions of information bytes per second, which make it possible to solve the fundamental equations determining the properties of complex materials in realistic device conditions and to train artificial intelligence to predict their complex emergent behaviour without even solving these equations in full detail.

Goals and Objectives

The spoke *Materials & molecular sciences* aims to
- **Strengthen** the Italian leadership on a global scale in the development, implementation, maintenance, and free dissemination of key enabling software technologies for the multi-scale simulation of materials and molecular systems;
- **Optimise** and enhance the scope of existing high-performance computer codes, to stay at the leading edge of the main scientific, industrial, and societal challenges in our country, in the fields of materials and molecular sciences;
- **Enable** the optimal access to the HPC infrastructure of the National Centre, for the good of the entire national research community, with emphasis on the transition to heterogeneous energy-savvy hardware architectures, based on specialised processing units and aimed at exascale performance;
- **Steer** the efforts of Italian materials and molecular modelling communities towards the *key enabling technologies* identified by the PNRR.

The scientific, technological, and societal benefits of our research will accrue to three levels of beneficiaries of increasing breadth:
- code developers will benefit from modular high-quality software components, optimised on a broad range of hardware platforms, ready to be embedded into third-party open-source and proprietary packages for materials and molecular modelling, and from state-of-the-art software-engineering best practices to be adopted in their projects;
- researchers in materials science, chemistry, and various branches of engineering with a varying degree of theoretical knowledge (ranging from the specialist down to the occasional user) will have a set of highly performing, ready-to-deploy, and easy-to-use software tools at their disposal;
- finally, the society at large will benefit from the technological advances that will be empowered by the ability to design new materials and molecular species with bespoke properties, for applications in such diverse fields as energy, environment, health, climate, fundamental research, or various branches of the retail sector.

The program of this spoke is being designed and will be implemented according to the guidelines of Open source software strategy, issued by the European Commission.

Research Topics and Lines

The activities of the spoke will be developed in coherence and synergy with the European strategic agenda of Horizon Europe implemented by EuroHPC - European High Performance Computing Joint Undertaking, and organised in 5 work-packages, along the following research lines:

1. **Flagship codes**: Extension, optimization, maintenance, and distribution of world-leading flagship codes for the high-performance numerical simulation of materials and molecular systems, developed by the national community (*i.e.* Quantum ESPRESSO, YAMBO, Plumed, Crystal, ...).
2. **Big data generation and harnessing**: Development and implementation of advanced methodologies (high-throughput computing, artificial intelligence, machine learning, and optimization techniques) for the generation and analysis of large volumes of data and for structural prediction, aimed at the discovery of new materials and drugs, and at the development of predictive models of molecular interaction in complex systems.
3. **Accuracy and reliability**: Evaluation of the accuracy of the theory levels implemented in simulation codes with respect to the chemical accuracy standard.
4. **Pilot applications**: Application of the software / middleware developed in points 1 and 2 to demonstrate their ability to solve selected and significant technological/scientific challenges, hitherto considered intractable.
5. **Materials foundry**: Development and enhancement of highly scalable, optimised codes and workflows towards an ecosystem of HPC applications aimed at:
    - Solving specific problems driven by **industrial demands**;
    - Assisting and supporting **experimental infrastructures** and research.

Spoke 8 – In-Silico Medicine & Omics Data

Spoke 8 will carry out two major activities, i) **In Silico Medicine** & ii) **Omics Data**.

In Silico Medicine

The term In Silico Medicine refers to the use of modeling and simulation technologies in healthcare. One of the most popular taxonomies divides In Silico Medicine's solutions into three categories depending on who the user is In Silico Medicine solutions into three categories depending on who is the user: Digital Patient solutions, where the models' predictions are used as clinical decision support systems; In Silico Trials solutions, where the computer models are used to evaluate the safety and efficacy of new medical products, and Personal Health Forecasting solutions, where the end users are the patients themselves.

In Silico technologies are widely used in the development of medical products. A possible list would involve medical devices, that are designed using computer aided engineering tools like finite element analysis of computational fluid dynamics. Simulation platforms (like agent-based systems), used to model and simulate network of biological processes. Systems biology models, used to identify druggable targets. Molecular dynamics models employed to optimize the potency of new molecular entities. Software as a medical device i.e., software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device.

Using a taxonomy introduced for methods alternative to animal experimentation, we can categorize the use of In Silico Trials technologies to reduce, refine, or replace in vitro, in vivo, or human experiments. Human experimentation is the most ambitious target for In Silico Trials, with a growing difficulty going from the refinement of clinical trials to their reduction, and last to their replacement.

Omics Data

Precision medicine aims to enable clinicians to act quickly, efficiently, and accurately in deciding the most appropriate course of action for each specific patient. This is achieved by analyzing big data produced (and stored) on each patient, from omics sciences. However, to be exploitable by clinicians, these data need to be fully integrated into the clinical workflow, using modern information-technology healthcare platforms.

Notably, omics data are broad and quantitatively massive, including genomics, proteomics, metabolomics, lipidomics, metagenomics, and transcriptomics analyses. The human genome, the complete genetic information contained in every single cell of the human body, is per se big data as it is generated by sequencing

the DNA composed of over 3.4 billion pair bases. As a matter of fact, human genome sequencing represents a major revolution in medicine and biology. It requires critical hardware and software to be securely stored and adequately analyzed, with the promise to precisely determine clinical care. Each genome may occupy up to 100 Gigabytes and requires HPC to be promptly examined. Additionally, the other omics above mentioned may rapidly produce additional massive data (e.g., untargeted lipidomics may provide quantitative data for over 120 lipids with a single UPLC-MS/MS run, proteomics may produce data for more than 1000 proteins, metagenomics generates genetics data for gut microbiota (billions of microorganisms), etc.).
The systematic and massive production of such data requires adequate computational infrastructure, including high-capacity storage (with particular attention to cybersecurity issues) and HPC for data analysis. For instance, variant calling pipelines applied to genomics sequence may require over 50 hours of simulations with traditional CPU-based software. The simulation burden can be reduced to slightly more than 5 hours with GPU-compliant software (or novel codes for FPGA). Furthermore, the systematic investigation of drug-target interaction with physics-based approaches requires repetitive simulations and large computing infrastructures for achieving reliable outcomes and predictions.
Against this scenario, we aim to provide a full proof-of-concept of data production, storage, and analysis, and to investigate drug-target interaction to improve diagnoses and therapies in different pathological conditions (e.g., oncology, neurodegenerative diseases, neurodevelopmental disorders, etc.). Through collaboration with clinics and hospitals (mainly IRCCS), we plan to build a complete pipeline for omics data management and analysis, as follows:
Patients → Biological (blood) samples → Omics data generation → Omics data analysis → Precision medicine (i.e., improved diagnoses, better pharmacological therapies, drug repurposing, etc.).

Goals and Objectives

In silico medicine aims at the following objectives:

1. To develop a fully integrated digital data flow between the HPC centre and the healthcare centres.
2. To use HPC solvers to develop vertical solutions for In Silico Medicine (both Digital Patient and In Silico Trials, which provide sufficiently large-scale or the necessary computational performance)
3. To expose these solutions through automated pipelines that hide the complexity to clinical and industrial end-users and allow automated workflows that execute and return into the Electronic Health Record of biomarkers predicted in silico as soon as the necessary input data are made available for that patient.
4. To conduct comparative studies on selected in silico medicine applications to quantify the performance and cost benefits of using an advanced HPC platforms and not medium-range clusters or cloud computing.

Omics data aims at the following objectives:

1. To design and develop a fully automatized workflow, in which all the steps can benefit from ad hoc software, graphical interfaces (i.e., LIMS), etc.
2. To deploy bioinformatics pipelines capable of performing exceptionally on next-generation hardware, including GPU-based HPC or FPGA.
3. To develop machine learning algorithms (including machine vision) to analyze omics data (including radiomics) efficiently, with the final goal to provide reliable correlations and associations.
4. To stratify patients for improved diagnoses and more efficacious pharmacological treatments.
5. To calculate thermodynamics and kinetics of drug-target binding through HPC-optimized physics-based methods (portability to quantum computing will also be investigated).
6. To deploy user-friendly and ready-to-use software packages for the medical and biological community country-wise.

The possibility of sharing omics data with the broader scientific community, in compliance with FAIR principles though protecting the privacy of patients' data according to the "as open as possible, as closed as necessary" paradigm, will enable new scientific insights on the molecular mechanisms underlying diseases and, in turn, increase over time the predictive power of the data collected from new patients. These activities are aimed to bring precision medicine into the clinical practice in Italy and Europe.

This virtuous circle will be made possible thanks to the integration of the Omics Data platform within the broader ELIXIR and ELIXIR-IT services ecosystems for biomedical data that will ensure the adoption of internationally recognized best practices for human omics data and metadata handling, processing, and sharing. For example, services like the Federated European Genome-Phenome Archive (EGA) Node that is currently in its development phase at ELIXIR-IT will provide the natural long-term data repository interface for the HPC services the Omics Data platform will develop and provide.

RESEARCH TOPICS AND LINES

Spoke 8 will deal with two research lines: i) **In Silico Medicine** aimed at designing and developing modeling and simulation platforms for in silico trials; ii) Production of **Omics Data** and their analysis through the development of next-generation bioinformatics pipelines and machine learning codes.

Spoke 9 - Digital Society & Smart Cities

HPC and big data - combined with suitable models, methodologies, and algorithms - offer new opportunities to solve key challenges in smart cities and in digital societies. Indeed, these are characterized by the confluence and interaction of different systems in the social, organizational and technological domains, thus impossible to solve through decomposition into easier problems, and requiring new approaches able to overcome this "complexity wall".

The spoke intends to face this complexity by investigating novel approaches that build upon - and extend - the concept of "digital twins". The aim is to create a faithful digital representation of social and organizational structures of cities and communities and of their citizens, and of the physical and virtual contexts where they operate and interact, by exploiting available "big data" digital tracks, powerful data analysis and Artificial Intelligence (AI) techniques and advanced simulation opportunities unlocked by HPC infrastructures.

Through digital twins, the spoke intends to improve our capability to (i) replicate and understand the functioning and behaviors of our cities and societies, (ii) forecast future evolutions, also in response to changes, and (iii) support the experimentation and the evaluation of the effects of policies, protocols and scenarios aiming to change the behavior of cities and communities.

GOALS AND OBJECTIVES

In the context just described, the spoke intends to address three objectives:

1. **Research** on the foundational, methodological, and technological aspects relevant for the challenges of the spoke.
2. Identification of the relevant **application domains** in the wide area of smart city and digital society, of the research challenges and of the exploitation opportunities offered by these domains.
3. **Experimentation and evaluation** of the research results with "beneficiary" partners.

RESEARCH TOPICS AND LINES

Three research lines will be pursued by the spoke: (**M**)odelling, (**C**)omputing, and (**S**)oftware systems. Regarding **(M)**odelling, the spoke plans to develop innovative AI approaches, with a focus on deep learning methods and technologies such as computer vision, speech recognition, natural language processing and generation, conversational agents and dialogue systems, etc. Additionally, methodologies from computational social science, complex networks and complex systems will be adopted to model and predict societal phenomena. Finally, strong attention will be devoted to data governance and privacy aspects.

The spoke will also develop, within the (**C**)omputing research line, architectures and enabling technologies for distributed IoT sensing systems and for the integration with cloud infrastructures and edge systems as well as architectures for twin systems and hybrid AI-based/physics-based modeling of large-scale systems and interconnected infrastructures. Finally, the research line on (**S**)oftware systems will be devoted to the definition of system software platforms, to software specification and maintenance, to the prototyping and scaling of AI algorithms and applications, etc.

Application domains

The spoke is focusing its activities in the following 5 application domains.

- **Health and lifestyle**: our spoke plans to develop HPC and IoT technologies for e-Health, digital twins for the management of smart hospitals and personal health records, big data platforms for the advanced management of territorial medicine as well as for the monitoring of behavioral reactions and infodemic outcomes associated with COVID-19 pandemic, etc.
- **Mobility**: our focus is mainly devoted to the modeling of citizens' mobility patterns and large-scale crowd dynamics in order to design more efficient traffic management systems, sustainable multimodal public transportation, etc.
- **Socio-economic analysis**: we will focus on models for individual/collective human behavior, learning and adaptation, relationships (including diversity and inequality), analysis of economic flows, movable and immovable asset markets, labor market gaps, as well as novel approaches for the identification and tracking of misinformation, disinformation and hate social media, new algorithmic frameworks and tools for monitoring urban crime using multiple sources of data (e.g. video/audio from surveillance cameras,..), etc.
- **Infrastructures and utilities**: the spoke will investigate models and techniques for smart grids, including innovative applications such as drone-based maintenance, radio coverage planning for wireless networks, real-time monitoring, risk analysis, and forecasts for interconnected infrastructures.
- **Environment**: the spoke will address monitoring and forecast for environmental conditions, regeneration of city biodiversity and ecosystems, water cycle management, etc.

Key enabling technologies

Enabling technologies and infrastructures for the activities of the spoke include:

- **HPC**: availability of suitable computing and storage infrastructures;
- **IoT**: integration with networks of sensors (including video/audio devices) deployed in the city / territory;
- **Big Data**: availability and accessibility of relevant, diverse and interoperable data sources;
- **Simulation**: specifically, simulation models and technologies based on "twin" concept;
- **AI**: availability of suitable models and algorithms for data analysis, forecasting and risk estimation.

Spoke 10 - Quantum Computing

The spoke deals with the exploitation of Quantum Computing technology which could radically transform business in every industry, through the resolution of complex problems in the field of optimization, simulation and machine learning. Research and development efforts to create a large and scalable quantum computer, able to support a practical use case, are accelerating with the goal of overcoming technical challenges related to error and system reliability.

Goals and Objectives

The development is conducted from two perspectives: The first perspective is to identify quantum algorithms characterized by a speedup, even limited, if compared with the corresponding classical algorithm. Working on this kind of software applications, eventually using different hardware approaches (e.g. Gate Model Quantum Computing, Quantum Annealer), could accelerate quantum solutions industrialization. The second perspective is to use the quantum approach to perform calculations on state-of-the-art classical computers (quantum-inspired algorithms, emulators, tensor network computations) in order to gain significant benefits in terms of algorithms efficiency without using a quantum hardware.

In both cases, the burdensome challenge is to redesign and redevelop existing algorithms in the quantum environment, which is itself under development from different points of view: the programming language, the middleware platforms connecting quantum solutions to the hardware and the usable toolkits for developers. It is necessary to find a compromise in order to benefit from Quantum Computing when the classical computation fails while managing the trade-off generated by the effort to develop the system. This implies the identification of effective design and programming abstractions that allow people skilled in computer science to take advantage of the enormous power of quantum computing still keeping the complexity of design and programming activities under control and enabling analysis and testing of the developed code.

Identifying use cases, identify performance measures to evaluate potential benefits, changing the systems architectures and developing quantum algorithms are future challenges and require strong technical skills.

Research Topics and Lines

The research activities carried out in the spoke embrace different levels of abstraction that contribute to building a quantum computing system:

Layer 1. Applications. High-level quantum applications for the solution of special-purpose research and industrial use cases (chemistry, biology, high energy and condensed matter physics, data-science. industrial optimization, etc.).

Layer 2. Algorithms. General purpose algorithms (e.g., linear algebra, variational eigensolvers, machine learning, hard optimization, etc.). The research challenge is to identify general-purpose library algorithms best suited to be accelerated on near-term quantum computers and that can be used as building blocks for vertical applications.

Layer 3. Emulation. Software for the emulation on classical computers of particular quantum architectures. Benchmarking and verification of quantum computations.

Layer 4. Compilation. Tools for compilation and optimization of algorithms. Toolchains for hw/sw codesign of special-purpose quantum accelerators.

Layer 5. Firmware. Low-level software for the physical operation of specific quantum computers: control of physical operations, optimization of the operations, measurement protocols, scheduling of the operations, automatic calibration, etc.

Layer 6. Hardware. Quantum computer hardware components. Here, the research challenge is to play a role in the international production chain.

Different work packages take main responsibility for different layers:

- WP1: layers 1 and 2
- WP2: layers 3 and 4

- WP3: layers 5 and 6

The spoke will concentrate on the following research lines:

- development of quantum applications for scientific and industrial vertical domains;
- development of general-purpose algorithms, libraries and frameworks to access quantum accelerators;
- development of software for benchmarking and verification of the quantum computations;
- development of software for optimization and compilation of quantum algorithms on specific target architectures;
- development of low-level software for the physical operation of specific quantum computers;
- development and support of the quantum computer hardware chain.

We will develop an experimentation phase with use cases and demonstrators, that will be defined based on:

1. Cooperation with other spokes of the National Centre, that may want to experiment their application-driven algorithm on the quantum platforms;
2. Use cases developed in cooperation with the companies participating to the National Centre, that have indicated their interest in exploring the application of quantum computing, in particular Thales, Banca Intesa, ENI, Fincantieri, and may be others that may specify their interest in later stages.

SII - Transversal Research Group on Societal Implications and Impact

The research group on Societal Implications and Impact is an inter-spoke group that operates within the Spokes' Coordination Board.

Goals and Objectives

The research group covers issues on the social, legal, ethical, scientific, political, and economic implications and impact of big data and HPC, such as the following ones: (i) *Scientific Implications and Impact* – creating high-quality new knowledge, strengthening human capital in R&I, fostering diffusion of knowledge, Open Science, and transformation in research systems, developing models and skills for data-driven science; (ii) *Social and Ethical Implications and Impact* – improved efficiency and innovation, improved awareness and decision-making, participation, equity, fairness, social trust and acceptance, sustainability, transparency, anti-discrimination, access to education/quality of education; (iii) *Economic Implications and Impact* – improved efficiency, innovation, changing business models, employment and job opportunities, dependency on public funding, fintech, socially inclusive economic growth; (iv) *Legal Implications and Impact* – fundamental right, intellectual property issues, adverse effects on workers, privacy and data protection issues, liability and accountability, cybersecurity, unfairness, bias and discrimination, contestability, compliance; (v) *Political Implications and Impact* – private, public and non-profit sector, control of actors abroad, improved decision-making and participation, e-democracy and digital citizenship, political abuse and surveillance, digital sovereignty.

Research Topics and Lines

The research group provides a science and knowledge service for the governance of the CN; it advises all spokes and industrial partners regarding their research programmes; it develops independent research activities and training programmes for the CN; it publishes policy recommendations and other similar documents and enables the CN to contribute institutionally to policy discussions, at the national and international level, and to European public consultations, such as the ones on Data Act or AI Act; it promotes public awareness and trust through dissemination and co-creation activities for the public engagement in the science, technology and innovation practices and policymaking of big data.

A.3 The methodology

Spoke 0 – Supercomputing Cloud Infrastructure

The methodologies that will be adopted are described in the following according to the main planned actions.

Network

GARR, the Italian Research & Education Network, will deploy the new GARR-T infrastructure able to scale to multiples of Terabits per second (Tbps). GARR-T(erabit) is the new generation network of GARR which will seamlessly evolve from the current infrastructure to an innovative network based on the state of the art of transmission technologies and will be future proof. The main advantage of this new network architecture will be the possibility of adaptation to the increasing requirements of users in terms of capacity, capillarity and services. A larger and pervasive use of automation will be one of the major characteristics of the new network that will thus be more dependable and measurable thanks to sophisticated monitoring systems.

Technically GARR-T is a fibre optic network (currently comprising more than 16.000 km) owned in IRU (Indefeasible Right of Use) using a high-performance transmission system based on Open Line System technologies (i.e. a combination of flexgrid spectrum and super channels opening the way to a new open and programmable optical layer in networking, that enables more effective sharing the spectrum among a wider number of users) with an orchestration based on software and hardware under the full control of GARR and able to scale to several tens of Terabits.

The new infrastructure will follow state-of-the-art paradigms that will allow to manage the network in a flexible way, giving access to the lower optical level also to users not on the same premises of the PoP (Point of Presence). It will also implement new solutions in terms of security and Edge Computing.

GARR-T has been designed and will be implemented to fulfil the needs of Research and University in Italy for next decade in complete synergy with the evolution of the European Research and Education Network GEANT aiming to build a common and open optical network landscape in Europe. As part of this process, the international connectivity between the Italian and European HPC centres will be increased to address the challenges of the next decades.

With regard to the enhancement of the GARR network infrastructure, the integration of the HPC data canters of the private partners will also be envisaged, with particular reference to Eni's Green Data Centre where the installation of an HPC system of the infrastructure of the National Centre is planned and in order to also integrate Eni's HPC system into the national federated HPC infrastructure. Leonardo company HPC infrastructure will be integrated in the GARR-T infrastructure and SOGEI system cloud facility will be integrated as well.

Data Centre Facilities

The main data centre is part of the *Tecnopolo (Bologna)* and its realization represents one of the most important convergences of European, Italian, and Regional investments in favour of HPC and its applications. This new facility will begin operations in the second half of 2022 and will be able to accommodate HPC resources up to 10 MW of power consumption (“Phase1”, 2022-2024). At least two world-leading infrastructures will be hosted in the same site: Leonardo supercomputer, the EuroHPC Tier-0 pre-exascale system, and the INFN

Tier-1 system, part of the WLCG grid and cloud infrastructure, for processing data produced by the LHC experiment at CERN, as well as by other scientific collaborations. At the Tecnopolo, within the present PNRR research program, it is also planned to: (1) upgrade Leonardo supercomputer up to the maximum permitted by EU regulation (30% of the total acquisition cost), with 35% of co-funding by EuroHPC according to the recently published "Call for expression of interest for upgrading the EuroHPC JU Supercomputers" (EUROHPC-2022-CEI-UPG-01), (2) host and operate a European quantum computer co-funded by EuroHPC according to the recently published "Call for expression of interest for the hosting and operation of European quantum computers integrated in HPC supercomputer" (EUROHPC-2022-CEI-QC-01), (3) participate to a European exascale initiative according to an upcoming call by EuroHPC, (4) integrate Leonardo supercomputer with a tier 1 system, also managed by CINECA, to support the CNR and INAF needs.

On a slightly longer time scale, the Tecnopolo infrastructure will be evolved to serve up to 24/30 MW by the end of 2025, prepared to host a second-generation post exascale EuroHPC Tier-0 system. This upgrade of the data centre will not be funded by the CN; the operation will anyhow be executed with great attention to the energy aspects, employing advanced solutions to maintain the power usage effectiveness at 1.1 or better.

The other centres are operational and expected to participate to the national HPC and Big Data network; some of them, though, have facility infrastructures (power distribution, cooling, space for HPC resources) which were not renewed in the last 10 years or more, and may therefore lack the capability to guarantee appropriate levels of flexibility, power consumption and carbon footprint appropriate to the current challenges. The tow of the spoke, which foreshadows widespread investments for the renewal of the computing systems, it will have to induce further economic mobilization in order to renew such infrastructures and meet the most advanced standards in terms of requirements, footprint and future sustainability.

In addition, the creation or the upgrade of thematic data centres based on existing competence centres is planned: a data centre devoted to "Natural Disaster Resilience" (HPC4NDR) in L'Aquila, an HPC facility on climate change in Lecce managed by the CMCC, a data centre devoted to "Space science" in Frascati, the set-up of a Tier-1 HPC infrastructure in Naples and a few smaller centres that will be definitely selected before the end of this year among some possible sites already proposed by Spokes, according to an internal assessment based on scientific, technical and feasibility requirements. These data centres will operate in conjunction with the HPC network.

Computational Resources

The European HPC (EuroHPC) Joint Undertaking co-funded three pre-Exascale systems in Europe. One of them, Leonardo, will be hosted in the *Tecnopolo di Bologna*, managed by *CINECA* and start its operations in 2022. In order to further develop and provide access in Italy to a world leading computing service and data infrastructure - by the means of high-end supercomputers indispensable to run the most demanding and strategic applications - as part of the actions for this National Centre it is foreseen to further increase the operational performance of Leonardo upgrading the system with another investment, economical equivalent to the 30% of the initial investment of 120 millions of Euro, as also foreseen in the strategic plan of EuroHPC, which will cofund such an upgrade. This will provide new functionalities to address the evolution of Italian research system. Moreover, the integration on the Leonardo platform of a quantum computing, simulator or physical system, is foreseen, to combine digital and quantum technology in advanced use and scientific cases, that also in line with the strategic plan of EuroHPC. CINECA, beside the deployment and the exploitation of the Tier0 HPC national supercomputer system, will manage and deploy also the main Tier1 HPC systems at national level. Indeed Galileo 100, the CINECA Tier 1 system already in production in the Casalecchio data Centre will be integrated in the National Centre HPC infrastructure; one more Tier1 system will be installed in the Tecnopolo data centre in Bologna, and one in Naples. It's also foreseen the installation of a Tier1 system in Milan, hosted in the Eni green data centre, and one more as a development process of the data centre of the CINECA point of presence in Rome. The storage system HPC system infrastructure will be also upgraded as well, moving from the current capacity in the order of 100 PB to the level of some hundreds of Petabytes of online capacity.

INFN is operating a system of data centres spread over Italy, able to manage and elaborate the data coming from several (~50 currently) scientific experiments, and in particular from those at the CERN LHC collider. The system currently offers about 200 PB of storage (as on-line storage or long term archival) and 100 thousand CPU cores. The main INFN data centre will be located at the Tecnopolo. By 2025, the computing and storage capacity will be increased by at least a factor 2 to offer services also to research domains beyond those directly connected to INFN.

The processing capacity of several HPC and Big Data infrastructures participating to the Italian network will be enhanced to increase the offer for all scientific domains, thus facilitating the development and migration of HPC optimised scientific applications toward state-of-the-art, federated and scalable solutions.

In order to fulfil the needs of as many domains as possible, it is foreseen to procure and operate heterogeneous resources (e.g., conventional, accelerated, FPGA, neuromorphic...) for both R&D and production purposes. Furthermore, it is foreseen to create systems to manage data at different level of confidentiality, e.g. sensible data coming from medical applications, genomics, etc.

To provide to Italian and European researchers with quantum computing resources to match a growing demand from European industry and academia and to foster the emergence of real use case applications and mature large-scale quantum computing in Europe, Quantum Computing systems and simulators will be procured and be part of the computing network, possibly in collaboration with EuroHPC. This will also contribute to the development of an ecosystem of quantum computing algorithms, application libraries and skilled workforce.

Summarizing, the methodology in this context develops along the following two main lines:

- HPC, Big Data and Quantum resources: increase the state-of-the-art computational, data storing resources and acquisition of a quantum simulator and accelerator to improve research competitiveness and serve the needs of scientists across all domains.
- Cloud infrastructure and services resources: Increase the state-of-the-art computational power, data storing resources and scalability capabilities. This could be achieved by the acquisition of “commodity hardware” to compose several cloud clusters.

Middleware

The overall architecture underpinning the National Centre is based on federation services enabling the aggregation at multiple levels (IaaS, PaaS or SaaS) of existing data centres managed by the partners of the National Centre through open-source middleware. The architectural model is dynamic, so that both the resources and the application domains can be volatile. Next Figure shows the proposed layered architecture

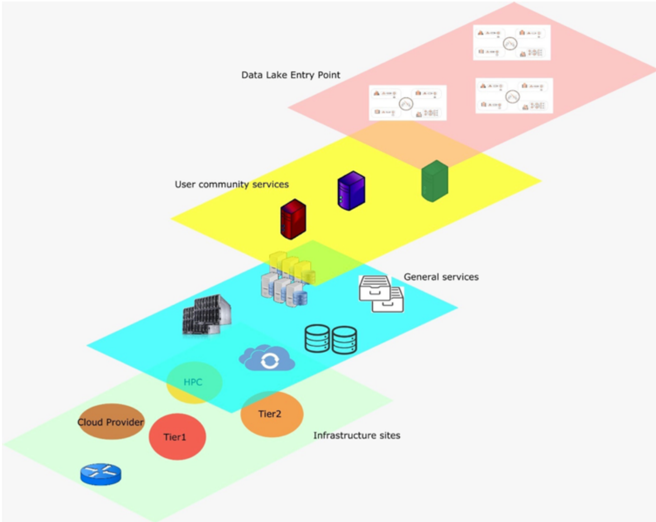


Figure 5 Layered architecture of the National Centre resources, connected by the National Centre middleware at different layers

The high-level federated architecture will be compatible with EuroHPC, the European Open Science Cloud (EOSC) and with the principles and implementations of GAIA-X. It will be based on a clear separation between the physical and logical layers. In particular the federation will easy the researchers to leverage their home institution credentials in order to access to an easy-to-use and transparent compute resource federation. This will be implemented in particular using OpenIDConnect based solutions. At the physical level, the high-performance network (GARR-T) makes the system independent of the actual location of the resources. At the logical level, all the resources are seen as part of an Italian research data-lake. Each spoke will have a unique entry point to its own data-lake that is connected to the top-level, national entry point.

Each physical data centre can support one or more application domains, by publishing specific capabilities and running application-specific services. The middleware will make it possible to create, evolve and share a service portfolio composed of both core components and possibly reusable application-based, composed services. Each application domain data-lake is defined as the sum of the services (e.g., portals, software services, resources) available for that domain. A dynamic match-making mechanism connects the physical and logical domains allowing to fulfil the high-level requests coming from end users. This matchmaking will maximize resource utilization across the National Centre and facilitate a reduction of energy consumption, through the orchestration of workloads according to customizable policies.

The middleware will be modular and will be based on production-level services and experience of the National Centre partners, adhering to de facto and de jure standards of Cloud best practices. A limited amount of development effort to integrate and extend existing middleware technologies for the specific needs of the National Centre is foreseen, while the main effort will be devoted to the development of SaaS and application stack solutions.

Spoke 1 - Future HPC & Big Data

Embracing Open Standard. FutureHPC relies on Open standards to support EU digital sovereignty and foster user adoption. Also, researchers are committed to participating in international bodies and actions driving the evolution of open standards to maximize the research impact.

A national umbrella HPC community. FutureHPC aims to create an umbrella community of HPC researchers as a capillary distributed national research companion for national supercomputing centres focused on operation rather than research. We know that there are no software development resources to spare. There is a global talent shortage for developers who understand HPC engineering challenges and parallel and distributed systems (HW and SW). Given the duplication of effort we see today on the same core functions, we need to constantly look for opportunities for developers to work on shared code and roadmaps. We are not only building consensus and integration mechanisms; we need to live them as a development priority, and we sustain existing codes by integrating them without interfering in individual development trajectories.

The natural selection of prototypes. The umbrella community naturally selects the best prototypes (HW and SW) and integrates them into a unifying national stack of prototypes representing the output of the "technology development" step in the HPC value chain. The national supercomputing centre helps validate, improve and promote the prototype hardware-software stacks by adopting them and supporting national and international users in their use. Thanks to the tight interaction with the umbrella community created by the FutureHPC spoke, the national supercomputer centre can broaden its portfolio of services with highly innovative pre-commercial prototypes that precede the broader market and match the research capability with potential industrial exploitation of prototypes.

The living labs. FutureHPC will incubate and promote two national living labs acting as permanent workshops aimed at 1) supporting the integration of scientific activities, hardware and software prototypes across all the spokes; 2) contaminating university and industry researchers by making them work together in the same place, matchmaking industry needs with academic capabilities; 3) training early-stage researchers and industrial practitioners raising the baseline of skills in HPC and BigData; 4) fostering the inception of spin-offs and the integration of SMEs into the broader industrial market.

Spoke 2 - Fundamental Research & Space Economy

Fifteen (15) Italian research institutions participate in the activities of Spoke 2, by dedicating personnel and effort to the planned activities. A total of almost 200 scientists have committed a part of their research time, for a total of more than 1,000 months within the 36 months of the project lifetime.

The institutions participating in Spoke 2 are among the major stakeholders in the scientific activities at the hearth of the Work Package plans. INFN and INAF are the Italian Research institutions (EPR) devoted to base and applied research in the fields of nuclear, subnuclear, astroparticle physics, with direct commitments to the construction and operations of the most important experiments in today's big-science landscape (e.g., CERN and FNAL experiments in particle physics, gravitational-wave detectors, ground-based telescopes and satellites for space observation). All the Universities, and in particular their researchers exposed in the Spoke 2 activities, collaborate with INFN and/or INAF via at least decennial agreements and have a solid tradition and organization of interinstitutional teamwork.

Among the background of the scientists, all of them high-profile, we wish to highlight a number of past experiences and responsibilities relevant to the spoke activities:

- physics and computing coordination for the most computing-intensive experiments;
- experience in obtaining large computing grants on HPC systems from international bodies;
- project coordination of multi-million research projects, including design, construction, operations and analysis of Exabyte-scale datasets;
- design and operation of global scale e-Infrastructures serving science and industry, also via the coordination of multi-million Europe funded projects.

The project goals are expected to have a huge scientific impact in the selected domains, where it has been recognized since long[7] how a simple scaling of the current computing methodologies is expected to fail within a decade. In particular, the project aims to be a key player in developing and testing solutions which ensure the sustainability of computing for the next generation of scientific experiments. Sustainability needs to follow at least 3 distinct directions:

[7] A Roadmap for HEP Software and Computing R&D for the 2020s, Comput Softw Big Sci (2019) 3, 7

1. Economic sustainability, where the computing infrastructure (including resources, codes and algorithms) needed to extract physics knowledge from the collected / produced data can be procured within a reasonable cost envelope. According to the currently adopted standard criterion adopted for the computing of high-energy physics experiments, economic sustainability means that the yearly budget for computing must at most be constant at today's level. Solutions such as the use of heterogeneous computing with accelerators promise to decrease the cost per operation, while data-lake infrastructure designs point towards a reduction of the total storage deployment;
2. Environmental sustainability. This includes many aspects covered by the Spoke 2 activities:
   a. the design of algorithms with a better cost per operation. This includes brand new solutions that may be designed, but also general strategies like a more pervasive utilization of Machine Learning[8] technologies, which are known to lower the processing time in situations such as heavily combinatorial algorithms.
   b. the use of heterogeneous technologies with a better operation/Joule cost. This includes the use of mobile-derived CPUs like ARM, known to perform at least 3x[9] better energetically, and the use of GPGPU solutions, which scale similarly[10] by using a more parallel architecture;
3. Scientific sustainability, in which the algorithms increase in complexity when analyzing data from more and more precise experiments. This is crucial to the advancement of the scientific domains and implies that the resource needed will increase faster than a naive extrapolation from data rates would suggest.

All the activities in Spoke 2 will be executed in strict collaboration with the major players in the respective scientific domains; it is thus important to maintain active and frequent communication channels with them.

Examples are:

- activities on porting / modernizing high energy physics codes need to be coordinated with the experiments selected as use cases. In particular, an active communication channel needs to be maintained with their Computing and Software coordination, with constant participation to meetings and sharing of plans;
- activities on porting / modernizing computing models towards the data lake must be coordinated with the Spoke 0, responsible for its development and operations, and at the same time with the Computing coordination of the experiments selected as use cases, in a manner similar to that described in the previous point;
- Similar activities are carried on by a number of boards and committees on the national and global level, like the Hep Software Foundation (HSF[11]), the Worldwide LHC Computing Grid (WLCG[12]), Excalibur in the UK and IRIS-HEP in the USA.

In all these cases, the Spoke 2 will enable active communications as soon as the project starts and will be kept active during all its duration; communications and shared plans are also important since the experiments are the owners of the data and tools we plan to improve.

---

[8] Computing models in high energy physics, Reviews in Physics, Volume 4, 2019, 100034, ISSN 2405-4283, https://doi.org/10.1016/j.revip.2019.100034

[9] Explorations of the viability of ARM and Xeon Phi for physics processing, https://arxiv.org/abs/1311.1001
[10] Performance studies of Goofit on GPUs vs roofit on CPUs while estimating the statistical significance of a new physical signal, J. Phys. Conf. Ser. 898 Track 5: Software Development.
[11] https://hepsoftwarefoundation.org
[12] https://wlcg.web.cern.ch/

While Spoke 2 is initially focused on scientific domains, and develops solutions and ideas for their advancement, we claim that such solutions will be useful outside basic science. The expectation is corroborated from previous experience: basic science, and in particular High-Energy Physics, has been historically anticipating needs that later have been seen in other domains. The solutions developed internally, together with research personnel flowing from academia to the productive sector, have been seeding similar solutions in the same direction; classic examples are the World Wide Web in the 80s, and the Grid in the first decade of the 2000s.

In this same direction, the Spoke will work towards a dissemination of the solutions towards sectors like the Space Economy and in general the productive sector, with the ambition to show, via shared testbeds, proofs of concept, testing and benchmarking campaigns, their capability outside basic research. A more direct and strict contact with the productive sector is expected to have a flow of ideas and solutions in the opposite direction, where tools and ideas from the productive sector serve as seeds for the development of new ideas in research. The participation to the Spoke of two of the largest companies in Italy ensures this bidirectional siphoning of techniques and technologies.

Spoke 3 - Astrophysics & Cosmos Observations

The Objectives and Actions introduced in (Goals/Objectives) will be addressed according to the methodology described below, consisting of the following pillars: *A)*: *ACO-S will implement a fertile environment* within which researchers, technologists and experts from private enterprises involved in the Spoke activities can confront each other on domain-specific common issues and challenges, increasing the incisiveness of code-specific actions by sharing knowledge, outcomes and possibly software technologies and implemented solutions. *B)*: *Co-design approach for exascale-enabling of ACO-S applications and algorithms*. This calls for a continuous interaction among the science partners and the HPC and hardware experts in XC, that bring in the project cutting-edge hardware, technologies, expertise, and knowledge. The latter items are also being refined and strengthened in other XC Spokes and more in general within the HPC European projects. In particular, in EPI, FENIX, EUPEX, DEEP-SEA, RED-SEA, eProcessor, where ACO-S and XC partners already collaborate with private enterprises. *C)*: *The ecosystem of codes selected in the spoke represents the state of the art of computing in the AAA community*, covering a large fraction of the HPC simulations and data analysis codes and methodologies adopted in Italy. *D)*: *The enhancement of data analysis workflow based on innovative algorithms*. Advanced approaches to post-processing and data analysis will be adopted: *(i)* the integration of advanced in-situ and in-transit visualization tools, *(ii)* the integration of ML techniques to address tasks like automated regions of interest extraction or dimensionality. *E)*: *Improving energy efficiency at software level.* This will be implemented at co-design level providing feedback to Spoke 0 and 1 and working on top of the optimized hardware and therefore even further improving the FLOPS per Watt ratio.

Partnership – The Spoke partnership is constructed to provide the minimal, essential and complementary group of partners, who have the knowledge, the expertise and the experience to address the targeted objectives. It is balanced so as to involve research institutions, universities and industries. It brings the key partners that *(i)* have direct connections with relevant international projects as EPI, POP, EOSC, ESCAPE, EuroHPC, IVOA; *(ii)* have developed expertise in the co-design at exascale level; *(iii)* have proven capabilities in systems, software, hardware, data management and storage, energy efficiency and security; *(iv)* have proven capabilities in visualization, data analysis and cloud to develop the ACO-S ecosystem; *(v)* have a consolidated experience in higher education and training; *(vi)* contribute or lead large international observational projects involving the latest and future generation of observatories.

Gender dimension – The project will contribute to the European objectives of promoting participation and equality of women in sciences, as highlighted by the Helsinki Group on Gender and Research and documented in the EU document "She figures 2015". All partners have policies that dictate equal opportunities for applicants irrespective of gender, race or religious beliefs. Moreover, the responsible persons will fully follow these policies when incorporating young researchers into the project. Project planning and activities are compliant with Council Resolution 1999/C 201/01 which states that "the gender mainstreaming of research policy is not limited to promotion of women as research workers but should also ensure that research meets the needs of all citizens and contributes to the understanding of gender-relevant issues".

ACO-S integration with ICSC (XC) activities and resources: Crossbreeding Activities

ACO-S activities are fully integrated into the overall XC and a strong interaction between the different Spokes will be implemented at the HUB level. This will maximize the sharing of resources, knowhow and RI products (when possible), while minimizing the duplication of efforts. Several actions will be implemented by the ACO-S to improve cross-fertilization with specific Spokes; in particular the following activities will be carried out during the development phases of the ACO-S WPs.

Spoke 0 and XC. All the main activities of the ACO-S will be accomplished by exploiting the enabling infrastructure provided by the Spoke 0, deploying computing power, storage and archive resources for the next two decades. Such activities will focus on code enabling toward exascale and post-exascale architectures, for both numerical simulations and Big Data Analysis, addressing, in particular, the huge data volumes coming from the next generations of large-scale experiments (SKAO, CTA, EUCLID, LOFAR, MEERKAT etc).

Spoke 1, Future HPC. The Spoke addresses the exploration and adoption of innovative software and hardware solutions, with specific focus on co-design activities and energy efficiency. Hence, it represents the ideal framework to extend the work of the ACO-S, exploiting the experience and competence of the Future HPC experts toward the exploration of novel technologies capable of overcoming the limits of current code implementations. On the other hand, ACO-S experts will provide specific sets of applications and algorithms, essential testbeds for the innovative solutions experimented in Spoke 1, promoting an effective cross contamination among complementary scientific areas.

Spoke 2, Fundamental Physics and Space Economy. New algorithms for data reconstruction, analysis and simulations of astroparticle experiments, also based on AI techniques. Portability of applications on GPU/CPU and on heterogeneous architectures. Development of data interpretation tools, and statistical analysis, also based on AI, and high-performance techniques for accessing high-energy and astroparticle physics data. Development of algorithms and methodologies for integrating heterogeneous data (e.g., from different instruments / observers, and multi-messenger). INAF and INFN are participating actively in Spoke 2 RI WPs: this will guarantee cross-fertilization and maximize the sharing of RI products and know-how.

Spoke 10, Quantum Computing. The exploitation of high-end, ground-breaking computing technologies is essential for observational and theoretical research in AAA. The ACO-S will experiment the application of Quantum Computing in Astrophysics by probing the feasibility and the suitability of representative problems by experimenting with quantum technologies on selected algorithms, implemented to effectively exploit such solutions.

Long Term Sustainability

The key mechanism that will secure the long-term sustainability of the actions promoted by the ACO-S is represented by the essential role it will play in the major observational projects and observatories the Italian astrophysical community will contribute to in the next decade and beyond. Projects like ESFRI SKAO, ERIC LOFAR, CTA, EUCLID, MEERKAT+, are already funded by Italy and considered as cornerstone applications for the development of science and knowledge in our country. The success of these projects and instruments depends dramatically on the availability of outstanding computational resources, as those expected within the XC-NC, and suitable software tools and, even more critical, highly qualified researchers, capable of optimally exploiting such hardware and software infrastructure, as those specifically addressed by the ACO-S initiative.

This has been already recognised at the National level and the related activities have already been included in the long-term plans of the main research institutes involved in the ACO-S. As an example, INAF will significantly contribute to the novel Italian SKA-Regional Centre with an investment of about 2 million Euros per year for the next 10 years. The SKA-Regional Centre will be integrated and will progressively grow within the HPC infrastructure deployed by XC. An additional example is provided by the CTA Data Analysis Centre with a large financial and scientific involvement that will run far beyond the end of the ACO-S project and will last for at least 10-15 years. Similarly, the ASTRI-MiniArray Data Centre, hosted by INAF, will be set in place in 2022 and will gather the data, the analysis chain and the ICT infrastructure that will benefit from the new analysis algorithms and methodologies developed in the ACO-S and in the XC.

In addition, ACO-S achievements will become sustainable by consolidating effective synergies with stakeholders' lines of business (be it research or business per se), involving private companies, industries, and PMIs, using open call and demonstrators & use-cases applications, on the whole support of their adopters.

Finally, from a technical point of view, the long-term sustainability of the data products, of the software components and of the services developed by the Spoke will be facilitated by the broad adoption of the FAIR principles and of the interoperability standards promoted by the International Virtual Observatory Alliance.

Spoke 4 - Earth & Climate

The software and data infrastructure delivered by the Earth & Climate spoke will enable the development of the next generation Earth system models workflows where numerical simulations and data post-processing, analysis, learning and AI tasks will be part of the same end-to-end scientific workflow. In this respect the simulation-centric paradigm and the data-centric one will be seamlessly integrated into the same scientific discovery process, thus addressing challenges at the intersection of HPC, big data and data science. The Spoke will follow an agile methodology through early and continuous delivery of valuable software (releases). Additionally, besides the initial design and setup, two tasks will proceed in parallel on Earth Simulation Models (ESM) optimization and parallelization as well as on the data science and learning platform. A fourth one on the overarching end-to-end ESM workflow (including a multi-level provenance management support) will ensure that the different components will be properly managed and orchestrated on the underlying hardware infrastructure. A co-design approach will be considered in the overall development cycle thus integrating a wide spectrum of competences available within the Earth & Climate spoke (ICT, domain-specific, etc.) as well as from other relevant spokes of the Centro Nazionale ICSC (i.e., Spoke 0). Additionally, the proper level of interaction with other relevant (to the Earth & Climate) spokes will be addressed through dedicated meetings. So, besides a very focused, domain-oriented, and *vertical* development of the spoke activities foreseen in the 5 WPs, a regular cross-spoke communication will ensure, at the same time, the proper level of *horizontal* interaction with other spokes of the CN to share/re-use relevant outcomes, exploit common services, avoid duplication, align on the architectural vision, fully exploit the capabilities offered by the underlying infrastructures.

While the fundamental physics underpinning the main dynamical features of the climate system is reasonably well established, the processes at sub grid-scale, i.e., occurring at spatial scales not explicitly resolved by numerical models, and/or involving biophysical and biogeochemical interactions rely on a set of approximate empirical or theoretical equations that depend on parameters, whose values are often poorly constrained by observations. In this respect, inverse modeling and error-minimization techniques will be applied to the global-scale observations being provided by the latest satellite campaigns in order to obtain a substantial development/revision of these parameterizations. For instance, new observational records of land variables (such as land cover, green vegetation fractional cover and state, albedo, and soil moisture content) will be used to update equations of the parameterization as well as better constrain related parameters. Furthermore, the parametrization development will be also pursued by applying machine learning approaches, which try to blend AI with physical knowledge to achieve solutions that are physically more consistent.

The implementation of a digital platform to facilitate the combination of a range of regional models with different global models' boundary conditions, will provide a variety of regional dynamical downscaled ensembles through a smart and transparent process. The expected benefit of such a strategy is that it will allow analysis of a representative ensemble produced with multiple models, with different characteristics, providing a more reliable assessment of the models' uncertainty.

Finally, the methodological approach and the range of investigation tools adopted will be further extended through a series of open calls that will allow the engagement of other partners and competences, both from the research and academic world and from the relevant industrial sectors, which are not already involved as affiliated to the ICSC National Centre.

Spoke 5 - Environment & Natural Disasters

Making Italian science an international leader and advancing the understanding of our environment requires that computational infrastructure, research data, digital services, and advanced facilities be retrievable, accessible, interoperable, and reusable (FAIR) for researchers across scientific disciplines and national boundaries. In Spoke-5, understanding the nature of complex dynamic phenomena requires the integration of diverse "multi-messenger" data (physical, chemical, economic-social, etc.) obtained from multiple sensors and sources, as well as relevant datasets generated by national and international observatories and/or stored in third-parties distributed repositories. An additional barrier today to more and better knowledge about the environment and natural disasters is the availability of increasing the use of scientific computing capabilities so that researchers, as well as citizens and policymakers, can participate in accessing and reusing scientific products.
For the modelling and analysis of the different phenomena and processes, HPC will be exploited both for intensive simulations, which allow high-resolution modelling of single complex processes (capability-type), and for extensive ones, in order to model massive multi-processes (capacity-type).
Availability Spoke-5 data in a FAIR manner is critical to ensure reproducibility of research results, to nurture emerging scientific practices, and to enable researchers to fully achieve project goals.

The proposed methodology is developed according to the WPs already as described in detail the following. They can be summarized in three main activities:
1) Hazards definition & Modelling of phenomena;
2) Assessment of Damage, Losses and Resilience;
3) Risk Mitigation Strategies and Solutions.

In the first phase of the CN activity, the eight research institutions participating in Spoke-5, in collaboration with the four participating companies, will proceed to an assessment of available data and of the methodologies already developed for the analysis of such data, for the different WPs. The computing and storage requirements will be evaluated, in order to size the service demands required to the CN, as well as the TRL of the services/products already developed. After this first phase, activities will continue as planned in each WP.

Spoke 6 - Multiscale Modelling & Engineering Applications

This Spoke will exploit the Centre infrastructure according with the following structure of work packages

- *WP1: Complex multiscale and multiphysics modelling*
    - Participants to this WP have a very deep knowledge about several fundamental problems in multiscale modelling of complex systems. For example, in mesoscale modeling with quantum accuracy, multiscale numerical methods (from the atom to the device) to model charge/heat/mass transport properties and phase transitions, multiscale numerical methods for the investigation and the design of multifunctional materials, multiscale numerical simulations of strongly disordered systems, from neuronal networks to multi-agent interacting systems. Some of these problems will be chosen (D.1.1) as a challenge to stress and test our current capabilities of scaling up simulations. We plan to benefit from the broad competences of our team members to identify key problems and challenges.

- Multiscale modeling is very demanding from a computational viewpoint, so it is essential to have efficient, scalable and reliable algorithms as fundamental building blocks of the simulations and combine them through well-defined interfaces. WP1 aims at providing solutions that exploit at their best the possibilities offered by modern computing platforms (multi node systems equipped with accelerators and connected by high performance networks) that make it possible to overlap computation and communication. However, many existing codes and numerical libraries barely exploit this possibility, so a rethinking of some basic numerical kernels is required. Another technique that will be employed to improve the performance is the mixing of different levels of numerical precision. The WP will pay special attention to the potential issues of Uncertainty Quantification (UQ) using novel methods for surrogate modeling and sensitivity analysis. In a multiscale/multiphysics context, surrogate models can be used to replace expensive single-scale model components, thereby speeding up computations and enabling UQ tasks that would require too many model evaluations. However, the approximation errors made by surrogate models can amplify over time so fidelity to the original problem must be carefully controlled. Multiscale modeling often requires a mix of numerical methods, in particular when surrogate models are employed. WP1 will work with flexible and meaningful interfaces among the different components so that the risk of losing precision and efficiency in crossing the different scales is reduced as much as possible. This is a key requirement to simplify Verification and Validation (VV) steps of multiscale simulations. WP1 will provide i) a detailed analysis of the issues that currently hampers scalability and reliability of multiphysics simulations; ii) a selected set of fundamental building blocks for addressing some of those issues (e.g., scalable solvers for large scale sparse linear systems); iii) a sample multiscale-multiphysics application for the study of a complex system that will be defined in the initial part of the project and that will be based on the computational kernels of step ii).

- *WP2: Simulation of large systems via multiscale and AI methods*

  Real-life applications typically involve the closely coupled interaction of different physical phenomena occurring at multiple spatio-temporal scales. Examples include, but are not limited to, environmental problems (e.g., establishing the safety of soil exploitation activities), advanced manufacturing (e.g., improving the design of products and/or processes), aerospace, naval and automotive applications (e.g., efficient design and green propulsion for aerospace, ships and automotive), advanced micro/nanostructured material modeling (e.g., evaluating how microstructure affects the desired macroscopic properties), material science (e.g. synthesis of advanced molecular materials), and process engineering (e.g. design of new energy-efficient and clean processes). The work package aims at taming such complexity by developing innovative numerical schemes in an HPC environment as well as integrating AI and physics-based modeling to improve the simulation of large systems. Participants to this WP are recognized researchers with experience in numerical modeling and simulation of multiscale and multiphysics complex systems. The research tasks are organized as follows:

  - <u>T2.1 Development of multiscale multiphysics theory, numerical methods, and modeling strategies.</u>

  Multiphysics phenomena can be modeled via coupled systems of heterogeneous Partial Differential Equations (PDEs) set in complex (possibly moving) domains and involving the mutual interaction of mechanical, thermodynamical, and chemical kinetics effects. Such PDEs systems are typically characterized by problem coefficients (and/or solutions) that vary across multiple space and/or time length scales (from the macroscale up to the atomistic scale). From a multiscale perspective, such models can be coupled, when needed, with molecular models used to describe the thermochemical and kinetic aspects responsible for the system evolution. From the numerical viewpoint, such numerical methods boil to assembling and solving large (linear or nonlinear) algebraic systems, for which HPC-dedicated algorithms are required, aimed at the reduction of their overall costs (not only in terms of computational costs, but also in terms of energy consumption).

This task aims at developing innovative numerical schemes having robustness with respect to the many parameters involved, high-order accuracy, and algorithmic efficiency in an HPC environment. Algorithms will be made more efficient by developing adaptive approaches (in time and/or space), and algebraically efficient solution techniques will be by ad-hoc massively parallel preconditioning strategies.

- T2.2 Tools for orchestration and co-simulation

The digital twin modeling of complex systems and of systems of systems usually requires the use of multiple tools, operating at different scales, having different interfaces and data structures. This is a barrier to a widespread adoption of community tools and is a key issue when transferring low level TRL codes developed by academia or R&D centres to higher TRL levels for industrial exploitation. To overcome these issues this task will also investigate orchestration tools (i.e., tools of tools) that allow the easy integration and co-simulation of different community-specific codes ensuring data consistency and convergence of the multi-scale, multi-physics and multi-domain models and simulations. Examples of this approach are the FMI (Functional Mock-up Interface) and the GAIA-X European initiatives for a federated and standardized management of data, models, and simulations.

- T2.3 AI & Big Data analytics for complex/large systems.

Integrating data and physics-based modeling is driving a paradigm shift across the engineering, natural, and physical sciences, known as the fourth paradigm of scientific discovery. Combining high-fidelity models, data with machine learning (ML) algorithms allows for the efficient extraction of meaningful spatio-temporal patterns. The benefits of hybridizing and/or enhancing PDEs solvers with modern ML approaches are increasingly apparent. The objective of this task is the integration of a variety of artificial intelligence algorithms in conventional numerical methods to face large-scale simulations. We will investigate advanced algorithms in this framework to enhance PDEs solvers' accuracy and accelerate their performance. The typical scenario is an offline-online paradigm: a computing intensive offline training phase is executed first and then multiple queries to the solver can be quickly run (online phase). This paradigm will be applied in several scenarios:

- to enhance algebraic efficiency;
- for data assimilation/calibration of the PDE model, within e.g., upscaling approaches;
- to accelerate the description of dynamical processes;
- for the development of grey-box / explainable models to successfully combine physics-based models with data-driven approaches;
- for system control, optimization, and dimensionality reduction schemes.
- for reliability/sensitivity/uncertainty analysis, to provide decision-makers with robust information about the outputs of the solution and their dependence on the model inputs

The research carried out within this WP will lead to new simulation strategies and open-source computational platforms for Engineering and Applied Science applications in the fields of environmental, advanced manufacturing, energy and power systems, aerospace and automotive applications, advanced micro/nanostructured materials, and process engineering. The expected deliverables are the benchmarking of available tools at the end of first year and identification of routine tools for WP3 towards the end of 2nd year.

- *WP3: Digital twin modelling and tools for engineering disciplines*
  - The work package aims to define innovative methodologies and raise the TRL of emerging methods to support the definition and fine-tuning of digital twins for different engineering sectors. Digital twins are synthetic replicas of real systems and processes that combine virtual prototyping approaches with physical data to define faithful behavioral models, which accurately reproduce functional behaviour and respond quickly (or even in time -real) to queries. Digital twins can be used both to support design, with particular reference to optimization and as an expert control solution for systems and processes, with vast potential. Building accurate digital twins for major engineering problems requires multidisciplinary skills and knowledge that help define the complex behavior of simulation models.
  - In the work package, we will specifically explore the synthesis methodologies of multi-scale systems in which the model requires the description of the behaviour on two or more dimensional scales. In particular, regarding the aspects related to numerical fluid-dynamics, the research will be focused on developing optimized algorithms for the dissection of fluid flow inside and around obstacles with complex morphologies. The study will consider both solid and compliant structures, schematized through the Immersed Boundary Method and 2nd-order accurate wall boundary conditions, to dissect the underlying dynamics and propose drag-reduction strategies for aero-naval and civil construction applications. Machine learning algorithms will be employed to span the phase-space of both fluid and structural parameters, thus providing fast and accurate solvers for industrial applications. Possible applications span from complex structures for drag reduction to lubrication and dynamics of compliant vesicles in flows (red blood cells, drug delivery). Nucleation problems will also be investigated as a multi-scale phenomenon in which the very small, fundamental scales trigger non-linear phenomena with macroscopic repercussions, such as cavitation. In this case, high-performance simulations will provide evidence of the effects of the fundamental phenomena on large scale non-linear dynamics, and machine learning algorithms will be trained with the results of the super simulations to provide a fast and reliable tool for the simulations of complex, multiphase flows. Possible applications span from spray evolution and wall impingement (painting, printing, coatings) to optimized combustion processes, cavitation in air and water propellers, emulsion, and soft-glassy material dynamics in micro-/nano-channels. Other relevant studies will be focused on the development of multiphysics and multiscale computational modeling for thermal transfer turbulence simulations and for energy and power systems on an open source multiphysics and multiscale computational platform. Open-source platform will be adopted that allows the transfer of data among different multiphysics and multiscale codes in collaboration with the work developed in WP2. Another study will be focused on the development of high-fidelity digital twins of the wind field over the built environment and structures to assess resilience to climatic actions and human hazards. In particular, detailed Computational Fluid Dynamics simulations of the flow in the Atmospheric Boundary Layer based on Scale-Resolving turbulence models (e.g., Large Eddy Simulations), will be used to evaluate the performance of structures and infrastructures in extreme wind conditions, potentially leading to alert criteria to be coupled with real-time monitoring systems
  - Aspects of electronic and photonic phenomena will also be investigated, especially as regards multi-scale electronic device simulation, multi-scale coupling, atomistic/continuum modeling, optoelectronic device simulation and device modelling for energy harvesting.

Electromagnetics will also be at the centre of the investigations regarding the transmissions of contactless magnetic micro- and macro-drives that could represent the future of high-efficiency mechanical transmissions.

- As regards mechanics and structures, advanced fluid-structure interaction models will be developed based on multi-scale ROM (reduced-order models) approaches obtained from initial studies with a high computational burden but then condensed with the deduction of response surfaces. This approach will allow having computationally efficient but accurate digital twins, capable of responding to queries in real-time, managing to estimate the structured geometry modifications, the fields of stresses, deformations and displacements of very complex structures. Both mechanical, biological and environmental interactions between them will be faced, a significant issue for the industrial sector and the environment and medical sectors. This study will allow meeting efficient product and process optimization, real-time control interfaces and integration with the Internet of Things (IoT). Another study will be focused on the development of digital twins for the evaluation of the vulnerability to extreme wind conditions of infrastructures, strategic structures and energy production systems, with the aim to establish alert criteria based on the combined analysis of the actions exerted by the wind and the structural response, eventually taking into account their interaction. Moreover, further research will be focused on the development of detailed digital twins of cultural heritage structures and infrastructures will be developed accounting for the texture and quality of the masonry, with the aim to guide cultural heritage management and maintenance. This will pave the way to a model-based decision-making approach to decrease the vulnerability of cultural heritage against environmental actions such as earthquakes and human hazards.
- As regards optics and optoelectronic devices, atom-to-device simulative models of next-generation devices based on semiconducting nanostructures and novel 2D material will also be part of the research. The goal of this proposal is to give an important scientific and interdisciplinary contribution to application-oriented, efficient multiscale simulations of optoelectronic devices governed by nanoscale materials and interface properties. The core consists in joining high-precision quantum mechanical atomistic calculations with semi-classical bulk transport models like drift-diffusion, implemented in a multiscale framework.
- Last but not least, implementation methods of digital twin will be studied, able to predict and estimate the reliability of production plants, actively monitor the state of efficiency, predict failures, and plan the optimized maintenance interventions. In particular, the research will be focused on the development of digital twins of complex industrial production plants at the micro-, meso- and macro-levels to study the behavior and (re-)design functional modules (micro/local-level), manufacturing/assembly stations, CNC machines and workplaces (meso-level) and production areas as job-shops and highly reconfigurable cellular systems (macro-level) with specific focus on high-value production activities such as scheduling, maintenance, setup optimization, logistics, etc.
- The results of the work package will be harmonized by using a novel paradigm as the combination of advanced artificial intelligence routines merging information from digital twins and extended reality (virtual and augmented reality) to build interactive tools including neuromodulation, multisensory biofeedback (audio, video, haptics), and subjective perception of time-space. These paradigms will be designed and tested on digital twins, engineered and modelled specifically to reproduce all relevant sensory systems from a multi-scale simulation of biological subunits and natural interfaces.

- *WP4: Engineering applications: verticals and methods/tools integration*
  - WP4 bridges the gap between methods and applications, theory and practice, academia and practitioners. With a large focus on the engineering sector and considering relevant industrial problems, through the development of verticals, this WP aims at building, testing and optimizing tools to tackle industrial problems requiring multiscale modelling, simulation,

HPC and Big Data. WP4 aims at integrating the methods developed in WP1/2/3 following a final user-perspective, i.e. the developed tools will face relevant and practical engineering issues, and, then, exemplifies the tool application to a set of use cases, i.e. verticals, to debug, improve and demonstrate the potential benefit from a joint multi-target perspective (technical, economic, environmental, social, etc.). A wide spectrum of applications within the engineering sector are considered, according to the pervasive use of data, data management, and HPC, typical of engineering problems. For each application, i.e., vertical, a common methodological framework is adopted, stating the input data flow and the expected target. Then, for each case, the methods/tools selection and integration are done and exemplified. In case of alternatives, a quantitative based critical assessment of each alternative is pursued to provide, as a key WP output, a roadmap for users to select and best apply the most suitable solutions to handle the engineering problem under investigation. According to these goals, the following tasks and milestones will be developed:

- T4.1 Integration of methods and tools developed in WP1/2/3 ;
  - A successful integration requires a survey of the most relevant problems of the industrial partners, focused on the entire lifecycle of the product, to provide inputs to WP1/2/3 and to define a first list of industrial study cases, to be shared among the partners of the spoke.
  - An investigation of the existing tool chain associated with each process will also be performed with a twofold objective: i) develop interfaces to integrate the tools developed in WP 1/2/3 in existing industrial processes, ii) provide guidelines to industry to update obsolete processes to accommodate the newly developed tools.
- T4.2 Engineering verticals in industries ranging from ground and aerospace vehicles to green energy, manufacturing.
  - *Vertical on product design.* Multi-physics and multi-scale modeling will support the design of components, systems, devices and tools, thanks to the simulation of complex/large systems, taking advantage of newly developed algorithms and methods and of the CN infrastructure, aiming at the development of innovative products, the optimization of the design loop and at the reduction of the time-to-market.
  - *Vertical on advanced approach for the maintenance of complex/production systems based on the Prognostic Health Management (PHM) method*. HPC can leverage management and processing of data from multiple machinery/systems to train more accurate diagnostic and prognostic models on a huge amount of data, whose results can be used to predict the health status in real-time. In this vertical also spacecraft health management will be deepened based on AI/ML-based on-board algorithms/solutions aimed at leveraging large dataset generated on board for supporting space operations, FDIR (Failure Detection, Isolation and Recovery) and applications such as anomaly detection and prediction.
  - *Vertical on AI-based data processing for aerospace vehicle motions*. In particular applied to rover vision and docking-grasping operations. Security monitoring on EO systems, aiming at analyzing, detecting and predicting attacks and intrusions in constellations of satellites
  - *Vertical on advanced manufacturing for 'mass customization' production industries.* Twofold focus on HPC applications for (a) machine/production line reconfiguration strategies and (b) product families formation problem; facing the trade-off between high production performances and high product personalization to meet the market requests.
  - *Vertical on supply chain design for sustainability and industrial symbiosis.* Focus on techniques to optimize logistics reducing costs and environmental burdens enhancing renewable energy and secondary resources exchange among industries belonging to different sectors. Optimized data exchange, shared demand/offer platforms to

circumvent informative barriers, vehicle routing and supply plans, scheduling, non-traditional distribution solutions (crowd), etc.

- *Vertical on green energy systems for engineering.* Focus on methods and tools for multiscale modelling, simulation, HPC and Big Data adoption applied to the design, prototyping, scaling-up, testing, and industrial use of low-carbon energy components, modules, and plants toward the so-called 'zero-emissions' target. Integration of these energy systems into sustainable energy networks, smart grids, and circular energy chains with a focus on their forthcoming managerial implications and data flow optimization.

- *WP5: Proof-of-concept PoC demonstrators*

The goal of this WP is the following: starting from the analysis of the research challenges indicated by the industries in the CN governing board (e.g. Leonardo, ENI, Fincantieri, TAS-I, Terna), and after an assessment of the TRL of the results developed by the spoke partners in previous WPs (digital twin models, community codes and tools), the spoke will select a sub-group of the most promising solutions suitable for the creation of Proof-of-Concept demonstrators at high TRL.

In WP5 the spoke will also set up Living Labs arranged as a network of federated laboratories - physically placed at the spoke partner premises - to implement the PoCs and other AM methodologies (rapid prototyping, test before invest, training as defined in WP6, tech transfer as defined in WP6) .

The PoC demonstrators and the living labs will be funded through the demonstrator/innovation budget of the spoke and through a large part of the Open Calls, that will allow the involvement of mostly start-ups and SMEs.

The main targets of the PoC demonstrators and of the living labs are list hereafter:

- to define the technologies that in WP6 will be part of the *last mile* technology transfer to industry and to the market.
- to secure the sustainability of the spoke operation after the expending of the ministerial funding of the 3 year PNRR project thanks to "services" (training, technology transfer, test before invest, rapid prototyping) provided by the Spoke to third parties
- to find synergies with European initiatives on digital skills and techs such as the European Processor Initiative, European Digital Innovation Hub, EuropanChips Act, EuroHPC

- *WP6: Training, Technology transfer activities and Outreach*

This task will foresee specific activities for knowledge and technology transfer of the results of the spoke to maximize its social and economic impact on the italian societies, with a specific focus in reducing the gender gap and the territorial gap in the italian society.

<u>T6.1 Training and knowledge transfer</u>

As far as the transfer of knowledge is concerned the spoke will address two main lines:

i) training of young and talented people through the establishment of a novel national PhD program in *Multiscale Modeling and HPC Engineering Applications* with a close link between academia and industry;

ii) re-skilling/up-skilling of workers already employed in the public administration and in the private industry to sustain the digital transformation of Italy and the industry 4.0 initiative.

The novel national PhD program in "Multiscale Modeling and HPC Engineering Applications" aims at involving as founding members the Universities and research centres that are affiliated to the spoke together with the industries (e.g., TAS-I, Leonardo, ENI, Fincantieri, Terna) that are co-founders of the CN and expressed specific interest in the spoke. The target is to finalize the preparation of young PhDs/graduates in the field of HPC to the different productive sectors and, specifically for what is of interest to the industry and society at large. These young resources could be allocated on projects of industrial interest and can spend a significant period at the industries themselves. To foster the rise of close collaboration across the italian territory, this PhD program will foresee that each PhD student should have at least 2 tutors from different research/academic entities (located in different Italian regions) and at least 1 industrial advisor. Each PhD student will have access to the massive HPC infrastructure of the Hub, thus interacting with Hub institutions like Cineca.

The novel PhD program will be economically sustained with the specific CN funding devoted to PhD positions, and, at the end of the period covered by the NRRP, with resources coming from co-funding of affiliated companies/industries (e.g. industrial PhD positions), or affiliated institutions (PhDs working in public administrations, or in research centres like ENEA, INFN, CNR, INGV, INAF...), or from Italian government PhD future funding.

The "re-skilling/up-skilling initiative", instead, will target the sustainability of the digital transformation of both the Italian public administration and the Italian companies. Both SMEs and big companies will be involved. The idea is to define the first year of the spoke a professional plan, agreed with stakeholders like Confindustria, CNA, industries involved in the CN, institutions (MUR, MITE, MPA, Regional Governments), in which the spoke members define professionalization courses to be held on-line and/or in person. Different levels (basic or advanced) of courses, with different duration (ideally from 1 day to 1 or few months) will be defined and the specific excellence in each spoke member will be exploited. This training initiative will be economically sustained with the funding for innovation of the spoke and also with co-funds from the beneficiaries (public institutions, companies) exploiting the many economical instruments (e.g., Fondimpresa, European Social Fund FSE to name a few) that are specifically available for the training of people already employed or needing a re-conversion to new jobs.

T6.2 Technology transfer

For technology transfer of the most promising results achieved in the spoke, the spoke members within the first year, in close collaborations with the companies Leonardo, TAS-I, Fincantieri, ENI and the hub of the CN, will define a "tech transfer plan" specifying the rules, IP right management policy, co-funding mechanism to cover "the last mile" needed to transfer to industry and integrate in market-ready solutions at TRL 8-9 the tools, digital twin models and community codes developed at high TRL (TRL6-7) in the spoke, and particularly in the Proof of Concept demonstrators in WP.

For this technology transfer activity, the spoke will rely on:

- living labs, funded by the demonstrato/innovation budget of the spoke, as defined in WP5
- high technology start-up companies, generated as spin-off from the Universities and research centres of the spoke, funded through the open calls

T6.3 Outreach

For Outreach the spoke will pursue several channels including:

- Communication and dissemination channels coordinated with the hub
- Scientific outreach through scientific publications, organization of special issues in journal and special session in conferences on multi scale modeling and HPC engineering applications
- At least 3 multi scale modeling and HPC engineering workshops (1 per per year) involving all research and industrial spoke partners (1 hosted at Rome La Sapienza, 1 at Pisa, 1 in a site in Southern Italy).
- Disseminations events by the different partners coordinated by the spoke, e.g., *café scientifique*, interacting expositions and open labs initiatives. The goal is to make people aware of the many aspects of our life where HPC and multiscale modeling play key roles

Spoke 7 - Materials & Molecular Sciences

The methodology adopted to pursue the goals of this spoke will adapt to the three-fold scope of its programme, targeting *i)* code developers, *ii)* scientific end users, and *iii)* society at large, and addressed by different WPs in mutual synergy.

DEVELOPERS

Materials modelling and discovery stands on the simulation of properties and processes at the molecular scale, which is performed thanks to powerful computer codes that operate as *quantum engines* of the simulation and that are extremely CPU- and energy-greedy. Present and future HPC systems, particularly those racing for exascale, are expected to be more and more heterogeneous (GPU-accelerated) in order to reduce power consumption, while enhancing the performance, and also enabling advanced data-oriented operations. Within this scenario, WP1 will develop, maintain, and distribute CN flagship codes along the following lines: *i)* enhance the performance portability and intra-node optimisation of the codes, including support of multiple GPU hardware (as expected to become relevant in the EuroHPC context); *ii)* improve on the parallel performance and scalability of the codes (especially addressing fast communications amongst accelerators); *iii)* take advantage of established best practices in scientific software engineering, and develop new ones, in order to avoid code duplication and disruption while enabling long-term maintainability and engagement with the scientific community; this will include the implementation of SW strategies such as modularization, HW/SW separation of concerns, and exploitation of data locality, to name but a few; *iv)* implementation of new scientific features enabled by or motivated by the exploitation of modern HPC capabilities. A special attention will be paid to code integration and reusability. Flagship codes will be refactored into a number of independent highly performing modules and libraries to be built and integrated with each other and as components of third-party software and of advanced workflows. WP1 will feature a task entirely dedicated to the development of auxiliary libraries to perform specific high-level tasks, not bundled into any flagship codes, but designed to operate with all of them and with third-party software.

END USERS

The operation of the quantum engines described above and the exploitation of the large amount of data generated by them requires the development of specific user-oriented methodologies. First and foremost, the implementation of the equations describing the fundamental laws of nature in the quantum engines requires different levels of approximations that need to be carefully benchmarked against accurate reference data. Quantum engines typically rely on density-functional theory and many-body perturbation theories: the first stands on the use of so-called *energy functionals* whose existence and uniqueness are mathematically guaranteed, but whose exact form is unknown; the latter requires careful manipulations of diagrammatic expansions whose effectiveness can only be assessed by experience and extended benchmarks. A thorough validation of the energy functionals and diagrammatic expansions currently in use will be performed in WP3 against accurate quantum-chemistry and quantum-Monte-Carlo data. Second, in order to deploy advanced modelling techniques to search the highly dimensional structural and chemical spaces of materials for specific custom-designed properties, in WP2 we will develop novel methods in materials informatics, multi-scale simulations, and artificial intelligence, which will permit to explore portions of this space that are currently not accessible. Our approach will empower end users to focus on the application-, rather than software- or data-, oriented, aspect of their research. By employing advanced workflow managers, such as AiiDA, as operative systems for computational studies in materials and molecular sciences, end users will be enabled to use, build, and share FAIR-compliant datasets, automation strategies and thoroughly tested computational protocols in an user-friendly software environment. This will allow end users to focus on the application of the most suited computational methods to solve the scientific problems of their interest, while the materials informatics infrastructure will take care of handling the simulations on HPC facilities and exporting the data to easy-to-search databases that will be available and browsable online. We will embed automation strategies and rigorous computational protocols into automated workflows, based on flagship codes, and distribute them as ready-to-use open-source Python code. Open Data requirements will be fulfilled by uploading and browsing our data on the materials repositories of our partners, which will ensure long-term storage and accessibility. In addition, WP1 will also design and implement specific tools to assist users in the task of building optimised executables for specific platforms, ranging from laptops to exascale systems. For advanced users, we will provide utilities for the automated resolution of dependencies among libraries, drivers, and architecture-specific installation flags. For less skilled users, we will distribute black-box widgets based on containers (built *e.g.,* with Docker or Singularity) and/or virtual machines fully equipped with preinstalled flagship codes and other scientific tools. Containers will also be tailored to run on cloud based HPC infrastructures. Finally, advanced user interfaces will be designed and implemented, also aiming to provide quantum simulation tools "as a service" using web and cloud technologies.

SOCIETY AT LARGE

Materials and molecular simulations require a substantial training of scientific personnel. Also, first- principles methods address nano-scale phenomena and thus provide little insight by themselves onto the macroscopic scale that is often more relevant for industrial research. Finally, although computational studies of materials, molecules, and drugs are way faster and cheaper than experimental research, still they often require extensive HPC resources that might go beyond the reach of SMEs and the corporate policies of even big companies. The concerted actions of the five WPs of this spoke aim at lowering the barrier that industrial research has to overcome in order to benefit from advanced simulation methodologies, by providing: *i)* a toolbox of sophisticated simulation tools, finely tuned to run optimally on a broad range of compute platforms, from laptops to exascale systems; *ii)* a set of ready-to-use workflows targeted to "vertical'' applications, allowing R&D personnel to address specific problems of direct industrial relevance without prior specialistic training in computer simulations; *iii)* large datasets of calculated properties to be used to train data-driven methods (materials intelligence) that can be deployed with limited infrastructure costs; *iv)* a set of innovative tools to facilitate the approach to molecular and materials simulations, ranging from advanced user interfaces, to cloud-based web platforms able to provide "materials simulation as a service", thus freeing the end user from the need of installing, maintaining, and managing a complex and expensive simulation platform. Finally, the spoke will qualify as a permanent reference desk to provide guidance and consultancy to enterprises of all sizes, willing to take advantage of the R&D opportunities offered by digital technologies for materials research and discovery.

Spoke 8 – In-Silico Medicine & Omics Data

In silico medicine will employ the following methodologies:

1. A techno-legal infrastructure which allows to elaborate with HPC resources large volumes of clinical data that cannot be irreversibly anonymized but only pseudo-anonymized.
2. A collection of Open Source and commercial in silico medicine solvers optimized for the next-generation HPC systems, for all three key patterns used in biomedical computing:
    a. Monolithic - Single computational job spread over a large fraction of a single computer, e.g., highly parallel CFD and FEA codes;
    b. Coupled - Multiple, communicating subcomponents each assigned to a sub-section of a single computer, e.g., FSI codes, agent-based models, constrained optimization problems;
    c. Ensemble/Workflow - Multiple instances of one or more applications and analysis launched concurrently with performance properties, Monte Carlo, Particle tracking, etc.
3. Technical solutions to simplify computationally intensive VV&UQ activities necessary for the certification of in silico medicine solutions.

Omics data will employ the following methodologies:

1. Informatics pipeline (LIMS) for managing the entire workflow from bedside to laboratory and back.
2. Next-generation bioinformatics tools fully optimized for GPU/FPGA HPC infrastructures.
3. Algorithms and codes for clinical machine learning for efficient analyses of omics data.

4. Computer vision algorithms for dimensionality reduction and machine learning analyses of radiomics data.
5. Physics-based methods and codes (optimized for HPC) for drug discovery and development (including statistical algorithms (QSAR, QSPR, etc.)).
6. The above informatics tools will be included into the national cloud system for rapid and secure access by a broader scientific/medical community.

Spoke 9 - Digital Society & Smart Cities

The research work of the spoke is organized in three research lines, i.e. (M)odelling, (C)omputing, and (S)oftware, and in five application domains, i.e., health and lifestyle, mobility, socio-economic analysis, infrastructures and utilities, and environment. Dedicated work packages are foreseen for each research line (WP1-WP3) and for each application domain (WP4-WP8), in order to facilitate reaching critical mass, knowledge sharing and collaboration among partners on these topics.

The research approach followed in the spoke ensures a coherent, synergistic progress of the activities across the work packages. In particular, the work on the core research lines of the spoke (WP1-WP3) is not only integrated across the three lines; it will also be driven by the needs and challenges identified in the five application domains (WP4-WP8). On the other hand, the novel approaches implemented in the three research lines (in particular in (M)odelling and (C)omputing) will be evaluated with the real-world challenges identified by the application domains. The timeline of the activities and the spoke milestones are defined to facilitate this mutual interaction among core research lines and research in the application domains, implementing an interplay of specification, development and validation activities in the research lines and in the application domains.

Considering the expected impact of the activities of the spoke, we remark that the identified application domains not only provide a diversified range of contexts to drive and validate the research undertaken by the spoke; they also serve a broader set of expected outcomes, including: (a) advancing knowledge and stimulating the research foreseen in the spoke; (b) exploiting the computational resources made available by the "Centro Nazionale"; (c) building on top of available data to experiment in real/realistic socio-technical settings; (d) capturing the interest and participation of large communities and key stakeholders; (e) open up exploitability opportunities through value-added services offering long-term sustainability opportunities.

Spoke 10 - Quantum Computing

The research activities performed in the spoke require to coordinate a wide range of competences, spanning from quantum physics to high level application domains. The research methodology adopted within the spoke combines results from the different domains through the development of a number of layered hardware and software components, with an increasing level of abstraction, with the dual objective of hiding the complexity of each underlying layer to the upper ones and allowing the different teams within the project to design components able to interact with each other.

The challenge is to design and develop algorithms in the quantum environment, which is itself under development from different points of view: the programming language, the middleware platforms connecting quantum solutions to the hardware and the usable toolkits for developers. It is necessary to find a compromise in order to benefit from quantum computing when the classical computation fails while managing the trade-off generated by the effort to develop the system.

This research is by nature experimental, both on the hardware side, where researchers must explore the design of frontier quantum computing technologies, and on the algorithmic side, where researchers need to test and comparatively benchmark high-performance quantum and classical applications. These lines of research will be based - on the theoretical side - on a class of development and simulation methods that will allow to design, test and verify new quantum algorithms or adaptation to the specific scenario and use cases of existing ones (Grover algorithm for database searches, phase estimation algorithm, adiabatic or optimal strategies, etc.). The technical tools used to support such activities ranges from well-maintained commercial software such as, e.g., Qiskit and Cirq and cutting-edge research emulation and benchmarking software – also to be developed within the centre activities – based on tensor network methods, quantum optimal control, classical and quantum or quantum-inspired machine learning. On the experimental side, the activities will be based on existing programs in active laboratories and long-standing experience of the scientists identified by the CN to develop a coordinated Italian quantum computing industry on the different hardware platforms.

National research in this area is limited by difficulties in accessing appropriate quantum computing resources and for this reason the spoke research activities will benefit from access to both on-premise and in-cloud infrastructures as planned at the CN level. This will lead to the design and testing of hybrid systems configurations, that integrate quantum and classical high-performance computing in order to solve complex problems in a sustainable time.

A.4 The data management

The activities of the CN build on the capability to store, access and process large data repositories, either created internally during the project, or acquired from external sources. Data handling, and in particular its aspects of sensitivity, rights and patterns of access and lifetime, is an essential part of the project, and needs to be covered by a specific Data Management Plan (DMP).
In general terms, the CN and its spokes will follow the "Guidelines on Data Management in H2020", "Horizon Europe Data Management Plan Template", in addition to the EU General Data Protection Regulation (GDPR): these documents offer a good operational context and are already known and adopted by most of the CN partners. Recognizing the importance of ethical and privacy concerns, the adopted approach will address ethical and privacy requirements, including explainable and trustworthy AI. Additionally, the CN will follow the implementation guidelines of the National Plan for Open Science from PNR 2021-2027, which is currently still in draft status.

The DMP is not to be considered a static document; it will be prepared early in the project (M12), and will be updated constantly during the project lifetime, to reflect changes in the relevant EU or National Regulatory frameworks, and to include modifications from a better understanding as acquired while performing the research activities.

The DMP will include information on (at least):

- the handling of research data during & after the end of the project
- what data will be collected, processed and/or generated
- which methodology & standards will be applied
- whether data will be shared/made open access and
- how data will be curated & preserved (including after the end of the project).

The DMP will base the data management according to the FAIR principle to maximize the effectiveness and reproducibility of the research undertaken; in this respect, data will be Findable (defining the methodology and standards for data collection and storage), Accessible (defining how data are deposited, preserved and accessible during and after the project lifetime), Interoperable (defining the sharing policy and associated ethical issues) and Reusable (defining conditions and modalities of data re-use). The Data Management Plan will be primarily oriented towards describing the standards, conventions, documentation and tools being developed and deployed, and how the spoke is delivering the mechanism by which the data producers can deliver on the FAIR principles for their data.

It must be specified that FAIR principles do not imply that data will be openly available in all cases, as impossible for (e.g.) the sensitive data in Spoke 8. The paradigm to be followed is "As Open as Possible, as Closed as Necessary", which implies a clear definition of the access patterns for each data object. The implementation of this paradigm requires an underlying technical infrastructure able to guarantee data safety and correct access granularities. These will be guaranteed by specific infrastructures as deployed and managed by the CN (e.g., ISO 27001 certified repositories and centres), and a state-of-the-art Authentication & Authorization Infrastructure (AAI), based for example on the AARC Blueprint architecture.

In accordance with article 37[13] of the GDPR, a Data Protection Officer will be appointed by the project; he/she will work in strict contact with the DPO of the single institutions' members of the CN, with data controllers and processors designated per each data source, and will be responsible for:

- coordinating with the DPO of the institution's members of the CN;
- informing and providing advice to the data controller or the data processor as well as to the employees who carry out the processing on the obligations deriving from the GDPR, as well as from other national or European Union laws related with data protection;
- monitoring the compliance of the data controller or the data processor to the GDPR requirements, including the assignment of responsibilities, awareness-raising, and training of personnel involved in the processing and related control activities;
- providing an opinion on the impact assessment on data protection and monitor its performance according to art. 35 of the GDPR (if specifically requested);
- cooperating with the privacy authority (Garante per la Protezione dei Dati Personali) for any issue related with data processing and data protection, including early consultations (art. 36);
- supporting the data controller in keeping the processing register, follow the data controller's instructions.

A single DPO will be designated for the whole CN; he/she will be assisted by experts from each spoke to cover the specificities in each domain.

Aspects specific to each scientific domain are detailed in the following relevant spoke sections.

Spoke 1 - Future HPC & Big Data

The key partners of the proposed national centre, including the FutureHPC leader and co-leader institutions, are directly involved in the "Open-IT" PNRR proposal (under research infrastructure call), which directly addresses all aspects of the open science data lifecycle. Thanks to shared partners, both initiatives will synergistically benefit from novel technologies and tools for data management provided by the FutureHPC spoke of the HPC centre and the data management practices, data steward training, and ethical processes developed within Open-IT. On the one hand, Open-IT adopts a FAIR-by-design approach to overcome barriers and move toward open science as the default approach across the research lifecycle. On the other hand, FutureHPC Spoke embraces open standards as a method and conceptualizes the open-access research laboratory through its living labs. FutureHPC will extensively benefit from data steward skills formed within the partner universities by the Open-IT action to tailor a data management plan for each specific research project developed in the flagships and the living labs.

The Open-IT methodology and expected outcomes align with the objectives of Missions 1 (in part. M1.C1 Digitization, innovation, and competitivity in the Productive System and M1.C2 Digitization, innovation and security in the PA) and 4 (Education and Research) of the National Recovery and Resilience Plan and with all EC's Key Impact Pathways (Regulation EU2021/695 Establishing Horizon Europe, Annex V). In terms of societal impact, Open-IT will

[13] https://gdpr-info.eu/art-37-gdpr/

Address EU policy priorities & global challenges through R&I: Open-IT addresses the widespread adoption of OS by Italian researchers in line with the EC strategy for research and innovation funding programs. By achieving its objectives, Open-IT will empower Italian researchers to improve the quality and reusability of their results and participate in EU projects to address global challenges.

Deliver benefits & impact via R&I missions: Open-IT objectives and activities will concur to achieve the "Europe fit for the digital age" mission; additionally, OS can act as a transversal enabler for the advancement of other R&I missions in Europe and at the national level.

Strengthen the uptake of R&I in society: Open-IT will provide researchers with tools and workflows FAIR-by-design for transparency, reproducibility, dissemination (WP2, WP4) and can be an effective tool to transfer scientific results to society (policymakers, public protection agencies, the public health system, the education system, and citizens in general). OS practices can also impact research evaluation exercises, indirectly fostering greater societal impacts for its outcomes.

Overall, the two actions will lay a solid foundation for implementing the National Plan for Open Science. By making scientific research more transparent and collaborative through the possibility of combining large amounts of research outputs and accessing and reusing publications, software, and educational material. They will empower national research communities to transform how they carry out their research work and exploit research outputs, leading to better quality and more productive research, encouraging innovation and growth in research and development workers, in line with the UNESCO Recommendation on Open Science.

Spoke 2 - Fundamental Research & Space Economy

The management of data and products of interest in the activities of Spoke 2 will follow a clearly defined Data Management Plan, to be finalized by the end of the first year of the project. It will follow the FAIR principles, declined for the stakeholders in the project; we can anticipate the general directions what will be followed in the rest of this paragraph.

Concerning research publications, INAF and INFN, the two public research institutions in the spoke and the links between universities and the specific research activities, support a policy of open access to the results and data obtained from public funds. INFN is part of the cOAlitionS[14] and has signed the PlanS, initiative for Open Access publishing that was launched in September 2018. PlanS requires that, from 2021, scientific publications that result from research funded by public grants must be published in compliant Open Access journals or platforms. INAF approved an open access regulation[15] in 2019. As to the management of experimental data, the main and most valuable output of experiments, the landscape varies not only between domains, but also within the same domain. In general, data from publicly funded research is expected to be open after an embargo period, in which only the scientists directly involved are able to access and analyze the data. Portals allowing for direct data access, for research and educational purposes, are becoming widespread (see CERN Open Data Portal[16] and the efforts to connect the astroparticle research infrastructures to the European Open Science Cloud (EOSC) project through the Virtual Observatory (VO) framework). Code and algorithms are a third important product from research. While there is no overarching decision from the scientific domain, the largest institutions and experiments (see for example CERN experiments CMS[17] and ATLAS[18]) are committed to releasing their codes as open, with different flavors of licenses.

---

[14] https://www.coalition-s.org/
[15] https://openaccess-info.inaf.it/policy
[16] https://opendata.cern.ch/
[17] https://github.com/cms-sw/cmssw
[18] https://gitlab.cern.ch/atlas

It has to be noted that neither the CN nor the Spoke 2 will be the *owners* of the processed / created data: their ownership will remain to the institution or the experiments, and the capability to access it will be, also within Spoke 2 activities, limited to the scientists and collaborators which are allowed by the data original owners.

Given the complex mixture of ownership and access permissions to the data considered in Spoke 2, including private data from the productive sector while performing shared tests it is essential to have a fine-grained and omni-comprehensive infrastructure for the Authorization and Authentication of accesses.

The centre, and in particular the Spoke 0, in charge of the infrastructure, will deploy a state-of-the-art Authorization and Authentication Infrastructure (AAI), based on the AARC Blueprint Architecture, ensuring full granularity of access to data, for individuals and groups.

Spoke 3 - Astrophysics & Cosmos Observations

The data and research output handled by the ACO-S will be managed in line with the FAIR principles. Indeed, the management of the research outputs during the Spoke activity is part of WP0, partly dedicated to the Project and Data Management Plan (PMP/DMP), FAIR data services and to the assessment of the compliance with the FAIR principles.

This PMP/DMP will cover: a) privacy/ethical processes for the management of data/research outputs collected, processed or generated, and b) modalities for the publishing of documents. This will ensure that all documents produced in the course of the activities are properly shared and published, as this will affect future decision-making. It will define the methodology and standards to adopt to handle any data/research outputs; how this data/research outputs will be shared and/or made open; and how the data/research outputs will be curated and preserved during and after the project, in particular within the definition and scope of the EOSC vision.

ACO-S will promote the FAIRness of the research outputs and services across research communities involved in the project. WP0 and WP4 will ensure that the FAIR principles are fully endorsed and implemented in the Spoke activity. In particular, WP4 will evaluate the services proposed in the technical WPs and propose guidelines to improve the FAIRness of the data and research output of the project.

All the knowledge developed during the ACO-S activities will be open to access when not covered by copyright or included in patents. Knowledge management will be carried out towards two main directions: the exchange of knowledge among the partners within the consortium and the exchange of knowledge with stakeholders including XC partners and Spokes of the CN, and the general public outside the consortium. The former case is rather straightforward, the latter may be more elaborate, since it needs a careful evaluation of each result and a customized user-friendly formatting in order to maximize the project impact.

ACO-S partners will follow a proactive strategy for internal management of both data and knowledge. A project restricted area will be set up by the coordinator to store internal common information, results, reports and deliverables.

Spoke 4 - Earth & Climate

The Earth &Climate spoke will contribute to the implementation of the EU Data Strategy to support the Green Deal, through the development of a data science and learning software infrastructure in WP1 which will follow a Data Space approach.
To the maximum extent possible an Open Data Policy will be applied to data and other research outputs as outcomes of the Earth and Climate Spoke activities. A Data Management Plan (DMP), will be developed in the early stages of the project (and consolidated at M18) to define how research data will be handled during the project's lifetime and after it is completed.
The DMP will base the data management according to the FAIR principle to maximise the effectiveness and reproducibility of the research undertaken; in this respect, data will be Findable (defining the methodology and standards for data collection and storage), Accessible (defining how data are deposit, preserved and accessible

during and after the project lifetime), Interoperable (defining the sharing policy and associated ethical issues) and Reusable (defining conditions and modalities of data re-use). The Data Management Plan will be primarily oriented towards describing the standards, conventions, documentation and tools being developed and deployed, and how the spoke is delivering the mechanism by which the data producers can deliver on the FAIR principles for their data.
The application of FAIR principles will ensure open data access and interoperability wherever possible, and it will allow us to rely on continuously updated standards and best practices for metadata and data harmonization. The DMP will be thus managed as a dynamic document with an update (M18) to take into account possible changes in the relevant EU or National Regulatory frameworks.
Finally, the data collected and produced within the spoke will be managed within a data space infrastructure designed in agreement with the ICSC National Centre and in compliance with standards and best practices defined in the context of European initiatives, such as the European Open Science Cloud (EOSC) and GAIA-X, as well as (from a domain-specific perspective) Destination Earth.

Spoke 5 - Environment & Natural Disasters

Spoke 5 will rely on an infrastructure for standardized data collection approaches, common ontologies, and sharing protocols to allow both Spoke 5 researchers and stakeholders to make use of powerful methods to curate and integrate diverse sources of data to promote a better understanding of complex and dynamic systems.

The envisaged architecture follows the 3 layer structures: (L1) data provider; (L2) federated data management & processing; (L3) Big Data Analytics & Visualization.

The cloud infrastructure design activities will cover the configuration of the modular architecture, the management of the infrastructure lifecycle and functionality, the control and orchestration of the exposed services, and the data derived from the machine learning and artificial intelligence models.

In particular, the platform will ensure real-time processing without data movement and interruption between individual AI modules, maintaining data consistency with fault tolerance and resilience.

It will have to provide for the ontological engineering of a virtual interface that unifies the data buses present in the individual WPs platforms with an integrated approach. This system will have to allow dynamic access to the data and its processing (AI processing) without generating an external repository and guaranteeing high fault tolerance, remaining GDPR compliant.

Spoke 6 - Multiscale Modelling & Engineering Applications

A structured Data Management Plan will be prepared and will include the best standards for the generated data and assess their suitability for sharing and reuse in accordance with official guidelines. The GAIA-X initiative in Europe will be taken as reference.

Data Summary

The objective of the project is to develop knowledge advancement in the computational aspects of multiscale modelling and new improved algorithmic solutions to several real case applications.

Data will be divided in two main parts:

1) data generated for purposes of investigating the classes of algorithms which can be solved via the implemented hardware,
2) data for the development and usage of the sampling machines.

Types of data will include: Software, reports. In addition, reports on financial analysis and end user analysis.

Formats will include:

- Raw data: binary and numerical data.
- Software and codes.
- Reports in PDF and Office documents.

Origin of data: All software and reports originate with the partners developing the system.

Expected size of data: ~ 100GB (preliminary estimation, mainly from raw data to be exchanged)

FAIR data

Making data findable, including provisions for metadata

- Data will have a unique identifier.
- Publications will have appropriate search keywords.
- Clear version numbers will be provided.

Metadata includes: Digital Object Identifiers (DOI) for published papers, preprint identifiers, cross-reference between repository preprint and published version, institutional repository identifiers.

Making data openly accessible

Three levels of accessibility for the data will be employed:

1) Data and software generated for will be for internal use only. Specifications and manual will be for internal use. Data developed will be stored locally and shared between the involved partners.
2) Gold open access will be preferred for publications. If not possible, green open access will be granted for all publications through institutional repositories and/or arXiv repository (https://arxiv.org). Links redirecting to publications will be posted in the dedicated section of the project website.
3) Open access to data and software supporting publications, or additional material generated within the project, will be eventually made available through specialized portals (e.g. GitHub) upon agreement between the involved partners,

Making data interoperable

Compliance with currently available open software applications (e.g. Python, or other open-source programming languages) will be preferred for software development.

Data shared will be formatted by using standard methodology and formats.

Reports and publications will be prepared by using standard terminology and methodologies.

Increase data re-use (through clarifying licenses)

- Re-usage of specifications of the developed system will depend on the project outcomes.
- Data produced for the development of the prototypes will be for internal use only.
- Re-usage licenses for data referring to publications will be evaluated within the course of the project.
- Re-usage of open source software made available via dedicated portals will be granted.

Allocation of resources

Costs for making data FAIR will be mainly due to the cost of gold open access publications.

Costs will be covered from the project budget.

Person responsible for data management: Each beneficiary will be responsible for their part of the data.

Resources for long term preservation (costs and potential value, who decides and how what data will be kept and for how long) are still being discussed.

Data security

Provisions for data security (including data recovery and secure storage and transfer of sensitive data) will be based on standard university procedures.

Safe storage of data in certified repositories for long term preservation and curation: Under consideration.

Spoke 7 - Materials & Molecular Sciences

The scientific value of materials intelligence crucially depends on the quality of the data produced or harvested, and on the ability to reuse pre-trained algorithms and curated datasets for different applications. All the methods, software, and data produced in this spoke will be disseminated following the principles of Open Science and Open Data, targeting the Findability, Accessibility, Interoperability, and Reuse (FAIR) of digital assets. Our data governance policy will stand on two levels. The first will be enforced for all the activities, for which we ensure that data are produced and stored complying to the FAIR principles. In collaboration with CINECA, we will establish formal partnerships with at least one of the two official repositories for materials data recommended by the EU Commission in Open Research Europe, Materials Cloud (MC) and NOMAD, with whom preliminary discussions are already taking place. The curated data will be uploaded to the MC or NOMAD platforms, and stored exclusively or mirrored on the CINECA data centre, ensuring data sovereignty. A second layer will be enforced for selected datasets, characterised by the readiness to automation, wide interest, and broad reusability: here the simulation pipeline will be orchestrated by the AiiDA open-source infrastructure, which gives access to the full provenance of all parents and children calculations and data nodes.

Spoke 8 – In-Silico Medicine & Omics Data

The EU General Data Protection Regulation (GDPR) came into force in May 2018. It aims to strengthen and update data protection and governance across the EU and respond to novel technologies' challenges in the digital age. Genomics and other omics data and clinical and health data are very sensitive. The Spoke 8's participants have remarkable experience in processing human-related data, and this expertise will be exploited here to fully implement the most appropriate cyber-security and anonymization measures for the Spoke's data.

All the data will be generated, managed, and stored in full compliance with the legal, organizational, and technical requirements resulting from the GDPR. We will employ wiping techniques for temporary storage (warm storage) of raw data during the data collection; then the data will be pre-processed and stored in a petabyte storage system. The final (cold) storage is endowed with encryption at rest procedure (AES256). The system will also be equipped with a disaster recovery storage located more than 50 km from the cold storage.

Next-Generation Sequencing (NGS) data may require special care, and this Spoke will only work with Whole-Genome Sequencing (WGS); therefore data protection is extremely important. The file fastq, generated by the NGS machine, will be compressed to fastq.gz, whereas the bam files should be stored unaltered. We plan to store fastq and bam for at least two years, whereas smaller VCF files may be stored for 20 years. VCF files will also be stored along with the diagnostic reports of medical doctor.

In the following, the list of measures employed by the Spoke to ensure compliance with the GDPR is reported:

- adopt a procedure for the pseudonymization of personal data;
- adopt a strategy for the management of data breaches;
- perform a Privacy Impact Assessment (PIA) and report on the adopted methodological approach;
- update and validate all relevant privacy-related documents (e.g., reports, informed consents, etc.);
- adopt and update a processing register;
- identify a person responsible for the protection of personal data or data protection officer, with the assignment of specific tasks for the effective implementation and effective application of compliance according to the GDPR (for example, providing day-by-day advice on the issues concerning data protection, performing risk assessments);
- define Standard Operating Procedures (SOP);
- carry out information and training activities on the requirements and obligations resulting from the GDPR;
- schedule periodic audit activities to monitor correct SOP application;

Since the approach is risk-based and requires an impact assessment, it is important to adopt a solid risk management framework and perform a Data Protection Impact Assessment (DPIA) The Spoke will:

- identify the architecture, the parties involved, the roles and responsibilities for risk management;
- assess the confidentiality of personal data processed;
- implement the risk management through risk identification, assessment, prioritization, and mitigation strategies;
- review of the risks (e.g. data loss, data integrity, unauthorized access, unauthorized or non-compliance of the processing, continuity of service, etc.) and related risk avoidance and risk mitigation strategies;
- assessment of the residual risk and review of the risk avoidance and mitigation strategies;
- monitoring and review of the risk management model for a continuous improvement of the data protection and risk management system (monitoring for continuous improvement).

Following the European and national rules and regulations regarding personal data protection, it is always required to define minimum data retention time. Any processing of personal data carried out by employees, collaborators, and all those who, although not employed by but operate in various roles in the activities involving data processing, must respect privacy by design and by default principles. In full compliance with regulations in force and following the principles of necessity, correctness, pertinence, and minimisation, data processing takes place solely for determined, explicit, and legitimate purposes.

The information and documents will be managed and preserved on registered and approved systems for their technical and contractual elements, respecting the current applicable rules and regulations and the guidelines referring to the protection of personal data. The security measures adopted must be in line with the level of risk that has been identified.

In accordance with article 37 of the GDPR, a Data Protection Officer must be appointed who will be responsible for:

- informing and providing advice to the data controller or the data processor as well as to the employees who carry out the processing on the obligations deriving from the GDPR, as well as from other national or European Union laws related with data protection;
- monitoring the compliance of the data controller or the data processor to the GDPR requirements, including the assignment of responsibilities, awareness-raising, and training of personnel involved in the processing and related control activities;
- providing an opinion on the impact assessment on data protection and monitor its performance according to art. 35 of the GDPR (if specifically requested);

- cooperating with the privacy authority (Garante per la Protezione dei Dati Personali) for any issue related with data processing and data protection, including early consultations (art. 36);
- supporting the data controller in keeping the processing register, follow the data controller's instructions.

DPOs are involved in the definition of the risk avoidance and mitigation measures and in general in all the data protection-related activities in compliance with the requirements of the GDPR.

Spoke 9 - Digital Society & Smart Cities

The spoke builds on the concept of exploiting big data - together with suitable models, algorithms, software architectures, and HPC infrastructures - to investigate new approaches for solving key challenges in smart cities and in digital society, covering a broad range of application domains. The capability of the spoke to process a wide range of extremely heterogeneous data sets - including personal data and data of a sensitive nature - is fundamental for the successful achievements of its goals; this, in turn, requires a rigorous approach to data management.

In general terms, the spoke will follow the “Guidelines on Data Management in H2020”, "Horizon Europe Data Management Plan Template", in addition to the EU General Data Protection Regulation (GDPR): these documents offer a good operational context and are already known and adopted by the partners of the spoke. Recognising the importance of ethical and privacy concerns, the adopted approach will address ethical and privacy requirements, including explainable and trustworthy AI. The activities of the spoke are articulated **around FAIR data principles**: making data findable, including provisions for metadata; making data openly accessible; making data interoperable; increasing data re-use.

In order to favor open accessibility to research data, the spoke will use open data sets available to the public domain and data accessible through open services (e.g. open APIs), hence accessible also to third parties. These datasets will be preferred whenever possible. New datasets will be defined by the project and will be identified as a result of project activities. Whenever possible, these additional datasets will also be made available as open data or via open services.

For sensitive datasets or datasets concerning personal information, such as health information or mobility tracking records, which cannot be made publicly available, an evaluation will be performed to assess if, after suitable aggregation, anonymisation and/or pseudonymisation, they can be made available to external stakeholders. At the same time, attention will be devoted to the generation of realistic data on human characteristics and behaviors. More in general, we will adopt the following principles for all datasets concerning personal information that are collected, processed and/or managed by the spoke: collection of informed consent; processing personal data lawfully, fairly and transparently; specific attention to aspects related to AI explainability; data protection by design/default; data security, integrity and confidentiality; data accuracy; and storage limitation.

The appropriate monitoring, management and compliance maintenance of research data will be guaranteed by the implementation of the **Data Management Plan (DMP)**: this plan will be defined in the early stages of the research on (M)odeling (WP1), and will then be regularly updated and continually assessed. Each data provider will be appointed as joint data controller after the signature of a Joint Controllership Data Processing Agreement (JCA). The DMP of the spoke enables all partners that signed the JCA to understand the different actions that handling data of different nature, origin and size imply for anyone wanting to stay in a safe harbor while actively contributing to the successful achievement of project outcomes. The JCA will also define the terms and conditions for data usage, both by the partners of the spoke and by third parties; for what concerns the partners, specific attention will be given to data usage by the industrial partners that are directly and indirectly involved in the spoke, taking into account the whole range of activities from foundational research to feasibility studies and support to innovation.

Spoke 10 - Quantum Computing

The research activities of the spoke do not devise the use or production of sensitive data. Still, the activities of the spoke will be guided by the usage of appropriate data ethics frameworks to guide data usage within the project. The building blocks will be those of transparency, accountability and fairness.

Special care will be given to ensure effective governance of data: organisations within the spoke will have the responsibility to keep both personal data and non-personal data secure in compliancy with Italian and European data governance policies. Special care will also be given in understanding how the findings from the spoke activities should be used with effective accountability mechanisms which could extend beyond the duration of the program.

B. Characteristics, feasibility and control

Composition of the CN – Centro Nazionale

The National Centre for HPC, Big Data and Quantum Computing (CN in the following) has the ambition to provide a solid reference point for large scale computing to the Italian scientific community and industrial system. It fosters computing related activities towards the three directions of the "knowledge triangle[1]": research, innovation and education.

The CN will deploy top-class resources and make them available to internal and external users based on policies to be defined according to the European Charter of Access for Research Infrastructures[2]. The total value of the resources will largely exceed the budget funded on the PNRR, with roughly 50% of the total costs to be provided by the hosting institutions or by EU initiatives.

The CN is organized as a Hub with 11 Spokes, with Spoke 0 ("Supercomputing Cloud Infrastructure") having the special role of designing, procuring, and managing the resources to be acquired and shared for the benefit of the users from the scientific Spokes (1-10), the industrial partners, and beyond, and providing high level support for the deployment and exploitation of the developed infrastructure. Societal implications and impact are addressed by a transversal research group (SII).

The Hub, located in Bologna at the **Tecnopolo**, is the implementing body for the project, oversees the management of the research program and the accounting; it fosters innovation, education, and training activities, with the support by specific professionals like the Innovation Manager and the Education Manager. The research activities are developed in the Spokes, which define the research programs, hire personnel, and procure and operate the infrastructure; they are the responsible towards the Hub for the scientific program, carried out by their shared personnel, the personnel they hire with contracts or PhD grants, and by Open Calls and Innovation grants. For a more resilient organization, a Spoke co-leader institution has been selected for each spoke among its affiliates based on the scientific track record and critical mass. The research activity of each spoke is coordinated by a Leader, nominated by the Spoke (leader) institution, together with a co-Leader, nominated by the Spoke co-leader institution. Spoke Leader and co-Leader are top-class scientists. Private companies are part of the CN as founding members; they participate to the activities by sharing technological solutions, use-cases and real test environments for scale-up activities and proof of concepts. The scientific goals of the CN are described in Section A. In this Section, the structure of the CN project is described, with an explanation of the composition, of the workplan, of the procedures and of the budget.

Figure 1 Members of the CN

[1] https://cordis.europa.eu/article/id/86306-fostering-the-knowledge-triangle
[2] https://op.europa.eu/en/publication-detail/-/publication/78e87306-48bc-11e6-9c64-01aa75ed71a1/

A large and high-quality **public-private partnership of 49 partners** has been selected to implement the research program of the CN, including: (a) 25 Statal Universities and Research Institutes supervised by MUR (Dlgs 218), (b) 9 Public Research Institutes and (c) 15 Private Partners, as shown in Figure 1. The partnership is well distributed all over Italy, with Northern, Central and Southern regions equally represented. The largest research institutions in Italy (CNR, INFN, INAF, INGV …) offer a further level of uniformity, with their operational units ensuring the possibility to cover the few areas not covered by academic members. Moreover, a well-balanced pool of leading companies covering all the spokes' domains are involved, with a similar level of pervasiveness as well as multiple units and subsidiaries. IFAB Foundation will also play the role of engaging more companies, including SMEs and innovative start-ups. Finally, 3 additional public research institutions are planned to be involved into the Hub only, without budget on the Research Program, based on their strategic role in connection to other initiatives. The overall selection process is described later in this document and more details are given in the Governance Section.

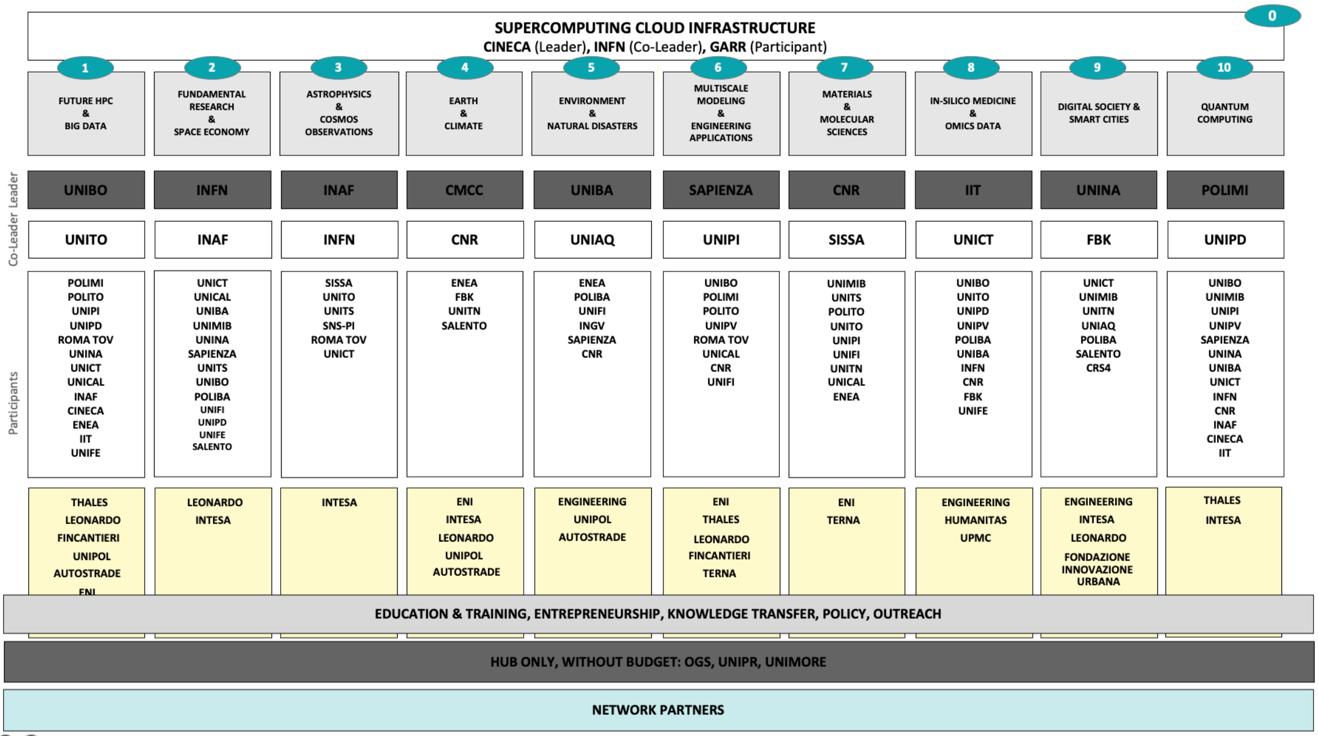


Figure 2 The full composition of the CN

The involvement of institutions in the **11 Spokes** is shown in Figure 2, including the current snapshot of the affiliation of private companies to the spokes.

The figure shows a structural difference between the Spoke 0 and the Spokes 1-10.
The **Spoke 0** handles the process of creation and management of the shared infrastructure, distributed on the Italian territory. It will procure resources, ready, upgrade or modernize facilities, and federate them via a middleware layer to ensure a global view and access to the various type of resources to all the users, researchers (and beyond) involved in the CN, also via the upgrade of the networking handled by GARR. The same Spoke will operate most of the CN infrastructure and provide continuous support to the user communities. While most resources will be located at properly modernized existing centres, Spoke 0 plans to deploy new additional centres especially in the Southern regions, as leaves of the same federation.
The **Spokes 1-10**, already presented in Section A, represent the scientific domains, chosen for their relevance to the PNRR planning. Their goal is to support and enhance top level research groups in a vast landscape of sciences, by using the e-Infrastructure provided by the CN. The workplan in each Spoke includes the formulation of new use cases made possible by the enhanced computing capabilities, the availability of new computing paradigms (like Quantum Computing), and the collaboration opportunities from the increased exchange of information and solutions within the CN, including the companies present. The number of thematic Spokes was chosen to support a wide range of scientific domains, from material sciences to life sciences and engineering to fundamental research, while, at the same time, guaranteeing a critical mass of qualified participants and effort in each domain, as explicitly required by the PNRR call.

The composition of the research partners has been carefully chosen among the strongest institutions and research groups, active on large scale scientific computing either as providers of state-of-the-art infrastructures or as users (organized within scientific domains), with a proven portfolio of applications; additionally, a selected yet omni-comprehensive number of companies will directly participate as full members, guaranteeing the fulfilment of the PNRR goals on digital transition and growth of productivity.
The selection process has proceeded via a bottom-up collection of candidate institutions, in which the candidates were asked to demonstrate their experience in scientific computing related research and in particular related to the themes in the Spokes. They were also asked to present and qualify the research groups proposed to join the CN as quality (composition and number) and via the advanced use cases they could propose on the CN infrastructure.
Among the large number of answers, the proponent has selected a strong partnership, based on the quantifiable metrics such as:

- Previous participation to successful large-scale projects in scientific computing, and their success (the relevant metrics are total budget, scientific results and proven resource deployment capability – CPU/GPU cores, PB of storage, already established participation to networks with peer institutes…);
- a critical mass of research groups led by world-class PIs (identified by metrics such as the h-index and the participation and coordination to regional / national/ EU projects);
- the demonstrated capability to develop, deploy and maintain in operations large scale multi domain e-Infrastructures, also from the organizational point of view;
- the adherence of the proposed research programs with the requests from the PNRR call: to foster national cooperation between all actors, "*to address major social concerns and global challenges such as climate change, zero hunger, good health and well-being, affordable and clean energy, the preservation of life below the water and on land",* to *"improve productivity and the low investment in human and physical capitals",* and to *"operate towards an ecological transition";*
- an appropriate geographical balance to account for the Southern regions and allow the deployment in these regions of 40% at least of the new acquired resources, as required by the call;
- the adherence to the objectives of Next-Generation EU and the integration with ongoing and future projects and initiatives such as EuroHPC, EOSC Data Cloud, EPI, and the EU Data Strategy.

The procedure has selected the research partners as shown in Figure 1 and in Figure 2.

The list of research partners reflects the two main needs of the CN: e-Infrastructures need to be procured and maintained by the most experienced institutions in large-scale scientific computing and, on top of this, champions from the scientific domains relevant for the PNRR themes, which have been selected to bring use cases that can scale up (both from scientific and computing points of view) by using the CN infrastructure.
In the first category (corresponding to the Spoke 0), the selected partners are INFN, which has the largest number of distributed GRID and Cloud centres and the largest storage deployment, GARR, which handles the Italian Research Network (NREN), and CINECA, the Italian PRACE Tier-0 node that is going to host one of the three pre-exascale systems as co-funded by the EuroHPC JU.
In the second category, selected partners were selected having in mind scientific excellence: each of them brings specificities to the CN, via groups at the forefront of their research, and with established links with the state-of-the-art global e-Infrastructures on the continental and international levels.
Table 1, Table 2, Table 3 and Table 4 show the critical mass of participation from the research institutions, with breakdowns for Spokes and Affiliates. Equivalent information from private companies is not presented since they do not expose personnel costs in the PNRR budget. The details include:

- the number of staff personnel shared with the CN (per project year);
- the number of staff personnel ready to dedicate to the CN at least 3 months per project year;
- the fraction of women in the shared personnel;
- the fraction of shared personnel with a Ph.D. obtained by no more than 10 years.

In total, 1517 research staff will be involved in the activities of the CN, dedicating a fraction of their research time. Of particular importance are the 645 researchers that are ready to dedicate a consistent (≥25%) effort – far higher than the minimum required and a proof of their interest on the CN activities; these are the professionals who guarantee the strength of the CN plan and build trust on the positive realization of the project.

Legal Entity	Year 1		Year 2		Year 3		% Female
	Tot. Resources Involved	Resources Involved for at least 3 months/year	Tot. Resources Involved	Resources Involved for at least 3 months/year	Tot. Resources Involved	Resources Involved for at least 3 months/year	
0 - SUPERCOMPUTING CLOUD INFRASTRUCTURE	43	26	43	26	43	26	32,5%
1 - FUTURE HPC & BIG DATA	194	69	193	69	193	69	17,0%
2 - FUNDAMENTAL RESEARCH & SPACE ECONOMY	194	80	194	80	193	80	15,1%
3 - ASTROPHYSICS & COSMOS OBSERVATIONS	124	41	124	41	124	41	24,2%
4 - EARTH & CLIMATE	108	39	108	39	108	39	26,9%
5 - ENVIRONMENT & NATURAL DISASTERS	103	42	103	41	103	41	24,3%
6 - MULTISCALE MODELING & ENGINEERING APPLICATIONS	191	45	181	45	191	45	22,0%
7 - MATERIALS & MOLECULAR SCIENCES	131	59	131	59	131	59	17,6%
8 - IN-SILICO MEDICINE & OMICS DATA	116	86	116	86	116	86	30,2%
9 - MATERIALS & MOLECULAR SCIENCES	127	49	127	49	127	49	18,9%
10 - QUANTUM COMPUNTING	186	109	186	109	186	109	20,8%
TOTAL	**1517**	**645**	**1506**	**644**	**1515**	**644**	**21%**

Table 1 Personnel shared by Public Research Institutions per Spoke

Legal Entity	Year 1		Year 2		Year 3		% Female
	Tot. Resources Involved	Resources Involved for at least 3 months/year	Tot. Resources Involved	Resources Involved for at least 3 months/year	Tot. Resources Involved	Resources Involved for at least 3 months/year	
CINECA	27	25	27	25	27	25	41%
GARR	7	4	7	4	7	4	29%
INFN	76	41	76	41	76	41	23%
UNIBO	76	23	76	23	76	23	22%
ENEA	48	27	48	27	48	27	27%
IIT	22	15	22	15	22	15	0%
INAF	78	22	78	22	78	22	28%
POLIMI	45	19	45	19	45	19	20%
POLITO	40	12	40	12	40	12	23%
ROMA TOV	44	12	44	12	44	12	11%
UNICAL	43	17	43	17	43	17	14%
UNICT	62	30	61	30	62	30	24%
UNIFE	30	14	30	14	30	14	27%
UNINA	77	38	77	38	77	38	13%
UNIPD	50	31	50	31	50	31	22%
UNIPI	76	22	76	22	76	22	22%
UNITO	56	24	56	24	56	24	25%
POLIBA	43	17	43	17	41	17	25%
UNISALENTO	44	13	44	13	44	13	25%
SAPIENZA	72	18	62	18	72	18	17%
UNIBA	58	20	58	20	58	20	24%
UNIFI	53	17	53	17	53	17	11%
UNIMIB	48	17	48	17	48	17	10%
UNITS	23	14	23	14	23	14	35%
SISSA	23	12	23	12	23	12	4%
SNS-PI	9	8	9	8	9	8	11%
CMCC	34	5	34	5	34	5	35%
CNR	128	55	128	54	128	54	27%
FBK	28	21	28	21	28	21	21%
UNITN	35	12	35	12	35	12	6%
INGV	7	7	7	7	7	7	43%
UNIAQ	21	11	21	11	21	11	24%
UNIPV	25	18	25	18	25	18	20%
CRS4	9	4	9	4	9	4	11%
TOTAL	**1517**	**645**	**1506**	**644**	**1515**	**644**	**21%**

Table 2 Total shared personnel per Public Research Institution

Legal Entity	Year 1		Year 2		Year 3		% Female
	Tot. Resources Involved	Resources Involved for at least 3 months/year	Tot. Resources Involved	Resources Involved for at least 3 months/year	Tot. Resources Involved	Resources Involved for at least 3 months/year	
0 - SUPERCOMPUTING CLOUD INFRASTRUCTURE							
CINECA	**13**	**13**	**13**	**13**	**13**	**13**	**46,15%**
GARR	7	4	7	4	7	4	28,50%
INFN	23	9	23	9	23	9	26%
1 - FUTURE HPC & BIG DATA							
UNIBO	**26**	**7**	**26**	**7**	**26**	**7**	**15%**
CINECA	7	5	7	5	7	5	28,57%
ENEA	11	7	11	7	11	7	36%
IIT	8	4	8	4	8	4	0%
INAF	7	4	7	4	7	4	14%
POLIMI	15	4	15	4	15	4	20%
POLITO	15	4	15	4	15	4	33%
ROMA TOV	14	4	14	4	14	4	7%
UNICAL	8	4	8	4	8	4	13%
UNICT	10	4	9	4	9	4	20%
UNIFE	10	4	10	4	10	4	10%
UNINA	11	4	11	4	11	4	9,09%
UNIPD	11	4	11	4	11	4	0%
UNIPI	15	5	15	5	15	5	7%
UNITO	26	5	26	5	26	5	27%
2 - FUNDAMENTAL RESEARCH & SPACE ECONOMY							
INFN	**20**	**17**	**20**	**17**	**20**	**17**	**10%**
INAF	14	8	14	8	14	8	14%
POLIBA	11	4	11	4	9	4	33%
UNISALENTO	16	4	16	4	16	4	19%
SAPIENZA	15	4	15	4	15	4	7%
UNIBA	10	4	10	4	10	4	10%
UNIBO	10	4	10	4	10	4	30%
UNICAL	12	5	12	5	12	5	8%
UNICT	12	4	12	4	13	4	31%
UNIFE	10	4	10	4	10	4	10%
UNIFI	13	4	13	4	13	4	0%
UNIMIB	14	4	14	4	14	4	0%
UNINA	19	5	19	5	19	5	16%
UNIPD	11	4	11	4	11	4	27%
UNITS	7	5	7	5	7	5	29%
3 - ASTROPHYSICS & COSMOS OBSERVATIONS							
INAF	**47**	**5**	**47**	**5**	**47**	**5**	**36%**
INFN	13	4	13	4	13	4	23%
ROMA TOV	15	4	15	4	15	4	7%
SISSA	12	8	12	8	12	8	8%
SNS-PI	9	8	9	8	9	8	11%
UNICT	8	4	8	4	8	4	25%
UNITO	11	4	11	4	11	4	18%
UNITS	9	4	9	4	9	4	33%
4 - EARTH & CLIMATE							
CMCC	**34**	**5**	**34**	**5**	**34**	**5**	**35%**
CNR	31	13	31	13	31	13	32%
ENEA	10	8	10	8	10	8	20%
FBK	7	4	7	4	7	4	14%
UNISALENTO	16	5	16	5	16	5	19%
UNITN	10	4	10	4	10	4	10%
5 - ENVIRONMENT & NATURAL DISASTERS							
UNIBA	**20**	**5**	**20**	**5**	**20**	**5**	**20%**
CNR	15	7	15	6	15	6	20%
ENEA	16	5	16	5	16	5	19%
INGV	7	7	7	7	7	7	43%
POLIBA	9	4	9	4	9	4	44%
SAPIENZA	12	4	12	4	12	4	8%
UNIAQ	12	6	12	6	12	6	33%
UNIFI	12	4	12	4	12	4	25%

Table 3 Full matrix of shared personnel per Spoke and per Affiliate (first part)

6 - MULTISCALE MODELING & ENGINEERING APPLICATIONS							
SAPIENZA	**35**	**5**	**25**	**5**	**35**	**5**	**29%**
CNR	21	4	21	4	21	4	19%
POLIMI	17	4	17	4	17	4	18%
POLITO	15	4	15	4	15	4	13%
ROMA TOV	15	4	15	4	15	4	20%
UNIBO	15	4	15	4	15	4	33%
UNICAL	14	4	14	4	14	4	14%
UNIFI	15	4	15	4	15	4	20%
UNIPI	33	8	33	8	33	8	21%
UNIPV	11	4	11	4	11	4	27%
7 - MATERIALS & MOLECULAR SCIENCES							
CNR	**30**	**9**	**30**	**9**	**30**	**9**	**16,67%**
ENEA	11	7	11	7	11	7	36,36%
POLITO	10	4	10	4	10	4	20,00%
SISSA	11	4	11	4	11	4	0%
UNICAL	9	4	9	4	9	4	22%
UNIFI	13	5	13	5	13	5	0%
UNIMIB	7	4	7	4	7	4	14,29%
UNIPI	11	5	11	5	11	5	36,36%
UNITN	12	4	12	4	12	4	0%
UNITO	10	8	10	8	10	8	20%
UNITS	7	5	7	5	7	5	43%
8 - IN-SILICO MEDICINE & OMICS DATA							
IIT	**7**	**6**	**7**	**6**	**7**	**6**	**0%**
CNR	11	11	11	11	11	11	55%
FBK	8	8	8	8	8	8	13%
INFN	10	5	10	5	10	5	50%
POLIBA	8	5	8	5	8	5	0%
UNIBA	10	6	10	6	10	6	40%
UNIBO	12	4	12	4	12	4	17%
UNICT	13	10	13	10	13	10	23%
UNIFE	10	6	10	6	10	6	60%
UNIPD	11	11	11	11	11	11	36%
UNIPV	7	7	7	7	7	7	14%
UNITO	9	7	9	7	9	7	33%
9 - MATERIALS & MOLECULAR SCIENCES							
UNINA	**27**	**10**	**27**	**10**	**27**	**10**	**7,41%**
CRS4	9	4	9	4	9	4	11%
FBK	13	9	13	9	13	9	31%
POLIBA	15	4	15	4	15	4	20%
UNISALENTO	12	4	12	4	12	4	42%
UNIAQ	9	5	9	5	9	5	11%
UNICT	11	4	11	4	11	4	27%
UNIMIB	18	5	18	5	18	5	22%
UNITN	13	4	13	4	13	4	8%
10 - QUANTUM COMPUNTING							
POLIMI	**13**	**11**	**13**	**11**	**13**	**11**	**23%**
CINECA	7	7	7	7	7	7	42,90%
CNR	20	11	20	11	20	11	30%
IIT	7	5	7	5	7	5	0%
INAF	10	5	10	5	10	5	20%
INFN	10	6	10	6	10	6	16,70%
SAPIENZA	10	5	10	5	10	5	0%
UNIBA	18	5	18	5	18	5	28%
UNIBO	13	4	13	4	13	4	23,10%
UNICT	8	4	8	4	8	4	12,50%
UNIMIB	9	4	9	4	9	4	0%
UNINA	20	19	20	19	20	19	20%
UNIPD	17	12	17	12	17	12	23,50%
UNIPI	17	4	17	4	17	4	29,40%
UNIPV	7	7	7	7	7	7	14%
TOTAL	**1517**	**645**	**1506**	**644**	**1515**	**644**	**21%**

Table 4 Full matrix of shared personnel per Spoke and per Affiliate (second part)

On top of the shared research personnel, the CN will use a fraction of its budget for the **recruitment of more than 200 fixed term researchers and technologists**, with another allocation budgeted for **more than 300 PhD grants**. The recruitment procedures will satisfy the criteria of the call, and will aim at selecting top quality personnel, with a correct balance of age and territory (as explained in the last part of this document)..

A total of **15 companies will join as full members, at the limit of what is allowed by the call**. Their role is to ensure a link to the productive system, and to foster a reuse of technologies towards an improvement of its productivity and capability in creating value; their personnel is self-funded and not part of the CN budget, adding value to the proposal. The CN proponent is a founding member of the GAIA-X Hub, and as such already in strict connection with the most important realities among Italian companies.

The selection process for companies has started from a top-down preselection by the CN proponent, followed by a request of formal engagement. Additionally, the activities of the CN have been presented in various public events, resulting in additional bottom-up expressions of interest. The selection, among large number of interested candidates, has considered four main criteria:

- Excellence in their productive domain;
- Experience in computing-related R&D activities
- Being stakeholders in applications and use cases of national interest, which will make leverage on CN infrastructure to scale up the impact
- Dimension and/or proven capacity for funding research projects or innovation processes.

Selected companies are shown in Figure 3.

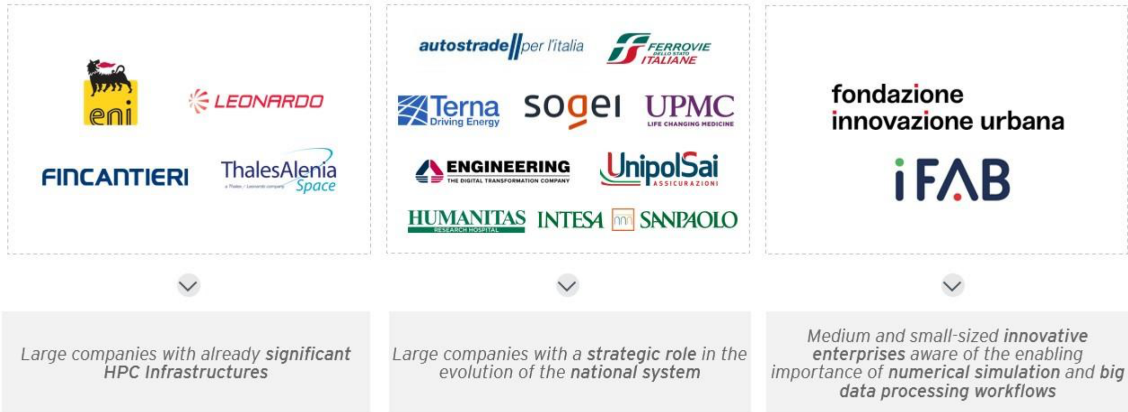


Figure 3 Selected companies

They are large multinational enterprises, with long standing interests on digital themes, and with a large portfolio of units and subsidiaries to match the activities in the Spokes. The CN private company members with a cumulative yearly turnover of nearly 60 B€ (as shown in the figure below) represent a relevant part of the Italian GDP, taking into account also the subsidiaries and controlled companies that belongs to this industrial ecosystem:

Company	Turnover (€M)
Autostrade	2.730
Engineering	880
ENI	18.000
Ferrovie dello Stato	188
Fincantieri	3.360
FIU	0,13
Humanitas	423
IFAB	1
Intesa San Paolo	13.170
Leonardo	7.225
SOGEI	625
Terna	2.037
Thales	635
Unipol Sai	10.130
UPMC	43
TOTAL	**59.447**

Table 5 Industrial partner business dimension

The large companies that already have significant HPC infrastructures such as Eni, Leonardo Company, Fincantieri, ThalesAlenia, have joined with the interest to the latest generation computing systems and emerging technologies, as well as to be part of the CN national federated infrastructure, capable of responding to extreme situations, as in the case of support to scientific research for the design of new drugs effective against the COVID-19 pandemic, or in perspective to respond to extreme natural events. In addition to the infrastructural aspects, the large companies have also shown interest in sharing methodological vertical skills, such as those expressed by the

research and academia, and advanced professional skills for the engineering of application solutions as part of their production workflows and support for the exploitation of new systems, as those that will soon be based on newly accelerators technology such as FPGA, neuromorphic or quantum computing accelerators. This need regarding high-level specialist support has proven to be relevant not only for the industrial system with a consolidated tradition in the computing field, but also for the service and infrastructure companies, called to face the needs of strategic structural evolution for the national system. Autostrade, Ferrovie dello Stato, Terna, Sogei, Engineering, Humanitas, UPMC, UnipolSai, Intesa San Paolo have joined the CN because they are interested in sharing their expertise for creating value from data, taking up and integrating new technologies, developing new services, elaborating the use cases of the future, supporting innovation and new entrepreneurship.
Finally, in the sector of PA Urban Innovation Foundation has been engaged as a key partner to develop and test digital twins for cities. Secondly, a structured collaboration with the CN of other companies, including medium-sized and small-sized innovative enterprises, is planned through the pivotal role of IFAB Foundation, that will operate similarly ETP4HPC, BDVA or QUIC for EuroHPC. The ecosystem of SMEs is aware of the enabling importance of numerical simulation and big data processing workflows, but not immediately able to profit from them due to skill limits and economic barriers to exploit new technologies. So, IFAB as a CN Founder will play an aggregator role for structured partnerships, including ISV's application providers and technology providers. Thirdly, strategic partnership agreements to create joint living labs with the main technology providers are planned (some of them have been already engaged) and network partners will be involved to jointly support and promote innovation (see innovation network partners).
On top of above, the open calls and innovation grants will further broaden the involvement of private partners and increase the overall impact.
The willingness of private companies to actively participate in the CN is also demonstrated by the fees they have committed to pay over a period of 5 years, ensuring about 3,5M€/year as capital. Private partners will participate actively at the CNMS governance being part of the BoD and in the industrial board.

The **CN is designed to be a stable and permanent Centre**, with a sound commitment to support the Italian scientific community and industrial system in the long run, far beyond the PNRR project. To this extent, great care is put on the sustainability issue, by asking precise commitments to the partners; in particular, it was decided to define an annual financial contribution for participation for different categories of partners and level of engagement, collected by the Hub and used to sustain the operations of the infrastructure after the project end date. All the CN partners have agreed on that and are committed for at least 4 years, upon joining the proponent for this call of interest; the annual contribution amount is set to:

- 100,000 Eur per year for public research institutions promoting founders,
- 25,000 Eur per year for public research institutions supporting founders,
- 250,000 Eur per year for private institutions promoting founders,

for a total of more than 6 MEur per year, which constitutes a solid basis for the sustainability beyond the project funded by PNRR budget. In addition, the CN may implement some services upon payment, such as development of tailored and modular solutions.

Governance introduction and description

The Foundation is the Legal Entity of the CN, based in Italy – Tecnopolo di Bologna. According to its purposes, it has the legal nature of a private non-profit foundation which satisfies the principles and the legal form of the "Fondazione di Partecipazione" (Participatory Foundation). The majority of founders is composed by State Universities and Public Research Institutions supervised by MUR (DLgs 218).

Hereafter the main Elements:
Name – National Centre for HPC, Big Data and Quantum Computing (ICSC)
Location – Bologna in a dedicated Centre; in the future it will be hosted in the Big Data Technopole, a former tobacco plant owned by Emilia Romagna Region. The area of 120.000 square meters is currently undergoing a requalification process and, in addition to hosting the ICSC, it will be a Technological and Scientific pole for big data and high performance computing. The overall Masterplan of Bologna Technopole realization is indicated in Figure 4; the overall project started in 2015. The intervention was divided into lots and, to date, all tenders have been published. The work progress varies in the individual lots. The ECMWF Data Centre is already operational. In September 2022, the first portion of the Supercomputing Centre will come into operation: it will be managed by CINECA and it will host the Leonardo supercomputer as well as the other facilities envisaged by the investment

program in Bologna and hosted by CINECA. In mid-2023 the second portion of the Supercomputing Centre will also come into operation: it will be managed by INFN and it will host the CNAF as well as the infrastructure upgrades envisaged by the investment program on Bologna and hosted by INFN. In mid-2023 the Technopole Labs will also be operational, with spaces for conferences and events, exhibitions and educational paths, civic engagement activities and data visualization labs. The building that will house the headquarters of the CN will be delivered in mid 2024. Until then the provisional headquarters for the start of the activities, including the Living Labs based in Bologna, will be the building in Figure 5, located in via Magnanelli 2 in Casalecchio di Reno (BO), already free and completely renovated, made available to the CN by CINECA and located next to its headquarters.

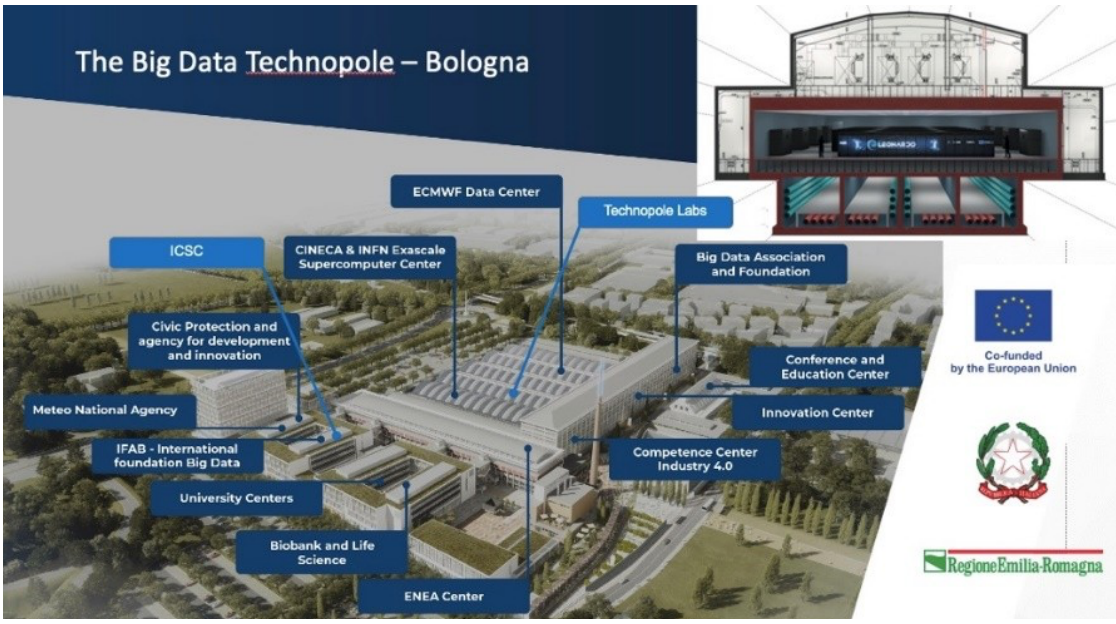


Figure 4 The Big Data Technopole – Bologna

Figure 5 Temporary ICSC building (Bologna)

Governance Structure – Hub and Spoke
Hub purpose – management and coordination as well as transversal support to spokes and affiliates
Spokes purpose – Execution of all CN activities (research, development, infrastructures and research material hosting, etc.) in collaboration with all spokes affiliates; one of the affiliates for each spoke will take the role of Spoke "co-leader Institution"; the coordination of scientific activities will be performed by the Spoke Leader and Spoke Co-leader, nominated by their respective institutions, chosen among top scientists with high scientific quality and experience in the coordination of complex research projects. Spoke Leaders and Co-leaders, whose names are listed in Table 6, serve a 4-year term which can be renewed only once.

SPOKE	Spoke Leader	Institution	Spoke Co-leader	Institution
Spoke 0	Sanzio Bassini	CINECA	D.Bettoni / C.Grandi	INFN
Spoke 1	Luca Benini	UNIBO	Marco Aldinucci	UNITO
Spoke 2	S. Malvezzi / T.Boccali	INFN	Antonio Stamerra	INAF
Spoke 3	Ugo Becciani	INAF	P. Lubrano / G. Carlino	INFN
Spoke 4	Silvio Gualdi	CMCC	Cristina Facchini	CNR
Spoke 5	Roberto Bellotti	UNIBA	Paola Inverardi	UNIAQ
Spoke 6	Mauro Valoran	UNIROMA1	Sergio Saponara	UNIPI
Spoke 7	Stefano Fabris	CNR	Stefano Baroni	SISSA
Spoke 8	Andrea Cavalli	IIT	Francesco Pappalardo	UNICT
Spoke 9	Nicola Mazzocca	UNINA	Marco Pistore	FBK
Spoke 10	Donatella Sciuto	POLIMI	Simone Montangero	UNIPD

Table 6 Current Spoke Leaders and Co-Leaders

Hub and Spokes relationship – collaboration agreement between Hub and each Spoke Institution
Spoke and Affiliates relationship – collaboration agreement among Spoke Institutions and Affiliates; the current spokes configuration will last until the end of the PNRR research program, at which point it will be possible to reconsider it following an assessment aiming to evaluate the research activities, the results achieved and the economic/stakeholder context.
Members and partners
The Foundation will involve two categories of founding members (or simply members) as shown in Figure 6:
A) **Promoting members** - the following public and private institutions can choose to be promoting members:

- A1) State Universities and Public Research Institutions supervised by MUR (DLgs 218)
- A2) Non-statal Universities and other Research Institutions
- A3) Private Companies

B) **Participating members** - Only Universities and Public Research Institutions can choose to belong to this category. This category does not include Private Institutions.
The two categories of Members have different governance rights, as described in the Hub Governance Structure Section. The extent of the contribution due annually by the Members of the Foundation is determined, separately for the various categories, by the Board of Directors. For the first four years it is fixed in:

- Promoting Members A1 and A2 – 100.000 Eur/y
- Promoting Members A3 – 250.000 Eur/y
- Participating Members – 25.000 Eur/y

After the first 36 months (CN start-up phase) it is possible to admit new Founding Members.
As shown by Figure 6, all public and private institutions involved into the research program of the CN are Founding Members. The majority of them is composed by State Universities and Public Research Institutions supervised by MUR (DLgs 218). All the Spoke Leader and co-Leader Institutions are Promoting Members, as well as all the private Institutions.

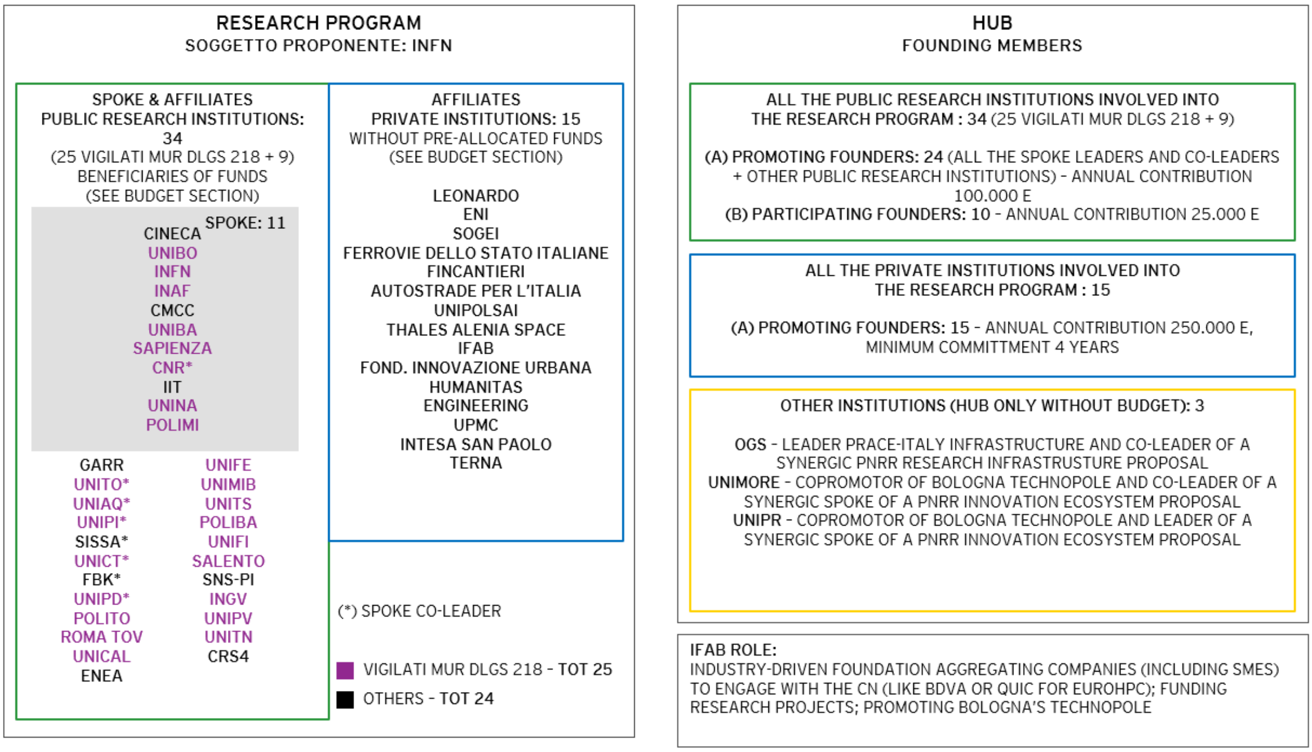


Figure 6 ICSC founding members

Public or private institutions which undertake to contribute to the mission of the CN for at least 4 years are called the Partners of the Foundation. Partners are not by obligation Foundation Members.
There are three categories of Foundation Partners, depending on their role in the CN and on the different level of involvement in the Advisory Committees, as described in the Hub Governance Structure Section.:
Associate Partners have full access to scientific strategy and activities. They are required to pay an annual contribution equal to that paid by Promoting Members A3 (fixed for the first 4 years at 250,000 Eur / y). Public and private institutions and companies can choose to belong to this category;
Network Partners. This category includes various types of subjects which can be considered as stakeholders of the CN and which are interested in setting up collaborations at various levels with the CN:

- Institutional Network Partners (e.g.: Emilia-Romagna Region, Municipality of Bologna ...)
- Scientific Network Partners (e.g.: CINI, HPC4NDR, Italia Meteo, Fond. Links, ESA, ECMWF, ECCOMAS…)
- Trade and Professional Associations
- Innovation Network Partners (Bi-Rex - Industry 4.0 Competence Centre on Big Data, EDIH ...)

No financial contribution is required from Network Partners.
Supporting Partners. A further form of partnership with the CN is offered by IFAB, which is an Industry-driven International Foundation on big data and artificial intelligence for human development and has the role of aggregating companies (including SMEs) interested in collaborating with the CN (such as BDVA or QUIC for EuroHPC), to finance research projects as part of the activities of the CN and to animate the Bologna Technopole. IFAB, as Promoting Member of the CN, acts as a bridge between its members and the CN. The annual contribution of an IFAB Member is 60,000 Eur / y with a commitment of at least 3 years. This collaboration between the CN and IFAB will be formalized as soon as the CN is launched and the IFAB-Members will have the qualification of **Supporting Partners** of the CN. The modalities for Partners to get involved in addition to Open calls and Innovation grants are summarized in Figure 7.
Partners are admitted upon presentation of an application for admission addressed to the Board of Directors; the Board of Directors approves admission with the favorable vote of the majority of its members.

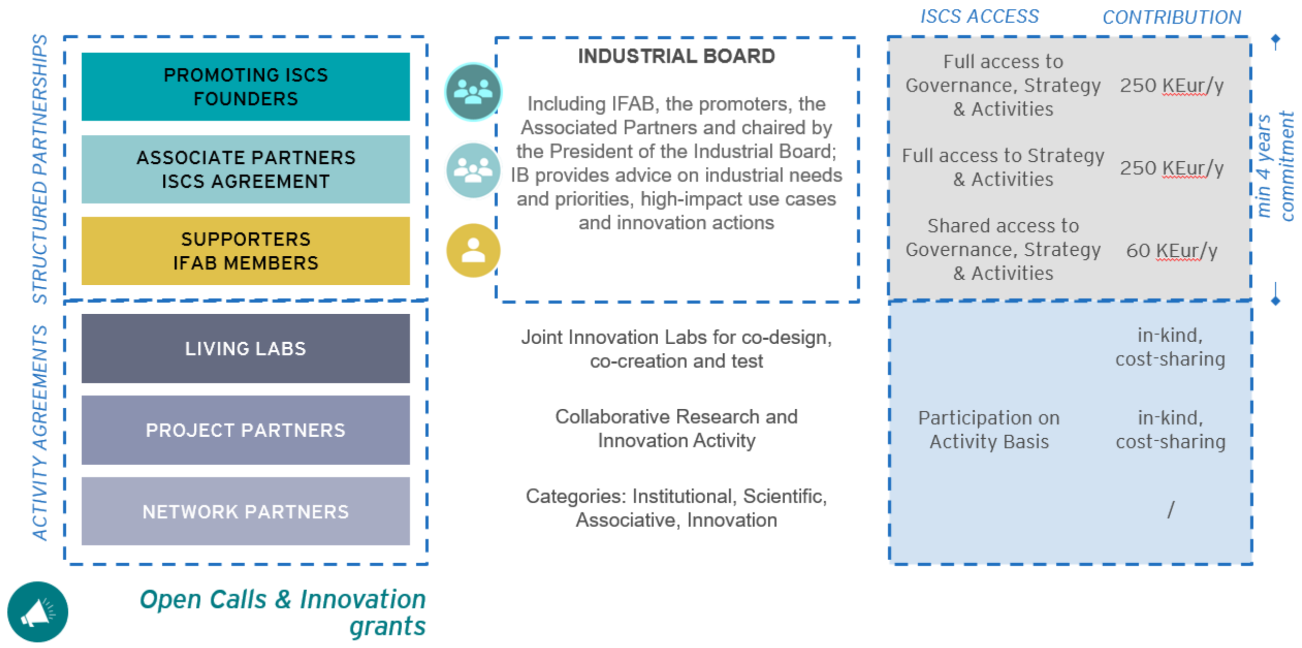


Figure 7 Partners involvement

Hub governance structure

The ICSC Governance is sketched in Figure 8.

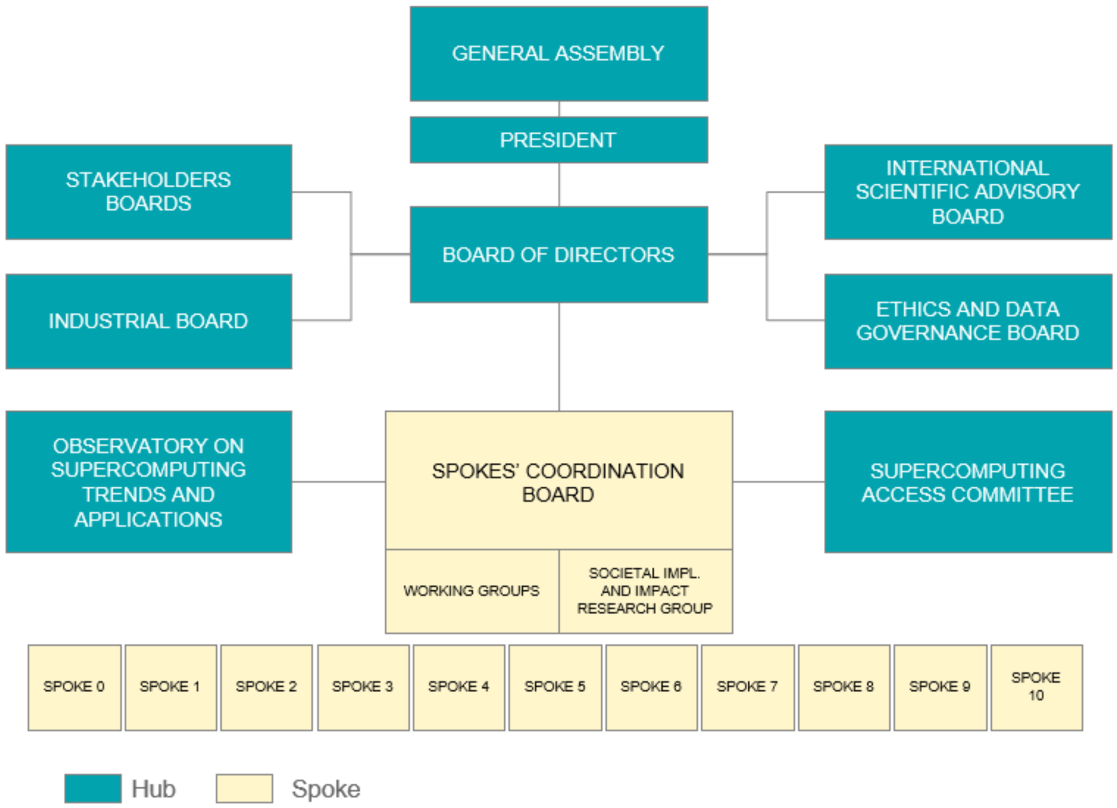


Figure 8 ICSC Governance (Hub)

The core bodies are: ICSC President, Board of Directors and General Assembly. There will be 5 additional boards with specific tasks: Spokes Coordination Board, International Scientific Advisory Board, Ethics and Data Governance Board, Industrial Board, Stakeholders Board. The structure is completed with the ICSC Observatory and the Supercomputing Access Committee.
The Management structure is led by ICSC Director and is described in the Management Section.
All elected offices have a duration of 4 (four) years and election / appointment mechanisms will be provided that respect gender balance.
ICSC President - The ICSC President is the President of the Foundation: he/she is the legal representative of the CN and the chairperson of the Board of Directors and of the General Assembly. The president is proposed by INFN as CN Proponent and is appointed together with the other members of the Board of Directors by the General Assembly, as described below.
Board of Directors - The Board of Directors is composed of an odd number of members varying between 7 and 11, as determined by the General Assembly, in compliance with the following conditions:

- 1 member is expressed by INFN, as CN Proponent, and assumes the position of Chairperson of the Board of Directors and president of the Foundation;
- Out of the remaining members, one half must be an expression of Promoting Members A1 (State Universities and Public Research Institutions supervised by MUR-DLgs 218)
- The other half must be an expression of Promoting Members A2 and A3, with at least 1 member of Promoting Members A2 (Non-statal Universities and other Research Institutions) and at least 2 members of Promoting Members A3 (Private Institutions)
- at least 1/3 (one third) of the members of the Board of Directors must be of the least represented gender.

Each member of the Board of Directors must: (i) possess high professional skills in matters relating to the activities of the Foundation and (ii) have gained significant management experience in Boards of Directors or similar areas.
The Board of Directors appoints a Vice-President among its members, who replaces the President in case of impediment. Within one year of the approval of the statute, the Board of Directors issues a regulation governing conflicts of interest, based on:

a) principle of membership: the members of the Board of Directors undertake to act in the national interest and undertake, in the event of conflicts of interest, to favor the general interest over the obligations of corporate or associative loyalty;
b) principle of transparency: the members of the Board of Directors are required to report annually to the Foundation any potential personal advantage deriving from projects or actions, falling within the competence of the Foundation, in which they are involved in any capacity;
c) principle of responsibility: the members of the Board of Directors are required to answer for the correctness and effectiveness of the activities carried out and declare themselves available to report on the results by subjecting the activities to evaluation processes.

The members of the Board of Directors are elected by the General Assembly with a voting mechanism by lists and one vote per Member. Each member of the Foundation has the right to present or to compete for only one list and can cast only one candidate. The candidates from the list that obtained the highest number of votes among all the lists presented are elected members.
The Board of Directors holds all the powers necessary for the ordinary and extraordinary administration of the Foundation. It is the responsibility of the board to approve the documents relating to the strategic and operational guidelines of the body and the program of scientific activity, taking into account the non-binding recommendations expressed by the Foundation's bodies.
The Board of Directors is expected to activate 3 committees within it: Program monitoring Committee, Nomination and remuneration Committee, Risk and long-term sustainability Committee.
The Scientific Coordinator of the Spokes Coordination Board and the ICSC Director are invited to the meetings of the Board of Directors.

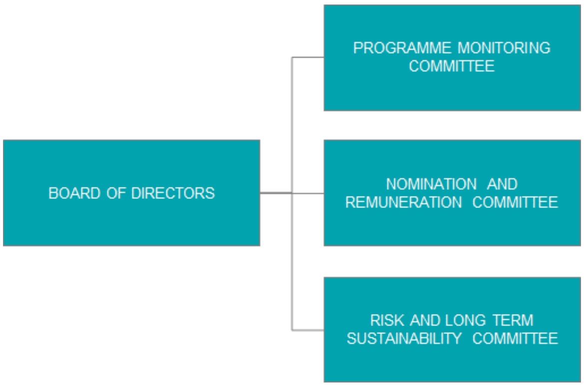


Figure 9 Board of Directors Committees

General Assembly - The foundation President is the Chairperson of the General Assembly. All Foundation Members participate in the General Assembly. Each member of the Foundation has the right to one vote and can be represented by written proxy. The assembly can be convened by the President or at the request of the majority of the members. In the absence of the President, the assembly is chaired by the Vice-President.
The assembly:

- approves the financial statements;
- approves any amendments to the Foundation's statute;
- expresses its preventive and non-binding opinion on initiatives undertaken by the Foundation in pursuit of its institutional purposes;
- formulates consultative opinions and proposals on the activities, programs and objectives of the Foundation;
- proceeds with the determination of the number of members and the appointment of the members of the Board of Directors.

The assembly validly meets, on first call, with the participation of the majority of the Foundation's Members; on second call, the assembly is validly constituted whatever the number of attendees. Resolutions are taken with the vote of the majority of the Foundation Members.
Spokes Coordination Board - The Spokes Coordination Board is the advisory body that: (i) interacts with the Board of Directors, (ii) coordinates the activities of the Spokes, (iii) interacts with other Hub bodies. It is composed by the spoke leaders and co-leaders and chaired by a Scientific Coordinator selected among its members. It addresses and monitors the scientific activity plan and the overall performance indicators, shares best practices among spokes and promotes cross-contamination through inter-spoke or multi-spoke initiatives. In carrying out its functions it will make use of 6 Working Groups that will operate across the various Spokes (as shown in Figure 10): WG1 - Outreach & Community Engagement, WG2 - Education & Training, WG3 - Building Capacity, WG4 - Competitive Projects, WG5 - International Initiatives, WG6 - Innovation Initiatives and Ecosystems. In particular, WG3 will be responsible for promoting and coordinating initiatives aimed at unlocking the research and innovation potential of Academia, Industry (including Smes) and Public Administration and at planning and promoting actions aimed at strengthening the less advanced areas of the country or those that have less capacity, both as disciplinary fields or productive sectors and as geographical location.

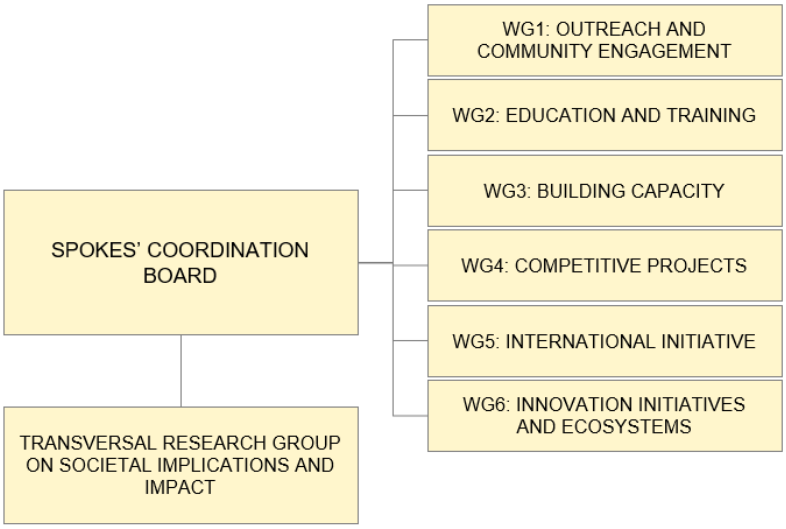


Figure 10 Spokes Coordination Board: Working Groups and SII Research Group

Furthermore, a **Research Group on Societal Implications and Impact (SII)** will be established, which will operate across the Spokes, as described in Section A. The Group, whose leader and co-leader are Luciano Floridi and Nino Rotolo, both from Bologna University, is composed of nine outstanding experts and nine PhD students or post-docs. The Spokes Coordination Board will regularly organize meetings and workshops on topics of general interest and aimed at the entire CN to promote community building, internal networking and spokes synergies and to share new ideas and proposals.

International Scientific Advisory Board - It is an independent body of 7-11 members with outstanding scientific profile, selected by the Board of Directors via an open call. It is chaired by a Chairperson selected among its members. It provides advice on the Research Program, the scientific plans and goals, on the impact and protocols of scientific research activities. It also assesses the correlation between scientific activities and multi-year plans.

Ethics and Data Governance Board - The Board is composed of 7 outstanding experts in the field and is appointed by the Board of Directors. It is chaired by a Chairperson selected among its members. The Ethics and Data Governance Board provides independent advice, guidance, and feedback to the CN on ethical and data governance issues. Details are given in the next section. From a policy and strategic perspective, the CN aims to maximize the positive societal impact of its activities. In particular, it adheres to Open Science standards (e.g., following the guideline "as open as possible, as closed as necessary") and to the principles of the upcoming EU Data Act in maximizing "the value of data in the economy by ensuring that a wider range of stakeholders gains control over their data and that more data is available for innovative use, while preserving incentives to invest in data generation".

Industrial Board - It is an industry-driven board, composed by one person nominated by each Promoting Member of A3-type (Private Institutions) and by each Associate Partner, with proven specific expertise on economic and sector vision and on high-impact projects. It is chaired by the President of the Industrial Board, appointed by the Industrial Board itself. The Board provides advice on industrial needs and priorities and proposes high-impact use cases and innovation actions.

Stakeholders Board - It is a board formed by one representative from each Network Partners. It is consulted by the bodies of the Foundation on strategic decisions and can make proposals or recommendations to the Board of Directors.

ICSC Observatory - An Observatory will be created on Supercomputing Trends and Applications in order to: (a) provide evidence on the latest innovative trends and the related potential socio-economic impact, (b) identify the dynamics of good practices, with a special focus on small and medium-sized enterprises (SME), (c) understand the barriers to innovation and propose policy tools to overcome them, (d) nurture 'win-win' relationships between entrepreneurs, policy makers, innovation facilitators and researchers, (and) enhance awareness of the latest innovation trends, success stories, and related opportunities and barriers, through case studies, trend reports and workshops / conferences, (f) distribute the CN / Spokes findings via the website and social media tools, (g) support policy makers. The Observatory, together with the spokes, will produce some MOOCs on the topics of the CN.

Supercomputing Access Committee - The Committee has the task of evaluating the requests for access to the infrastructure connected to Open Calls and allocating the computing resources in accordance with the Access Policy of the CN, in analogy to the European model Prace and the Cineca ISCRA Panel.

Details on Ethics and Data Governance Board

The Ethics and Data Governance Board (henceforth, the EDGB) provides independent advice, guidance, and feedback to the Centre on ethical and data governance issues. The EDGB is expected to elaborate guidelines to ensure that the Centre conforms to national and international standards, and, upon request, monitor ongoing activities of the Centre. It can also be consulted to provide policies and opinions on ongoing ethical issues in research. The primary responsibility of the EDGB is to enable the Centre to protect all participants in the research and all stakeholders and consider potential risks and benefits for the community within which the research will be carried out. The EDGB independently operates also in support of the Board of Directors.

Thematic Scope of the EDGB

The EDGB provides expertise and guidance on the following themes/values:

- equity (avoiding disparities)
- ethical alignment with regulatory compliance
- fairness
- privacy and data protection
- legal risk analysis, jurisdiction, negligence, liability
- risk assessment, management, and governance
- intellectual property
- social trust and acceptance
- sustainability
- data management and sovereignty
- transparency
- human factors
- authentication, authorisation & accountability.

Legal matters and cases are not directly in the scope of EDGB insofar as they involve organisational and legal responsibilities. However, it is strongly recommended that the EDGB provides some advice on the ethical alignment with regulatory compliance, e.g., policies for promoting legal compliance by design, or legal risk assessment. Both issues are central in the GDPR and the current discussion in Europe of the Data Act and AI Act. Regarding data protection and privacy, the primary oversight comes from the GDPR (with the DPO, working in collaboration with the DPOs in the partner institutions, and the other figures in the Centre as required by the law). Still, the EDGB can deal with privacy issues by offering guidelines specific to big data.

EDGB Structure

The EDGB is appointed by the Board of Directors and comprises seven outstanding experts in the field, including a Chairperson appointed among its members. Each appointment is for three years, renewable.

If needed, the EDGB can create ad hoc and temporary working groups focusing on specific themes and topics

The EDGB can rely on external experts for specific scientific consultancy services or when in support to ethical reviews of projects. All ordinary costs incurred by the EDGB for its activities (travelling, meetings, etc.) are supported directly by the Centre. The EDGB is provided with a small, discretionary budget of € 10,000 a year to cover extraordinary costs (expert fees, open access publication of a report, etc.). Members of the EDGB are not renumerated (*pro bono*). The EDGB and the Chairperson are supported by a staff.

Functions of the EDGB

The EDGB meets at least three times a year.

All resolutions by the EDGB are based on a majority vote.

The EDGB draws on best practices established nationally and internationally.

The EDGB will

- examine and report on the themes and values specified above to promote responsible scientific innovation and advancement of knowledge, social values, environmental sustainability, and public understanding and engagement;
- identify, define, and address ethical and data governance issues related to current activities of the Centre;
- advice on the ethical impact and acceptability of the future projects and activities Centre;

- in the light of the outcome of its work, and in agreement with the Centre, the EDGB may publish reports and provide other advisory input and guidance as the Centre may judge appropriate;
- serve as a last-instance Commission for evaluating violations of ethical standards within the Centre;
- operates as an independent body in support of the Board of Directors according to the following procedure:
    - with the preliminary advice of the International Scientific Advisory Board and the Spokes Coordination Board, the Board of Directors deliberates on the development goals and strategic activities of the CN;
    - the EDGB elaborates standards, high-level operational guidelines, and binding principles that are expected to contribute to the achievement of the goals of CN concerning data governance; the deliberations of the EDGB become effective once adopted by the Board of Directors;
    - these standards, guidelines, and principles are primarily addressed to the CN spokes regarding the following issues on human, organisational and regulatory aspects.

The EDGB provides its ethics advice independently. The EDGB does not provide any legal advice.
The EDGB avails itself of external experts in supporting the routine ex ante and ex post ethical review of the projects of the Centre. This activity is operationally ensured by the Ethics and Data Management Unit.

Spokes governance structure

The Governance of each Spoke is sketched in Figure 11.

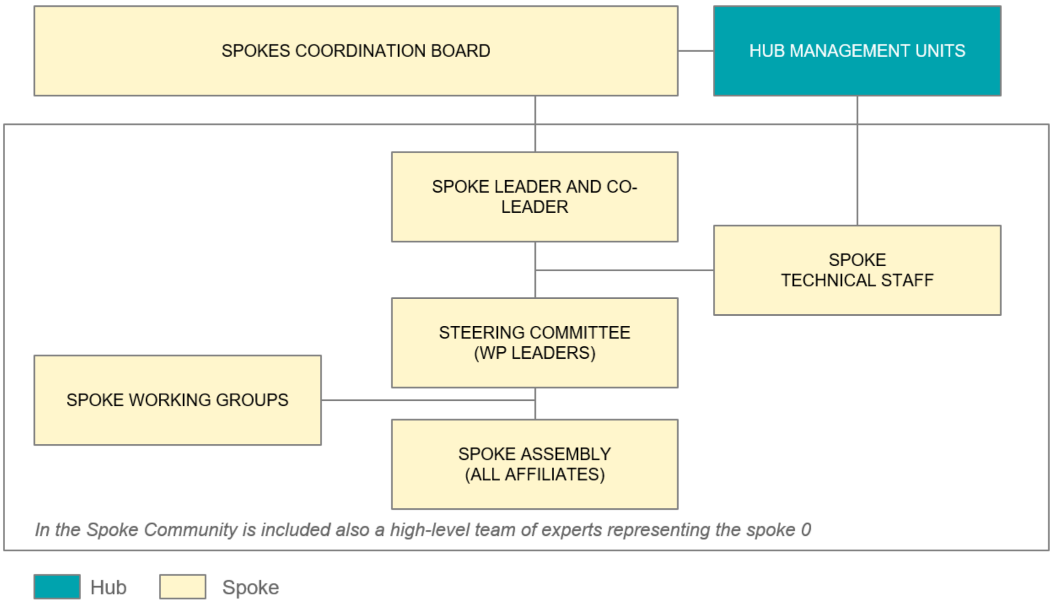


Figure 11 Spoke Governance and Management

The activity of the Spoke is directed by Spoke-leader and Spoke co-leader (names shown in Table 6), selected from figures of high scientific qualification and experience in coordinating research programs and proposed respectively by the two subjects: Spoke Leader Institution and Spoke Co-leader Institution. The governance duration is 4 years. The Steering Committee organizes its activities through thematic Working-groups, which, when necessary, are integrated with teams of experts from Spoke 0 to ensure effective coordination with Spoke 0 - Supercomputing Cloud Infrastructure. The Steering Committee regularly organizes initiatives to engage the entire community of researchers involved in the activities of the Spoke and its Stakeholders. The spoke makes use of a dedicated Technical Staff (see the next Section) which also has the function of liaising with the Management Team of the CN on the one hand and with the administrative contact of the Spoke Institution. Each Spoke Affiliate will identify an administrative referent of his / her Institution who will report to the Administrative Referent of the Spoke Institution.

ICSC management structure

The ICSC Management structure is sketched in Figure 12.

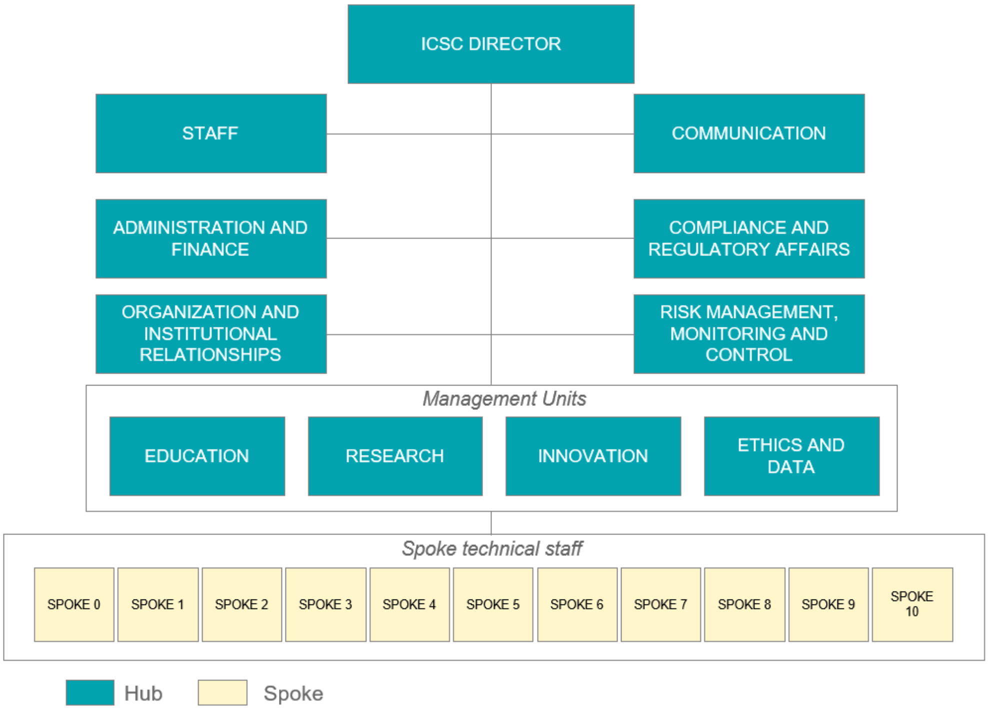


Figure 12 ICSC Management

The **ICSC Director** is responsible for the implementation of the planned activities and the decisions of the Board of Directors. The ICSC Director is appointed by the Board of Directors among highly qualified professionals with experience gained in the management of complex structures and knowledge of the operation of public structures. The ICSC Director is supported by the Hub operational units, each assigned with specific functions. The Director is assisted by 4 managers, each of whom is entrusted with the management of a specific unit: Research Management Unit, Innovation Management Unit, Education Management Unit and Ethics & Data Management Unit. The first three units manage the implementation of the activity plan in the 3 domains: Research, Innovation and Education. The fourth unit has a management and support function in the area of data and ethical-legal implications. The other units shown in Figure 12 are in staff to the ICSC Director. The CN Management Team is composed of: ICSC Director, who chairs it, Research Manager, Innovation Manager, Education Manager and Ethics & Data Manager. Finally, the management of the CN includes a team of Technical secretariats, one for each Spoke, which guarantees technical support for the governance of the Spokes and effective coordination with the Hub and the various administrative structures of the institutions affiliated with the spokes. Overall, the organizational structure described will involve around 30 people from the Foundation when fully operational, plus dedicated staff from Spokes / Affiliates.
More in detail all operational units and functions are described in Table 7.

Operational Unit	**Functions**
Communication	• Ensuring implementation of the ICSC Communication & Dissemination Plan • Guaranteeing the coordination with the communication structures of the Spokes • Supporting the ICSC Observatory • Ensuring visibility of the Union funding, by providing coherent, effective and proportionate targeted information to multiple audiences
Administration and Finance	• Providing administrative and financial management and support for the CN • Overseeing budgeting and accounting • Ensuring traceability of expenses and avoidance of double funding of operations through specific accounting and information coding procedures • Responsible for the procurements

	• Developing calls for tenders/proposals containing selection criteria and requirements for participating entities, in compliance with national and EU legislation • Identifying any factors that may affect the timing of implementation and expenditure as defined in the time schedule
Organization and Institutional Relationships	• Developing personnel policies and administering current personnel regulations and rules with a view to ensure high-qualified and motivated staff • Recruiting ICSC staff • Ensuring effective coordination with all the institutions involved in the CN • To offer support to the other Units in the preparation of agreements • Support to the bodies of the Foundation
Compliance & Regulatory Affairs	• Ensuring compliance of all activities with relevant national and EU legislation, including state aid regulation • Guaranteeing compliance with the "Do no significant harm" (DNSH) principle in the assessment and implementation of activities • Adopting measures to ensure compliance with the principle of sound financial management, in particular concerning the prevention of conflicts of interest, fraud, corruption and the recovery of unduly assigned funds • Ensuring compliance with any regulatory provisions, guidelines and technical instructions issued by relevant national and EU entities involved in the implementation of the Research Program
Risk Management, Monitoring& Control	• Managing and mitigating the risks associated with the implementation of the Research Program • Carrying out the managerial, administrative and financial controls of the operations as required by law • Facilitating verifications by the authorized bodies, including on-spot verifications, on the beneficiaries of the loans, including on the Hub, the Spokes and the affiliated entities • Guaranteeing a sound conservation of project documentation and of supporting documents relating to expenses incurred to ensure traceability of operations
Research Management	• Managing all the research activities of the CN and ensuring the implementation the Research Program as approved • Supporting the Spokes Coordination Board • Producing, on a bimonthly basis, reports on the activities of the National Centre, highlighting the achievement of intermediate and final milestones and target, and reports of expenses incurred by the National Centre • Producing a final technical report detailing the implementation of the Research Program and ensuring transmission of data relating to the lines of activity in order to allow for the preparation of annual reports • Ensuring the correctness, reliability and consistency of the financial, physical and procedural data of the operations and compliance with the relevant information system requirements • Overall responsibility for quality of ICSC research projects • Monitoring research opportunities at national, European and international level and support the CN in the preparation of research projects • Ensuring effective coordination with the Spokes research structures
Innovation Management	• Managing all the innovation activities planned by the CN • Managing IP policy, agreements and exploitation • Integrating marketing and communication, aligning the objectives of the entire Foundation • Working closely with experts from the different areas of the Foundation, detecting improvements, and new opportunities

	• Monitoring, analysing, and communicating innovation metrics to the management to seek opportunities to improve the Foundation's innovative performance
Education Management	• Managing all the Education and training activities planned by the CN • Identifying present and future needs within the educational system, and plan, develop and modify facilities and programs • Networking and building professional relationships with members of the institution and the community • Guaranteeing an effective coordination among Spokes
Ethics and Data Management	• Coordinating the data management routine procedures of the CN, in which also specific data management figures operate within each spoke: it is responsible in the CN for the implementation of policies, supported by a team of stewards (at least 10) representing data domains, to ensure the quality and integrity of the data lifecycle • Implementing executive decisions and ensuring adherence to procedures • Enforcing the data governance policies of the Ethics and Data Governance Board • Ensuring the routine ex ante and ex post ethical review of the projects of the CN; in doing so, the Unit avails itself of the advice of the Ethics and Data Governance Board

Table 7 Main functions of each operational unit

Operational Units

In this Section the operational units of the Public Research Institutions involved in the Program are described. As far as the private partners, as mentioned above, they are mainly large groups and their interest and engagement tend to spread over various spokes with different operational units. Since they have no costs allocated, for the sake of brevity, their overall involvement is described in a specific section of the document.

The research institutions are structured internally as Departments, Operational Units or laboratories, often covering multiple scientific domains of interest for the CN.

Table 8 indicates the units expected to be involved in the deployment of the research activities.

CINECA	Dip.	HPC Bologna e Napoli
INFN		Bologna, Cagliari, Catania, Firenze, Ferrara, Frascati, Genova, Bari, Perugia, Pisa
		L'Aquila, Lecce, Legnaro, Milano, Napoli, Padova, Pavia, Roma, Torino, Trento, Trieste
GARR		Roma
INAF		All the Structures, Cagliari, Catania, Firenze, L'Aquila,
		Milano, Napoli, Padova, Palermo, Roma, Torino, Trieste
CNR	Dipartimenti	Scienze del Sistema Terra e Tecnologie per l'Ambiente
		Ingegneria, ICT e Tecnologie per l'Energia e i Trasporti
		Scienze Fisiche e Tecnologie della Materia
		Scienze Chimice e Tecnologie dei Materiali
		Scienze Biomediche
ENEA	Divisioni	TERIN - ICT
		TERIN - SEN
		SSPT-MET
		TERIN - BBC
INGV	Dip.	Terremoti
		Vulcani
CRS4		Cagliari
CMCC		Lecce
IIT		Genova
FBK		Trento
POLIBA	Dipartimenti	Fisica
		Ingegneria Civile, Ambientale, del Territorio, Edile e di Chimica
		Scienza dell'Ingegneria Civile e dell'Architettura
		Ingegneria Elettrica e dell'informazione
		Energia Elettrica e dell'Informazione
		Meccanica, Matematica e Management
POLIMI	Dipartimenti	Matematica
		Bioingegneria
		Ingegneria Civile e Ambientale
		Chimica, Materiali e Ingegneria Chimica
		Elettronica e Telecomunicazioni
		Energia
		Ingegneria Meccanica e Aerospaziale
		Meccanica
		Scienze e Tecnologie Aerospaziali
POLITO	Dipartimenti	Automatica e Informatica
		Elettronica e Telecomunicazioni
		Interateneo di Scienze, Progetto e Politiche del Territorio
		Energia
		Scienze Matematiche
		Scienza Applicata e Tecnologia
		Ingegneria Meccanica e Aerospaziale
ROMA TOV	Dipartimenti	Matematica
		Fisica
		Scienze e Tecnologie Chimiche
		Ingegneria Civile ed Informatica
		Ingegneria dell'Impresa
		Ingegneria Elettronica
		Biomedicina e Prevenzione
SALENTO	Dipartimenti	Matematica e Fisica
		Ingegneria dell'Innovazione
		Scienze dell'Economia
		Scienze e Tecnologie Biologiche ed Ambientali
		Studi Umanistici
SAPIENZA	Dipartimenti	Ingegneria dell'Informazione, Elettronica e Telecomunicazioni
		Ingegneria Strutturale e Geotecnica
		Fisica
		Ingegneria Chimica, Materiali e Ambiente
		Ingegneria Informatica, Automatica e Gestionale
		Ingegneria, Meccanica e Aerospaziale
		Matematica
		Scienze di Base e Applicate per l'Ingegneria
SISSA	Area/Dip	Fisica
		Scienze Matematiche Informatiche e Fisiche
SNS		Classe di Scienze
UNIAQ	Dipartimenti	Ingegneria Civile, Edile-Architettura e Ambientale
		Ingegneria e Scienze dell'Informazione e Mat
		Scienze Fisiche e Chimiche
		Ingegneria Industriale e dell'Informazione e d
UNIBO	Dipartimenti	Ingegneria dell'Energia elettrica e dell'Informa
		Informatica - Sienza e Ingegneria – Bologna
		Matematica
		Fisica e Astronomia – Bologna
		Scienze Statistiche
		Ingegneria industriale
		Chimica
		Scienze Politiche e Sociali
		Ingegneria civile, chimica, ambientale e dei M
		Chimica industriale
		Farmacia e BioTecnologie
		Medicina Specialistica Diagnostica e Sperime
		Scienze Mediche e Chirurgiche
		Informatica - Sienza e Ingegneria
UNICAL	Dipartimenti	Matematica e Informatica
		Culture, Educazione e Società
		Ingegneria Meccanica, Energetica e Gestional
		Ingegneria Informatica, Modellistica, Elettron
		Scienze Politiche e Sociali
		Fisica
		Chimica e Tecnologie Chimiche
UNICT	Dipartimenti	Ingegneria Elettrica, Elettronica e Informatica
		Matematica e Informatica
		Fisica e Astronomia
		Medicina Clinica e Sperimentale
		Scienze Biomediche e Biotecnologiche
		Scienze del Farmaco e della Salute
		Scienze Mediche, Chirurgiche e Tecnologie Avanzat
		Ingegneria Civile e Architettura
UNIFE	Dipartimenti	Fisica e Scienze della Terra
		Ingegneria
		Matematica e Informatica
		Scienze dell'Ambiente e della Prevenzione
		Scienze Chimiche, Farmaceutiche e Agrarie
		Medicina Traslazionale e per la Romagna
		Scienze della Vita e Biotecnologie
		Scienze Mediche
UNINA	Dipartimenti	Ingegneria Elettrica e delle Tecnologie dell'Informaz
		Matematica e Applicazioni
		Ingegneria Elettrica e delle Tecnologie dell'Informaz
		Fisica
		Scienze Economiche e Statistiche
		Ingegneria Civile, Edile e Ambientale
		Scienze Politiche
		Scienze Sociali
		Scienze Chimiche
		Economia, Management, Istituzioni
UNIPD	Dipartimenti	Fisica e Astronomia
		Ingegneria Civile, Edile e Ambientale
		Ingegneria dell'Informazione
		Scienze Chimiche
		Matematica
		Biologia
		Medicina Molecolare
		Scienze Biomediche
		Scienze Statistiche
UNIPI	Dipartimenti	Informatica
		Ingegneria dell'Informazione
		Matematica
		Chimica e Chimica Industriale
		Fisica
		Ingegneria Civile e Industriale
		Ingegneria dell'Energia, dei Sistemi, del Territorio e delle Costruzioni
		Farmacia
UNIBA	Dipartimenti	Scienze della Terra e Geoambientali
		Interateneo di Fisica
		Scienze Agro Ambientali e Territoriali
		Farmacia - Scienze del Farmaco
		Bioscienze, Biotecnologie e Biofarmaceutica
		Interdisciplinare di Medicina
		Scienze Mediche di Base, Neuroscienze ed Organi di Senso
		Informatica
		Chimica
UNIFI	Dipartimenti	Fisica e Astronomia
		Scienze e Tecnologie Agrarie, Alimentari, Ambientali e Forestali
		Biologia
		Ingegneria dell'Informazione
		Statistica, Informatica, Applicazioni
		Scienze della Terra
		Ingegneria Civile Edile e Ambientale
		Chimica
UNIMIB	Dipartimenti	Fisica
		Scienze dei Materiali
		Informatica, Sistemistica e Comunicazione
		Economia, Metodi Quantitativi e Strategie di Impresa
		Matematica e Applicazioni
		Medicina e Chirurgia
		Psicologia
		Scienze dell'Ambiente e della Terra
		Statistica e Metodi Quantitativi
UNIPV	Dipartimenti	Ingegneria Industriale e dell'Informazione
		Ingegneria Civile e Architettura
		Matematica
		Chimica
		Biologia e Biotecnologie
		Fisica
		Medicina Molecolare
		Scienze del Sistema Nervoso e del
UNITO	Dipartimenti	Fisica
		Informatica
		Matematica
		Chimica
		Scienze Agrarie, Forestali e Alimentari
		Biotecnologie Molecolari e Scienze per la Salute
		Neuroscienze
		Oncologia
		Scienza e Tecnologia del Farmaco
		Scienze Cliniche e Biologiche
		Scienze Mediche
		Scienze Veterinarie
UNITS	Dip.	Fisica
		Ingegneria e Architettura
		Scienze Chimiche e Farmaceutiche
UNITN	Dipartimenti	Ingegneria e Scienza dell'Informazione
		Fisica
		Ingegneria Civile Ambientale e Meccanica
		Matematica
		Sociologia e Ricerca Sociale

Table 8 Public research units involved in the CN

Research groups

The 11 Spokes have each several affiliated research members, as shown in Figure 2, varying between 3 and 15. In many cases, still, single affiliates participate to each Spoke activities with more than one organizational unit (for example, different departments from the same university). The shared personnel, as demonstrated by the CVs and in the description of each affiliate (in Section A), includes the champions in each scientific domain, with a proven log of results and participation to successful projects. Most importantly, the affiliates to each spoke have at least a decennial history of collaboration among them, which guarantees the capability to co-operate in a large and structured project, while maintaining and make the most of different experiences and roles. The national research institutions INFN, INAF, CNR and INGV, play an important role in the harmonization of activities, given their well-established links with university departments and regional entities; indeed, their institutional activities already happen in a strong synergy with academia that will be replicated in the CN. The affiliates CINECA and GARR are structured as consortia or associations of research entities and are hence by construction in contact with all the Italian research institutions. Companies participate to the activities of the CN, either by exposing use cases of national interest, or by contributing directly to the spokes' activities, for example in the testbeds and PoCs needed to show the fulfilment of increased TRL levels. The selected companies have strong interests and investments in digital infrastructures, with R&D efforts which match well the needs of the scientific domains. Indeed, collaborations between the companies and the research partners of the CN already exist at multiple levels, via direct collaborations ad previous common projects.

Research plan

The 11 Spokes have been asked to provide a **detailed workplan, including the definition of Work Packages and of the main milestones**, based on the assumption of a 3-year project.
For all the spokes, a "Management" WP0 has been added, in charge of the overall coordination, of the administrative steps towards the affiliates and the Hub, and of compulsory activities like the preparation and the maintenance of a Data Management Plan. Its description and activities are similar in each spoke and are detailed only once here unless relevant specificities are present.

WP0: Management
This WP implements the coordination and management of each Spoke, covering technical, administrative, legal and financial issues.
Task 0.1: **Overall coordination, governance, monitoring, communication flow and methods**. The task will oversee the activities of the Spoke, from the legal, financial, administrative, and technical points of view. It will be responsible for receiving the accounting reports from the affiliates and transmitting them to the Hub. It will organize day-to-day activities, oversee the advancement of milestones and the preparation of deliverables. It will organize Spoke coordination meetings. WP leaders will remain responsible for the coordination of their correspondent WPs and task leaders will remain responsible for the coordination of their respective tasks.
Task 0.2: **Data Management Plan.** A Data Management Plan (DMP) is needed to describe in detail how the data products are handled, including reports, publications, data and code repositories, in compliance with national and European laws (e.g., the GDPR). The DMP will include information related to: data collection, documentation and metadata (description of data), ethics and legal compliance (intellectual and copyrights issues), storage and backup plan and access and security issues, selection and preservation policy, data sharing (retrieval and restrictions, authentication and authorization methods), responsibility and resources (who will handle the DMP). The DMP is a live document, which needs to be updated with evolving regulations and a deeper understanding of the use cases. A final version of the DMP is expected at the end of the project.
Task 0.3: **Training, dissemination plan and outreach activities**. Activities in the Spoke will include the organization of training events, in order to internally bring up to speed the hired personnel, and to foster a wider adoption of the technologies as tested. Training, dissemination and outreach activities are organized at the Hub level; the task will participate to central activities for the specific parts of the Spoke interest.
Task 0.4: **Open Calls, Innovative Grants and funding and management.** The activities in each Spoke include the publication of calls to CN external and internal entities, in order to foster co-design and co-development of solutions, the procurement and operations of services and resources as needed, and the inclusion in the work plan of the industrial CN partners.

Specific spoke-by-spoke work plans are described in the following paragraphs. The activities of the Spokes are aggregated in CN-wide Milestones, named "MX-MY" in the milestone list attached to this document.

Milestones and related deliverables are described in the present document and summarised in the CN milestones plan according to MUR requirements as transmitted in Phase 2. Some of the technical milestones, for example those about the opening of recruitment positions are generally common to different spokes, while those about scientific topics are directly associated to their spoke.

Spoke 0

Work Packages

WP0.1 Network (GARR): Upgrade of the national research network infrastructure, for interconnecting the national data centres and data repositories to the European system through GÉANT and its global international connectivity
WP0.2 HPC and Big Data Infrastructure (INFN, CINECA): Upgrade of the facilities of data centres and increase of computational and data storing resources of the HPC and Big Data Infrastructures. The list of interventions is the following:

- Leonardo upgrade with digital components (CINECA)
- Leonardo upgrade with quantum components (CINECA)
- Upgrade of CNR and INAF resources (Tier-1) at Tecnopolo (CINECA)
- Participation to exascale EU projects (CINECA)
- The CINECA – CNR Tier-1 in Naples (facilities and resources) (CINECA)
- Cloud access to quantum resources (CINECA)
- Upgrade of the Big Data distributed infrastructure (INFN)
- Upgrade of the Cloud services infrastructure (INFN)
- Data Centre for Disaster Resilience at LNGS (INFN)
- Data Centre for Space economy at LNF (INFN)

WP0.3 Middleware and Resource Federation (INFN, CINECA): Deployment and operation of middleware supporting composable federation services, providing easy and transparent access to the distributed resources of the HPC and Big Data Infrastructures
WP0.4 high Level Support, Training and Dissemination (CINECA, INFN): High level support and cross-sectional activities for software portability and optimization, co-design, software deployment, code development, data visualization and access interfaces and resource access.

Workplan

M1-M8
CINECA: Signature of the contractual agreement with EuroHPC for the upgrade of the Leonardo supercomputer with digital components (Expression of interest to the call **EUROHPC-2022-CEI-UPG-01**); Signature of the contractual agreement with EuroHPC for the upgrade of the Leonardo supercomputer with quantum components (Expression of interest to the call **EUROHPC-2022-CEI-QC-01**); Signature of the contractual agreement with exascale hosting Entity consortium (with reference to the call **EUROHPC-2021-CEI-EXA-01**); Definition of the requirements of the Tier1 supercomputer; Signature of the contractual agreement for the site location of the new data centre in Naples. Collection of user requirements for libraries and tools. Signature of the contractual agreements with selected PhD scholarship hosting Universities for 1st year (5 positions funded 100%). Publication of calls for 20 technologists.
GARR: Network design
INFN: Setup of activities for resource acquisition and upgrade; Collection of user requirements for software infrastructure

CINECA: Publication of calls for 20 technologists
INFN: Publication of calls for 11 technicians and 10 technologists

CINECA: Definition of innovation activities with industries. Publication of calls for 17 technologists. Consolidation of programming environments, mathematical libraries, optimisation tools.
INFN: Publication of tenders for the upgrade of the data centres facilities of the distributed infrastructure and for the acquisition of a fraction of the IT resources (network devices, about 140 servers and 38 PB) in the northern southern regions. Collection of user requirements for software infrastructure, identification of middleware stack
GARR: PHYSICAL NETWORK: Long term Fibre Acquisition (IRU) in the southern regions (> 4000 km); Include 20+ point of presence, 40+ amplification sites; Are also considered the local-tail acquisition to extend the infrastructure into the user premises.

M9-M15
CINECA: Procurement by the EuroHPC JU of the digital upgrade of the Leonardo pre-exascale supercomputer. RFP by the EuroHPC JU of the quantum upgrade of the Leonardo supercomputer. Procurement by the EuroHPC JU of the exascale system, advanced payment. RFP by CINECA of the Tier1 Supercomputer to be hosted at Bologna Tecnopolo, Prefinancing RFP for site preparation of a new data centre in Naples Prefinancing. Cloud access to quantum resources. Signature of the contractual agreements with selected PhD scholarship hosting Universities for 2nd year (5 positions funded 66%). Consolidation of programming environments, mathematical libraries, optimization tools. Preparation of summer and Specialist Schools.
INFN: Publication of tenders for the acquisition of a fraction of the IT resources of the data centres of the distributed infrastructure (network devices, about 55 servers and 15 PB) in the northern regions. Identification of the software stack and primary services starting from the user requirements.
GARR: OPTICAL NETWORK: Optical network Upgrade extending the open line system to the south regions, including HPC sites. Possibility to access 1T+ speed via coherent DWDM transmission systems.

M13-18
CINECA: Installation of the digital component of the Leonardo supercomputer, advanced payment. Selection of the quantum components of the upgrade of Leonardo supercomputer technology provider. Selection of the Bologna Tecnopolo Tier 1 system technology provider. Selection of the operator for the site preparation for a new data centre in Naples. Terms of reference for a Tier1 system to be installed in Naples. Activation of innovation activities with industries. Work on codes of scientific community and workflow consolidation for research and innovation
GARR: PACKET NETWORK: IP/MPLS network Upgrade to provide uniform network packet layer. State of art access capacity: 100G+,400G+. Acquisition HPC sites equipment to provide a dedicated packet network for a multi-Terabit communication between HPC sites.
INFN Design of the of PoC of the federation services of distributed HPC and Big Data systems

M17-M22
CINECA: Open in production of the upgrade of the Leonardo supercomputer. Installation of a quantum components of the upgrade of the Leonardo supercomputer. Acceptance test of the exascale system. Prefinancing 10% of the Capex. Installation of the Bologna Tecnopolo Tier 1 system. RFP for a Tier1 system to be installed in Naples. Works on codes for the scientific community and workflow consolidation for research and innovation
INFN: Publication of tenders for the upgrade of the data centres facilities of the distributed infrastructure and for the acquisition of a fraction of the IT resources (network devices, about 15 servers and 4 PB) in the northern and southern regions. Activation of the of PoC of the federation services of distributed HPC and Big Data systems.
GARR: Progressive activation of the Terabit band connectivity services

M22-M26
CINECA: Open in production of the quantum components of the upgrade of the Leonardo - final payment. Open in production of the exascale system. Open in production of the Bologna Tecnopolo Tier1 system final payment. Completion of the site preparation of a new data centre in Naples final payment. Selection of the Tier1 to be installed in Naples technology provider. Work on codes for the scientific community and workflow consolidation for research and innovation. Signature of the contractual agreements with selected PhD scholarship hosting Universities for 3rd year (3 positions funded 33%).
INFN: Publication of tenders for the acquisition of a fraction of the IT resources of the data centres of the distributed infrastructure (about 225 server and 62 PB) in the northern and southern regions. Execution of the of PoC of the federation services of distributed HPC and Big Data systems.
GARR: Progressive activation of the Terabit band connectivity services.

M25-M36
CINECA: Beginning of installation of the Tier1 system to be installed in Naples. New simulation codes and new applications
GARR: Progressive activation of the Terabit band connectivity services
INFN: Activities for the facilities renovation and the deployment of procured IT resources. Start implementation of the federation services of distributed HPC and Big Data systems

CINECA: Open in production of the Naples Tier 1 System - final payment. New simulation codes and new applications.
GARR: Progressive activation of the Terabit band connectivity services
INFN: Deployment of procured resources. Implementation of the federation services of distributed HPC and Big Data systems

GARR: Completion of the terabits band service for the identified points of presence
CINECA: All the facilities renovation activities and deployment of procured IT resources completed. Work on new calculation codes and new applications completed.
INFN: All the facilities renovation activities and deployment of procured IT resources completed. Work on federation services of distributed HPC and Big Data systems completed.

Spoke 1

Work Packages activities

WP1.1: Flagship on non-functional properties and design exploration (leader POLITO)
Modern HPC systems must be designed taking into account not only the key parameters in terms of cost, performance, and throughput but also non-functional properties, such as power/energy consumption and reliability, which are becoming increasingly important while the complexity of HPC systems increases and the semiconductor technology they use becomes complex. This requires the system designer to explore the design alternatives, considering all the different parameters to identify the best trade-off, early estimating energy, power, and reliability characteristics of each solution (in terms of HW and SW modules).
Main activities This project addresses the above issues by combining the partners' expertise and background in a holistic and complementary manner. In particular, partners will focus on the most advanced design techniques for devising energy- and power-efficient reliable parallel architectures, possibly based on Open source models for CPUs (e.g., RISC-V), GPUs, memory hierarchy, and on-chip interconnect. They will also define design space exploration techniques and tools, customizing the available solutions to the HPC scenario. Effective solutions for flexible reliability evaluation and enhancement, power and energy monitoring and management, thermal/power modelling, and control will also be devised, also considering the ramping-up approximate-computing paradigm.
Moving to the software aspects, the partners will work on suitable programming models and tools for energy efficiency, portability and performance portability, runtime resource management, autotuning mechanisms, and energy-aware hardware-dependent software.

WP1.2: **Flagship on heterogeneous acceleration, architecture, tools, and software (leader POLIMI)**
To design energy-efficient HPC applications, we plan to exploit custom accelerators based on a variety of approaches, ranging from reconfigurable devices, such as FPGAs and CGRAs to ASICs that decouple computation from memory accesses and specialize the memory subsystem to processing-in-memory designs for the specific workload characteristics. In this scenario, energy-efficient heterogeneous supercomputers need to be coupled with a radically new software stack capable of effectively exploiting the benefits offered by heterogeneity.
Main activities: In WP2, we plan to address these challenging problems by proposing the design, modelling, and simulation of heterogeneous accelerators for HPC systems, cloud systems, and edge servers. We also propose a holistic framework spanning all the decision layers of the supercomputer software stack based on tools and libraries for heterogeneous HPC-cloud-edge systems. The framework includes:

- High-Level Synthesis (HLS) approaches to reduce custom accelerator design time by automatically generating hardware descriptions of accelerators for regular and irregular workflows.
- A runtime system to autotune applications and dynamically monitor and manage the resources for energy-efficient heterogeneous HPC-cloud-edge systems at the runtime.

In WP1.2, we also plan to investigate how to apply disruptive acceleration technologies, such as processor-in-memory computing, neuromorphic computing, chiplets, 3D integration, silicon nanophotonics networks, and even the integration of quantum compute kernels "as accelerators" into conventional applications.

WP1.3: Flagship on workflows, I/O, and HPC-cloud convergence (leader UNIPI)
The modernization of HPC applications requires addressing different aspects such as *portability* across systems and new processors, *scalability* matching end-user expectations, *modularity,* and *composability* to accommodate different programming needs and styles while enabling the reuse and integration of legacy code. The main aim of the WP is to establish an umbrella community contributing, developing, and maintaining an integrated set of SW tools implementing a unifying software stack for HPC-oriented workflows, integrating tools and methodologies

for cloud-HPC and high-performance storage & IO, and providing suitable design and operation tools for digital twins.
Main activities: The main activities in the WP will be aimed at providing an integrated set of tools suitable to support the development, deployment, operation, and tuning of HPC applications targeting different kinds of systems and providing efficiency and portability along with maintainability beyond the existing state of the art tools. In particular, the integrated toolchain should include support for workflow management, resource management, deployment, and specific high-performance storage and I/O management. Reproducibility and FAIR-by-design workflows will be also addressed. The results achieved will be demonstrated through "application use cases," where several independent and integrated tools will cover the different aspects related to the efficient development and deployment of different HPC applications. Specific requirements and suggestions will be gathered from the industries participating in the WP activities and use cases from industries will be considered for the final assessment of the integrated tools developed in the WP. Tier-II platforms are envisioned as development and testing platforms to bring the software stack up to TRL 6 (full-scale pilot). At this stage, the software stack can be tested onto a pre-exascale facility and moved to the technological transfer path together with industrial partners. This latter stage does not necessarily involve the full software stack; rather, it focuses on specific parts of the stack that are of interest to the business plan. The parts that are transformed into products should maintain the integration within the stack. Multitenancy and data segregation will also be considered to address critical data challenges.

WP1.4: Flagship on trustworthiness, security, privacy (leader UNINA)
Information security is today emerging as a crucial dimension for new classes of HPC and Big Data applications, which increasingly involve massive amounts of privacy-sensitive data handled by the computing facilities. This Flagship project will address innovative technological solutions enabling multi-tenancy HPC/Cloud platforms with strong security and data privacy guarantees. In line with the objectives of the Future HPC Spoke, the Flagship will target open-source processor and platform architectures, establishing comprehensive support for a Trusted Execution Environment (TEE) based on RISC-V and experimenting with federated learning in the compute continuum (HPC-cloud-edge). The activities will also address software building blocks and infrastructures, including security services enabled by the availability of a TEE and scheduling issues in multi-tenancy HPC/Cloud platforms.
Main Activities: The Flagship will involve many multi-faceted research challenges and technical objectives, ranging from the development of hardware-level security primitives and the study of crypto-based solutions, e.g., oblivious memory and homomorphic encryption, to the definition of a reference architecture for a trusted execution environment targeted at HPC and associated protocols for confidentiality, attestation. The activities will also involve software support and tool flows, provisioning infrastructures, and multi-tenancy support in HPC/Cloud environments, emphasizing schedulers allowing migration and task placement policies for improved energy efficiency and resiliency and integration of privacy-preserving analytics (PPA) modules in distributed workflows.
The results will be evaluated in several application domains, from privacy-preserving data analytics in biomedical applications to Internet and social media network analysis, predictive maintenance, and Federated Learning (FL), a promising approach for improved AI systems that do not compromise the privacy of final users and the legitimate interests of private companies. In particular, the project will extend the FL paradigm and allow it to work in a black-box setting, where the federated model is built by ensembling local models. These techniques will be applied to real-world settings, including learning about big data and complex systems (e.g., geospatial networks).
The participants have already identified a range of industrial players potentially benefiting from the Flagship results. Furthermore, some of the large companies in the CN have expressed their interest in the activities related to the Flagship, involving resource optimization, federated learning, fault recovery, and on-premise and on-cloud resource management, as well as hardware/software-level security primitives for trustworthy computing, security algorithms and protocols for confidentiality and attestation.

WP1.5: HW-SW co-design, benchmarking, patterns, and microkernels (leader UNICT)
One of the key aspects of the future generation of HPC systems is to exploit heterogeneous systems to be customized-tailored by considering different issues. There is a real need to assess and validate through benchmarking activities: HW accelerators, heterogeneous architectures, especially for software optimization, microarchitecture efficiency, performance analysis and portability, and profiling of HPC codes.

Main Activities: Activities of codesign HW Development to be investigated will regard: VLSI and FPGA-based architectures; CPU-GPU algorithms, VLSI and FPGA-based architectures; CPU-GPU algorithms, data-driven parallelism, data affinity, and data locality, streaming computation. To guarantee disruptive impact on vertical fields in WP5, we plan to address mini-applications and benchmarking from multiple domains by algorithmic prototyping and algorithmic co-design, including specific issues related to in-depth simulation, modelling, and optimization tools. A possible not-exhaustive list of application domains is the following: AI/ML, Big Data, fluid dynamics, multi-scale simulations, data analysis for astrophysics, N-body dynamics, NN for large scale image processing (e.g., SAR, Multispectral), atomistic simulations, simulation of systems at the mesoscale, design of optical dielectric and plasmonic devices, materials modeling, social-media network analysis and graph analytics multi-particle, long-range interacting systems; computational geometry; light-matter interaction, low dimensionality systems, quantum materials, numerical analysis, N-Body Dynamics & Astrophysics, high-performance linear and nonlinear solvers.

WP1.6: Deployment of Spoke living labs (leaders UNIBO, UNITO)
FutureHPC is one of the two technology-centric foundational spokes of the CN proposal. Through its living labs, it aims to define permanent workshops populated with industrial experts to forge a new generation of early-stage researchers and practitioners with a clear vision of what could be the industrial needs and the opportunities offered by HPC key technologies and tools in 5-10 years that could address them.
Main Activities: Incepting two Spoke living labs on “Hardware and System Prototyping” (HSW) and “Software and Integration” (SWI): set up two physical facilities; hire resources; set up research and technology transfer activities.

Milestones

- M1-M8: Federated learning: surveys on the state of the art and gap analyses for RISC-V, Accel-HW-Spec, HPC dev tools, TEE, mini application benchmark and accelerated applications; procurement for laboratory setup (HWS & SWI).
- M5-M8: Selection of candidate prototypes. Procurements for laboratory full operation; Specification of first-year prototypes and demos for power management and energy/reliability monitoring platform, software framework for acceleration Accel-SW-spec, frameworks and development tools for HPC, Federated Learning and Trusted Execution Environment, benchmarking tools and mini-applications.
- M9-M15: Conclusion of Specification collection for Federated Learning, procurement phase for laboratory operations. Procurement for laboratory operation, Preliminary report on P&D for the first year and initial planning of P&D for the 2nd year.
- M13-M18: Initial Dissemination & exploitation activities. Federated Learning: report about: the development of the RISC-V platform; the development of acceleration platform: Hardware & Software side; Dissemination actions report & Interoperability WS proceedings; Hardware/software TEE building blocks; the development of mini-applications.
- M17-M22: Interim releases and integration of prototypes. HWS & SWI Labs: Expenses for laboratory operation, Report on execution P&D-Y2, Release of Y2 prototypes and demos. Federated Learning report about: Interim release of Power/Reliability management and monitoring platform; interim release of acceleration platforms; interim release and integration planning of dev tools for HPC; interim release of the RISC-V TEE and ML architectures; release of mini-applications and profiling and performance analysis tools.
- M22-M26: Pilot and implementation phase activities. Report on Federated Learning: Power/Reliability management and monitoring platform, development of acceleration platform, TEE and FL results, profiling and performance analysis tool, benchmarking, and profiling;
- M25-M36: Final demonstrators, integrations and reports of all the spoke activities.

Spoke 2

The Spoke 2 activities, divided into the 6 Work Packages (WP), involve almost two hundred (200) collaborators.

WP2.1: **Design and development of science-driven tools and innovative algorithms for Theoretical Physics** (INFN, INAF, UNIBA, UNIMIB, UNINA, ROMA1, UNIBO, UNIFI, UNICT, UNIPD, UNICAL, UNISALENTO, UNIFE): T1.1 Development of algorithms and codes for Exascale architectures; T1.2 Tools and Algorithms for Lattice Field Theory; T1.3; Tools and Algorithms for Collider and Nuclear physics phenomenology and theory; T1.4 Tools and Algorithms for multi-messenger and Cosmology on Gravitational Waves; T1.5 Tools and Algorithms for Complex Systems; T1.6 Tools and Algorithms for condensed matter; T1.7 Tools and Algorithms for Quantum Systems.

Main Activities: development of algorithms, codes, and computational strategies for the simulation of physical theories and models, towards pre-Exascale and Exascale architectures. Theoretical research projects in domains already using HPC solutions.

Milestones: M9-M15: Landscape recognition of the state-of-the-art and technological investigation on the opportunity of the CN infrastructure - report submitted with detailed plan of work and selection of specific case studies; M22-M26: report on first implementations and “ready for tests”; M25-36: results from testbed and benchmarking activities; final report and evaluation

WP2.2: **Design and development of science-driven tools and innovative algorithms for Experimental High Energy Physics** (INFN, UNIBA, POLIBA, UNIMIB, UNINA, ROMA1, UNITS, UNIBO, UNIFI, UNICT, UNIPD, UNICAL, UNISALENTO, UNIFE): T2.1 Innovative algorithms for Experimental High Energy Physics simulation, selection, data reduction, reconstruction and analysis; T2.2 AI inspired techniques for Experimental High Energy Physics.

Main Activities: selection, data reduction, simulation and reconstruction algorithms (either via explicit programming or large scale Machine Learning solutions) for HEP experiments (LHC, Future Colliders, KEK, IHEP, neutrino experiments...), with applications ranging from innovative triggers to distributed analysis techniques.

Milestones: M9-M15: Landscape recognition of the state-of-the-art and technological investigation on the opportunity of the CN infrastructure - report submitted with detailed plan of work and selection of specific case studies; M22-M26: report on first implementations and tests; M25-36: results from testbed and benchmarking activities; final report and evaluation

WP2.3: **Design and development of science-driven tools and innovative algorithms for Experimental Astroparticle Physics and Gravitational Waves** (INFN, INAF, UNIBA, POLIBA, UNIMIB, UNINA, ROMA1, UNITS, UNIBO, UNIFI, UNICT, UNIPD, UNICAL, UNISALENTO, UNIFE): T3.1 Innovative algorithms for Experimental Astroparticle Physics and Gravitational Waves simulation, selection, data reduction, reconstruction and analysis; T3.2 AI inspired techniques for Astroparticle Physics and Gravitational Waves.

Main Activities: data reduction, reconstruction and time cross-correlation algorithms, data selection and simulations of astroparticle and gravitational waves experiments, tools for cross-correlations and pattern recognition in multi-messenger physics, including novel implementations using techniques like Machine Learning.

Milestones: M9-M15: Landscape recognition of the state-of-the-art and technological investigation on the opportunity of the CN infrastructure - report submitted with detailed plan of work and selection of specific case studies; M22-M26: report on first implementations and “ready for tests”; M25-36: results from testbed and benchmarking activities; final report and evaluation

WP2.4: **Boosting the computational performance of Theoretical and Experimental Physics algorithms** (INFN, INAF, UNIBA, UNIMIB, UNINA, UNIBO, UNIFI, UNICT, UNITS, UNIPD, UNIFE): T4.1 Tools and guidelines for developing and porting heterogenous codes and algorithms on modern architectures; T4.2 Competence and training centre for heterogeneous computing.

Main Activities: porting of applications to GPUs and heterogeneous architectures (e.g., scalability of scientific codes and applications on GPU/CPU many-cores clusters, local and remote offloading, mission-critical algorithms on FPGAs, ...). The solutions and tools implemented during the project will be easily extendable to other scientific domains of the Centre and to the industrial partners in the Spoke; moreover, the personnel trained within the

Centre will help to spread and boost the application of HPC methodologies to Italian academic and industrial fields, for a comprehensive advancement of the Italian system.

Milestones: M9-M15: report on best practices for heterogeneous computing; M22-M26: first training opportunity; testbeds ready for users (initial handshake); user support in place; M25-36: results from testbed and benchmarking activities; final report on technologies, training and support system; white paper for use cases external to the CN

WP2.5: **Architectural Support for Theoretical and Experimental Physics Data Management on the Distributed CN infrastructure** (INFN, UNIMIB, UNINA, ROMA1, UNITS, UNIBO, UNIPD, UNIFE): T5.1 Support of the adaptation of existing applications on the data-lake distributed infrastructure, and via innovative computational models; T5.2 Competence centre for the design, implementation, and test of computing models.

Main Activities: support for the adaptation of existing applications on the data-lake distributed infrastructure, and via innovative computational models (for example sharing of gauge configurations in lattice field theories, long-term data preservation, streaming access to data, tiered storage solutions, …). The solutions implemented will be tailored to the needs of the scientific fields, easily extendible not only to the nearby scientific domains in the Centre, but also to all academic and industrial realities where needs to access distributed computing and large amounts of data exist. In particular, the industrial partners in the Spoke have expressed interest in using the same technologies for their specific use cases.

Milestones: M9-M15: documentation and best practices for data lake compliance; M22-M26: first training opportunity; virtual machines ready for use; user support in place; implementation of solutions for science-driven use-cases; M25-M36: results available from testbeds; final report on technologies, training and support system; final report with recap and white paper for use cases external to the CN

WP2.6: **Cross–domain Initiatives** (INFN, INAF, UNINA, UNIBO, UNIFI, UNIFE): T6.1 Optimization and adaptation of widely used cross-domain software package; T6.2 Techniques and tools for high intensity analysis (techniques for fast data access, AI-based tools, data interpretation tools). The same tools are expected to be exploited for the analysis of the data streams from the Mirror Copernicus constellation of satellites. An effort at this scale needs a proper methodology to ensure a positive conclusion of the activities.

Main Activities: optimization and adaptation of widely used software packages on the national Centre infrastructure, like Geant4 or FLUKA or generic high-performance techniques for data access/analysis; statistical and AI-based tools; data-interpretations tools. In the context of the Space Economy Italian Strategy, develop and deploy techniques to access, analyse and process the data from the Mirror Copernicus program, creating the conditions to enable radically innovative services. In particular, enable thorough and continuous observation programs for global and local processes, allowing external partners to operate a large variety of services, including the planning for emergencies, risks and resources.

Milestones: M9-M15: investigations and identification of package(s) to be modernized; report and detailed plan; landscape analysis of solutions for high intensity analyses; choice of proof(s) of concept to be realized; M24: pilot implementation and first evaluation of performance for the selected package(s); pilot implementation of the high intensity solution(s); M25-M36: benchmarking and testing activities executed; final report including performance gain assessment; white paper produced to the larger scientific and industrial community; showcase of the analysis techniques and white paper for external users

We plan to carry out the 3 years' research following a standard approach, in which 4 phases have been identified. For the "Scientific Work Packages" 2.1, 2.2, and 2.3 this means:

1. planning and identification: the first year of the project is dedicated to a landscape recognition for interesting state-of-the-art use case; its outcome is a work plan identifying the activities on which the core part of the project will be focusing - in particular, algorithms and services to be accomplished;
2. a realization phase, in which the actual development is performed via the staff/ hired personnel; the outcome is usable algorithms / services, documented (alpha/beta level) and ready to be tested on a larger scale;
3. a validation phase, in which the outcomes of the realization phase are verified in testbeds and proofs of concept, and benchmarked in order to assess their adherence to the specifications;

4. a wrap-up phase, in which results are analysed and consolidated in reports and white papers to be used as guidelines for similar use cases.

The three "technical Work Packages" WP 2.4, 2.5 and 2.6 follow a similar approach:

1. planning and identification: landscape recognition for best solutions for the realization of heterogeneous and portable code (e.g. software frameworks, compilers, programming models, …), for the integration of services into a data-lake infrastructure; cross domain software and services will be identified if appropriate. Moreover, solutions for handling user support, user fora, and training opportunities will be identified;
2. a realization phase, in which the services and the support systems are put into place, at least in alpha/beta phase. These include the testbeds to be used for benchmarking of scientific and industrial solutions, the user support system, the training opportunities.
3. a validation phase, in which experience on the supported services and codes are reported, to be used as a touch base before the end of the project.
4. a wrap-up phase, in which results are reported for executed activities, and are disseminated via white papers for future and external use cases.

The testbed and benchmarking ("validation") phase, in all WPs, will be executed partially via the "innovation grants" available via the project, in which the solutions developed in alpha/beta level will be tested together with industrial partners, on hardware either provided by the CN, or acquired via the same grants.

The Spoke 2 activities need a number of support and ancillary services, like a web portal, a ticketing / support system, and help in organizing activities like benchmarking and training. We intend to use a part of the "Open Calls" to this purpose, selecting professionalities from companies or academic institutions with experience on the subjects from previous projects.
The largest fraction of "Open Calls", still, is expected to be dedicated directly to academic institutions. This will allow it to extend the project beyond the direct involvement of partners and will allow it to operate in co-design and co-development mode. The current plan is to devote 11 open calls (~200kEur each), or almost 4 per scientific WP.
A summary of the Open Calls we aim to open, very early in the project (within month 6) is:

- 11 open calls for co-design and co-development, explicitly directed to academic partners; almost 4 per each of the scientific Work Packages 1, 2, 3;
- 2 open calls for engineering and development of domain scientific services;
- 1 open call for Spoke services, like a web portal, ticketing and support system;
- 1 open call for helping organizing training activities;
- 2 open calls for support to benchmarking and testing activities.

The last 6 open calls are tentatively directed towards private parties, but we expect a possible collaboration also with academic partners with a proven experience.

Spoke 3

The Spoke 3 (named internally as ACO-S) follows a detailed work plan encompassing and coordinating all the activities of the project. That is required in order not only to drive the technical and scientific results foreseen, but also to ensure the quality of those results and the achievement of maximum impact and sustainability, via the engagement of stakeholders (i.e. the largest international projects where the partners are involved) and the introduction of innovative business perspectives.

The implementation of the ACO-S will have a duration of 36 months and will be organized around 5 Work Packages.

WP3.1: HPC Codes Enabling and Optimization. This WP selects a number of codes that require intensive computational resources to face the next generation of scientific challenges and performs their redesign, reimplementation and optimisation in order to effectively exploit state-of-the-art HPC solutions. [Leader: INAF; Participants: INFN, UniTOV, UniTO, UniTS, SISSA, SNS]

T1.1: Selection, Analysis and testing of codes, algorithms and programming models; development plan. Based on interactions with the scientific community initiated prior to the start of this project addressing the identified key research areas, a number of codes will be selected at the beginning of the project. Possible candidates among the astrophysics and astroparticle community are PLUTO, OpenGADGET, RAMSES, AREPO, XMAP, TOAST, Pinocchio, COMPLETE and 21cmFAST, among others. The performance-relevant software components and libraries will be identified. Performance models of the selected codes will be developed and validated. Representative parts of these software components will serve as prototypes that will be later (T1.2) rewritten in the form of pilot applications, in order to test out possible performance improvements and support to large data volumes, due to software refactoring and algorithm reengineering.

T1.2: Software Development, Refactoring and Optimisation according to the plan setup. Computational intensive and performance critical parts of the codes will be rewritten according to the plan setup and the design developed in T1.1. An "agile" working methodology – continuously reviewing the achieved algorithmic performance and scalability – will be adopted, in order to promptly overcome difficulties, further improve the developed software, adopt new available solutions and support possible novel scientific requirements. The development teams will be composed of domain scientists, computer scientists and HPC experts, who will work synergically. Rewritten codes and software components will be documented, and managed using versioning systems (tools managed and deployed by WP6).

T1.3: Integration, Verification and Validation. The newly developed performance critical software components will be integrated with the respective community codes. This will take place upon successful optimization of the pilot application, by re-inserting original code components while adhering to the newly developed optimized framework. The community code will be tested and verified by checking results against reference solutions and/or known use cases with the supervision of the domain scientists. The achieved performance will be reported.

WP3.2: Design of innovative Algorithms, Methodologies, Codes toward Exascale and beyond. This WP identifies innovative algorithms and methodologies upgrading their capability to exploit, and scale on, the exascale and post exascale architectures, reintegrating the resulting improved features in codes, workflows and pipelines. The energy impact will also be specifically considered. [Leader: INAF; Participants: SISSA, SNS, UniTS, UniCT, UniTO, UniTOV, INFN]

T2.1: Science cases definition, algorithms identification, parallelism level assessment and profiling. Based on interactions with the scientific community, a number of algorithms to be innovated will be selected at the beginning of the project. The first activity of this task will be to define a complete set of scientific cases that will be used to prepare the relevant stress tests (e.g. soft and hard scaling), in order to have standard benchmarks for the various classes of the ACO-S algorithms. The second activity of this task will be to assess the parallelism level of each ACO-S algorithm (multi-threading at single-node parallelism and multi-node communication). The third activity of this task will be to profile the ACO-S algorithms and identify their bottlenecks and the modules that need improvements and re-design to perform efficiently on upcoming Exascale architectures.

T2.2: Algorithms Co-design and methodologies to scale-up the capabilities of the algorithms and to find new innovative solutions. Selected portions of the ACO-S algorithms will be rewritten according to the plan designed in T3.1. This task is devoted to scale-up the capabilities of the algorithms and to find new innovative solutions also in terms of methodologies. We will investigate innovative workflow and data models to efficiently run and scale on exascale platforms. We will improve the implementation of each critical algorithm to eliminate bottlenecks and inefficiencies, to enhance the fine-grained (i.e. at node-level) and coarse-grained (i.e. multi-node level) parallelism. Algorithms and data layout different than the legacy ones may be in order or can be tested and profiled, depending on individual cases.

T2.3: Design of new architectural solutions aimed at the exploitation of post-exascale infrastructures (GPUs, FPGAs, Vector accelerators, NVM, HBM, ARM). This task aims to: the development (or, in some cases, the enhancement) of the offloading for computation-intensive kernels to diverse accelerators (GPUs, FPGAs and vector accelerators); the enhancement of code quality and data structures to boost the vectorization of computationally-intensive loops; the development of experimental kernels that profit of emerging HBM and NVM memory (e.g. to reduce the imprint of checkpointing) technologies; testing and development of a shared library with the best implementations of common algorithms, when possible (e.g.: direct/tree/fastmultipole and multigrid Poisson solvers for gravity computation in Lagrangian and Eulerian codes, respectively).

T2.4: Algorithms and methodologies integration into new big-data analysis applications. The aim of this task is to integrate algorithms and methodologies into new big-data analysis applications and numerical simulations to enhance their capabilities. Define specific metrics to verify the improvements. Test the codes on exascale platforms to verify the improvements and the capabilities of the new applications. Extrapolate a path towards post-exascale infrastructures. The community codes will be tested and validated against known use cases. The achieved performance will be reported.

WP3.3: Big Data Analysis, Machine Learning and Visualization. This WP develops a prototype framework of data analysis, based on Machine Learning (ML) and Visualization tools exploiting diverse computing platforms and combining them with exascale applications. The framework will be tailored to the ACO-S community, identifying the use case best suited to tackle high-performance visualization tools and ML techniques. Furthermore, it copes with observational data coming from challenging large experiments (e.g. LOFAR, MeerKAT, SKAO data challenges, Large Simulations Datasets, CTA etc). Leader: INAF; Participants: SISSA, SNS, UniTS, UniCT, UniTO, UniTOV, INFN, ISP]
T3.1: Requirements from AAA community. It will assess the requirements of the ACO-S and assess corresponding solutions in terms of Machine Learning and Visualization, exploiting the functionalities available in selected tools and/or the development of new functionalities. It will also assess the requirements of big simulations and observational data, focusing in particular on the use case of the SKAO and its precursors and pathfinders (in particular LOFAR, MeerKAT and ASKAP) to address the comparison between theoretical and observational data. The SKAO use cases have been identified as a prime example of outstanding science, producing enormous volumes of complex data requiring innovative solutions to be processed.
T3.2: Innovative Machine Learning. This task is in charge of designing, implementing and evaluating ML components in ACO-S pipelines. Targets include off-line processes for the transformations and enrichment of data, such as classification, segmentation, reduction, and emulation. For each component targeted, the task will provide an adequate ML model, together with an efficient and scalable implementation. Finally, its performance (both in terms of task outcome and efficient computation) will be assessed, to produce a final evaluation on the advantages and disadvantages of the produced ML solutions. Together with industrial partners the task will address *a)* search of patterns using network analysis in very large, noisy datasets, *b)* anomaly detection, *c)* techniques of privacy preserving, deployed in federated learning services (Cybersecurity)
T3.3: HPC/Cloud Visualization Services. The main emphasis will be on fast and interactive visualization, using services at the HPC facilities near the data, but with high-speed delivery of the results to the user's desktop. Such tools will be tested and integrated on the existing open-source database as https://smart-turb.roma2.infn.it/ which will be further developed during this project such as to host a large quantity of geophysical and astrophysical datasets. The ultimate goal is to provide visualization capabilities that are fully interactive and can access and visualize large data sets at a remote repository.

WP3.4: Big Data Management, Storage and Archiving. WP5 will analyse, explore, standardize and store data of different collections, characterizing them in the appropriate way in order to facilitate the respect of the FAIR principles and enable users to benefit from an innovative storage and archiving platform. A distributed archive infrastructure with hot and cold storages and proper access tools will be implemented or customized by existing ones, respecting the interoperability directives (RDA, IVOA, ...). The requirements of the storage in terms of identification of data collection characteristics (data models, data dimension, accessibility methods, cold or hot storage etc...) will be analysed, defining and implementing suitable technical solutions. [Leader: INAF; Participants: INFN, UniTS, UniTO, UniTOV, ISP]
T4.1 Data management, Standardization and Interoperability. This task is intended to gather all the information related to collect, organize, analyse, preserve, curate and share data. It will be documented in light of the analysis of relevant use cases. The general information related to the data collection content will be exposed using the most current and updated standards available in the international fora like RDA, IVOA, Open Access, so they will be interoperable with similar resources in order to reach the data FAIRness.
T4.2 Local and Distributed Long and Short-term Storage Optimization. This task will be carried on in synergy with WP4 and will prototype the protocol stack for data transfer both externally (from data providers) and internally (from the storage space to the data-intensive computation area), by using as study implementation, the existing hardware infrastructures. Transfer protocols and data management systems will be tested and configured to optimize the band occupancy and parallel file systems will be tested and configured to guarantee writing on device and computing performances.
T4.3 Archive and Repository Definition and Implementation. This task is related to the software integration of the Distributed Data Archive Framework and the access tools already implemented or customized by the project. It will compose a framework where users can perform data storage, retrieval and exploitation. The User Space (storage part) implementation will be fully VO compliant. A web interface will give access to remote resources that will adopt appropriate libraries and tools for data exploitation. The access to the storage resources

will allow authentication of users for private data and will be performed with standard and well consolidated authentication and authorization (AAI) tools, fully compliant with the Single Sign On paradigm.

WP3.5: HPC Services and Access. The WP is in charge of managing, maintaining and deploying an integrated environment providing the tools for the efficient development of the work described in WPs 1-4. [Leader: INFN; Participants: INAF, UniTOV, ISP]
T5.1 Collaborative software development, management and continuous integration platform. The implementation of a fully automated Continuous Integration/Continuous Delivery solution significantly speeds up the process of software delivery and, above all, allows its repetition. Further, the delivered software version is always being built and implemented in the same, fully automatic way which minimizes the risk of making mistakes. Support to the future community will also be taken into consideration with tools for ticketing, bugs/features requests and discussions groups.
T5.2 Design, implementation and validation of an interoperable service architecture. The task aims at developing a number of data services that all follow the same rules, as defined by the FAIR and VO standards, implementing interoperability. These data services are software tools and web portals that follow such common rules and can fetch data from these services and know what to do with the data when they get it. The advantage of standardized services is that they deploy a common, single interface to data and tools and exploit standard formats, defined in tasks T1.1 and T4.1 to access data products.
T5.3 ML and Visualization enabling services deployment and HPC/Cloud integration. This task will focus on the efficient and optimized deployment of ML and Visualization services developed in WP4 for operational delivery to the end users. In particular container-based delivery will be adopted to exploit the integration of the HPC backend facilities mainly devoted to computing demanding processing with on-demand cloud resources for frontend, including web-based modules. The task will develop prototypes of new services or new components/functionalities of existing services, prototypes of virtualized services, and investigate federated learning for large AI workflows. Demonstrators and use cases will be developed according to the Spoke industrial partners.

Milestones

- M1-M8: Start of the activities for the definition of the science cases in WP1-T1.1, WP2-T2.1; activity plan definition for different WPs
- M5-M8: Advancements on different WPs; definition of science cases
- M9-M15: Definition of science cases and of the relevant algorithms, models and codes. Review and finalization of the WPs tasks requirements.
- M13-M18: 3 Demonstrators development for testing enabled codes on innovative platforms, innovative codes and algorithms green computing test, Machine Learning and Visualization prototyping on Exascale systems applied to industrial applications; 1 Demonstrator development for activities related to WP3 (Machine learning/visualization).
- M22-M26: 1 Demonstrator development for activities related to Machine Learning for industrial applications; 1 Demonstrator for GPU+CPU synergy into an hybrid high-bandwidth clusters.
- M25-M36: WP3 - design and implementation of energy-saving algorithmic solutions in the codes for astrophysical simulations; WP4 - Finalization and delivery of tools based on Machine Learning techniques for CMB simulation and component separation data analysis. WP5 - Preliminary implementation of an evolutive prototype for Long and Short term Storage Optimization and data transfer from and to the computational clusters WP6 Implementation of an interoperable service architecture. Spoke 3 - Completion of all WPs activities

Spoke 4

The Spoke's activities are structured into six Work Packages: the first one (WP4.0, with a description specific for the spoke and hence reported here) devoted to the management and scientific coordination of the activities, including the administrative and financial management, and the internal and external communications; the other five work-packages (WP4.1–WP4.5) are dedicated to conducting the specific scientific activities aimed at achieving Spoke's objectives. Next Figure provides a schematic representation of the interlink between the WPs, while the rest of the section describes the WPs, their objectives and tasks in more detail, including a concise description of the methodological approaches.

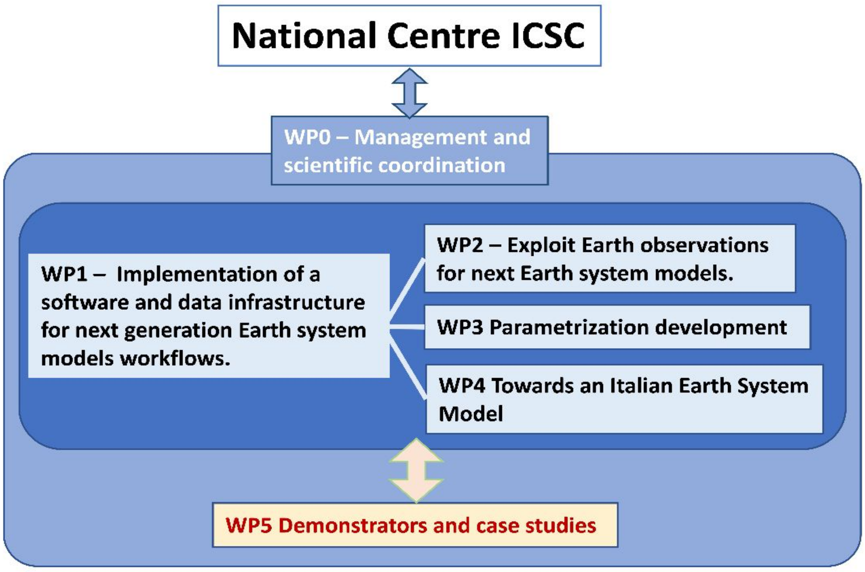


Figure 13 Interlinks between the WPs in Spoke 4

WP4.0 Management and scientific coordination Lead: CMCC, Co-Lead: CNR–ISAC; Participants: CNR-IMATI, CNR–ISMAR, ENEA, FBK, OGS, UniSalento, UniTrento.

Objectives: ensure sound, effective and efficient management and coordination towards achieving its objectives, in compliance with the Grant Agreement (GA), Consortium Agreement (CA), and best management practices.

WP4.0 will assure:

- Intensive, flexible, and open dialogue among partners
- Effective decision-making on technical and organizational issues
- Coordination of the management structures and effective risk management
- Compliance with the PNRR and CN–HPC administrative and reporting requirements

***Task 0.1*: Administrative, data and financial management.**

This task will implement an efficient exchange of administrative information and in particular financial documents and reports among partners and with the ICSC National Centre board. Before the start of the project, the Spoke Coordinator will draft a Consortium Agreement (CA). Financial guidance will be provided to all partners through a document, synthesizing the rules, which will be made available to all partners through a project wiki. This task will also ensure the preparation and the issue of the tender for purchase of a supercomputing infrastructure to be installed at CMCC (M5-M8), of the open calls for contributions from third parties (Milestones M1-M8 and M9-M15) and it will coordinate the preparation of the implementation plans for the innovation grants aimed at raising the readiness level of the most promising Spoke's developments. It will monitor the effective and timely issuing of the PhD bursaries and calls for staff recruiting as foreseen in Milestones M9-M15 and M1-M8.

Contributes to Milestones M9-M15, M5-M6, M5-M8, M9-M15, M13-M18.

***Task 0.2*: General management and risk management.**

Management of the Spoke will be provided, ensuring a continuous control and coordination of the various activities through the following organizational structure: a Spoke's Steering Group (SSG) will be formed by the representatives of the affiliated institutions, and it will be chaired and co-chaired by the Spoke leader and co-leader respectively. This task will ensure the timing release of all milestones and deliverables in the required form, be they documents or reports, models or data documentation. The task will organize frequent (monthly) review meetings of the SSG to assess the project status, monitor the activity progress and actions. ensure the coordination among the different WPs and the achievements of goals and milestones. The task will, furthermore, organise regular meetings with the ICSC National Centre board and the annual Spoke assembly. Regarding Risk Management, the process will be based on identification of best practices and experience and, through a circular lessons-learned process, will include (i) Monitoring to identify emerging issues that will be collected in a risk register; (ii) Identification of corrective actions: according to the gravity and urgency of the identified problems, actions will be identified to tackle the different issues, and will also be recorded in the risk register; (iii) Internal

audits to monitor if the management system matches its processes and weather new and improved techniques become normal practice. All risk monitoring tasks will be addressed within the regular internal review meetings.
Contributes to Milestones M1-M8, M09-M15, M13-M18, M22-M26.

***Task 0.3*: Dissemination and Communication Activities.**
This task will deal with project communication, including internal communication (that will rely on an interactive wiki web-tool), external communication (conducted via a public website). The wiki will provide areas to discuss issues, exchange information and documents necessary to achieve the Spoke's objectives. Templates, manuals and guidelines needed for an efficient project reporting will be made accessible through this wiki, together with a contact list. A website will be designed to satisfy the communication requirements and the dissemination of the project results, accessible and providing quick response with different devices: desktops, tablets and smart-phones and according to the standards required by the most used Operating Systems. The Content Management System will be realized with open-source software. The content of the website will be organized in sections dedicated to: (i) institutional information about the project, according to the guideline provided by the funding institutions; (ii) aims and outcomes of the project; (iii) news and information about project-related events; (iv) results of user engagement meetings and feedbacks; (v) contacts and relations with media and stakeholders. E-training material, including tutorials to disseminate and exploit the Spoke's tools, and communication products will be made publicly available through the website. A logo will be designed to give a defined and recognizable image to the project. The task will also guarantee the Industrial partners engagement in the project activities by means of dedicated meetings. Finally, dissemination and Communication activities will also benefit from the involvement of some partners of the Spoke to the training program "HPC Training and Research for Earth Sciences" (HPC-TRES), promoted by OGS and CINECA and co-funded by the PRACE-Italy national research infrastructure.
Contributes to Milestones M1-M8, M9-M15, M22-M26, M22-M26.

WP4.1 Implementation of a software and data infrastructure for next generation Earth system models (ESMs) workflows [M1-M36] Lead: UniTrento; Participants: CMCC, CNR-IMATI, CNR–ISAC, CNR–ISMAR, ENEA, FBK, OGS, UniSalento, UniTrento)

Objectives: Development of an integrated HPC-based and FAIR-enabled digital infrastructure able to integrate into the same environment a wide spectrum of tools, ESM components, and applications, but also data and workflows, to support climate and Earth System scientists in terms of modeling, as well as handling, processing and analysis of large volumes of climate data. The work package will address data and computational challenges (from code optimization to data management), including end-to-end aspects (workflow and provenance management) towards a sustainable ESM and Data Science (ESM&DS) software stack. The activity will benefit from and will be strongly integrated with the resources and software platform made available by the Centro Nazionale ICSC.

Task1.1: **Design and setup of the software and data infrastructure.** This task will design a software and data infrastructure able to merge/integrate the most appealing and promising software tools able to shape the proposed vision. The activity will be performed in close synergy with the architecture design defined centrally by the Centro Nazionale ICSC. Besides, the task will also be responsible for the initial setup of the digital infrastructure on the resources made available by the ICSC National Centre. This activity will cover preliminary setup/deployment tasks and will leverage a close collaboration with the system admin staff of the Tier0 resources.
Contributes to Milestones: M5-M8

Task 1.2: **Code harmonization and release management [M6-M36]**. This task will integrate existing packages for the coupling, set up tools for collaboration and versioning of the codes. This task will also be devoted to software release management. It will leverage on services provided by the Tier0 to ensure the proper software lifecycle management.
Contributes to Milestones: M13-M18, M22-M26, M25-M36.

Task 1.3: **End-to-End workflow and provenance management.** The main goal of this task is to develop an end-to-end support to automate the execution of the ESM workflow on the resources made available by the Centro Nazionale ICSC. This task will allow an effective, and optimized orchestration of a large set of tasks, thus efficiently streamlining the management of the ESM workflows. The management of these workflows will be mainly conducted by means of graphical user interfaces (GUI). Such support will seamlessly combine model and data components and it will be able to manage provenance and system logs to keep track of all the activities performed in the workflows by the users, for both in-flight and retrospective analysis.
Contributes to Milestones: M13-M18, M22-M26, M25-M36.

Task 1.4: **Data Science and learning software infrastructure.** This task aims to develop a data science and learning software infrastructure, by following a Data Space approach according to the vision promoted by the

European Data Strategy. It will integrate and provide solutions at the intersection of HPC, big data and data science able to support the analysis of large-scale datasets (both data-driven and data-intensive workloads). State-of-the-art solutions will be considered as the starting point and properly adapted and extended to fulfil the requirements coming from climate scientists. Both batch and interactive data analysis will be supported; programmability will be a key requirement to support application development as well as exploratory data analysis.
Contributes to Milestones: M13-M18, M22-M26, M25-M36.
Task 1.5: **Optimization of code, data streaming and diagnostics to fully exploit hybrid pre–exascale HPC machine.** This task will deal with the optimization of codes, support for data streaming and diagnostics to fully exploit hybrid pre–exascale HPC machines. Data streaming and in-situ approaches can enable enhanced High Performance Data Analytics (HPDA) strategies supporting the definition of more structured diagnostics. In-situ/in-transit techniques will be designed and applied to allow on-the-fly diagnostics computation and visualization of ESM results during the simulation run. A software interface will be developed to address (ESM agnostic) data analytics features and quicklook capabilities to support scientists for fast and real-time evaluations. Well known or newly developed HPDA solutions will form the analytics backend to provide diagnostic computation capabilities on hybrid HPC architectures. Moreover, within this task, the unstructured ocean model component and the computational demanding diagnostics will be optimized to efficiently exploit the HPC architectures by means of the OpenMP/MPI parallel approach, with the adoption of parallel I/O solutions and with the exploitation of high-performance numerical libraries.
Contributes to Milestones: M13-M18, M22-M26, M25-M36.

WP4.2 Exploit Earth observations for next generation ESMs. [M1 - M36] Lead: CNR–ISMAR; Participants: CMCC, CNR-IMATI, CNR–ISAC, ENEA, FBK, UniSalento, UniTrento
Objectives: Improve exploitation (including AI) of large observational datasets to identify and analyse gaps in ESMs at different time and spatial scales; improve data assimilation for climate applications. To this aim, the activity of this WP will be strongly linked to the WP1 workflows regarding the access/exploitation of shared data and computational resources as well as the integration of new software components provided by WP2 into the overall infrastructure.
Task 2.1 **Observational data collection**. This task will be responsible for the collection, quality control, organization and sharing of the observational data required for data assimilation, process parametrization development and model validation. In particular, it will ensure that the collected data will fulfil the requirements identified in Milestone M2.1 and that they will be stored and made easily accessible to all project partners in a proper format. In this task the data will be managed following the principles and recommendations defined and adopted by the Spoke and by ICSC National Centre Data Management policy.
Contribute to Milestone M1-M8, M17-M22, M22-M26.
Task 2.2 **Development of novel, including AI, algorithms to optimize the exploitation of comprehensive observational datasets** (including satellite and in situ observations). This task aims at interfacing advanced statistical and AI techniques with Earth system geophysical observations in order to extract relevant information for climate and model applications. The activities of this task include: (i) observation-based statistical downscaling techniques (ii) data-driven methods for downscaling, based on Generative Adversarial Network (GAN) considering physics-informed constraints (iii) features extraction generation, including machine learning approaches to address most of the processing steps required to assimilate the observed data into numerical models. Examples include: the use of observation errors to represent varying levels of uncertainty, observation operators to map from regular model grids to irregular, sparse observations, and the use of physical model components or layers to impose physical constraints on machine-learned networks. (iv) Identification and characterization of poorly observed extreme events (e.g., intense convective events, hail, marine heatwaves). Finally, a machine learning based approach for predicting sea-level in the short term by exploiting Recurrent Neural Network (RNN) and a huge amount of data derived from a network of sparse and distributed tide gauge devices will be investigated. Contributes to Milestone M1-M8, M17-M22, M25-M36.
Task 2.3 **Exploit new and existing observations to identify and analyse gaps in climate/Earth system models at the different time and space scales**. This task will identify processes that are poorly modelled in the ESMs with the aim to develop observational constraints to help improve the simulation of these processes. The present-day mean climate state and historical variability will be evaluated in the most advanced Earth Observational datasets and as simulated by state-of-the-art ESMs using a range of existing diagnostic tools (e.g. ESMValTool) and where necessary developing new tools. Particular focus will be in the analysis and quantification of couplings and feedbacks among components of the climate system that could lead to abrupt climate transitions. The analysis

will include features and change detection, including critical thresholds characterizing regime shifts, also using novel methodologies (e.g. A.I.).
Contributes to Milestone M1-M8, M17-M22, M22-M26, M25-M36.
Task 2.4 **Enhanced data assimilation for climate applications: towards coupled assimilation algorithms**. This task will be responsible for advancing and applying coupled data assimilation schemes in the context of climate predictions and Earth System model applications. More specifically, coupled data assimilation systems among the atmospheric, land and ocean components able to simultaneously control all the resolved scales, to propagate the net positive information across the climate system components and create an initial state that minimises the coupled initialisation shocks will be developed. In particular, we envisage two main developments: (i) coupled land-atmosphere data assimilation, exploiting LAI and soil moisture satellite observations; (ii) coupled atmosphere-ocean data assimilation in both a global and a regional Earth System model, with a coupled variational scheme that uses hybrid ensemble-variational covariances. Besides, The full and efficient exploitation of the HPC infrastructure requires a revision and optimization of the current data assimilation codes. In this task, the Ocean-VAR data assimilation model, enhanced with the barotropic operator, will be optimized for pre-exascale, GPU-based parallel architectures. The methodology used will be based on the adoption of OpenMP/MPI parallelization which ensures performance portability among heterogeneous architectures.
Contribute to Milestone M1-M8, M17-M22, M22-M26, M25-M36.

WP4.3 Parametrization development [M1 - M28] (Lead: CNR–ISAC; Participants: CMCC, CNR–ISMAR, ENEA, FBK, OGS, UniSalento, UniTrento)
Objectives: The principal aim of WP3 is to develop a set of novel parameterizations able to realistically represent physical and biophysical processes that are required for Earth system modelling. WP3 will ensure: (i) exploitation of the observational constraints from WP2; (ii) design of parameterizations that can be seamlessly integrated at the different time and spatial scales; (iii) flexible characteristics adaptable to different modelling frameworks; (iv) adhere to the optimization procedures implemented in WP1.
Task 3.1 **Land-surface Parametrizations.** Design and development of land parameterizations based on observational constraints: the unprecedented observational knowledge from WP2 will be exploited to overcome limitations in modelling hydrological and biophysical processes over land and to obtain well constrained land parameterizations. This set of improved land-surface parametrization will include a soil hydrology and land cover (including urban texture) representation and a realistic parameterization of vegetation effective-cover, surface emissivity and albedo. The potential benefit of using additional external parameters available from satellite platforms for a more detailed description of the urban areas' features (e.g. H/W, percentage of impervious area in the grid cell) will also be investigated.
Contribute to Milestone M1-M8, M9-M15, M13-M18, M25-M36.
Task 3.2 **Atmospheric and oceanic Drag**. Design and implementation of new parameterizations for different sources of earth surface drag processes affecting the atmosphere and the ocean. Develop a scale-aware orographic drag parametrization for the atmosphere (which will include both turbulent frictional effects, orographic and non-orographic gravity wave drag and form drag due to orographic flow-blocking). Non-elliptical mountain shape and flow directional effects will be taken into account, overcoming present-day limitation of subgrid orography representation. For the ocean, a parameterisation will be implemented of the over-water drag mediated by sea-spray aerosols generated under strong wind conditions. The results will be tested against state-of-the-art reference datasets and will be included in at least one GCM taking part in WP4.
Contribute to Milestone M1-M8, M9-M15, M13-M18, M25-M36.
Task 3.3 **Machine learning applied to physical processes.** Design of machine learning approaches for the parametrization of physical processes, leveraging explainable AI and physics-informed machine learning and HPC-enabled large scale data understanding and processing, which try to blend machine learning with physical knowledge to achieve solutions that are physically more consistent. For example, development of a new machine-learning based radiative transfer (RT) parametrization for use in climate models.
Contribute to Milestone M1-M8, M9-M15, M13-M18, M25-M36.

WP4.4 Towards an Italian Earth System Model. [M1 - M28] (Lead: CMCC; Participants: CNR–ISAC, CNR–ISMAR, ENEA, OGS, UniSalento, UniTrento)
Objectives: development of community global and regional ESM components; multi-scale, structured and unstructured, ocean modelling from the open ocean to the coast; development and implementation of modular regional ESMs interchangeable in their components and coupling with global ESMs.

Task 4.1: **Towards an Italian ESM: development of community Earth system components.** This task will identify and implement modeling components to be integrated to build new ESMs that will be shared by the Italian climate modeling community, with the aim of improving its ability to produce high quality climate data and information and strengthen its international competitiveness. The ESMs components will exploit the improved parametrizations developed in WP3 and will include new components such as river routing schemes, ocean waves, ocean biogeochemistry, atmospheric chemistry. The coupling interfaces will be implemented in the different components to allow integration in the ESMs using state-of-the-art couplers, such as, for example, OASIS (https://oasis.cerfacs.fr) or platforms such as the Common Infrastructure for Modeling the Earth (CIME, https://github.com/ESMCI/cime). Furthermore, the new ESMs and model components will be shared through the digital infrastructures and tools developed in WP1 and will exploit the new diagnostic tools developed in WP2. A detailed description of the global ESMs that will be developed, including the identification of the set of components, design of the physical coupling and specification of the coupling environment will be provided in Milestone M6-M8. Contributes to Milestone M5-M8, M17-M22, M25-M36.
Task 4.2: **Development of a global version of a community ocean model, including the coupling with a sea-ice model**. This global ocean–sea ice model will represent the oceanic component of a possible future Italian Earth system model. The global ocean–sea ice components of the Earth system model will be developed exploiting the capabilities of unstructured meshes, with the aim to operate on different spatial scales and serve the need to represent global coastal scales in continuity. The final goal is to develop an ocean-sea ice model able to operate from very large open-ocean scales to the coastal regime. The unified ocean model implementation will be based on the open-source community model SHYFEM, addressing 3 main characteristics: robustness and accuracy of the numerical formulation, efficiency of the program, and usability. The development of the global SHYFEM will benefit from the combination of (i) the long and robust experience gained by the task partners with the NEMO model used with global and regional configurations, and (ii) the effort devoted by the task partners during the last years to improve and optimize the unstructured-grid SHYFEM model, including its parallelization. The sea-ice model component will be chosen in such a way to be compatible with the coupling interface and match the mesh of the ocean component. The sea-ice (thermos-)dynamics model development will build on the long experience gained by the task's partners with sea-ice models such as CICE and SI3, the current sea-ice model developed within the NEMO Consortium. Contributes to Milestone M5-M8, M17-M22, M25-M36.
Task 4.3 **Design, development and implementation of modular regional ESMs easily interchangeable in their components and in their coupling with global ESMs**. This task will design and develop a new coupled Regional ESM (Reg-ESM) that will include several interchangeable model components, such as atmosphere (e.g. RegCM and WRF), river routing (including the presence of human infrastructures), ocean (e.g. MITgcm, NEMO and SHYFEM), wave, atmospheric chemistry, ocean biogeochemistry (BFM). The different components will be integrated in the RegESM using state-of-the-art couplers, such as, for example, the Earth System Modeling Framework (ESMF, https://github.com/esmf-org/esmf), the OASIS coupler (https://oasis.cerfacs.fr) or platforms such as the Common Infrastructure for Modeling the Earth (CIME, https://github.com/ESMCI/cime). The new Reg-ESMs will be implemented and tested in different configurations and over different domains, such as, for example: in the Northern Adriatic Sea area, with the aim to assess their feasibility to simulate the carbon cycle; in the Med-CORDEX (https://www.medcordex.eu) domain, following the recent phase-3 baseline protocol defined by the Med-CORDEX program, and in the Euro-CORDEX (https://www.euro-cordex.net) domain, with the aim to provide reliable CMIP6 downscaling for the Euro-Mediterranean region. These Reg-ESMs configurations will represent the basis of climate modeling tools that will produce the Italian contribution to the international MedCORDEX and EuroCoRDEX programs. Contributes to Milestone M5-M8, M17-M22, M25-M36.

WP4.5 Demonstrators and case studies. [M1 - M36] (Lead: Uni Salento; Participants: CMCC, CNR-ISAC, CNR-ISMAR, ENEA, OGS, UniTrento)
Objective: use of the tools developed in WP1-4 to provide new knowledge on the expected evolution of the relevant climate impact drivers with a focus on the Mediterranean region considering multiple time scales ranging from sub-seasonal predictions to long term projections. WP5 will provide information on the predictability of droughts, heat waves, extreme precipitation, storms and on their future frequency depending on time and projected socioeconomic global changes (Shared Socioeconomic Pathways). It will further explore the regional characteristics of sea level rise and coastal floods in climate projections. All the demonstrators and case studies will be codefined with the industrial partners involved in the Spoke.
Task 5.1: Simulations and predictions of the Italian and Mediterranean climate.

A series of demonstrative simulations will be designed and conducted following the protocol (see Milestone M9-M15) defined with the contribution of the industrial partners involved, in order to best meet the needs of users. In particular, these simulations will provide a new set of high-quality climate change projections for the Mediterranean region and the Italian Peninsula, allowing to explore the possible changes in the Mediterranean climate, improving the quality of climate related information and data for this region, reducing the uncertainty on the evaluation of important parameters such as the future sea-level rise in the Mediterranean basin. Furthermore, case studies will be defined to explore the predictability of the Mediterranean climate at different time and spatial scales. Contributes to Milestone M22-M26 and M25-M36.
Task 5.2: Evolution and shifts of extreme events.
In this task we will specifically focus on the analysis of the possible changes in the occurrence (frequency and intensity) of extreme events in the Italian Peninsula and in the Mediterranean basin. Using the simulations performed in Task 5.1, the characteristics of droughts and terrestrial and marine heat waves will be explored in the Mediterranean basin, investigating also their interlink and teleconnections with the general circulation and their possible predictability at different time scales. Furthermore, in this task a set of convection permitting models (CPMs), implemented over the Italian Peninsula domain, will be conducted and used to investigate: (i) the performance of these models over the historical period (using ERA5 as boundary conditions); (ii) the evolution and possible shifts of subdaily extreme patterns, including an assessment of the uncertainty associated with the results. This activity will demonstrate the added value offered by the CPMs for climate simulations in the Italian area and, in particular, their potential to reduce the uncertainty associated with the assessment of changes in extreme events. Finally, Machine Learning approaches for automating the detection of extreme events (e.g., cold and heat waves) in the Mediterranean region will be considered. Contributes to Milestone M22-M26 and M25-M36.

Spoke 5

For each WP, the Work Plan will be developed according to the activities and deliverables outlined below.
WP5.1 Dynamical hazard scenarios assessment
Activities
Task 1.1: assessment of the driving hazard factors for the Italian territory
Task 1.2: Modelling the impact of hazard factors on the Italian ecosystem
Task 1.3: High resolution dynamical hazard scenario maps of natural disasters threatening the Italian territory and ecosystem

Milestones
M9-M15 – 20% Release of the analysis concerning the hazard factors
M22-M26 – 40% Release of the hazard models
M25-M36 – 40% Release of the maps

WP5.2 Vulnerability inventory of elements at risk
Activities
Task 2.1 Assessment of the elements at risk
Task 2.2 Modelling the impact of risk factors
Milestones
M9-M15 – 20% Evaluation of the Italian situation and the elements at risk
M22-M26 – 40% Evaluation of the different risk factors
M25-M36 – 40% Integrated model of the different Italian risk factors on the elements at risk

WP5.3 Modelling of disaster-inducing processes
Activities:
Task 3.1 Gathering multi-source data and HPC numerical codes for the characterization of disaster-inducing processes
Task 3.2 Modelling disaster-inducing processes
Task 3.3 Integrating modelling and decision making
Milestones
M9-M15 – 20% Data collection
M22-M26 – 40% Data modelling of disaster-inducing processes

M25-M36 – 40% Elaboration of large-scale scenarios for the analysis of Italian resilience to disasters

WP5.4 Modelling of environmental disasters
Activities:
Task 4.1 Remote sensing and proximal sensing models for environmental disasters
Task 4.2 Models and tools for structural health monitoring and assessment of risks and impacts induced by environmental disasters
Task 4.3 Active surveillance for effective and rapid emergency management/response
Milestones
M9-M15 – 20% Data collection and management
M22-M26 – 40% Learning frameworks for vulnerability and risk
M25-M36 – 40% Decision support systems for emergency management/response

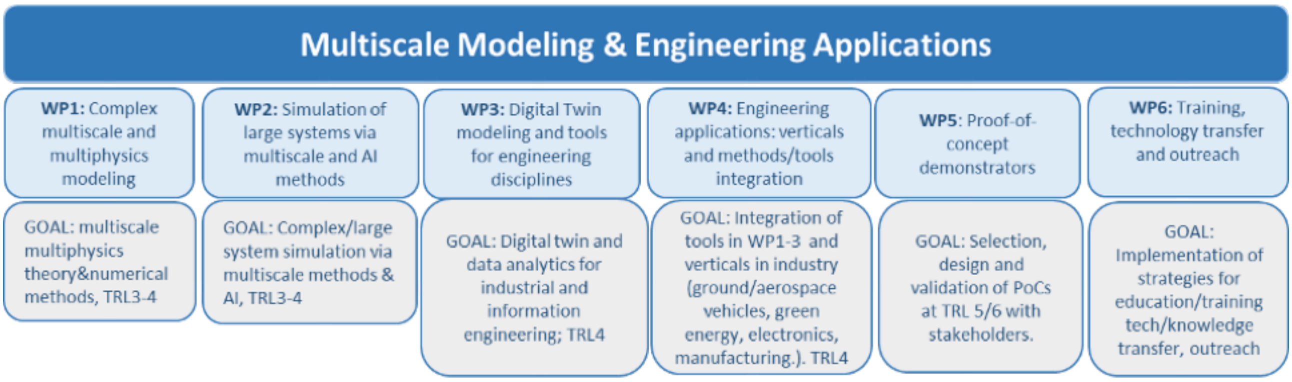
WP5.5 Multi-hazard modelling and analysis of damages and losses
Activities:
Task 5.1 Typological-mechanical, multi-scale analytical and numerical modelling and meta-modelling of response parameters.
Task 5.2 Multi-hazard probabilistic estimation of engineering and geophysical response parameters.
Task 5.3 Multi-hazard probabilistic analysis of losses (physical, economical, societal).
Milestones
M9-M15 – 20% Overview of models and methodological framework for computational workflows
M22-M26 – 40% multi-hazard models and analyses
M25-M36 – 40% Integrated models for physical, economical, societal losses of the Italian ecosystem

WP6 Analysis of citizen preparedness and ecosystems resilience to disasters
Activities:
Task 6.1 Gathering and integrate multi-source data
Task 6.2 Modelling with Big-Data strategies ensuring high throughput
Task 6.3 Assessment of citizen preparedness and ecosystems resilience to disasters
Milestones
M9-M15 – 20% Data collection and management with efficient database strategies
M22-M26 – 40% High-throughput strategies for fast management of disaster-induced emergencies
M25-M36 – 40% Models for citizen preparedness to disasters

WP7 Machine learning, Quantum Computing and AI platform to design and exploit Digital Twins
Activities:
Task 7.1 Efficient classical and quantum AI models for disaster analyses
Task 7.2 Scalable strategies for remote and proximal sensing analyses
Task 7.3 Computational infrastructure and integrated software as a service
Milestones
M9-M15 – 20% AI models for environmental data
M22-M26 – 40% Deployment of models for scalable applications
M25-M36 – 40% Release of IaaS/SaaS framework

WP8 Impact assessment and modelling of alternative risk mitigation strategies
Activities:
Task 8.1 Impact/risk assessment methodologies
Task 8.2 Definition and evaluation alternative risk mitigation strategies
Task 8.3 Multi-level performance-based framework for impact evaluation and risk mitigation
Milestones
M9-M15 – 20% Overview of methodologies for impact and risk assessment and mitigations
M22-M26 – 40% Comparison of alternative risk mitigation strategies for the Italian (built) environment and community
M25-M36 – 40% Development of a multi-level performance-based framework for an Italian-specific impact evaluation and risk mitigation model and plan

Spoke 6



WP Tasks & Deliverable Description

WP6.1: Complex multiscale and Multiphysics modelling (Leaders: CNR, Sapienza)

- **T1.1** High-precision simulation via multiscale approach
- **T1.2** Variable fidelity modelling from atom to devices
- **T1.3** Multifunctional materials
- **T1.4** Multiscale modelling of complex systems
- **M9-M15** Identification of key challenges (Year 1)
- **M22-M26** Benchmarking multiscale numerical methods (Year 2)

WP6.2: Simulation of large systems via multiscale and AI methods (Leaders: PoliMI, Pavia)

- **T2.1** Development of multiscale multiphysics theory, numerical methods, modelling strategies
- **T2.2** Tools for orchestration and co-simulation
- **T2.3** AI & Big Data analytics for complex/large systems
- **M9-M15** Benchmarking of available tools (Y1)
- **M22-M26** Identification of routine tools for WP3 (Y2).

WP6.3: Digital twin modelling and tools for engineering disciplines (Leaders: Tor Vergata, Firenze)

- **T3.1** Thermo-fluid-dynamics
- **T3.2** Electronics & Photonics
- **T3.3** Electromagnetics & Energy
- **T3.4** Mechanics & Structures
- **T3.5** Bio & Environment
- **T3.6** Industrial production plants
- **M9-M15** Twins selection (Y1)
- **M22-M26** Pilot implementation of twins (Y2)

WP6.4: Engineering applications: verticals and methods/tools integration (Leaders: PoliTO, Bologna)

- **T4.1** Integration of methods and tools developed in WP1/2/3
- **T4.2** Engineering verticals in industries ranging from ground and aerospace vehicles to green energy, electronics, manufacturing
- **M9-M15** development of tools integrated with CN infrastructure (Y1)
- **M22-M26** Tools optimization (Y2)
- **M25-M36** Applications and use cases of optimized tools (Y3)

WP6.5: Proof-of-concept PoC demonstrators (Leaders: Pisa, UniCalabria)

- **T5.1** PoC activities on modular HPC infrastructures
- **T5.2** Software: licensing and distribution policies intellectual property, open source
- **T5.3** Specifications and coordination with hub/stakeholders
- **M9-M15** Selection of PoCs (Y1)
- **M22-M26** Preliminary PoCs development (Y2)
- **M25-M36** Implementation and performance evaluation of PoCs at TRL5/6 (Y3).

WP6: Training, Technology transfer activities and Outreach (Leaders: Sapienza, Pisa)

- **T6.1** Implementation of education and training activities
- **T6.2** Implementation of knowledge and technology transfer
- **T6.3**: Strategies for education/training, technology and knowledge transfer
- **T6.4**: Definition of Spoke strategies for dissemination

- **M9-M15** Plan of education and training activities & knowledge and technology transfer (Y1)

Spoke 7

The five work packages of this spoke will carry out their R&D activities as described in the following sections.

WP7.1 Flagship codes
The goals of WP1 goals are the extension, optimization, maintenance, and distribution of world-leading flagship codes for the high-performance numerical simulation of materials and molecular systems, developed by the national community and selected to cover a large spectrum of materials and molecular properties (from electronic structure, to molecular dynamics, up to multiscale approaches) that can be computed by exploiting current and emergent HPC architectures. Importantly, all software developments produced by WP1 will be made publicly available via suitable open-source licensing, thus further strengthening the impact of the present activities.

Task 1: Quantum Espresso (SISSA, CNR, UniMORE, UniTrento)
Quantum ESPRESSO (QE) is a suite of applications for ab-initio electronic-structure calculations using plane waves and pseudopotentials. QE applications are also used as building blocks for advanced methodologies such as MBPT, QMC, DMFT, and others, as well as for complex workflows for materials design and discovery. QE is released under the provisions of the GPL licence and is the most widely used open-source software suite in its application domain. Within WP1, we will: *i)* enhance the performance of all the QE components, taking advantage of best practice in scientific software engineering and targeting specifically exascale systems; *ii)* consolidate GPU support in all the components and extend it beyond NVIDIA hardware (notably including INTEL and AMD GPUs); *iii)* implement more property calculators, targeting e.g. electron-transfer reactions relevant to photochemistry; *iv)* implement advanced and friendly users interfaces, also aiming at easy integration within complex workflows managers; *v)* periodically distribute updated releases of the codes to the developers and end-users communities.

Task 2: Yambo (CNR, UniMORE, UniTS)
Yambo is an open-source community code enabling the calculation of excited-state properties of materials and molecules from first principles. The code implements MBPT, DFT and Non-Equilibrium Green's Function Theory (NEGF) in order to allow the user to calculate a wealth of physical properties including reliable band gaps, band alignments, defect quasi-particle energies, optical properties, time-resolved and ultrafast spectroscopies. The hybrid MPI, OpenMP, and CUDA structure allows the code to run efficiently on large HPC machines. Within WP1 we plan to perform *i)* maintenance and optimization, taking advantage of best practice in scientific software engineering aimed at HPC; *ii)* extension of GPU support beyond NVIDIA hardware (notably including INTEL and AMD GPUs) and extension of GPU support to advanced parts of the code (beyond GW and BSE); *iii)* implementation of new theoretical tools including advanced self-energies (beyond GW), ultra-fast spectroscopy, coupled electron-ion dynamics; *iv)* periodic release of the code to the community. *v)* Alternative approaches to the calculation of quasi-particles based on Koopmans functionals will also be further developed and compared to GW data.

Task 3: CRYSTAL (UniTO)
CRYSTAL is a code for the ab-initio study of crystalline solids at the Hartree-Fock and DFT levels. It employs an atom-centred Gaussian basis set that allows for distinctive features, such as the computationally cheap use of hybrid functionals and the possibility of true low-dimensional periodicity: 2D (slab) 1D (polymer) up to 0D (molecules). It already features a massively parallel distributed memory implementation and full exploitation of point symmetry. Within WP1 we plan to perform *i)* maintenance and optimization of the code, according to best practices in scientific software engineering aimed at HPC; *ii)* enhancement of intra-node and inter-node optimization through hybrid MPI/OpenMP parallelism; *iii)* implementation of new features such as new composite DFT methods including system-dependent dispersion corrections; *iv*) release of code updates.

Task 4: PLUMED (SISSA)
PLUMED is a leading open-source software package for enhanced sampling and is widely adopted by both users and developers of enhanced sampling methods in the Italian community and abroad. It can be used both as a standalone tool for analysing molecular dynamics simulations and as a plugin that augments standard molecular dynamics software with collective-variable-based enhanced sampling methods. Within WP1, we will develop the

next generation of this software. This will require: *i)* integration of PLUMED with Python, including the possibility to develop collective variables in Python and integration of PLUMED with industry-grade machine learning libraries; *ii)* porting part of the software on dedicated hardware (e.g., GPU) using SYCL or similar languages. To facilitate the long-term sustainability of the project we will also work on: *iii)* construction and maintenance of regression tests validating both PLUMED and the coupled molecular dynamics engines; *iv)* improved integration between PLUMED and the PLUMED-NEST platform. This will make it easier for developers to document any new features that are added to the code and to thus promote faster adoption of these new methods by new users.

Task 5: Multiscale and common libraries (UniPI, UniTrento, UniTS, UniCalabria, CNR)
A set of libraries will be developed and interfaced with different flagship QM codes to enhance their diversity of features and their application to multiscale systems. One such library, LibEnv (developed at UNIPI), will contain a multilayer classical representation of a general environment (a solvent, a protein matrix or a more complex environment) in terms of polarizable atomistic (MM) and continuum models. LibEnv will be parallelized exploiting advanced and heterogeneous computational infrastructures. Another library, LibOpt (developed at UNITS) will allow one to describe optical properties of complex systems at the TDDFT level with efficient density fitting algorithms, massive parallelism and chemical accuracy, including photo-absorption spectra and Circular Dichroism. The Stochastic Schrödinger equation (SSE) will be employed for the study of dephasing. Additional libraries will focus on the simulation of various electronic and vibrational properties, including quantum decoherence. UniCalabria will further develop a package to compute the dielectric properties of crystals, while UniTS and CNR-ICCOM will implement the Hybrid Diagonal approximation, in widely used codes to enable TD-DFT simulations of large systems (hundreds of atoms) with hybrid exchange-correlation functionals. CNR-ICCOM will develop methods and codes (FCclasses) for computing the shapes of linear and non-linear electronic spectra, also accounting for non-adiabatic couplings, and interface them with other software.

WP7.2 Big data generation and harnessing
In this WP we will develop novel methods in materials informatics, multi-scale simulations, and artificial intelligence that will enable us to explore portions of the chemical and structural spaces of materials and molecules that are currently not accessible. We will empower our community with the capability to compute materials and molecular properties with automated workflows, that enforce rigorous computational protocols and streamline high-throughput studies targeting the discovery of new materials, molecules, and drugs. Our workflows will be accelerated by dedicated data-driven methods tailored to materials and molecular data, *i.e.* materials intelligence.

Task 1: Automated high-throughput workflows
- Automated training of machine-learned interatomic potentials (UniTS, CNR). We will develop automated workflows for building data-driven interatomic potentials, including the generation and dissemination of robust first-principles data sets, as well as the model training procedures.
- High-throughput many-body perturbation theory (CNR, UniTS). We will enable high-throughput excited-state calculations of electronic and optical properties via dedicated AiiDA workflow for Yambo, by extending the current prototype workflows for the GW and the BSE methods.
- Automated exchange coupling in magnetic materials (UniMORE, CNR). We will develop AiiDA workflows to streamline the calculation of exchange coupling parameters and single-ion anisotropy in magnetic materials from first-principles simulations.
- High-throughput thermodynamics of materials (SISSA). We will develop workflows for the efficient and automated computation of thermal properties of materials.

Task 2: Materials Intelligence
- Advanced machine-trained interatomic potentials (SISSA, UniMIB, UniTrento, CNR). We will extend the scope of ML interatomic potentials to complex materials and molecular systems by developing novel methodologies to incorporate long-range interactions (electrostatics, van der Waals), a large (arbitrary) number of chemical species, and spin degrees of freedom.
- Novel machine learning methods (SISSA, UniFI, UniTrento, PoliTO, UniTS). We will develop: *i)* ML methods to handle large datasets produced in molecular modelling, materials discovery, and drug design; *ii)* information-theory-based feature-selection and manifold learning; *iii)* Bayesian methods for molecular

models and unsupervised techniques to discover constitutive laws hidden in large volumes of data through inverse problems. *iv)* We will release open-source libraries for high-performance dimensionality reduction and manifold learning, with emphasis on maximising the efficiency on heterogeneous hardware architectures.

- Artificial intelligence for macro-scale simulations (UniMIB, CNR, UniCalabria). We will develop ML and DL methods for the simulation of *i)* the structural evolution of crystalline materials, training on phase-field methods (UniMIB), and of *ii)* the structure and elastic properties of polymer networks such as hydrogels and microgels (CNR).

Task 3: Multi-scale and coarse-grained molecular simulations.

- Biological and polymeric macromolecules (UniTrento, SISSA). We will develop novel in *silico* strategies to design and characterise biological and polymeric systems. The data production software will address the enhanced sampling of conformational transitions. Data-analysis techniques will be adapted to this specific class of biological and biomimetic materials.
- Continuum models (UniTrento). We will derive: *i)* suitable closure relations for continuum PDE models via the analysis of large sets of experimental equations of state and experimental stress-strain diagrams; *ii)* accurate equations of state for macro-scale continuum models via nonlinear nonconvex optimization and ML.
- Protein simulations (UniFI, UniTrento, UniPI). We will develop enhanced-sampling algorithms and non-equilibrium techniques for the simulation of molecular systems involving molecules/receptors interactions and membrane proteins, as well as coarse-grained methods for protein-protein interactions. We will deliver fast and accurate tools for computing drug-receptor/binding affinities and to design novel protein-based immuno-agents for immunotherapy. We will employ ML models trained with publicly available big data (e.g. ChEMBL) and receptor-based protocols.
- Colloids (CNR). In the field of fluid-mediated aggregation of colloidal particles and their breakup, we will first perform mesoscopic simulations with Lattice Boltzmann methods to create large databases of aggregates. We will then deploy AI methods to predict the size distribution of aggregates and their sedimentation.

Task 4: Applications to materials and molecular modelling.

- Large-scale atomistic molecular dynamics (UniMIB, CNR, SISSA, UniTS, UniMORE). We will leverage the methods developed in tasks 1 and 2 to study technologically relevant materials such as Ge-rich GeSbTe alloys for embedded electronic memories (UniMIB), hybrid perovskites for photovoltaics (CNR, SISSA), two-dimensional materials (UniTS) and silicate amorphous glasses with industrial synergies towards applications in bioglasses and nuclear waste confinement (UniMORE).
- Ferroics database (CNR). Leveraging the results of tasks 1 and 2, we will build a database of high-performance ferroics with calculated and experimental properties, including the optimization of figures of merit through ML methods.
- Advanced high-throughput characterization of 2D materials and heterostructures for next-generation enabling technologies (CNR, UniTS, UniMIB, UniMORE). Workflows developed in T1 will be exploited to investigate the electronic, magnetic, ferroelectric, topological, and optical properties of either isolated layers or van der Waals heterostructures, aiming at predicting the performance of new materials in a wide range of technologies.
- Nanowires design (ENEA). High-throughput workflows of task 1 will be adapted and customised to design novel nanowires for electronic and photovoltaic applications, leveraging ML tools developed in task 2. This computational environment will enable the optimization of nanowire properties through surface functionalization.
- Digital design of molecules (PoliTO, UniCalabria). We will design new molecular systems with specific properties–such as stimuli responsiveness, adaptivity, reactivity, and chemotacticity–using complementary AI-based strategies–such as: *i)* coarse-grained MD simulations, automated optimization, and genetic algorithms; *ii)* data-driven identification of molecular design rules; *iii)* sequence-based inference algorithms based on Direct Coupling Analysis combining global (homology-based), and local mutations information (from sequencing data of panning experiments).

WP7.3 Accuracy and reliability

WP3 is dedicated to evaluating the accuracy of the computational methods developed, implemented, and utilised within the Spoke, through hierarchical protocols designed to benchmark affordable lower-level approaches against state-of-the-art reference data. To this end, we will exploit some of the most accurate methodologies, like quantum Monte Carlo (QMC), multireference quantum chemistry and non-equilibrium green function methods, to produce reference data for DFT and TD-DFT methods. QMC and the validated (TD)DFT methods will be in turn employed as reference to machine-train accurate Force Fields for the investigation of complex systems along wider length/time scales.

Task 1: Validation of various DFT functionals against accurate reference calculations

- Quantum Monte Carlo. The GPU implementation of the TurboRVB QMC code, performed at SISSA, will boost our capacity of benchmarking widely used DFT approximations. A new class of AI-inspired methods will be developed, by training a neural network or Gaussian process to the *difference* between accurate QMC and approximated DFT data ("delta learning"), thus enabling simulations of QMC accuracy at a DFT cost.
- Standardised Gaussian basis sets for solids and implementation in Crystal Package. UniTO will develop a standardised Gaussian basis set library for solids, guaranteeing an easier access to non-expert users and a more consistent quality of the CRYSTAL package results, benchmarked against plane-wave codes (QE) on prototypical and realistic systems.

Task 2 – Validation of TDDFT approaches against accurate reference calculations

Novel reference approaches for excited-state properties. UniPI will combine CASCI with optimal active spaces and an efficient density-matrix renormalization group (DMRG) treatment together with a QMC treatment of dynamic correlation. Analogously, CNR-ISM and CNR-NANO will exploit automatic workflows to validate TD-DFT results with respect to real-time NEGF obtained with Yambo, implement advanced self-energies (vertex-corrected GW) and collaborate with UNIPI to produce highly accurate reference data using many-body perturbation theory.

Task 3 – Validation of FFs against accurate QM reference calculations

In this task we aim at devising machine-learning strategies to train high-quality force fields to high-accuracy reference data.

- Ground-state force fields. SISSA and CNR-ICCOM will exploit coupled-cluster, MP2, and QMC data to machine-train a new-generation of force fields to model weak intermolecular interactions.
- Excited-state force fields. The lack of readily available FFs designed for the excited states has hampered the use of classical MD to tackle photo-chemical properties and opto-electronic materials. CNR-ICCOM will exploit a newly developed software (Joyce) to build and refine specific and accurate FFs for simulating excited state properties, in synergy with WP2.

WP7.4 Pilot applications

WP4 aims at applying the software/middleware developed in WP1/2/3 to demonstrate their ability to solve significant technological/scientific challenges, hitherto considered intractable.

Task 1: Simulation of physico-chemical processes controlling materials functions for energy applications

In this task we address a few key problems in the field of energy materials, focusing on processes driven by light-matter interaction and relevant to thermoelectrics, photovoltaics, and electrochemistry (batteries and fuels cells). We will address: *i)* photoexcitation processes to control the thermoelectric flow in 2D materials (UniPI); *ii)* the optimization of molecular materials for solar energy transduction (UniTO); *iii)* the use of self-assembly effects to enhance their optoelectronic response (CNR); the photocatalytic water-splitting reaction in nanosystems (CNR, UniMORE) and the tuning of chemical reactivity through optical cavities (CNR). Multi-scale and multi-physics HPC methodologies will be adopted to study: *iv)* the effects of disorder and microstructure in chemically complex materials, relevant for energy applications, such as thermoelectrics, photovoltaics, and fuels cells (UniTrento), or innovative RRAMs (UniMORE); *v)* the mechanical properties of brittle and ductile materials (failure, fracture, and fatigue), as well as in modern bio-inspired materials and metamaterials (UniTrento). High-throughput and machine-learning strategies will be deployed to: *vi)* optimise next-generation battery components (improved electrodes, electrolytes, and combinations thereof: PoliTO, UniTO, UniMORE); identify promising electrocatalytic materials without critical raw elements for optimised electrochemical CO_2 conversion, hydrogen production, and other fuel-cell reactions (UniMIB, PoliTO); determine the structural conformations and interaction energies of nano-systems, such as functionalized nanowires, for the realisation of devices in the fields of photovoltaics, energy storage, and nano-electronics (ENEA).

Task 2: Design and discovery of smart materials and systems for biomedical applications

This task will address smart molecular systems and materials for biomedical applications, leveraging a multi-scale approach to tame the different time domains involved in drug delivery by nanocarriers, protein assembling, and protein-membrane interactions. This approach will encompass quantum chemistry, classical atomistic and coarse-grained models, up to continuum approximations, together with the use of workflows for an easy analysis of the large datasets produced. Applications will include: *i)* the design of shape-changing patient-tailored devices (4D printed artery stents: UniFI); *ii)* the study of host-guest complex as Drug Delivery Systems and soft materials for controlled Drug Delivery, (UniPI, PoliTO, UniTO); *iii)* a systematic first-principles study of inorganic nanoparticles of increasing size as realistic nanosystems for therapeutics and imaging (UniMIB); the modelling of the human cancer protein/protein interactome (PPI: UniTS); the atomistic study of photoreceptor proteins by identifying the molecular mechanisms driving the photochemical event and subsequent conformational change (UniPI); the multi-scale and AI-assisted design of biomaterials for tissue regeneration (UniMORE).

Task 3 – Complex systems for advanced technologies and applications

Task 3 addresses selected challenges involving complex systems for quantum-information technologies based on magnetic effects (spintronics and spin-valleytronics) and for applications in earth science and astrophysics. In the first area, we will address: *i)* the adsorption of molecular magnets on different surfaces and 2D materials (UniFI); *ii)* selected applications to molecular spintronic units based on molecular magnets (UniFI) and to spin-valleytronics based on van der Waals heterostructures of 2D materials with emergent functionalities (UniMORE). In the field of Earth sciences and technologies, we will *iii)* investigate complex materials and fluid-structure interactions for wind and offshore engineering (UNIFI). In planetary physics, highly accurate force fields machine-trained to quantum Monte Carlo data will be used to determine the equation of state of hydrogen and hydrogen-helium mixtures, aiming at improved planetary and astrophysical models (SISSA).

WP7.5 Materials foundry

WP5 will develop HPC optimised methods and multilevel workflows targeting the specific needs and requirements of experimental and industrial research.

Task 1: Multilevel Methods & Workflows for Response Properties

This task will develop theoretical methods and HPC workflows for the simulation of traditional and advanced spectroscopies, spectro-microscopy, diffraction, and scattering, with the aim to support the research at large-scale experimental facilities such as synchrotrons (Elettra, Fermi, Soleil) and X-ray free-electron lasers (Hamburg). Simulations of traditional (optical, Raman, NMR/EPR) spectroscopies via flagship codes be pushed toward large (>1000 atoms) and complex systems, exploiting methods of enhanced-structural-sampling as well as AI-regularisation/interpolation techniques, targeting defected and disordered materials, glasses, biological systems, hard (zeolites, nanoparticles, MOFs and COFs) and soft (vesicles, deep eutectic solvents and ionic liquids) matter. Code integration into commercial NMR/EPR software (e.g. Bruker) will also be pursued. Other third-party codes of common use for X-ray diffraction and scattering simulations (EXPO, SIR) will be empowered by software (MPI parallel computing, GPU-accelerated FFT) and algorithmic (global-optimization) advances, to address the structure of composite materials (macromolecules, natural fibres, etc.) with applications in biomedicine (drug carriers), organic synthesis, catalysis, and innovative and eco-compatible materials. New methods to simulate advanced core-hole synchrotron-based (EELS, XAS, IXS and RIXS, ARPES) and ultrafast spectroscopies (pump-probe, TD-ARPES, ultrashort x-ray pulses) will be developed, pursuing novel algorithms for post-DFT approaches (scalable Bethe-Salpeter, multiplet calculations, accelerated Time-Dependent DFT also exploiting methodological advances), strong spin-orbit coupling in strong magnetic field, Dynamical Mean-Field Theory for low-dimensional and quantum materials), core-hole pseudopotentials, and implementing advances from flagship codes. Applications to innovative materials for batteries, solar and photoelectrochemical cells, materials for renewable energy conversion and storage will follow.

Task 2: Advanced Methods for Single-Scale Structure Prediction and Dynamics

This task will develop theoretical methods, automatized algorithms, workflows, and advanced approaches (AI, ML, optimization) for the predictive simulation of (micro)structure and dynamics in support of experimental and industrial research, also leveraging machine-trained force fields, enhanced-sampling, global optimization techniques, empowered by the interoperability between DFT flagship codes and open-source molecular-dynamics software (*e.g.* LAMMPS). Target phenomena will include adsorption on proteins and their assemblies, metal-organic frameworks. as well as cancer protein/protein interactions and binding modes, with applications to virtual screening, drug design, sensing, molecular probes and binders, host/guest systems (properties and vibrational/electronic spectra for interpretative purposes), binding affinity/kinetics in cancer immunotherapy. We will address a number of technologies, including: (photo-) catalysis for energy production and environmental protection (metal/metal-oxide nanocatalysts for hydrogen production and utilisation, carbon dioxide conversion, biomass valorisation); drug delivery (nanoparticle carriers and biomarkers); the design of new antibodies and drug manufacturing; microplastic flocculation (adhesive forces and fluid stresses in complex flows); metal-organic frameworks for separation technology; molecular sensing; stability and ionic mobility of novel MAX-phases and MXene materials for Li- and Na-ion batteries; thermoresponsive polymer-networks.

Task 3 –Multi-Scale Models for Complex Materials and Processes

This task will address the properties and (micro-) structure on the time and length scales appropriate to hierarchical materials and/or real industrial processes/devices and enable the modelling of a broad range of 2D materials, thermoelectrics, superconductors, topological and artificial materials and their applications to energy and environment technologies (high-Tc superconductors, optoelectronic devices, quantum thermal engines), quantum sensing, quantum optics and photovoltaics, 3D structural analysis. Climbing the multiscale ladder bottom up, QM/Classical simulations will be enhanced, accelerated via linear-scaling fast-summation techniques, and deployed to advanced HPC infrastructures (GPU and multi-GPU systems). Interactions between polymeric surfaces and SARS-Cov-2 spike proteins will be analysed to extract simplified effective potentials and associated descriptors to model SARS-Cov-2 surface affinity dispersed in respiratory droplets or water solutions. Embedding methods will be developed for the predictive modelling of environmental effects, including: *i)* electron-phonon and quantum anharmonic phenomena in solids via the Stochastic Self-Consistent Harmonic Approximation, (superconductivity, thermal and electronic transport, electronic gap and optical absorption), and *ii)* effects of strong correlations, dissipation/decoherence and out-of- equilibrium conditions on topological materials via various complementary methods (density-matrix formalism, variational and Green's function Monte Carlo, density-matrix renormalization group). Quantum effects in functional nanophotonic materials will be modelled using TDDFT and its orbital-free extensions, and will be coupled with macroscopic FEM, FDTD, RCWA codes. An integrated simulation suite for modelling kinetics of atomic systems interacting with complex environments, including parallel and/or sequential coupling of DFT (*e.g.* BigDFT), atomistic (*e.g.* MulSKIPS) and continuum (*e.g.* FEniCS) codes will be extended and optimised. The goal is to achieve virtualization of experimental integration of novel materials/systems in high-tech manufacturing flows, a critical step in speeding up TRL evolution from basic research to industrial applications and the market. An open-source Discrete Element Method code for the coupled simulation of fluid/particle systems already applicable to process optimization relevant to the energy and the pharmaceutical industry will be extended to include QM-informed tribocharging phenomena in cohesive powders.

Task 4 – User Uptake, Technology Transfer & Dissemination

The ultimate goal of WP7.5 is to enable a broad base of scientists without prior training in materials and molecular modelling (experimentalists, Earth scientists, chemists, biologists, and engineers, possibly working in industrial labs) to take advantage of the limitless opportunities offered by modern numerical simulation technologies. To achieve this goal, this task will promote three distinct actions in mutual synergy:

- The development of dedicated web portals aimed at: *i)* disseminating ready-to-use workflows for the automated computation of specific properties and simulation of specific processes that would otherwise require a skilled concatenation of numerous intermediate steps, each involving the use of different codes and theoretical methodologies; *ii)* making the results of previous simulations publicly available and searchable through advanced AI-based methodologies; *iii)* providing "vertical simulation as a service": by this we mean a set of cloud-based dedicated services whereby the simulation of specific ("vertical") properties/processes is requested through a user-friendly web portal and carried out by a backend without the direct intervention of the users.
- The promotion of a series of training courses (both on-line and in-person) specifically targeted to experimentalists, industrial users, and other scientific communities.
- The promotion of a series of initiatives (both on-line and in-person: round tables, workshops, open days) targeting industrial needs and aimed at facilitating the exchange of information between researchers in

the academia and industrial stakeholders about the opportunities offered by the former and the business expectations of the latter.

IP protection and data-protection policies will be implemented as needed. Dissemination to the society at large will also be carried out through public conferences and lectures, also in collaboration with the national education system.

Deliverables

M9-M15:

- WP1 - code deployment on the Leonardo system at CINECA;
- WP1 - new release of all codes
- WP2 - Implementation of the methodological and software frameworks
- WP2 - Benchmark, demonstration, production of software
- WP3 - Benchmarks for ground-state properties
- WP4 - Validation of the developed software against the selected technological challenges
- WP5 - Release of computational codes for assisting experimental materials characterisation

M22-M24:

- all WP: National conference on Materials and Molecular Sciences
- WP1 - new release of all codes
- WP2 - Benchmark, demonstration, production of software
- WP3 - Benchmarks for excited-state properties
- WP4 - Validation of the developed software against the selected technological challenges
- WP5 - Release of computational codes for assisting experimental materials characterization
- WP5 Release of web-based GUI for computational codes
- WP5 Release of codes dedicated to industrial R&D

M25-M36:

- WP1 - new release of all codes
- WP2 - Benchmark, demonstration, production of software
- WP3 - Publication of reference benchmark datasets and assessment with respect to experimental data
- WP4 - Validation of the developed software against the selected technological challenges
- WP5 - Release of computational codes for assisting experimental materials characterization
- WP5 Release of web-based GUI for computational codes
- WP5 Release of codes dedicated to industrial R&D

Spoke 8

WP8.1 Implementation of modelling & simulation platforms (open source and commercial) through HPC solvers

Computer Modelling & Simulation (CM&S) application to life sciences span a variety, of black-box, grey-box, and white-box modelling methods. We see applications that go from pure data-driven, machine learning modelling methods to strongly mechanistic model where existing knowledge is expressed in terms of equations or rules. But purely mechanistic models rarely exist: to use a continuum mechanics terminology for any given characteristic space-time scale at which the phenomenon is described there are phenomenological closure for the upper scales (boundary conditions) and the lower scales (constitutive equation). Thus, most models are grey-box models, where some mechanistic knowledge is combined with some phenomenological knowledge, extracted from observational data. This is particularly true for Life Science CM&S, where there is almost always a significant data processing component in the computational needs.

With this in mind, the list of numerical methods relevant for the applications of Computer Modelling & Simulation (CM&S) to life sciences is quite long. Since the modelling of living organisms is characterised by a significant spread across space-time scale, a useful taxonomy can be based on the characteristic scale at which biological processes are described (i.e., atomic scale, single cell scale, cell networks scale, organ scale and organ system scale).

Examples of HPC solvers

For Molecular dynamics there are general-purpose codes likes GROMACS or NAMD, or specialised codes like TIES (a collection of software packages to calculate protein ligand binding free energies with physics-based alchemical methods).
For fluid dynamics, excellent scalability is achieved with Lattice-Boltzman solvers such as PALABOS, or HEMELB. For solid biomechanics the ALYA solver shows excellent scalability. Another HPC solver is PARAFEM, which OpenFPCI couples with the OPENFOAM CFD solver for Fluid Structure Interaction (FSI) problems. A specialised FSI solver is HEMOCELL, which simulates dense suspensions of deformable cells, blood cells in particular.
Biochemical network models solvers like COPASI do not usually have significant computational costs, but high-throughput computing implementations or multiscale models involving PDE-ODE coupling may require their deployment on the HPC systems. It is also worth mentioning single-cell modelling environments such as VCell or CompuCell3D.
With respect to commercial solutions, it is worth mentioning the Living Heart Project of SIMULIA, which provides a electro-mechano-fluidic simulation of the human heart, or the InSilicoTrials solutions such as NuMRis (Numeral Magnetic Resonance Implant Safety).
A lot of problems require the solution of stochastic formulations of ODE or PDE, which are handled with various forms of Monte Carlo simulations, possibly accelerated with Gaussian Processes.
Reasonably scalable implementations of ABMs use MPI or GPU architectures. To the first belongs the REPAST solver, to the latter FLAME-GPU.

Task 1.1 Investigation of methodologies for the implementation of HPC solvers (M01-M18)
At the outset a User Group will be established, which will be engaged in providing specifications for the vertical solutions to be supported by the National Centre in term of HPC solvers. For each type of solver required by the User Group, a survey of open source and commercial codes available, and their eventual evidence of scalability will be collected. Solvers will be chosen on the basis of a variety of factors including documented scalability, location of the development team and level of support provided, ease of use, preferences of the end users, integrability with the computational infrastructures, etc. Where there is contention between two or more solvers, a benchmark problem will be proposed by members of the User Group, and comparative scalability analyses will be conducted. A portfolio of selected solvers should be identified and installed by M06 (**MS1.1**).
Once a solver is selected, the User Group will provide a small collection of test models, representative of the variety of models to be solved. With these a solver optimisation will be conducted at multiple levels. First, for codes that run both GPU and CPU, the best performing architecture will be chosen. Then a fine tuning will be conducted on execution parameters such as cores x process, allocated memory, I/O parameters, etc. A report on the available HPC solvers and their optimisation will be produced by M12 (**D1.1**).
The end of this exploratory task will be the exemplary execution of one or more models using the largest possible volume of computational resources available. These exemplary HPC runs will be used to demonstrate the importance of using HPC solvers in In Silico Medicine, and to develop training materials for targeted hackathons where candidate users are taught how to best exploit the available HPC solvers for their own research problems. Are report on the first hackathon for Biomedical HPC solver users will be produced by M18 (**D1.2**).
Spoke milestones and deliverables (with matching CN Milestones)

MS1.1	M06 (M1-M8)	Portfolio of selected solvers should be identified and installed
D1.1	M12 (M9-M15)	Report on the available HPC solvers and their optimisation
D1.2	M18 (M13-M18)	Report on the first hackathon for Biomedical HPC solver users

Task 1.2 Development of modelling and simulation platforms through HPC solvers (M03-M36)
By M03, when the first survey of the user group is completed, the consortium will start the development of the modelling and simulation platforms for In Silico Medicine. Two platforms will be implemented from the outset. The first simulates the mechanobiology of the neuro-musculoskeletal system: and is composed by an ODE model of musculoskeletal dynamics, and a finite element solver to model the deformability of the bones coupled to an ODE/ABM solver to model the mechanobiology of bone tissue adaptation. The second platform is a complete implementation of the whole human immune system, composed by binding affinity model, a cell-to-cell and cell-to-tissue ABM model, and a diffusion-reaction-advection model of biochemical species for autocrine, juxtracrine, paracrine, and endocrine cells signalling. These two platforms enable a variety of In Silico Medicine applications including patient-specific prediction of the risk of osteoporotic fracture, In Silico Trials of antiresorptive drugs, personalised treatments for multidrug-resistant pulmonary tuberculosis, In Silico Trials for new vaccines, etc. The two HPC platforms should be running by M18 (**MS2.1**). In parallel available solvers will be used by the users to

develop and deploy additional In Silico Medicine platforms. A report on the deployment of all In Silico Medicine HPC Platforms will be produced by M36 (**D2.1**).

Spoke milestones and deliverables (with matching CN Milestones)

MS2.1	M18 (M13-M18)	Mechanobiology and Immune system HPC platforms running
D2.1	M36 (M25-M36)	Report on the deployment In silico Medicine HPC platforms

Task 1.3 Scalability and co-design (M13 – M36)

By M12 the optimisation of all HPC solvers should be completed. As the solutions are deployed, starting with the first two which should be in production by M18, a systematic activity aimed to improve the computational efficiency of the platforms will be conducted. This will articulate in two parallel directions. The first will aim to optimise the scalability of each platform using the available computational architectures. Here the focusing will be the partitioning between nodes and between processor types (CPU, GPU), scheduling strategies, multiscale and multiphysics coupling strategies, surrogate modelling accelerators, etc. A first report on efficiency gain will be produced by M24 (**D3.1**). The second direction of work will be on how changes in the computational architectures could benefit the computational efficiency and scalability of our In Silico Medicine HPC platforms. This includes co-design activities toward the next generation exascale supercomputers, as well as the testing for specific problems of unconventional computational paradigms, such as quantum computing. A final report on scalability and co-design will be produced by M36 (**D3.2**).

Spoke milestones and deliverables (with matching CN Milestones)

D3.1	M18 (M13-M18)	Report on efficiency gain
D3.2	M36 (M25-M36)	Report on scalability and co-design

WP8.2 Digital Twins and In Silico Trials

The use of a patient-specific model to predict quantities that are difficult or impossible to measure directly, but that are important to inform the clinical management of that patient was originally referred to as Digital Patient technologies. But because the concept resonates with very similar developments in other industrial fields, the term most used now to indicate this use case is Digital Twins in Healthcare.

While the concept of Digital Twin in Healthcare (DTH) is very powerful and potentially impactful, as some of the early solutions are demonstrating, the time and effort required to develop and certify for clinical use a DTH is very high and tend to grow exponentially with the number of organ systems and space-time scale involved. The In Silico Medicine research community has recently started to discuss a new data-driven approach to the development of digital twins in healthcare, which is referred to in some recent publications with the term Human Digital Twin (HDT); according to this definition, HDT is a technology that simplifies and speed-up the development of DTHs.

The development of In Silico Medicine solutions follow a well-defined trajectory, that can be linked to the Technology Readiness Levels of each vertical solutions (TRL Definitions according to the Insigneo Institute)

Task 2.1 Development of vertical solutions for In Silico Trials (M06-M36)

The development of vertical solutions, designed to target a specific regulatory or clinical need, will start in parallel to the deployment of the HPC solvers. Since the focus is to support only the development of vertical solutions that have a clear need for HPC, we expect all solutions that enter in the development pipeline for tasks 2.1 and 2.2 to have completed the design phase, and to have collections of experimental data that can be used for their validation (TRL4 or higher). Initially we will focus on vertical solutions made possible by the first two simulation platforms developed in T1.2 (neuro-musculoskeletal system, immune system). We will focus on the scalability and computational efficiency of two mature vertical solutions, developed in EU-funded projects: BoneStrength, an In Silico Trial solution to assess the efficacy of new drugs to treat osteoporosis, and UISS-TB, a solution to assess the efficacy of new treatments for tuberculosis. The two solutions will be running by the time the relative simulation platforms are ready, at M18 (**MS2.1**). BoneStrength simulates for each virtual patient millions of fall scenarios over ten years, but the real HPC challenge is the replacement of oversimplified data-driven model currently used to predict the effect of the treatment on the bone tissue properties with a fully mechanistic ODE/ABM model, coupled to the finite element model used to predict the fracture event. For UISS-TB the HPC challenge is to move from a model that is currently describing the infection's dynamics over one cubic millimetre of lung tissue to a much larger volume, ideally to the volume of a whole deflated lung (10^6 mm^3). As additional simulation platforms become available, additional In Silico Trial solutions will be added to the T2.1 pipelines. A report on all In Silico Trial solutions developed in the project will be produced by M36 (**D2.1**).

Task 2.2 Development of vertical solutions for Digital Twins in healthcare (M06-M36)
T2.2 will have a structure identical to T2.1, except here the focus is on DTH vertical solutions. Considerations of cost-effectiveness limit the opportunity to use expensive HPC systems only to two clinical scenarios: high impact problems, where the decision-support system provide a considerable clinical benefit, high enough to justify the extra costs the use of HPC systems; and urgent computing scenarios, where the timing of the clinical decision is critical. But there is a third use case, relevant here, which is that where DTH models are used in fundamental biomedical research to attempt the falsification of complex mechanistic theories. Using the two simulation platforms developed in T1.2, we will develop a vertical solution called BoBCAT-Onco that predicts the risk of unstable bone fractures in cancer patients with bone metastases. In these patients a fracture event may start a cascade of events that drastically shorten the life expectancy of these patients; we expect the use of HPC to be justified by such major clinical impact. UISS-MS is a vertical solution that predicts the progression of Multiple Sclerosis (MS). The aetiology of MS is still debated, and it is probably polygenic and multifactorial involving genetic, exogenous, and immunological factors. UISS-MS will be used to falsify concurrent mechanistic theories on the etiopathogenesis of MS, and in particular on the contribution that genetic, epigenetic, and immunological factors have in disease progression. As per T2.1, when these first two solutions are up and running, at M24 (**MS2.2**) more vertical solutions will be developed as additional simulation platforms become available. A report on all Digital Twin in Healthcare solutions developed in the project will be produced by M36 (**D2.2**).

Task 2.3 Specifications and prototypes for a Human Digital Twin infrastructure (M12-M36)
In this task we will explore the HDT concept and articulate its implications in term of computational infrastructure. We will contribute to the international effort to define specifications for the HDT, with particular attention to the problems of scalability and computational efficiency. Where necessary, proof-of-concept prototypes will be implemented to quantify the computational demand that specific operations might involve. A report on the HPC specifications for a future HDT infrastructure will be produced by M36 (**D2.3**).

Spoke milestones and deliverables (with matching CN Milestones)

MS2.1	M18 (M13-M18)	Two In Silico Trial solutions are running
MS2.2	M24 (M22-M26)	Two Digital Twin in Healthcare solutions are running
D2.1	M36 (M25-M36)	Report on all In Silico Trial solutions
D2.2	M36 (M25-M36)	Report on all Digital Twin in Healthcare solutions
D2.3	M36 (M25-M36)	Report on the HPC specifications for a future HDT infrastructure

WP8.3 Integrated digital data flow between clinics and HPC centres and Easy-to-use GUI for HPC solvers (hiding complexity for ultimate users)
After a phase of development and optimisation HPC-based In Silico Medicine solutions will be used only if two usability problems are solved.
The first problem is that most In Silico Medicine solutions are informed with clinical data. The handling of clinical data poses several issues. Most clinical data are generated within hospitals and stored in primary IT systems with limited interoperability. Health data are considered sensitive, so the most severe elements of the GDPR and of the Italian privacy law apply to them. Before a clinical dataset can be used for research purposes it is necessary (with a few limited exceptions) that the patients, who own their own clinical data, provide an informed consent. Even with that sensitive data can be treated without limitations only if they are irreversibly anonymised. However, there are some data that cannot be anonymised, and some use cases where the result of the simulation must be relinked to the patient identity, which prevent irreversible anonymisation. When only pseudo-anonymisation or limited anonymisation is possible, the handling of the data requires special attentions in terms of legal, ethical and cybersecurity aspects. But for an end user to use regularly In Silico Medicine HPC solutions, it is necessary to develop an infrastructure that automatically extract the data from the primary systems, organise / translate them as required by the simulation, perform all necessary operations to ensure the legal treatment of the data, securely transfer the clinical data from the hospital to the HPC centre, and return back into the hospital the results of the simulation, where necessary linked to the patient identity.
The second usability problem is related to the fact that most end users do not have a technical education, and thus the traditional user interfaces to HPC (command line, batch scripts, schedulers, etc.) are truly challenging for biomedical users. For reasonably standardised simulations, it is possible to provide web-based Graphic User Interfaces (GUI) that simplify job submission operations, and completely hide / automate the most gruesome aspects of running an HPC system.

Because of these reasons, an effective HPC solution in this domain cannot be only measured in term of scalability and computational efficiency, but also needs to be assessed in terms of usability, and cost-effectiveness.

Task 3.1 Digital data flow between the HPC centres and the healthcare centres (M07-M30)
With respect to the clinical data flow infrastructure between hospitals and the national centres, we will rely on the work being done in two Italian projects. The first, called AlmaHealthDB aims to enable the elaboration of clinical digital data for research purposes in the rest of the current Italian and European laws. Funded by UNIBO and the three National Research Hospitals (Istituti di Ricovero e Cura a Carattere Scientifico, IRCCS) located in the city of Bologna, AlmaHealthDB will provide a starting point for the clinical data flow infrastructure of the national centre. A second, more important project, called Health Big Data (HBD) is funded by the Italian Ministry of Health and coordinated by three national disease networks, with INFN as technical partner. HBD goal is to enable the exchange of digital clinical data between all 52 Italian IRCCS; AlmaHealthDB is part of HBD, as local node provider for the three Bologna-based IRCCS.
By M07, after the first solvers are installed (MS1.1) a pilot implementation based on the AlmaHealthDB infrastructure will be developed, which allows the transfer, elaboration, and transfer back of non-anonymised clinical data. The design will be presented and discussed at the Ethical Committee of Metropolitan area of Bologna; where necessary, we shall seek the advice of the Department of Legal Studies of UNIBO. First clinical datasets should be transferred from one of the Bologna hospitals to the HPC systems by M18 (**MS3.1**). The basic mechanisms provided by AlmaHealthDB will be revised in the light of the interoperability standards chosen in the HBD data, so to enable in principle all 52 Italian research hospitals to run simulations on their clinical data at the National Centre. Using cascade funding hospitals that are planning to do a regular use of the National Centre systems will be allowed to replicate locally the AlmaHealthDB systems, or to use them in service. A report on the clinical data flow infrastructure will be produced by M30 (**D3.1**).

Task 3.2 GUI implementation for HPC solutions (M13 – M30)
In parallel to the development of the first two vertical solutions in T2.1, we will develop two web-based GUIs to execute them. However, we will use a template-based approach, so that the adaptation of these GUIs to other similar solutions will require a moderate amount of work. By M30 all vertical solutions that require it will have the possibility to be run remotely via GUI. A report of user interfaces will be produced by M30 (**D3.2**).
Task 3.3 Conduct comparative studies on selected in silico medicine applications to quantify usability and cost-effectiveness (M24-M36)
As the vertical services are exposed through a web-based GUI, usability studies will be conducted to quantify the learning curve, and compute typical usability metrics including success rate of the simulation task, the time a task requires, the error rate, and the users' subjective satisfaction. In parallel we will evaluate for each vertical DTH solution the commercial cost per simulation, and in collaboration with the clinical users we will quantify the improvement that the use of the digital twin will have on the clinical effectiveness, so to conduct cost-effectiveness analyses.
In particular, this task will first select a limited number of in-silico medicine applications and will characterize them in terms of resource requirements, I/O needs (for example, data streaming requirements, or requirements related to specific storage classes), as well as of privacy needs (for example, related to the need to process sensitive data).
Subsequently, testbeds will be set up, where the applications can be run in traditional HPC centres or in Cloud-based HPC-capable environments, resorting to the resources offered by the ICSC infrastructure spoke. Whenever needed, centres offering proper ISO certifications such as ISO 27001, 27017 and 27018 will be used.
Then, comparative studies on the applications will be performed.
A complete report on usability and cost-effectiveness will be produced by M36 (**D3.3**).
Spoke milestones and deliverables (with matching CN Milestones)

MS3.1	M18 (M13-M18)	First clinical datasets transferred from hospital
D3.1	M30 (M25-M36)	Report on the clinical data flow infrastructure
D3.2	M30 (M25-M36)	Report of user interfaces
D3.3	M36 (M25-M36)	Report on usability and cost-effectiveness

WP8.4 Genome bioinformatics pipelines for GPU-based HPC infrastructures

Task 4.1 Benchmark of GPU-based codes for diverse genome sequences

Many new algorithms are continuously published to accommodate today's flood of genomic data, in this context an unbiased systematic assessment (benchmark) of software and tool performance is an essential task. In this project we will design benchmark studies based on real and in silico data to compare pipelines for the analyses of single cell data (task 4.2) and long-reads and structural variants (task 4.3). Datasets for structural variants and long-reads will be simulated by Sim-it (Dierckxsens et al 2021) while SPARSim (Baruzzo et al 2020) and ZIMB-WAVE (Risso et al. 2018), recently proposed by members of our team, will be used to simulate single cell data (Chao et al 2021).
Although simulated datasets allow the comparison in a controlled setting, they are always less complex than real data. Thus, we will complement the benchmark using real data for which a ground truth is known. Specifically, we will use datasets published by Tian et al (2019) for scRNAseq data, and the data from the Genome in a Bottle (GIAB) Consortium that recently developed a high-quality structural variant call set for the son (HG002/NA24385) of a broadly consented and available Ashkenazi Jewish trio from the Personal Genome Project (https://www.nist.gov/programs-projects/genome-bottle). Pipelines will be compared in terms of sensitivity, specificity, accuracy, and computational costs.

Task 4.2 GPU-based bioinformatics pipelines for single cell transcriptomics
Single cell RNAseq data is offering an unprecedented view on the heterogeneity and on the richness of cell dynamical states. The large dimension, the sparseness and the noise are three characteristics that make this data an ideal input for GPU based pipelines and deep learning analysis. Recently NVIDIA analysed a dataset of one million cells sequenced by 10X Genomics in 11 minutes on a single NVIDIA V100 GPU using RAPIDS workflow instead of three hours required for a CPU instance. A typical scRNAseq pipeline is based on several steps: quality assessment, data imputation, dimension reduction and clustering, cell cluster characterization. Leveraging on the expertise of the team members on scRNA data analyses and on the benchmark, results obtained in task 4.1 we will contribute to: (Aim 1) test and propose new methods based on variational autoencoders and other modern machine learning techniques as general solvers for scRNA-seq and (Aim 2) implemented on a GPU-based RAPIDS workflow the best performing pipelines obtained in Task 4.1.

Task 4.3 GPU-based bioinformatics pipelines for long-reads and structural variants
The huge amount of genomic data, including those from NGS platforms, requires novel and optimized algorithms to scale in HPC infrastructures and employ the massively parallel computing by GPUs. The best performing pipeline for long reads and structural variant obtained in Task 4.1 will be implemented using NVIDIA Clara™ Parabricks®, a recently proposed GPU-accelerated computational genomics framework used to implement GPU-based and HPC-compliant software pipelines for RNA and DNA sequencing analyses. NVIDIA Clara™ Parabricks® can be used also for long read analyses in terms of known and unknown variants - including SNPs, indels, copy number variations (CNVs) and other chromosomal aberrations. GPU-based and HPC-compliant workflow will be also used to decipher the epi-transcriptome landscape in long RNA reads.
Milestones and deliverables

MS4	M3 (M1-M8)	Novel genomic data through collaborations with clinics
D4.1	M3 (M1-M8)	Genome bioinformatics pipelines
D4.2	M12 (M9-M15)	Single-cell transcriptomics pipeline for GPUs
D4.3	M24 (M13-M18)	Bioinformatics pipelines for long-reads sequences

WP8.5 Development of clinical machine learning algorithms for EHRs and omics data (including radiomics)
Despite the great attention of the scientific community into clinical machine learning research, relatively few applications have been deployed at a large-scale in a real-world clinical environment. In this context, we will develop machine and deep learning algorithms to generate a catalogue of predictors of gene expression and co-expression, which are associated to different pathologies by employing very large datasets including genomics, transcriptomics and clinical data (behavioural, cognitive and psychiatric) available at our institutions, and generated through collaborations with clinical partners.
In parallel, we will develop new algorithms, rooted on the emerging field of high order in complex networks for relating transcriptomic and neuroimaging data and compare them with the state of the art, especially in terms of consistency and reproducibility of the results. Furthermore, we will develop entire computational workflows for investigating metabolic diseases and its complications using a multi-omics approach that includes transcriptomics, miRNomics, proteomics, phospho-proteomics and lipidomics/metabolomics datasets as well as anthropometrics,

clinical data and standard "clinical" biochemistry. The analysis will be performed on already available data as well as on novel data from patients that could be recruited by our clinical staff.
The Polytechnic team will design, develop and test intelligent algorithms for clinical data analysis and biomedical signal and medical images processing, with the aim of increasing the diffusion of sensors and measuring devices for real-time monitoring, of supporting robust precision medicine and target therapy of the health of population and individuals, without sacrificing measurement quality. Due to the high dimensionality characterizing omics data, and the necessity to integrate also heterogeneous data, e.g., clinical information and radiomics characteristics extracted from processing pipelines, such algorithms will be designed for being scalable and distributed in the training phase on the available HPC platform.

Task 5.1 Design and development of next generation EHRs for omics data
The goal of this task is to define and implement the structure of a new generation, holistic EHR (Electronic Health Record), capable of integrating and enriching the phenotypic data and the patient's clinical history with data from omics analysis. It will be a robust, reliable, interoperable, modular, secure and easy-to-use system capable to maintain in a structured way data and analytics results from high-throughput experiments in diagnostic and therapeutic relevant omics contexts. It will also offer advanced features to better support the diagnostic and drug treatment identification processes.
System for the collection, storage and processing of omics and clinical data. The core of this research activity will be the definition of a unified data model for the management of both clinical data and omics data resulting from bioinformatics pipelines developed in WP5.4, capable of maximizing the machine readability, interpretation and correlation of these heterogeneous information. Coding systems from literature or published by relevant organizations, including systems specified by national standards, most suitable for modeling different atomic information will be adopted. The integrated model will also include information entities to represent clinical recommendations, as produced by automated systems (CDSS, as explained below) or by human operators. The system will enable the secure integration, storage and access to the data in compliance with the current privacy regulation.
Support system for decisions relating to diagnostic investigations. It will be a flexible and open system of support for clinical decisions (CDSS), based on both clinical and omics data. The openness of the system is meant as the possibility to integrate new decision support algorithms as well as external systems designed to provide additional data or knowledge or to perform additional decision-making processes. In particular, the system will be sufficiently general to support the algorithms defined by the T5.2 task, to facilitate the identification of similar clinical cases and omics investigations to be performed on the patient. The system will be also open to the integration of both automatic recommendations and human experts, such as geneticists.

Task 5.2. Development of HPC-optimized pipelines and machine learning algorithms for omics data
In this activity, we will develop and deploy HPC-optimized pipelines and machine learning algorithms for omics data management, processing and exploitation. Raw genomics data is not particularly amenable to a direct exploitation via machine learning methods. Hence, it is necessary to move it from the raw format to something more easily intelligible and useful at the clinical and machine learning level. This activity is the variant calling one and we will develop pipelines that will take advantage of GPUs to allow a fast filtering and initial pre-processing. Along with this activity, we will develop HPC codes to move, in turn, from variant calling data to a matrix form to allow machine learning methods to work on it. At the same time, we will develop pre-processing algorithms for the other omics data formats. The end point is represented by machine learning methods. We will develop both shallow and deep methods depending on the application and on the empirical performance. To develop them we will take advantage of the proper machine learning software frameworks to take advantage of a HPC infrastructure. This may include but are not limited to: TensorFlow, PyTorch, Keras, Scikit-learn with Intel extensions (https://intel.github.io/scikit-learn-intelex/) and Spark. All these frameworks will exploit heavily HPC (both CPU and GPU) resources.

Task 5.3 Development of machine vision algorithms for radiomics
When dealing with clinical image data (RM, CT, US, PET, SPECT) there are several specific activities that can be developed and whose output can inform and simplify clinical decisions. These activities (algorithms) include chiefly tissue/image classification (often tumour) and segmentation. In this task we will develop and deploy (on HPC) algorithms to classify tissues. A key difference with respect to established approaches is that we will

combine these methods (often deep learners using Attention Mechanisms and Transformers) to the previously cited omics input. In this way, we will able to build learners (classifiers) that jointly take advantage of the full omics available data and possibly points to specific features. We plan to combine deep learning with interpretable models to shed light also on the possibly selected radiomics and other omics features. Also, we will evaluate quantitatively the feature stability among different datasets split, cohorts and models.
The aim is to address different clinically relevant tasks (e.g., staging, overall survival, event-free survival) developing task-specific tools running on HPC infrastructures. We will stress the reproducibility and the interpretability of the developed models, e.g., following well established guidelines explicitly designed for omics data by several US-FDA initiatives. To investigate interpretability, we will explore local (per sample) explanatory approaches like SmoothGrad and SHAP, compared with global feature selection methodologies like Lasso, Elastic Nets, Trees and Xgboost. Further, as a novel analysis strategy, we will test the effectiveness of coupling learning models with modern mathematical approaches dealing with the geometric structure of the data, namely Topological Data Analysis and Persistence Diagrams (TDA).

Spoke milestones and deliverables (with matching CN Milestones)

MS5.1	M3 (M1-M8)	Prototype of next-generation EHRs
D5.1	M3 (M1-M8)	Machine learning tools for omics (Month 3)
D5.2	M6 (M5-M8)	Machine vision for radiomics (Month 6)
D5.3	M12 (M9-M15)	Correlations between genomics and omics data (Month 12)
MS5.2	M24 (M22-M26)	Novel omics data through collaboration with clinics
D5.4	M24 (M22-M26)	Novel radiomics data in collaboration with clinics (Month 24)
D5.5	M30 (M25-M36)	Advanced version of EHRs (Month 30)
D5.6	M36 (M25-M36)	Bioinformatics pipelines ready for cloud (Month 36)

WP8.6 Drug-target studies and drug repurposing

The overarching goal of this WP is to develop general approaches for the discovery of drugs able to modulate the activities of novel molecular targets. The latter will be identified through the integration of information on sequence, structure, conformational dynamics, and interaction-networks of the biomolecules linked to disease states. Multiscale models of pathogenicity will be built, leveraging multi-omics data integration and computational methods. We will consider molecular targets related to different pathogenic mechanisms, ranging from proteins related to inflammation, to ion channels and transporters with their impact on the regulation of cell homeostasis and metabolism, to proteins and interaction networks involved in cancer development and/or neuromodulation. Experimental validation of the activities of discovered drugs will be pursued using a combination of *in vitro* biophysical and biochemical methods and *in vivo* cell-based assays. The tools and protocols developed will be included in the national cloud system for rapid and secure access by a broader scientific/medical community, through portable and scalable analysis pipelines.

Task 6.1 Identification of innovative drug discovery targets from genome bioinformatics analyses

This task entails the application and the development of advanced bioinformatics and computational genomics methods to (i) unmask patterns of genomic variation and (ii) unravel the structure/function of understudied, “dark”, genes which can be related to the onset of specific disease states with an inflammatory component, or characterized by tumour onset, imbalance in neurotransmission or neuro-degeneration. In this context, genomic data will also be processed with the use of machine/deep learning, thus focusing on broader and non-linear relationships to identify genomic and non-genomic parameters useful to predict how an individual’s profile will affect the evolution of a phenotype or disease. Data will also be processed with numerical and mathematical models to pin down the biomolecules determinant for disease. The latter will represent the targets for the design/discovery of drugs and drug-like molecules as new leads for treatment. These approaches will also be applied to unveil specific genomic patterns developed in response to drug treatments.

Task 6.2 Drug repurposing studies using novel target and the FDA-approved drug database

The biomolecules identified will be analysed in terms of structural and dynamic properties, using computational biophysical and simulative methods based, e.g., on molecular dynamics simulations. If already known, the structures of the targets will be retrieved from the Protein Data Bank. Alternatively, we will use AlphaFold-II to generate accurate models. The targets will be analysed to identify potential binding sites, with a combination of site identification methods, also based on innovative image analysis approaches. The stereochemical features of the binding sites will define the minimal complementary functional requirements necessary to a ligand for successful binding. Possible complementary interactions with the putative binding sites will be mapped with probe molecules, identifying conserved regions in ensembles of protein conformations where the same interactions are consistently made with the majority of structures. This information will be translated into pharmacophore models, recapitulating the relevant dynamic and stereochemical requirements for binding. The pharmacophores will be used to screen the FDA-approved drug database. In parallel, a biophysics-based virtual screening campaign will be carried out. The consensus of the two approaches will identify the best possible candidates. Screening on FDA-approved drugs and repurposing is expected to speed up the identification of leads close to clinical application, production, and market.

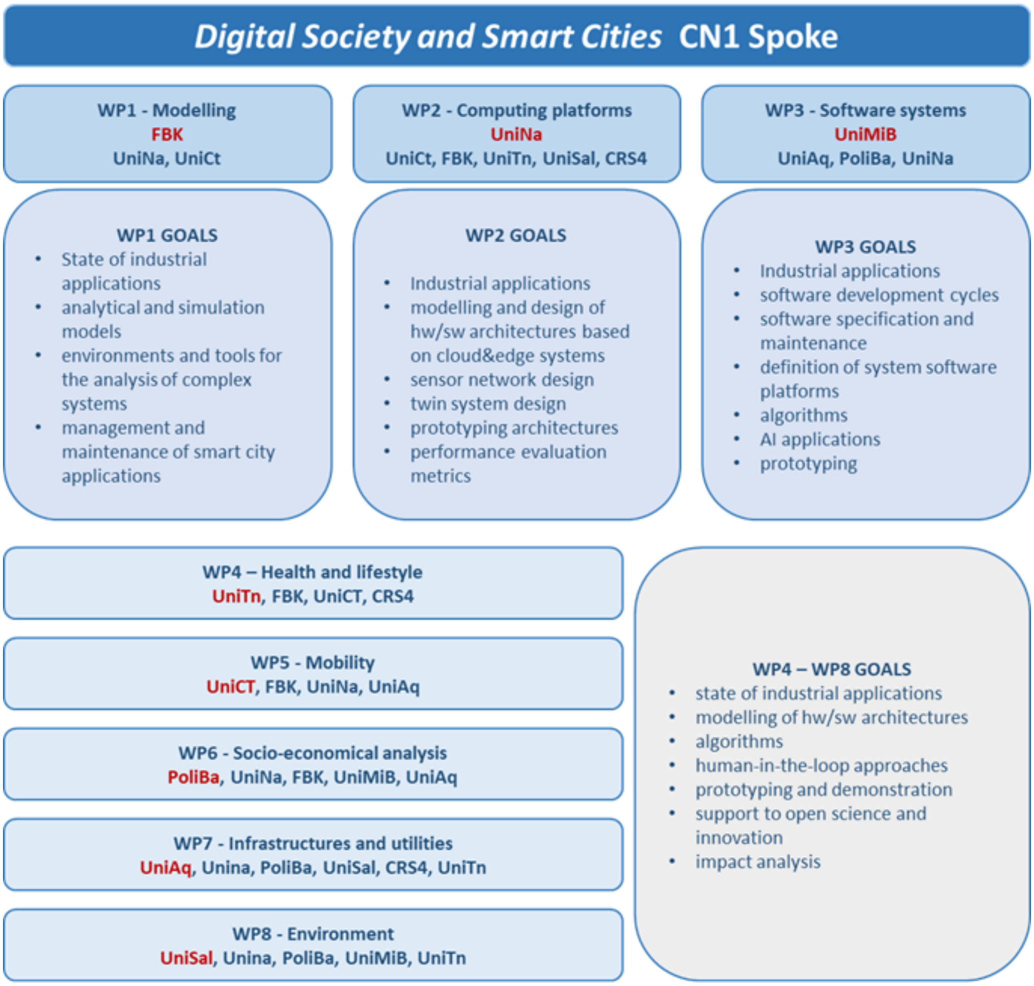
Task 6.3 Molecular medicine studies to link actionable genes, personalized treatments, and drug-target interaction model
One of the most challenging aspects of multifactorial pathologies such as cancer, neurological diseases, or the many conditions with an inflammatory basis is the heterogeneity of their phenotypes and of the factors causing them. These aspects also entail the response to drugs. Here, we will combine numerical tools, disease/patient-specific information, biophysical data on drug-target interactions to develop a model of drug response by those receptors endowed with specific sequences, interaction profiles and functional relationships. AI approaches will be developed to unveil drug-actionable genes and proteins. By focusing on pervasive mechanisms underlying patho-physiology, we aim at results that are generally applicable to a diverse range of diseases and patient cases.

Spoke milestones and deliverables (with matching CN Milestones)

MS6	M6 (M5-M8)	Drug repurposing for precision medicine
D6.1	M6 (M5-M8)	Novel targets for drug discovery (Month 6)
D6.2	M18 (M17-M22)	Models of protein-drug interactions (Month 18)

Spoke 9

Organization of the spoke
The spoke is organized in 8 work packages, as illustrated in the following figure:



The figure describes the key goals of the different work packages, as well as the research partners contributing to them.

Below, a brief description of the objectives of each WP is provided.

WP9.1 – Modelling

WP9.1 focuses on the development of innovative algorithmic solutions and modeling approaches for dealing with the challenges and needs of Smart Cities and Digital Society, identified by the five vertical domains. In particular, WP1 aims at developing innovative machine learning and deep learning methods (e.g. graph neural networks, neuro-symbolic approaches, causal and interactive machine learning, energy-efficient machine learning) as well as novel approaches and technologies in the fields of computer vision, speech recognition, natural language processing and generation, conversational agents and dialogue systems, predictive maintenance, etc. Additionally, methodologies from computational social science, complex networks and complex systems will be adopted to model, predict, and simulate complex societal phenomena on digital twin systems (e.g. human mobility, traffic, economic behaviours, daily routines and social interactions, social and epidemiological contagions, etc.). Then, being data and their high quality the basis of any knowledge and predictive system on human behaviour and society, attention will be dedicated to the development of learning models that favour the production and preparation of FAIR (Findable, Accessible, Interoperable and Re-usable) data by the members of a smart community. Finally, strong attention will be devoted to data governance, generation of synthetic realistic data, data collected from citizens and smart cities' infrastructures, and privacy and ethics aspects.

WP9.2 – Computing Platforms

The WP is focused on identifying and studying the key underlying building blocks for the computing platforms of interest in Digital Society and Smart Cities. Among other aspects, the WP will address the integration of cloud and IoT paradigms as well as the optimal allocation of tasks/software modules to edge/cloud resources under given cost, safety, security, performance, and latency constraints. In particular, the WP will study the integration and configuration of data acquisition devices for real-time monitoring of physical systems and human behaviours as well as the implications of underlying edge and IoT technologies on further non-functional properties, such as safety, security, and privacy. The WP will explore new methodologies and techniques for secure-by-design and privacy-by-design cloud / IoT / edge applications and for security assessment of critical systems, in the framework of modern software engineering processes. In addition, the WP will research AI solutions for inferring, from visual and/or other sensor data (depth, movement traces), structured 3D building models to be used for inspections and simulations, especially for security and energy management. Digital twin-based models and design principles of computing and communication solutions in the cloud and at the edge will be studied in order to enable the highest degree of flexibility in the deployment of services as well as to strike the right balance between performance and energy consumption / environmental impact. Models, techniques, and technologies for predictive maintenance will also be studied. Ultimately, the WP will deliver novel hardware/software architectures based on cloud/edge components, performance evaluation metrics, and prototypical implementations ready for demonstration with real-world applications of interest for the spoke.

WP9.3 – Software Systems
The aim of WP9.3 is to develop innovative software systems for the Digital Society and Smart Cities. In particular, software engineering techniques will be devised for the development and the operation of (possibly serverless) infrastructures and platforms for the management of all the layers between the Cloud and the IoT (Cloud-to-thing continuum). Further, techniques for creating highly reliable services for Smart Cities will be introduced, investigating approaches based on field-testing, failure prediction, self-healing and automatic program repair, methodologically conceived to be used in the DevOps field. All the phases of software development cycles will be covered, from software specification to prototyping, deployment and maintenance. Being the software systems and the platforms specifically conceived for Smart Cities, their structure and architecture will be designed with the aim of processing and analysing big amounts of data through HPC techniques, prototyping and scaling of AI algorithms and models (including deep learning, vision, natural language processing, recommendation systems, decision support tools, models of individual and collective behaviour), and their application to several domain fields.

WP9.4 – Health and lifestyle
WP9.4 covers the vertical domain of health and lifestyle in Digital Societies and Smart Cities. This WP focuses on the application of HPC, AI and digital twins to support citizen health and welfare. In particular, health scenarios where HPC technologies are key enablers include the development of fleet and resources localization management with customization to territorial medicine; of digital platforms for assisting therapeutics for mental health; of digital platforms for remote interactive coaching enabled by advanced HPC-based AI paradigms. Through the support of self-reports and tracking data from mobile devices, citizens with special health issues can be monitored, using the collected data to develop global models for health and lifestyle through AI. Digital Twins will be developed to study the interaction between citizens and the digital infrastructure of the Smart City, as well as for analysing diseases in order to drive personalized treatment and diagnostics. Quality and reliability of digital twins by facilitating the creation of machine-actionable structured metadata and by aiding the composability, interoperability, re-usability and reproducibility of datasets, computational tools and their results.

WP9.5 – Mobility
WP9.5 will cover the domain of sustainable mobility in Digital Societies and Smart Cities. This WP will focus on the application of HPC, AI, simulation techniques and digital twins to support suitable solutions aimed to introduce changes in the use of transportation systems inside cities, taking into account the important goal to achieve an improved quality of life and sustainable mobility for their residents. Many of these solutions are IT-based and include vehicle navigation systems, e-parking, e-tickets, signalling for information mobility, demand-responsive transport, car sharing, bike sharing, live monitoring of public transport, real-time monitoring, estimating and forecasting of traffic and citizens' mobility patterns. The WP9.5 will exploit collaborations with stakeholders in order to access data about traffic and environmental measurements acquired in real urban scenarios.

WP9.6 – Socio-economical analysis
We address the main challenges of sustainable development in local economic systems, taking advantage of the new analytical possibilities made available by big data and HPC. Territorial monitoring, modeling and control

from a socio-economic standpoint will be enabled by means of "digital twin" simulation frameworks integrating large-scale networks of sensors and data sources with analytical, AI-based and complex networks-based models as well as decision support systems for policymakers and economic operators. In particular, we focus on models for capturing individual and collective behaviours, analysis of economic flows, labour market gaps and urban segregation and inequalities, novel network science and AI approaches for the identification, tracking, and counteracting of misinformation, disinformation and hate in social media, disaster prevention, mitigation and management in areas vulnerable to natural or human-made calamities, and on how various forms of interregional and international socio-economic integration (migration, trade, supply chains, investments, international production networks, fight against urban crime and organized crime) can improve resilience and promote sustainable local development.

WP9.7 – Infrastructures and utilities

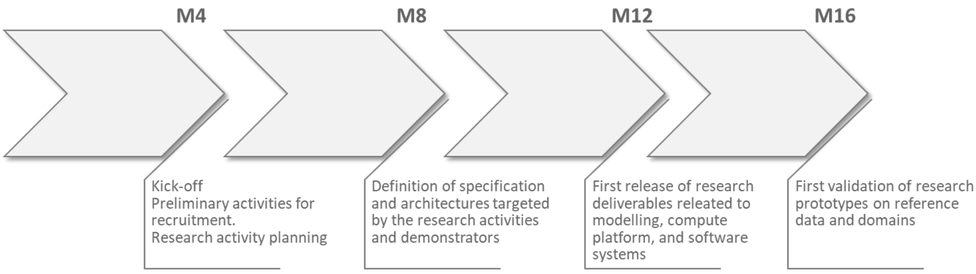
WP9.7 focuses on defining and realizing models and techniques for smart grids, control systems for autonomous building maintenance applications, and radio coverage planning approaches for outdoor and indoor wireless networks supporting digital services for smart cities. Research and development activities within WP9.7 concern probabilistic methods for forecasting electricity consumption and renewable generation for individuals and smart grids, and AI methods for analysing electricity consumption and non-intrusively disaggregating household electrical loads. The consumption and power generation potential in urban areas and the impact of construction techniques on consumption will be estimated by AI processing of satellite data, also in relation to different uncertain climate scenarios. Furthermore, Digital Twins related to building blocks of smart cities will be developed and integrated in a system-of-systems perspective, including public and private mobility, mobile cellular coverage, energy distribution, environment pollution, etc.

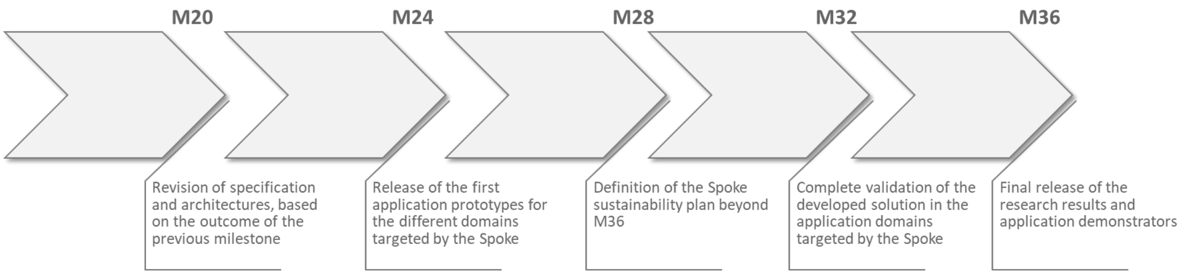
WP9.8 – Environment

WP9.8 is devoted to developing novel AI, learning and computational models for the management of indoor and outdoor pollution, coming from road, air and sea mobility and the impact of pollution on natural ecosystems and human health and wellbeing. Starting from data coming from living labs (i.e., traffic, buildings features, etc.), satellite images and weather forecast, WP9.8 will model, monitor and forecast air, noise, electromagnetic, water pollution, developing specific tools for analysing over time and space, an area carbon footprint and the change in the city biodiversity. WP9.8 will also deal with the development of decision support systems for the management of urban flooding risks, for improving the resilience of the urban and peri-urban area. Specifically, multi-Scale HPC-enabled models of smart city networks (water, energy, communications, wireless) and risk analysis exploiting advanced AI paradigms will be considered. Using approaches like the system of systems, WP9.8 will face the challenges to implement a Digital Twins city model (both anthropic and natural aspects), fed by data coming from the field, which will support the decision making related to the phenomena previously described. Specific attention will be paid to data centric AI and machine learning approaches, and to the compliance with the European environmental data space, as defined in the European data strategy and developed by GAIA-X.

Milestones

The planned activities, spanning three years, will be monitored by means of periodic milestones. The Spoke work plan includes nine milestones with an interval of four months, matched with the CN milestones. The detail of the milestones is displayed in the following figure. In general, the milestones (and the corresponding deliverables) encompass both administrative actions, e.g. the preparation and issuing of calls for hiring personnel or for cascaded funding, and scientific results, including the specification of the target architectures, the release of novel solutions involving modelling, compute platforms, and software systems, as well as their validation through real-world case studies.





Spoke 10

WP10.1. Software (Leader: INFN).
Development and application of high-level quantum software for algorithms solving general purpose problems, scientific and industrial applications.

- T1.1 New algorithms (Pavia, Bologna, IIT, Catania, CINECA, CNR, Pisa, Sapienza, Bari, PoliMI, Padova);
- T1.2 Applications and use cases (IIT, Bologna, CINECA, CNR, INAF, INFN, Pavia, Pisa, Bari, Bicocca, PoliMI, Padova)

WP10.2. Mapping, compilation and quantum computing emulation (Leader: CINECA).
Development of software toolchain for compilation, benchmarking, verification, emulation of quantum computers and algorithms.

- T2.1 Mapping and compilation (Bologna, CNR, Pisa, PoliMI);
- T2.2 Emulation (CINECA, INAF, Bari, Padova)

WP10.3. Firmware and hardware platforms (Leaders: CNR, Catania).
Development of low-level software for the physical operation of quantum computers. Development and support of the quantum computer hardware chain.

- T3.1 Photonic hardware (Sapienza, CNR, Bicocca, Pavia, Napoli);
- T3.2 Superconducting circuits (Napoli, INFN, Bicocca, CNR, Catania, Pisa);
- T3.3 Atomic hardware (CNR, Padova, Pisa);
- T3.4 Models and firmware (Catania, PoliMI, Bari, Padova, Bicocca, CNR, Pisa, Sapienza)

Milestones

- M9-M15: First Tender for Research Infrastructure
- M9-M15. Research activities on Software, Mapping, Compilation, Emulation, Firmware and Hardware at the end of Year 1: Design of quantum algorithms (SW); Classic emulator with 100+ qbits (MW); Report on architectural design of hardware platforms and tools (HW)
- M17-M22: Demonstrators: Use cases implementation and experimentation
- M17-M22: Second Tender for Research Infrastructure
- M22-M26: Research activities on Software, Mapping, Compilation, Emulation, Firmware and Hardware at the end of Year 2: Report on development and validation of quantum algorithms and applications (SW); Report on design of benchmarks for quantum computers and algorithms (MW); Report on design of quantum platforms (HW)
- M25-M36 Use cases: Report on use cases implementation and experimentation

- M25-M36 Research activities on Software, Mapping, Compilation, Emulation, Firmware and Hardware at the end of Year 3: Benchmarking quantum-accelerated applications against classical applications (SW); Test quantum supremacy in industrial setting (MW); Tools and methodologies for design automation and mapping (MW); One platform with 5+ qbits (HW); Photonic sampling machine with 5+ photons and 24+ modes (HW); Report on supporting tools for hardware platforms (HW)

Work plan feasibility

The **work plan feasibility** is ensured by the stepwise approach in which, for most of the activities within the spokes, the following phases are considered: a study phase, a pilot / implementation phase, and benchmarking / validation phase, eventually followed by a wrap-up phase.
The activities in the Spokes 1-10 are mostly independent, in such a way that a delay in one of these spokes has minor effects on the others; the scientific expertise of the research groups, at the base of their selection, builds trust in the feasibility of the specific workplans.
Spoke 0 activities are a prerequisite to the activities of the other spokes, since it provides the technological platform on which tests are executed. However, the participants in Spoke 0 (CINECA, INFN and GARR) are, among the partners, the institutions with the largest experience in handling large scale infrastructural projects, as demonstrated by the realization of previous Tier-0 HPC systems at CINECA (>100 M€), the largest Distributed Computing Federation (INFN) in Italy, and the current implementation of the Italian research networks at up to multiples of 100 Gbps (GARR). CINECA, INFN and GARR have units with personnel specifically formed and hired to manage all the phases in the realization of infrastructural projects, from the design, the procurement, the realization, up to the validation.
It should be noted that while the CN will procure and operate resources at a different scale than current deployments, the institutions coordinating Spoke 0 (CINECA, INFN, GARR) already today have world-class e-Infrastructures, like the Marconi 100, Galileo 100 and Leonardo HPC systems at CINECA, the 10 distributed centres for High Throughput Computing deployed by INFN, and the GARR-X infrastructure by GARR. These resources will be made available to the thematic Spoke affiliates for early tests, thus decoupling to a large extent the need to wait for the first CN procurements.
In order to monitor and control the execution of this complex program an industry-standard project management strategy will be adopted. We anticipate that the deployment of the structure will go through three phases:

- the norming phase, where major steps are
    - review of the organizational requirements;
    - hub's constitution;
    - appointing of the governance structure throughout all the program's components;
    - HUB personnel recruitment;
    - Detailed activity planning;
    - Setting common rules for managing and accounting;
    - Framework contracts definition and signing;
    - PMIS (Project Management Information System) is designed and starts operating;
    - Providing of knowledge management platform and a Data Management Plan;
- the capacity building phase, where the planned investments are executed;
- the operating phase, where the research projects run.

The PMIS will be the backbone for the flowing of all information regarding planning, monitoring and controlling and risk management. It will be implemented to set up a living model for scheduling and budgeting. According to modern industry standards, the **Earned Value technique** will be implemented to monitor project performances about time and cost management, the same system will provide data for financial reporting at different level according to the different needs of the stakeholder involved. They will find performance data published on intranet page for their own analysis. This will be particularly useful to monitor procurement activities. This set of information, continuously updated, is the base for risk management, the measurement of KPIs, the provisioning of reporting data to MUR and will allow the updating of the schedule model and the monitoring of the duly achievements of milestones and deliverables. Obviously, to run effectively, efficiently and smoothly this needs a professional design and a meticulous configuration management that will be run in collaboration with the stakeholders involved. The OBS (Organization Breakdown Structure), developed in accordance to the governance framework and the plans, will assign the correspondent roles and duties so that this platform can be continuously updated.

The OBS will allow to identify and assign roles and this will offer the opportunity not only to charge responsibilities but even to analyse communication channels so that will be clear who has to provide or receive data, by which report and when, really a fundamental to make the entire team perform successfully.
It has to be noted that the proponent has a specific operating unit to help the project management of large-scale projects, which will be involved in the realization of the detailed project.

Long Term Sustainability Beyond PNRR

Vision

In the aim of the call, the Centre has the role of federating and coordinating persistent development process for the country's most significant assets in the HPC, Big data and Quantum Computing domain, in order to unlock the discovery and innovation potential, build a globally attractive ecosystem, strengthen Italian competitiveness and contribute to Europe leadership.
The Centre intends to fulfill this strategic role by multiplying the opportunities for partners and stakeholders, by building capacity and maximizing the socio-economic impact in a competitive environment, by implementing something that does not exist today, without overlapping or competing with individual institutes and research organizations already present, instead creating added value for its partners and for society. The key to the success of the initiative relies precisely in the ability to create these conditions and the appropriate ramp up in the start-up phase (PNRR phase: 2022-2025): a unique ecosystem that is strategic for society and for the country and capable of creating value for their partners by playing a role that individually they could not afford or sustain.
On the basis of this vision, the Centre will have functions of direction, coordination and support, while the research and innovation activity will be entrusted to the Spoke and the partners, who will be able to benefit from the related funding directly or through the Centre. In consideration of this, spoke and affiliates will undertake when fully operational (from 2026 onwards) to support on their ordinary budgets some measures financed in the start-up phase by PNRR funds (contribution in kind) and to finance the centre with an annual contribution starting from 2022 (cash contribution).
During the first year of activity of the Centre, a rewarding mechanism will be identified to incentivize the most active and dynamic spokes, partners and researchers. At the end of the startup phase (2025), based on the monitoring of the activities and objectives achieved, it will be evaluated whether to confirm the current thematic subdivision of the vertical spokes or evaluating any aggregations, deletions or separations and the respective leading institutions.

Pre-conditions to be achieved in the startup phase (2022-2025)

- Ensuring ICSC at the forefront of scientific excellence
- Successfully positioning ICSC in the international landscape
- Being and being broadly perceived as the National platform for research and innovation in HPC, BD & QC
- Integrate ICSC into regional and thematic innovation ecosystems
- Building capacity and stimulating the potential of scientific communities by reducing skill, geographical and gender gaps
- Engaging with Industry, including SMEs, by actively involving the Industrial Board and Ifab foundation
- Involving a large community of network partners to unlock the innovation potential
- Engaging with Public Administration by supporting the digital transformation
- Vastly engaging stakeholders by sharing the socio-economic impact, accounting also indirect impact and intangible assets
- Broadening the range of potential users
- Creating permanent added value for partners beyond PNRR funds
- Ensuring an effective monitoring system and risk management
- Guaranteeing high-quality management team and effective governance, by actively involving the International Advisory Board and promoting regular management assessment

Budget

The ICSC Centre is very innovative, both in the concept, which brings together the most important supercomputing infrastructures and research groups in Italy in the various application fields through a public-private partnership, and in the organizational form based on a hub and spokes structure. This is unique at a European level, which will give the country a great competitive advantage, but which makes it difficult to identify benchmarks for comparison. The closest references, albeit with the necessary differences, are: Barcelona Supercomputing Centre (BSC) in Spain, GENCI in France and the Institute for Advanced Simulation (IAS) of the Julich Centre in Germany. From the available data emerge three realities in strong growth.
In addition to managing an HPC infrastructure, BSC develops research and innovation activities, only in some of the ICSC thematic domains, mobilizing just under 800 researchers. In 2021, BSC presents a balance sheet of approximately 50 MEur, with 38 MEur of revenues from research and innovation activities and an yearly infrastructure investment budget of 8.6 MEur.
GENCI focuses mostly on infrastructure management, in a system like the French one where operate several supercomputing centres, and in 2021 it has a budget of around 40 MEur.
Finally, the IAS is an Institute that develops research and innovation activities and manages the Julich Supercomputing Centre, without its own legal autonomy, being one of the 10 institutes of a large research Centre that mobilizes a total of 6,800 researchers and which in 2020 has presented the revenues for 812 MEur. Therefore, a disaggregated balance sheet figure by IAS is not available, but, assuming that the different 10 institutes in which the Julich Research Centre is organized have a comparable weight, it is possible to estimate a budget of around 80 MEur.

Overall, with all the caution in comparing different realities, this analysis presents a benchmark at European level between 40 and 80 MEur per year. In this context, the long-term sustainability plan was prepared on the basis of the following very conservative assumptions with the aim of showing its sustainability even in any unfavorable conditions:

- the start-up phase ends in 2025, at the end of the PNRR funding coverage, and from 2026 the Centre is fully operational
- in the simulation budgets the funding related to this call (PNRR) is not reported but only additional funding are considered
- any funding linked to other measures of the PNRR M4 (already referred to in the previous sections), to national strategic projects (such as Mirror Copernicus) and to research products that can be released to the market have also not been valued
- the current members confirm their commitment to participate beyond 2025
- the number of partners of the Centre involved expands (+ 23% in terms of contributions)
- IFAB partners (Supporters) is foreseen in increasing reaching the number of 30 and generating a financing capacity for projects equivalent to 2.0 MEur per year (IFAB-projects)
- as described above in the Vision paragraph, the research and innovation activity promoted by the Centre correspond of funding value that partly goes directly to the partners (the corresponding revenues are not reported in the budget because they are not related to the Centre's budget) and partly to the Centre (the corresponding revenues are reported in the budget); since the Centre does not develop research and innovation activities on its own but relies on spokes and partners, the revenues relating to research and innovation correspond to transfers to spokes and partners for the development of the corresponding activities (on average, transfers are equal to 85% revenue); the volume of revenues of the Centre for Research and Innovation is estimated as a whole when fully operational (2026) at 40 MEur, net of income from Living labs and IFAB-projects, which corresponds to less than 4.0 MEur on average for each spoke, with a progressive growth starting from 2023
- the income relating to the Living Labs relates to the contributions to the collaborations with external involved partners
- the revenue relating to the Observatory lab relates to services delivered outside the partnership
- the commitments undertaken by the spokes (spoke / affiliates) on their financial statements in support of ICSC are shown in a separate table

The long-term sustainability plan is summarized in Table 9 and Table 10.

Table 9 shows the normalized operating income budget (from 2026 onwards), including the Sensitivity analysis, and, in the box below, the cost items supported by Spokes / Affiliates on their financial balance sheet.
Table 10 shows the start-up phase (2022-2025) with reference to the additional activities with respect to the PNRR, both in terms of balance sheet and management income report.

FONDAZIONE ICSC (HUB)	MEur	MEur	SENSITIVITY ANALYSIS		Reduction	SENSITIVITY ANALYSIS		Increment
A) VALORE DELLA PRODUZIONE								
A1) CONTRIBUTI SOCI		**8,0**		**7,55**			**8,45**	
Soci attuali	6,5		6,5			6,5		
Nuovi soci (Associate partners)	1,5		1,05		-30%	1,95		+30%
A2) R&I ACTIVITY		**43,0**		**30,1**			**55,9**	
EU projects	10,0		7,0		-30%	13,0		+30%
National & Regional projects	10,0		7,0		-30%	13,0		+30%
Companies	20,0		14,0		-30%	26,0		+30%
IFAB - Projects	2,0		1,4		-30%	2,6		+30%
Living Labs	1,0		0,7		-30%	1,3		+30%
A3) EDUCATION & TRAINING		**5,0**		**3,5**	-30%		**6,5**	+30%
A4) OBSERVATORY		**2,0**		**1,4**	-30%		**2,6**	+30%
TOTALE VALORE DELLA PRODUZIONE		**58,0**		**42,55**			**73,45**	
B) COSTI DELLA PRODUZIONE								
B1) COSTI GENERALI HUB		**3,5**		**3,5**			**3,5**	
staff 30 persons								
e altre spese generali								
B2) TRASFERIMENTI SPOKE/PARTNERS		**40,8**		**28,56**			**53,04**	
EU projects	9,0		6,3			11,7		
National & Regional projects	9,0		6,3			11,7		
Companies	17,0		11,9			22,1		
IFAB - Projects	1,8		1,26			2,34		
Education & Training	4,0		2,8			5,2		
B3) COSTO AGGIUNTIVO LIVING LABS		**0,5**		**0,35**			**0,65**	
B4) COSTO AGGIUNTIVO OBSERVATORY		**1,0**		**0,7**			**1,3**	
B5) ADDITIONAL ICSC INITIATIVES		**3,0**		**3,0**			**4,5**	
a titolo di esempio:								
ICSC Fellowships (50% co-fund)	0,5		0,5			1,0		
ICSC PhDs (50% co-fund)	1,5		1,5			2		
ICSC Innovation grants	1,0		1,0			1,5		
TOTALE COSTI DELLA PRODUZIONE		**48,8**		**36,11**			**62,99**	
EBITDA (A - B)		**9,2**		**6,4**			**10,5**	
Capacità di sostenere ammortamenti nell'esercizio senza intaccare il patrimonio (pari a 16% del valore della produzione)								

SPOKES
Costi sostenuti direttamente da Spokes/Affiliates sui propri bilanci a supporto di ICSC
(i) shared researchers with ISCS
(ii) at least 10 permanent positions per spoke
(iii) operational costs for PNRR infrastructure investments
(iv) co-funding new infrastructure investments

Table 9 The normalized operating income budget (from 2026 onward)

STATO PATRIMONIALE IN ADDITION TO PNRR FUNDS	2022	2023	2024	2025
ATTIVO				
ATTIVO CIRCOLANTE				
CREDITI	0,0	2,4	5,3	7,9
DISPONIBILITA' LIQUIDE	6,5	13,7	22,8	31,5
TOTALE ATTIVO	**6,5**	**16,1**	**28,1**	**39,4**
PASSIVO				
PATRIMONIO NETTO				
FONDO DI DOTAZIONE	0,5	0,5	0,5	0,5
FONDO DI GESTIONE	6,0	13,5	21,6	31,3
AVANZO	0,0	0,1	1,6	1,2
TOTALE PATRIMONIO NETTO	**6,5**	**14,1**	**23,8**	**33,0**
DEBITI	0,0	1,9	4,4	6,5
TOTALE PASSIVO	**6,5**	**16,1**	**28,1**	**39,4**

RENDICONTO ECONOMICO DI GESTIONE IN ADDITION TO PNRR FUNDS	2022	2023	2024	2025
A) VALORE DELLA PRODUZIONE				
A1) CONTRIBUTI SOCI	0,5	0,0	0,0	0,0
A2) R&I ACTIVITY	0,0	9,3	22,1	33,0
A3) EDUCATION & TRAINING	0,0	2,0	3,5	5,0
A4) OBSERVATORY	0,0	0,5	1,0	1,5
TOTALE VALORE DELLA PRODUZIONE	**0,5**	**11,8**	**26,6**	**39,5**
B) COSTI DELLA PRODUZIONE				
A) COSTI GENERALI HUB	0,5	2,0	2,5	2,5
B) TRASFERIMENTI SPOKE/PARTNERS	0,0	9,6	21,8	32,3
C) COSTO AGGIUNTIVO LIVING LABS	0,0	0,0	0,0	0,0
D) COSTO AGGIUNTIVO OBSERVATORY	0,0	0,0	0,0	0,0
E) ADDITIONAL ICSC INITIATIVES	0,0	0,0	0,0	3,0
TOTALE COSTI DELLA PRODUZIONE	**0,5**	**11,6**	**24,3**	**37,8**
RISULTATO (A - B)	**0,0**	**0,2**	**2,3**	**1,7**
IMPOSTE	0,0	0,1	0,7	0,5
AVANZO	0,0	0,1	1,6	1,2

Table 10 The start-up phase (2022-2025) in terms of balance sheet and management income report.

When fully operational, the estimated production value is equal to 58 MEur (slightly higher than the BSC benchmark equal to 50 MEur). As can be seen from Table 9, the Centre is able to support its own initiatives for a total of 12.2 MEur per year (21% of the value of production), in addition to the 2 MEur of IFAB-projects indirectly related to the Centre. This annual capacity has been declined in the economic simulation into: 3.0 MEur for ICSC Initiatives (for example: ICSC Fellowships, ICSC PhD and ICSC Innovation grants) and 9.2 MEur (16% of the production value) for investments for the evolution of the infrastructure (supporting depreciation or co-financing investments of the spokes).

In addition, each spoke with its Affiliates undertakes to: (i) maintain the full time allocation of its researchers at the Centre, (ii) to consolidate indefinitely at least 10 positions per spoke to give continuity to the actions initiated with the fixed-term recruitment of the start-up phase, (iii) cover the operating costs of the infrastructures acquired in the start-up phase for their entire useful life cycle beyond 2025, (iv) contribute in co-financing to the renewal of the infrastructures. The total cost of items (i), (ii) and (iii) can be conservatively estimated at 20 MEur per year

(less than 2 MEur per spoke on average). Item (iv) relating to investments for the renewal of the infrastructure is discussed below.
When fully operational, the Centre is expected to have a staff of 30 employees, of which 19 to support all Hub activities, plus 11 units in support of the spokes (one per spoke). The overall costs in 2026, including personnel expenses, amounted to a total of 3.5 MEur. To these are added 0.5 MEur of additional costs for the Living labs and 1.0 MEur of additional costs for the Observatory. These costs are consistent with the expected revenue for the respective items.

The sensitivity analysis of the economic report (Table 9) was conducted assuming two scenarios: -30% and + 30% of assets, keeping only the contribution of current shareholders unchanged between revenues and costs. The two scenarios in terms of production value remain in the range of the benchmarks described above (40 MEur - 80 MEur). As can be immediately appreciated, the Centre is able to recover any deviations or delays in full operation: even with a 30% reduction in activities, the Centre can support its own initiatives for 9.4 MEur, divided in the simulation into 3.0 MEur for ICSC Initiatives and 6.4 MEur to support investments. If the activities were higher than 30%, the ability to promote own initiatives grows to 15 MEur, divided in the simulation into 4.5 MEur for ICSC Initiatives and 10.3 MEur to support investments.

In the start-up phase (Table 10) it is assumed that, given the constraints present in the call for expression of interest, a part of the operating costs are covered with the additional resources of the Centre (and therefore will not be charged to PNRR funds) and that , as a precaution, the first resources for additional activities occur only in the 2023 budget, to then grow progressively until 2026. For the latter, it has been assumed to collect in the current year on average 80% of the value and the remaining 20% in the following year. The costs related to the Transfers to Spoke / Partners were shared similarly. To give continuity to the Centre's action, it has been assumed that the ICSC Initiatives promoted on its own funds (not the PNRR) will be activated from 2025.

Overall, Table 10 shows that at the end of the start-up phase (31/12/2025) the net equity is estimated to be 33 MEur (of which 30 MEur from shareholder contributions) and a cash availability on the same estimated date to 31.5 MEur. This guarantees solidity to the Centre for the post PNRR phase.

The annual amount for investments is overall quite adequate, compared to the ambition of ICSC and the initial investment with PNRR funds, to keep the Tier1 systems constantly updated. In fact, the planned investment with PNRR funds on the Tier1 system, thus excluding the investment in Tier0 (Leonardo upgrade -digital and quantum- and participation in the exascale consortium) and the upgrade of the Garr network (which has a of long-term operations cycle), amounts to approximately 90 MEur. By applying the same approach as EuroHpc, the annual investment quota (9.2 MEur) used as a 50% co-financing to hosting institutions over a period of 5 years (average cycle of Tier1 operations) reconstitutes the initial investment.
The renewal of the Tier0 Leonardo system (financed 50% by EuroHpc and 50% by the Italian Government for a total of 240 MEur (CAPEX + 5 years of OPEX), which is expected to be upgraded for a further 50 MEur (CAPEX + OPEX funded at 35% by EuroHpc and 65% with funds from the PNRR), as has happened up to now and as it happens in other European countries, will necessarily have to provide for joint action between EuroHpc and the Italian government. The Centre and its members might be able to funds any update or evolution, similarly to the Leonardo upgrade (50 MEur), with a possible contribution scheme: 50% EuroHpc, 25% ICSC and 25% hosting institution / ISCS members. As shown in Table 10, the Centro will have a large net worth (33MEur) to support the investment (estimated at 16.2 MEur).

Based on the previous observations, if we consider the four-year period following the start-up phase (2026-2029), the Centre will directly and indirectly mobilize the following resources every year:

- value of production 58MEur
- costs incurred by the spokes (i), (ii) and (iii) 20 MEur
- co-financing of investments in the renewal of infrastructures 9.2 MEur

for an annual total of approximately 87 MEur, which corresponds to 348 MEur over the four-year period. Adding to this figure the net assets accumulated in the start-up phase, we obtain: 381 MEur, which substantially corresponds to the PNRR investment for the start-up of the Centre.
Finally, it should be emphasized that the described framework is extremely conservative because no activity linked to other PNRR initiatives has been considered as well as to national strategic projects in the start-up phase, for which the Centre will be a critical infrastructure asset.

Budget

The **CN budget (“agevolazione”) is 319.99 M€**, and it has been fully detailed by the Hub, the Spokes and by the Affiliates. It is categorized into budget for Personnel (either shared or hired specifically for the centre), budget for Infrastructure, budget for Open Calls, budget for Innovation (Demonstrators and Increase of TRL and Innovation, Education and training, Entrepreneurship, Knowledge Transfer, Policy, Outreach), budget for PhD grants and finally Indirect Costs. In phase 2, minimal changes have been made, that are not reflected here in new plots and tables.
The private partners participate with their self-financed personnel, bringing additional value to the CN, thus not appearing directly in the budget. The CN will benefit also from the annual economic contributions of the Founders and plan to attract matching funds from European initiatives and other PNRR initiatives, such as Research Infrastructures and Innovation Ecosystems, as well as to attract National and International innovation initiatives and Venture Capitals.
Almost half of the total budget is in the **Infrastructure** category, in great part in the Spoke 0 and in a more limited amount also in some of the thematic Spokes; the amount given in the budget represents roughly 50% of the whole planned Total Cost of Ownership (TCO). The other 50% will be granted either by the hosting entity, or by co-funding from European initiatives (for example, EuroHPC). The computing resources are made available to internal and external users through the Supercomputing Access Committee, based on access policy regulated according to the European Charter of Access for Research Infrastructures and sketched in Figure 13 and detailed in the governance section.

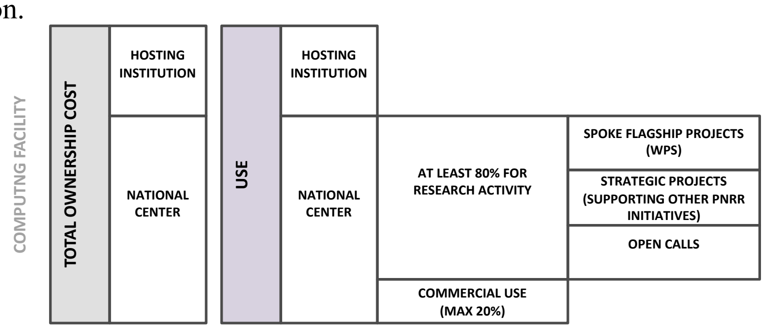


Figure 13 Policy for resource access

The **Personnel** budget amounts to about 15% of the total budget in the case of shared staff from partners, and 18% of the total budget for new personnel to be hired with the procedures explained in the last part of this document. Since the positions are budgeted for 3 years, it is expected that a part of the hired personnel will be accounted for only partially to the project, due to the misalignment of hiring and project dates; hence, a part will be co-funded by the partners, resulting in an actual number of researchers involved in the CN activities higher than what results from simple budget considerations. The same is valid for **PhD grants**, which will be co-funded. A direct and positive effect of co-funding is the natural extension of the positions beyond the 3 years of the project, hence helping in implementing the CN sustainability after the end of the PNRR funding.
The **Open Calls** are a key instrument for the success of the CN. The budget allocated to support Open Calls is 32 MEur, but CN in-kind resources will be made available as well, as described below. The goals of Open Calls are: (a) promoting access to computing resources of Academia, Industry and Public Administration, (b) stimulating the research potential of Academia, (c) stimulating the innovation potential of Industry, including SMEs and innovative start-ups, and Public Administration. Various targeted open calls are planned, with the following modalities: (a) open Research and Innovation, (b) free of charge (opportunely mixing depending on the target: computing resources, high level support, research support and training). They may include or not a financial grant. The activities supported by the Open Calls are: a) getting on advanced computing, b) optimization, scaling and testing, c) project use-cases, d) research and software development.
The **Innovation budget** is another key instrument for the success of the CN. Funds for Innovation (almost 31.8 MEur allocated 43% to the Hub and 57% to the Spokes) are intended for activities which foster the transfer of technologies, knowledge and people from the academia to the productive sectors; they include a budget for the deployment of research / industry shared demonstrators, grants to raise closer to production needs the TRL of research solutions, funds for training activities geared towards a spillover of knowledge to companies and to **address skill gaps**, grants for Proof of Concepts and pre-seeds to support new start-ups or spin-offs (with sizes

of 100 KEur, 200 KEur and eventually 500 KEur), as well as for actions to support entrepreneurship culture and call for ideas. Moreover, Innovation budget is planned to be used to support the start-up of Life Long Learning Programs delivered by partners and as matching funds to attract and engage top-class international scientists.
The budget for Innovation is currently not assigned directly to industrial partners, but part of the Spoke Leader Institution and Hub budgets; at the beginning of the activities it will be allocated to the partner companies.

Finally, a key role in this context will be played by the **ICSC Observatory on Supercomputing Trends and Applications** which is discussed in detail in the governance section.
The overall budget, built from the ground up via a detailed analysis of the activities, is shown broken down by categories in Figure 14. Figure 15 shows the same information per each Spoke.
It is noteworthy to discuss the high fraction of the budget assigned, in each Spoke, to the Lead. On one side, this is because some budget categories (Innovation, PhD, Open Calls) have been left undivided on the lead; on the other side, the Spoke lead will sustain more operating costs and hence was granted with a higher budget.

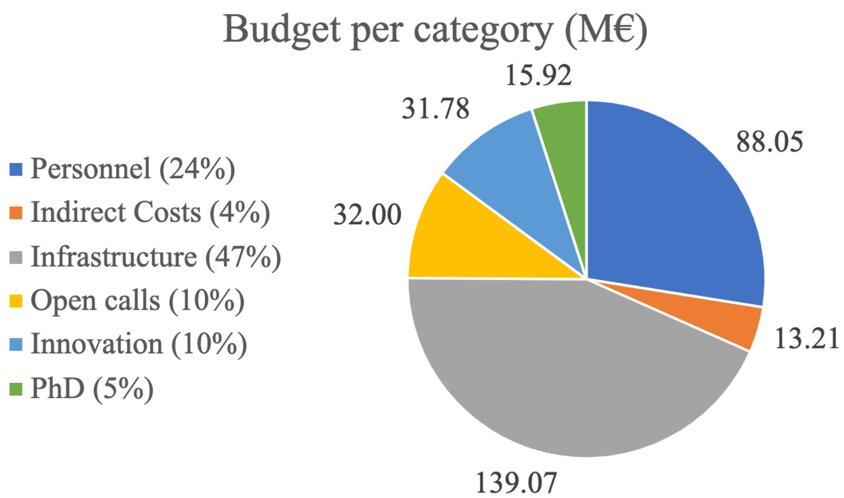
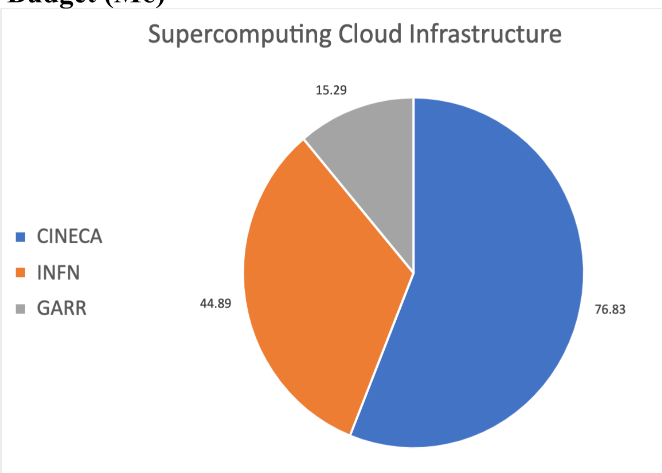
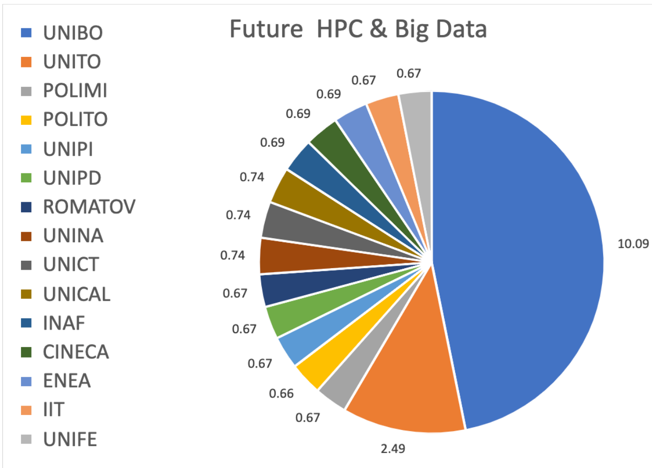
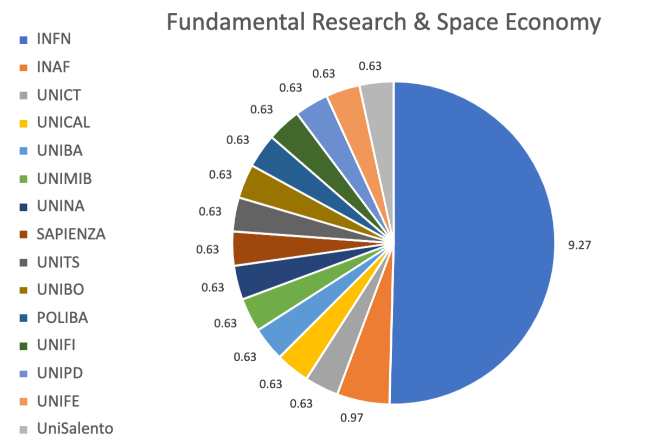


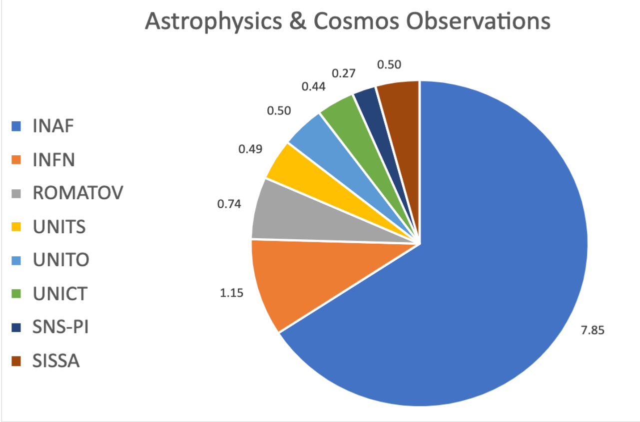
Figure 14 Budget breakout per expense category

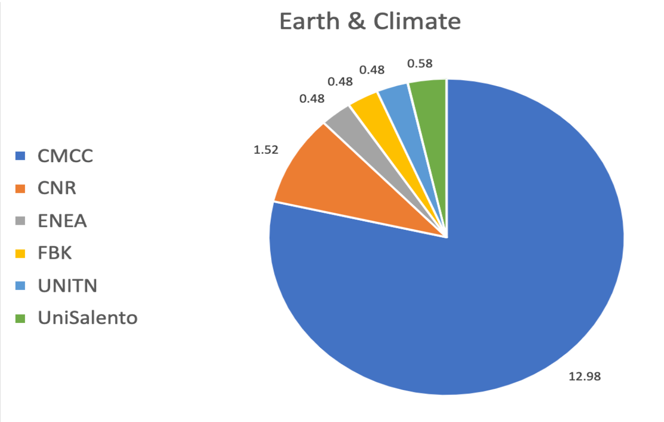
Budget (M€)

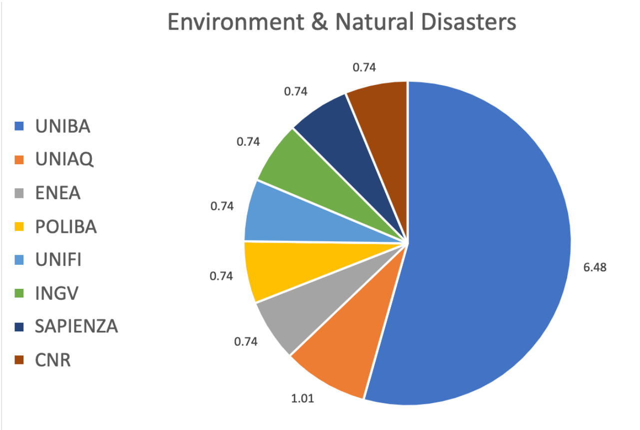


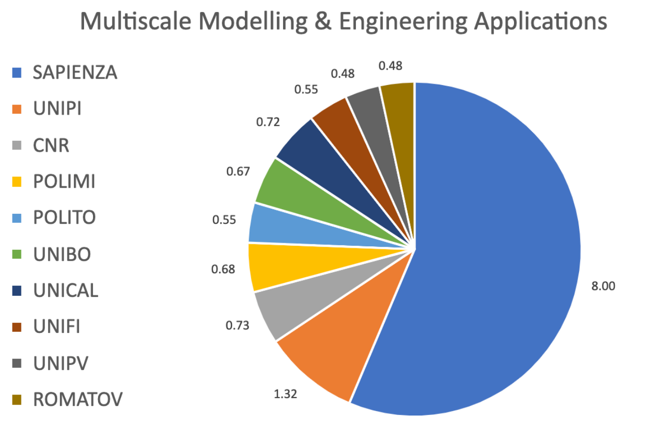


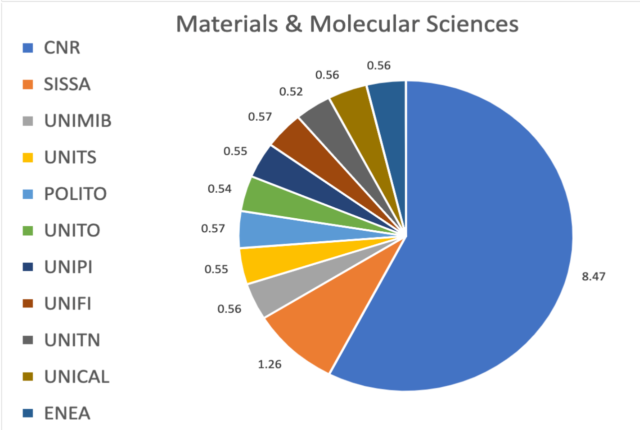

Fundamental Research & Space Economy
INFN
INAF
UNICT
UNICAL
UNIBA
UNIMIB
UNINA
SAPIENZA
UNITS
UNIBO
POLIBA
UNIFI
UNIPD
UNIFE
UniSalento
9.27
0.97
0.63
0.63
0.63
0.63
0.63
0.63
0.63
0.63
0.63
0.63
0.63
0.63
0.63

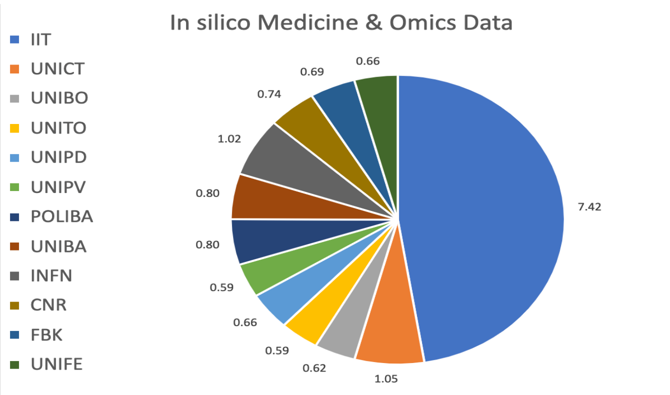

Astrophysics & Cosmos Observations
INAF
INFN
ROMATOV
UNITS
UNITO
UNICT
SNS-PI
SISSA
7.85
1.15
0.74
0.49
0.50
0.44
0.27
0.50

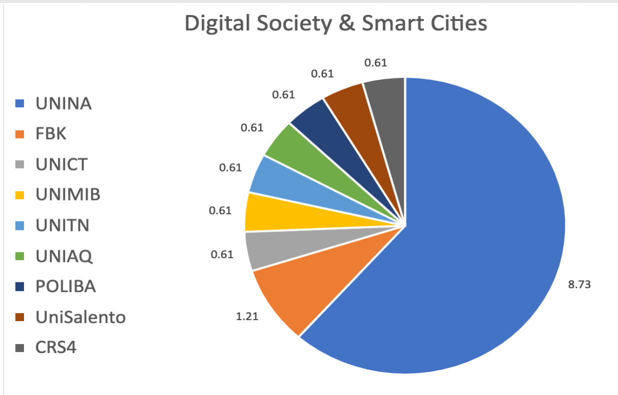

Earth & Climate
CMCC
CNR
ENEA
FBK
UNITN
UniSalento
12.98
1.52
0.48
0.48
0.48
0.58


Environment & Natural Disasters
UNIBA
UNIAQ
ENEA
POLIBA
UNIFI
INGV
SAPIENZA
CNR
6.48
1.01
0.74
0.74
0.74
0.74
0.74
0.74


Multiscale Modelling & Engineering Applications
SAPIENZA
UNIPI
CNR
POLIMI
POLITO
UNIBO
UNICAL
UNIFI
UNIPV
ROMATOV
8.00
1.32
0.73
0.68
0.55
0.67
0.72
0.55
0.48
0.48


Materials & Molecular Sciences
CNR
SISSA
UNIMIB
UNITS
POLITO
UNITO
UNIPI
UNIFI
UNITN
UNICAL
ENEA
8.47
1.26
0.56
0.55
0.57
0.54
0.55
0.57
0.52
0.56
0.56


In silico Medicine & Omics Data
IIT
UNICT
UNIBO
UNITO
UNIPD
UNIPV
POLIBA
UNIBA
INFN
CNR
FBK
UNIFE
7.42
1.05
0.62
0.59
0.66
0.59
0.80
0.80
1.02
0.74
0.69
0.66


Digital Society & Smart Cities
UNINA
FBK
UNICT
UNIMIB
UNITN
UNIAQ
POLIBA
UniSalento
CRS4
8.73
1.21
0.61
0.61
0.61
0.61
0.61
0.61
0.61

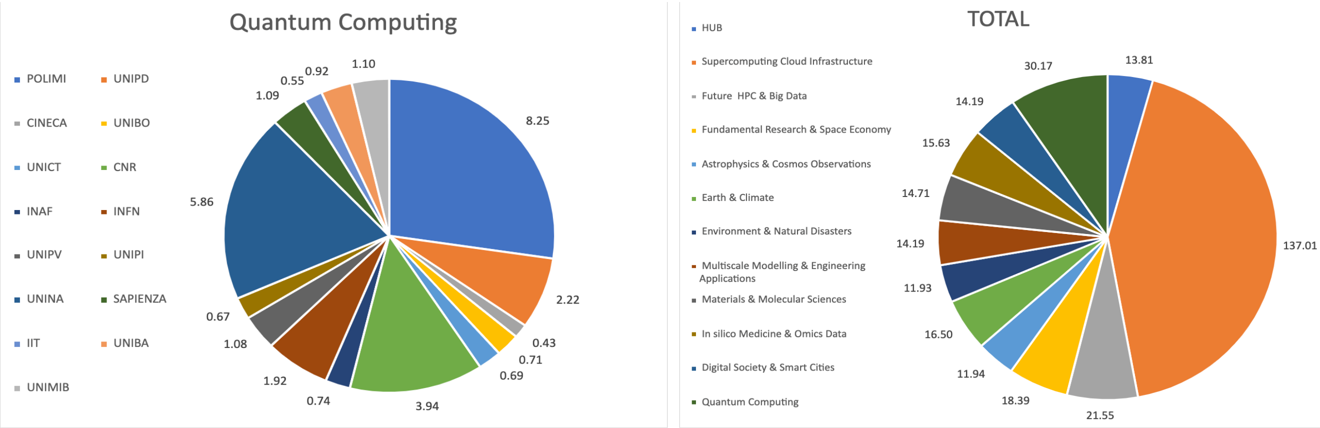


Figure 15 Budget per Spoke and per Affiliate

The budget in the Hub (14 MEur) is reserved for: innovation activities as described above (mainly to be implemented together with the spokes) and for further investments on small Partners' facilities.
The breakdown is also shown in Table 11.

	HUB	0 Supercomputing Cloud Infrastructure	1 Future HPC & Big Data	2 Fundamental Research & Space Economy	3 Astrophysics & Cosmos Observations	4 Earth & Climate	5 Environment & Natural Disasters	6 Multiscale Modelling & Engineering Applications	7 Materials & Molecular Sciences	8 In silico Medicine & Omics Data	9 Digital Society & Smart Cities	10 Quantum Computing	TOTAL	Fraction of total budget
Management (M€)	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0%
Investments (HUB will redistribute to spokes) (M€)	0.0	116.5	3.2	0.0	0.0	6.4	0.0	0.0	0.0	0.0	0.0	0.0	126.0	39%
Fraction of investments in southern regions (%)	0%	40%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	42%	
Staff (M€)	0.0	1.9	5.6	5.6	2.9	2.8	3.6	3.7	4.5	5.7	3.7	6.5	46.4	(15%)
New recruitment (M€)	0.0	11.5	4.3	4.3	2.8	1.8	2.2	3.2	3.0	2.4	2.6	3.5	41.6	(13%)
Indirect costs (M€)	0.0	2.0	1.5	1.5	0.9	0.7	0.9	1.0	1.1	1.2	0.9	1.5	13.2	(4%)
Total Personnel cost (M€)	**0.0**	**15.4**	**11.4**	**11.4**	**6.6**	**5.2**	**6.6**	**7.9**	**8.7**	**9.3**	**7.3**	**11.5**	**101.3**	**32%**
Fraction of personnel cost in southern regions (%)	0%	30%	27%	44%	29%	40%	57%	13%	23%	36%	67%	23%	34%	
Other expenses (M€)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.3	11.7	13.0	4%
Fraction of other expenses in southern regions (%)	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	43%	49%	
Open calls (M€)	0.0	0.0	3.2	3.2	3.2	3.2	3.2	3.2	3.2	3.2	3.2	3.2	32.0	10%
Fraction of open calls in southern regions (%)	0%	0%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	
Demonstrators & use cases (M€)	0.0	4.5	1.8	1.8	1.0	0.7	1.0	1.6	1.3	1.4	1.1	1.8	18.0	6%
Fraction of demonstrators & use cases in southern regions (%)	0%	40%	40%	40%	40%	40%	40%	40%	40%	40%	40%	40%	40%	
Other innovation initiatives (M€)	12.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	12.8	4%
Fraction of other innovation initiatives in the south (%)	35%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	35%	
Budget for PhD (M€)	0.0	0.6	2.0	2.0	1.2	1.0	1.2	1.4	1.5	1.7	1.3	2.0	15.9	5%
Fraction of PhD budget in southern regions (%)	0%	40%	40%	40%	40%	40%	40%	40%	40%	40%	40%	40%	40%	
TOTAL BUDGET (M€)	**13.8**	**137.0**	**21.5**	**18.4**	**11.9**	**16.5**	**11.9**	**14.2**	**14.7**	**15.6**	**14.2**	**30.2**	**319.9**	100%
Fraction of total budget in southern regions (%)	**35%**	**39%**	**29%**	**44%**	**37%**	**65%**	**52%**	**27%**	**32%**	**40%**	**62%**	**37%**	**40%**	

Table 11 Budget composition per spoke and per budget category

The work plan is bold and ambitious, and requires coordination between many subjects (for example, between the managers of the infrastructure and the scientific research teams), including a **strong administrative supervision** as needed to handle tender procedures for almost 150 M€. The affiliates have been evaluated and selected using metrics specifically designed not only to guarantee top scientific qualities and profiles, but also to ensure strong project and execution skills, as proven by the past successful handling of project of size (per affiliate) similar to those relevant for the CN.

Equal gender opportunities promotion

From a programmatic point of view that looks to the next three years and beyond, the actions and projects that the CN intends to promote in order to favour equal gender opportunities and the enhancement of other variables such as age, culture, physical ability, sexual orientation, are presented here, contributing to the creation of a frank, inclusive and effective research environment.

The presentation is divided into 4 sections and areas of intervention:

1. Gender balance in top positions and in decision-making bodies;
2. Gender balance in recruitment;
3. Integration of the gender dimension and intersectionality in research;
4. Actions to contrast stereotypes in the organisation's culture.

For each section, sheets will be presented with the objectives, actions, some targets for monitoring the actions themselves, the link with the SDGs of the UN 2030 Agenda and the Timeline.

1. Gender balance in top positions and in decision-making bodies

Objective	Achieve gender balance in top positions and in decision-making bodies.
Actions	• Setting up of elective mechanisms for governing bodies providing for balance in candidacies and the constraint for voters to express a double preference, in compliance of the gender variable. • Identification of gender balancing measures in management roles and research groups (e.g. gender alternation for management/vice-management or responsibility/co-responsibility of research groups).
Indicators for monitoring	At least 40% presence of the minority gender in governing bodies. Alternation in decision making roles.
Link with SDG Agenda 2030	SDG 5 Gender equality. SDG 10 Reduced inequalities. SDG 16 Peace and justice, strong institutions.
Timeline	2022

2. Gender balance in recruitment

Objective	Increase the percentage of women in the recruitment in the Super Computing and Quantum Computing areas (or alternatively in the STEM areas).
Actions	• Provide the members of selection committees with information material about unconscious bias in recruitment. • Foresee gender balance in selection committees. In case of draws foresee balanced lists. • Publication of a Call for Interest to attract female candidates in view of direct appointments. • Foresee direct appointments addressed exclusively to female scholars and researchers for sector with female population lower than 40%.
Indicators for monitoring	Recruitment of at least 40% of female staff.
Link with SDG Agenda 2030	SDG 5 Gender equality.

3. Integration of the gender dimension and intersectionality in research

Objective	Integration of the gender dimension and intersectionality in research.
Actions	• Organisation of dedicated sessions on gender equality, inclusion and well-being in Supercomputing and Quantum Computing research, during CN events. • Mentoring programs for female researchers (building relationships between senior and junior researchers) that identify suitable strategies for overcoming obstacles in career paths.
Indicators for monitoring	Organisation of an annual workshop. Setting up of mentoring programs.
Link with SDG Agenda 2030	SDG 5 Gender equality.
Timeline	Workshop 2022. Mentoring programs from 2023.

4. Actions to contrast stereotypes in the organisation's culture:
   Objective 4.1

Objectives	Strengthening a shared culture on the importance of gender equality and the effectiveness of gender balance.
Actions	• Tri-annual redaction of a gender balance document containing qualitative and quantitative data for the entire community of the centre, identifying indicators and indices for a context analysis, also in relation to the European panorama. • Publication of the document through the communication channels of the centre.
Indicators for monitoring	Tri-annual approval of the gender balance document by the governing bodies and monitoring of the indices.
Link with SDG Agenda 2030	SDG 5 Gender equality. SDG 10 Reduction of inequalities.
Timeline	Data collection: 2022. Redaction of the first document: 2023.

Objective 4.2

Objectives	Integration of the gender balance in the organisation of the centre and in the management of its activities.
Actions	• Examine the context analysis of gender balance and its indicator. • Form a working group.
Indicators for monitoring	Preparation and approval of the Gender Equality Plan (2024-2027) in 2024 by the governing bodies.
Link with SDG Agenda 2030	SDG 5 Gender equality. SDG 10 Reduced inequalities. SDG 11 Sustainable cities and communities.
Timeline	Gender balance analysis: 2023. Preparation of the GEP (2024-2027): 2024.

Involvement of young researchers

The Law Decree n. 77 of 31 May 2021 dedicates article 47 to equal opportunities and work inclusion, with a particular focus on the inclusion of disabled people, on gender policies, but also on generational ones. Furthermore, it is desirable that attention to these multiple aspects go hand in hand, so that professional contexts, and even more so those that aspire to have a future role in the circular economy, give an effective and satisfactory response to the new generations, in terms of enhancement of merit, inclusiveness, social sustainability and personal well-being.

The actions that the CN intends to take in order to involve scholars who have achieved their PhD since no more than 10 years (excluding maternity and parental leave) and to attract young talents from other countries (EU and non-EU), can be summarized around 4 thematic axes.

1. Strategies for attracting young researchers based on their curricula;
2. Recruitment methods that pursue the objective of youth employment also through the simplification of the procedures for participating in calls and policies for access to the scientific community;
3. Introduction of concessions for residential care, innovative working methods, actions that favor care activities and work well-being;
4. Monitoring of recruitment policies for young talents.

1. Strategies for attracting young researchers based on their curricula. The CN intends to publish, via its communication channels (Institutional Portal as well as Social media) "Calls for Interest" every six months, in order to allow young people from EU and non-EU countries to express their interest in working within the Centre. These calls will be published in English and will allow the creation of databases containing: The curricula of potential candidates, divided by areas of interest; Competitive research projects in which one has participated or for which the first selection phase has been passed; Availability to participate in thematic interviews in the presence of experts during which the candidates will present the lines of development of their research. At the end of each Call, and following the interviews, the candidates of greatest interest will be kept informed of the Recruitment Calls issued by the Centre. The Calls will also have the purpose of gathering the willingness of European and non-European scholars to spend a period of 6 months or one year as Visiting Researcher at the Centre, taking advantage of grants and fellowships.

2. Recruitment methods that pursue the goal of youth employment also through the simplification of the methods of participation in calls for tenders and access policies to the research community. The Centre intends to issue calls reserved for scholars who have obtained their research doctorate since no more than 10 years (except for maternity, parental or other leave). The notices will be issued in Italian and English and will include an IT platform for filing applications, as well as for carrying out the procedures. The selection committees will be created according to the principle of gender balance and will be made aware (with videos and information material created ad hoc) on the presence of unconscious bias (with respect to the gender variable but also other variables such as the starting culture of the candidates, physical ability, age, sexual orientation) which can affect recruitment and strategies to avoid them. In order to simplify the procedures for participating in the calls, standard parameters will be established that a doctorate must respect to be able to automatically recognize the title without having to go through the complex legal request for formal equivalence. The parameters for recognition may include the duration of the doctoral course, the selection methods, the activities envisaged by the doctoral curriculum, the methods of presentation of the final research work. A database of prestigious non-university research institutions will also be established within which the candidates may have carried out a research period equivalent to that of the doctorate and which may constitute a sufficient qualification to participate in the calls. As regards the selection, international standards of transparency will be promoted in the selection criteria and methods such as those provided for research staff by the EU actions (Marie Sklodowska Curie: "The overall approach to keep the call as inclusive as possible, in order to attract as many talented candidates as possible. For example, keeping a limited number of eligibility criteria, will lead to a higher number of applicants that, even if not selected, will receive comments on their CV and backgrounds that can be useful for personal improvement and better results in future selection processes "). A reception path for new researchers will also be developed aimed at immediately fostering a sense of belonging to the research community through paths aimed at getting to know the Governing Bodies, the Centre's Code of Ethics, and its Research Integrity rules. Team- and community-building initiatives will also be organized, as well as mentoring courses to create relationships between senior and junior researchers. Finally, a short course (possibly MOOC) will be offered to new hires on issues of inclusion and diversity and the impact they may have on working efficiency.

3. Introduction of concessions for residential care, innovative working methods, actions that favour care activities and work well-being. The attractiveness of young talents can certainly increase by providing a series of incentives that induce valuable scholars to participate in the calls issued by the Centre. It is therefore intended to facilitate:

   a. The search for housing through a dedicated service (possibly via agreements with estate agents);
   b. The enrolment of children in public schools and/or international school courses through a framework agreement with regional school offices;
   c. Learning of the Italian language through agreements with Italian language schools for foreigners.

It is also important to provide for flexible ways of carrying out one's research work and to allow simplified access to non-compulsory parental leave for both parents, promoting awareness policies also for paternity leave. In this sense, a mapping of the phenomenon of parental or other leave will be carried out annually to introduce further actions aimed at work well-being and facilitating the role of care.

4. **Monitoring of recruitment policies for young talents.** The Centre intends to carry out an annual monitoring of the generational variable by collecting data on the average age of the new hires and their subsequent career progress. This data collection will represent the starting point for subsequent reward measures to be introduced in order to lower the average age of access to the career and top research roles. On the basis of this monitoring, indicators will be studied that measure the progress of recruitment policies over time in a qualitative and quantitative manner.

Involvement of private parties

As previously mentioned, the objectives of the CN include enabling support for excellence in scientific research, technology transfer and persistent support for innovation, with a clear commitment to raise the level of TRL of the outcomes obtained from research and development in order to prepare solutions and applications of immediate industrial value. This action could not be completely effective without a direct involvement of the economic and industrial system, including the interaction processes with the system of SMEs, of technology providers and of technology ISVs. To this purpose, different actions are planned.

Firstly, the direct involvement as CN founders of a qualified group of large national production and service companies, that exhaustively represents the various industrial segments of strategic socio-economic impact at the national level. The large companies that already have significant HPC infrastructures such as Eni, Leonardo Company, Fincantieri, ThalesAlenia, have joined with the interest to the latest generation computing systems and emerging technologies, as well as to be part of the CN national federated infrastructure, capable of responding to extreme situations, as in the case of support to scientific research for the design of new drugs effective against the COVID-19 pandemic, or in perspective to respond to extreme natural events. In addition to the infrastructural aspects, the large companies have also shown interest in sharing methodological scientific skills, such as those expressed by the research and academia, and advanced professional skills for the engineering of application solutions as part of their production workflows and support for the exploitation of new systems, as those that will soon be based on newly accelerators technology such as FPGA, neuromorphic or quantum computing accelerators. This need regarding high-level specialist support has proven to be relevant not only for the industrial system with a consolidated tradition in the computing field, but also for the service and infrastructure companies, called to face the needs of strategic structural evolution for the national system. Autostrade, Ferrovie dello Stato, Terna, Sogei, Engineering, Humanitas, UPMC, UnipolSai, Intesa San Paolo have joined the CN because they are interested in sharing their expertise for creating value from data, taking up and integrating new technologies, developing new services, elaborating the use cases of the future, supporting innovation and new entrepreneurship. Finally, in the sector of PA Urban Innovation Foundation has been engaged as a key partner to develop and test digital twins for cities.

Secondly, a structured collaboration with the CN of other companies, including medium-sized and small-sized innovative enterprises, is planned through the pivotal role of IFAB Foundation, that will operate similarly ETP4HPC, BDVA or QUIC for EuroHPC. The ecosystem of SMEs is aware of the enabling importance of numerical simulation and big data processing workflows, but not immediately able to profit from them due to skill limits and economic barriers to exploit new technologies. So, IFAB as a CN Founder will play an aggregator role for structured partnerships, including ISV’s application providers and technology providers.

Thirdly, strategic partnership agreements to create joint living labs with the main technology providers are planned (some of them have been already engaged) and network partners will be involved to jointly support and promote innovation (see innovation network partners).

Finally, the open calls and innovation grants will further broaden the involvement of private partners and increase the overall impact.

Engagement with private companies

The current industrial revolution is characterized by an increasingly pervasive use of data, digital technologies, and interconnected components, as much as it is happening in data intensive research. R&D programs, testbeds and proofs of concept are at the heart of the efforts academia and industries are executing; while the specific domains appear from a distance different as methodologies and final goals, they are indeed very similar, to the extent that shared efforts exist since long for example in the context of EU projects.

Four different types of broad-brush activities have been identified for collaboration:

1. Industries will provide requirements and specifications drawn from industrial uses cases and application scenarios, guiding development choices and the design of validation strategies; the industrial use cases can be allowed, under data management plan policies, to access research data sources and platforms.
2. During the development phase, companies will be tracking the status of the research prototypes and the matching with their requirements as in the previous point; they will share technologies, design solutions (for example, AI solutions, software frameworks and data analytics techniques), and support and training opportunities.
3. During the benchmarking and validation phase, companies will participate to the assessment of the final prototypes from both academia and industry, via shared testbeds and proof of concept systems deployed either on the CN infrastructure, or made available in-kind, or procured by the spokes.
4. During the wrap-up final phase, companies will collaborate on the preparation of reports and white papers, showcasing general solutions beyond the use cases and the partners of the CN. Possibly, solutions with higher TRL will be guided through the first steps of pre-commercial and commercial exploitation.

The specific activity domains and use cases, when already agreed at a sufficient level, can be found in the specific spoke descriptions; they range from financial use cases, technological use cases in manufacturing, large scale data collections from multiple sensors, to analysis of heterogeneous data.

Funding for industrial involvement is expected via two instruments of the CN:

- Open calls, funded on the budget of each spoke, and directed to academic and industrial partners outside the CN;
- Innovation grants, directed prevalently to companies, and used to support co-development activities, co-design, shared testbeds and proofs of concept.

Detailed spoke-level involvement with private companies

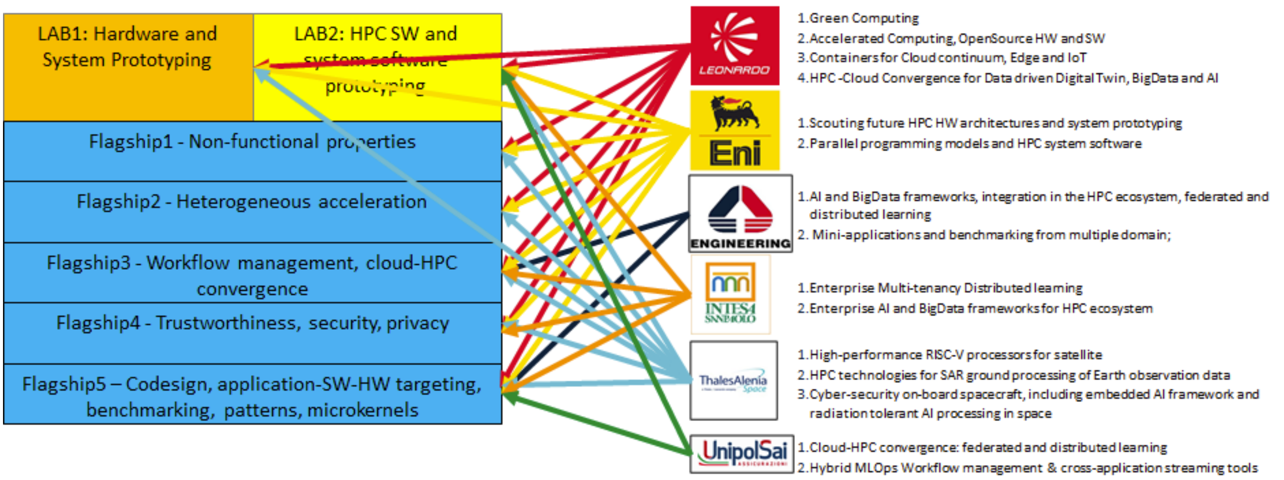
Spoke 0

The partnership includes qualified industrial entities, which not only joined the project by assuming an obligation of economic participation, but were also selected to cover the application spectrum represented by the thematic spokes, in order to associate some large private partners with each spoke: Leonardo Company and Fincantieri for industrial manufacturing and engineering applications; Eni for energy; Humanitas and UPMC for health and life sciences; Terna, Autostrade per l'Italia e Ferrovie dello Stato for complex networks and infrastructures; UnipolSai and Intesa San Paolo for technologies of finance and economic services; and finally Sogei and Engineering for complex information systems and big data management and processing services for the public administration. This specificity will be particularly enhanced in order to bring the solutions developed by the spokes to the appropriate industrial level of TRL.

With regard to the enhancement of the GARR network infrastructure, the integration of the HPC data canters of the private partners will also be envisaged, with particular reference to Eni's Green Data Centre where the installation of an HPC system of the infrastructure of the National Centre is planned and in order to also integrate Eni's HPC system into the national federated HPC infrastructure. Leonardo company HPC infrastructure will be integrated in the GARR-T infrastructure and SOGEI system cloud facility will be integrated as well. CINECA foresees the installation of a Tier1 system in Milan, hosted in the Eni green data centre.

Spoke 1

Several private partners are involved in the planned activities of the Future HPC spoke, as summarized by the figure above which provides a synthetic but quite detailed view of the key priorities and interactions. The spoke implementation actions (Spoke labs and Flagship projects) are depicted on the left side. The strategic priorities indicated by the companies are listed on the right site. Arrows indicate the thematic connections between strategic topics and activities, and they imply the involvement of the companies in technical collaborations with the research partners involved in the corresponding spoke implementation actions.



Companies will contribute to the milestones in the first phase of the spoke's work plan by providing requirements and specifications drawn from industrial use cases and application scenarios. This information will be included in the corresponding reports (first and second milestones, at months 4 and 8). During the implementation phase (all the milestones between months 12 and 28), companies will contribute in two main ways; (i) by tracking the status of research prototypes and the matching of the initially specified requirements, possibly running independent evaluations and trials, (ii) by benchmarking the research prototypes with state-of-the-art commercial solutions.

In the last phase of the project (the last two milestones, at months 32 and 36), the companies will be involved in a final assessment of the prototypes and engage in potential follow-up activities aiming at commercial development and exploitation.

The work plan of the spoke is based on the technical activities carried out in the living labs and flagships. Companies can be involved in these activities in three ways

Gap and requirement analysis, specification of key performance indicators. This activity will require the direct involvement of company employees and interaction with academic researchers. While it is expected that the bulk of this activity will be carried out at the beginning of the project, continuous monitoring and updates will also be required in later phases, depending on the typically fast evolution of hardware and software technology and the competitive landscape as well as the rapidly changing business priorities.

Feasibility studies, trials, and benchmarking of the hardware and software prototypes developed by the research partners. These activities will be typically performed in Spoke's living labs, which specifically aim at providing an open innovation environment and colocation space for running hands-on experiments in relevant deployment scenarios.

Collaborative research toward developing commercial products based on the prototypes developed by the spoke partners.

While funding productization is not part of the CN mission, the 2M€ budget in the Spoke linked to technology transfer and innovation is specifically earmarked for funding the three types of activities listed above, thereby preparing collaborative work between companies and research partners at higher TRL, to be funded through ad-hoc technology transfer agreements.

In addition to the involvement in the workplan and its technical activities, the companies will actively contribute to all the stages of the cascade funding pipeline (4M€ budget). They will first contribute to the definition of priorities for the cascade funding calls. Second, they will participate in the selection of cascade funding projects. Third, they will contribute to the continuous tracking and monitoring of the progress of these projects. Finally, they will be involved in integration and innovation actions that may follow after the successful conclusion of the projects.

Spoke 2

The activities in Spoke 2 are inspired by state-of-the art scientific goals, which are currently impossible to attain due to the complexity and size of the computing environment needed. In some cases, new ideas are required or welcome, but in most cases the use of modern technologies from the programming point of view (ML vs standard algorithms) or resource type (accelerators vs standard CPUs) seems the most promising R&D direction.
The work plan foresees the realization of scale-out tests, starting from the 3rd phase, on the solutions developed in the 2nd phase. During these periods a higher integration with the industrial partners is expected, to be developed in two directions:

- industrial partners can help providing the testbed platforms, by offering their infrastructure in-kind, or procuring in their centres new R&D platforms not available in the CN;
- industrial partners are expected to test their typical-use cases on the same platforms, under the paradigm that "*data are data*", and once one can abstract from the specific domain, the technology to treat them is similar. This second part is expected to siphon experience and technologies between the academic and productive sectors, and *vice versa*.

Interaction with private companies is expected to be executed mostly via the Innovation Grants, used by the Spoke as a mean to perform shared testbeds and proofs of concept, in which the TRL of the proposed solutions can be raised from R&D prototypes (TRL 3) to solutions tested in realistic environments (TRL 6).
At the time of writing, Intesa Sanpaolo and Leonardo have confirmed their participation in the Spoke activities; we foresee the involvement of additional industrial partners before and after the start of the project, within and outside the CN.
Great interest has been shown from private parties in the WP 6, which wishes (among the various activities) to serve as a link between basic research, private parties, and the huge potential coming from the data in the domain of Space Economy – with Mirror Copernicus data made available to the Spoke via already established collaborations (for example, ESA and INFN).
Intesa Sanpaolo, for example, in the context of this Spoke is interested in addressing and implementing a valid business model to sustain the correct deployment of the Mirror Copernicus program in Italy, aimed at analysing and delivering a huge amount of earth-observation data and information to the market. Interests are about:

- Market analysis: in-depth research of the market demand both from public and private sides.
- Analysis of innovative business models with a focus on the relevant aspects that may have an impact on the financial model.
- Private finance: identification/study of mechanisms and technical schemes to better support projects and investments in Space economy (e.g. SPV, Project Financing, PPPs, Securitizations, etc.) and the main financial instruments to employ within such structures (e.g. equity, quasi-equity, hybrid instruments, loans, etc.).
- Public funding: identification of the main European funding programs for R&D in Space, including grants, equity and guarantees, and the related synergies between these programs and private funding. This activity will even perform strategies to "crowd in" private financing to complement public funds (i.e. blending scheme).
- Identification of potential forms of international cooperation with a special focus on USA-Italy bilateral relations. Scouting and study of possible schemes of industrial and commercial cooperation among players in the Space Economy.

A further interest of Intesa Sanpaolo is the development of AI techniques for pattern recognition analysis. Many real-world interactions, such as banking transactions, can be modelled as a complex network with different levels of resolution, different channels of interactions and complexity.
Additionally, in the context of Space domain an area of interest relates studies and research to implement an accurate analysis of satellite data to assess and interpret risks in the agriculture field with the final aim to provide financial and insurance related risk assessment and risk monitoring in time, and advisory services for small to medium size farmers and corporate counterparties. In such a context it is of utmost importance to have access to algorithms and computing solutions capable of maximizing the potential of data processing to elaborate synthetic indexes of big bulk data.

Possible shared interests include the possibility to have common testbeds on technology and solutions. For example, there is interest in the spoke for:

1. **Shared testbeds and proofs of concept, using the infrastructure of the CN and other R&D platforms procured as needed, in order to test the processing of large, dispersed and heterogeneous data sources.** Solutions to be tested tentatively include
   - Tests on processing advantage using smart / dumb caches for remote / local data;
   - tests using tiered storage systems (from tape to rotating disks to solid state disks);
   - tests on remote streaming vs lazy download vs caches w or w/o prefetch.

Use cases:
- processing of O(10-100) TB of data from collider experiments, for example using workflows such as data processing in HEP;
- typical data-intensive-use cases from companies, such as the processing of agricultural data, as suggested by Intesa Sanpaolo.

Metrics to be considered for the testbeds are (at least) the processing efficiency / total time and power consumption using the various storage solutions, and the cost and scalability of the storage systems.

2. **Tests on the utilization of graph networks for pattern search recognition**, using the resources of the CN (and specifically the HPC-oriented resources). These will aim at testing the applicability of standard (Neo4j) or custom (via direct TensorFlow programming) techniques to heterogeneous data sources, generally big and without an intrinsic structure. Use cases to be tested include:
   - pattern search in financial data from Intesa Sanpaolo, for example to search for anomalous patterns in transactions, using multi-source data;
   - pattern search and recognition in High-Energy Physics and Astroparticle data (e.g. GW time-evolving patterns on noisy background and stochastic signals);
   - pattern search in Satellite data (from the Sentinel Program).

The tests performed aim at showing the capability of the selected techniques, and at showing the possibility to use solutions designed for the academic sector in industry, and *vice versa*.

With Leonardo, the interest is on the specific goal cited in the previous section: high-performance techniques for data access/analysis; statistical and AI-based tools; data-interpretations tools. In particular, in the context of the Space Economy Italian Strategy, develop and deploy techniques to access, analyse and process the data from the Mirror Copernicus program, creating the conditions to enable radically innovative services.

We plan to organize shared testbeds and proofs of concept in the activities of WPs 2, 3, 5 and 6 in order to test the developed services.

Spoke 3

The Spoke addresses topics that are key to the effective exploitation of the forthcoming and future HPC and Big Data technologies, paving the way to their efficient and broad adoption by industrial applications. The possibility of coupling HPC solutions to Machine Learning and federated learning approaches or to advanced visualization represent an outstanding opportunity for developing industrial solutions in critical areas like cyber-security, pattern identification and semantic segmentation, digital twins for prototyping, space technologies etc.
To this end, the Spoke has initiated an extensive action toward the Private Founding Members in order to identify common areas of interest and build synergies and collaboration targeting specific objectives, returning value to both the National Centre and to the industrial partners. *Intesa San Paolo* (ISP) has already formalized the interest in the Spoke activities. Other Private Funding Members (e.g. Leonardo, Sogei, Thales AleniaSpace, ENI etc.) have already informally expressed interest in the Spoke activities.
Further involvement of private entities will be addressed by exploiting the instrument of the Open Calls.

Spoke 4

Five industrial partners (Unipol Assicurazioni, Leonardo, Banca Intesa San Paolo, Autostrade per l'Italia, Eni) have expressed their strong interest in being involved in the Spoke activities. These partners represent different sectors of the economic world: financial (banking insurance), infrastructure, energy and technology, all differently, but substantially impacted by the effects of climate change. These economic actors pose a strong and growing demand for climate information and data, robust, accurate, timely and whose provenance is of

documented authority, but at the same time offer skills and knowledge that can help the science of climate modelling improve its tools.
At the starting of the Spoke activities, within the kick-off meeting, a specific workshop will be organized to start a discussion and proactive interaction with industrial partners, to identify the main mechanisms for their concrete involvement in the Spoke's activities.
During the preparation phase of the Spoke proposal, however, a number of meetings with the involved industrial partners have been made and during these meetings a number of needs and requests have already been identified, such as, for example: (i) a programmatic access to very high resolution, low delay, climate reanalysis data; (ii) creation of high-resolution short–and long–term climate projection data assets; (iii) access to a cloud infrastructure for sharing numerical modelling assets at national and Euro–Mediterranean level; (iv) ability to model and predict meteo–climatic phenomena with very local dynamics and severe impacting effects (e.g. microbursts, hailstorms, wind storms, intense convective events, coastal inundation, etc.). At the same time, the interest in the co-development of analysis tools based on the innovative investigation techniques offered by the Artificial Intelligence and Machine Learning methodologies has emerged.
Finally, the level of engagement of the industrial partners and their active involvement in Spoke's activities and interactions with affiliated research institutions will be further strengthened through the Innovation Grants (see Milestones M9-M15 and M17-M22). More specifically, these grants will be used to support co-development activities, carried out mainly but not exclusively in WP5, aimed at improving the Technology Readiness Levels (TRLs) of the Spoke's final products.

Spoke 5

The activities of Spoke-5 include the production, acquisition and processing of Big Data for monitoring, modelling of environmental data, civil infrastructures, natural environments. Downstream of the physical modelling, Spoke-5 will develop strategies, based on machine learning and artificial intelligence, for the mitigation of natural and man-made risks, related to the phenomena under study. For each of the phases described above, Spoke-5 foresees a fruitful synergy and collaboration with private entities.
Primarily we report the most significant collaborations envisaged.
1) Amazon Web Services (AWS) is the popular Cloud Service Provider that enables on-demand services like compute, storage, networking, security, databases, etc which can be accessed through the internet across the globe. As holders of advanced technologies in ICT - ML and AI in particular - the Spoke-5 collaboration will allow to accelerate the development of Spoke-5 proposed solutions in the envisioned research areas.

2) Engineering Ingegneria Informatica S.p.A., a company active in the field of software and IT services, specialized in digital transformation in the sectors of utilities, telco and industry, which impact and at the same time depend on the environment and the management of natural disasters. Through the collaboration with Spoke-5, Engineering will be able to improve its services and products, particularly in the areas of energy and industry, through quantitative assessments of the interaction between the environment and natural hazards and the man-made infrastructure managed by Engineering.
At the same time, AWS and Engineering, through collaboration with Spoke-5 will be able to develop meaningful testbeds, with proprietary data produced by Spoke-5 Research institutions.

3) Autostrade per l'Italia (ASPI), a national leader in the management of road transport infrastructures, will be able, through the collaboration with Spoke-5, to accelerate its transition towards sustainability, integrated mobility management and infrastructure digitalization. The collaboration with Autostrade will allow Spoke-5 to treat in an integrated way the data related to the freeway network, both from an environmental point of view and from the point of view of the management of the risks of degradation of the infrastructures.

4) UnipolSai Assicurazioni S.p.A. is an Italian group offering insurance services.
One of the most evident effects of climate change is undoubtedly the increase in natural disasters, in terms of frequency and intensity, with serious consequences for both quality of life and the economy.
Through the activities of Spoke-5, UnipolSai aims to improve several business processes, including policy pricing, claims management and settlement, anti-fraud, risk assessment, risk management, and prevention.

To achieve these goals, UnipolSai intends to rely on:

- Improved characterization of the hazard of the Italian territory, using homogeneous and high-resolution maps.

- Extensive mapping of the vulnerability of buildings, integrating existing databases available to local and national authorities with information derived from Earth Observation or at street level (e.g. Street View-like applications).
- Near-real time, programmatic access to meteorological and pollution data for the Italian country.
- Use of high-resolution models for simulating and classifying the impacts of catastrophic events, both in terms of damage estimation and spatial-temporal extension of the phenomena.

UnipolSai, together with the other insurance companies part of the Unipol Group, has a customer portfolio distributed throughout Italy. For this reason, the decision to focus on the creation and processing of data assets at the national or European level is perfectly in line with the company's needs.

5) e-GEOS S.p,A., an ASI (20%) / Telespazio (80%) company, is a leading international player in the geo-spatial business and one of the main industrial players in the European Copernicus Program.
e-GEOS leads the Copernicus Emergency Service - Rapid Mapping, the Copernicus Security Service – SEA and provides the European Ground Motion Service, through the Copernicus Land Monitoring Service.
e-GEOS develops and realizes application products through cutting-edge technologies, as AI-based satellite data processing, heterogeneous data integration (including drones) and 3D modeling. Existing solutions and new developments will be provided within Spoke-5 collaboration, in particular to address multi-hazard modelling and Digital Twin design.

6) Finally, the ARISTOTLE consortium provides a multi-hazard (MH) expert assessment service to the Emergency Response Coordination Centre of the EC DG ECHO. The service relies on hazard products aimed towards the quantification of the impact deriving from catastrophic events (e.g., earthquakes, tsunami, severe weather, floods, forest fires, volcanoes). Through the activities of Spoke-5, the ARISTOTLE consortium seeks to obtain improved quantification of MH impact by exploiting the abundant data available for geological hazard analysis, which includes diverse imagery (e.g., optical imagery, multispectral imagery, and radar imagery) and real-time in-situ monitoring information, within a scheme that includes physical phenomenon modelling and deep learning methodologies.

Spoke 6

This Spoke received the manifestation of interest from five Private Founding Members of the CN, namely the following Italian Industries: Thales Alenia Space Italia (TAS-I), Leonardo, ENI, Fincantieri, Terna.

Management of interaction with Private Founding Members

We envisage that the interaction between the Spoke affiliates and the industrial partners (Private Founding Members of the CN) will be pursued according to the following guidelines.

- The goal of the interaction will be to carry out a joint research and development activity on specific agreed thematics
- The actual project will rely on a prototypical use-case project to be jointly identified
- Identification of partnership with affiliates in Spoke 6
- Project cost and duration
- Evaluation of the project by the Management Board
- Negotiation
- Start of activities

Resource Typologies

- Centre resources
  - RTDA & AdR co-funding (for PhD type, consider 30K€ fellowship)
  - Computing time on the Centre Infrastructure
  - Reimbursement of worked hours provided by the industrial partner
- Industrial partner resources
  - RTDA/Doctoral co-funding (for PhD type, consider 30K€ fellowship)

Moreover, as described above the interactions with the industries will be essential to define:
- PoC demonstrators and structure of the living labs in WP5
- the national PhD program in Multiscale modeling and Engineering Applications
- the plans for training and for technology transfer in WP6.

In the following, we report the preliminary outcome of discussions with the industries.

Position from *TAS-I*
The possibility for the space industry to sustainably meet new market requests requires a re-thinking of the supply chain and product life cycle management, especially when low TRL is involved, and opening and making accessible the supply chain to new technological suppliers and partners without decades of space industry heritage.
An integrated, end-to-end, digitalization and virtualization of the full life cycle, methodologies and tools can provide a way to deal with the new challenges, create new opportunities and new sources of value (e.g. shop floor and field operation data) and simplify access to space manufacturing to new actors by removing current barriers linked to expensive (procurement, maintenance and know-how) facilities and to the complexity of the quality assurance process and manufacturing standards.
To pursue the afore-described objective, a key priority is to build the assets that will be used for developing and testing digital models that will follow the entire lifecycle processes of the product
CN HPC Spoke 6 foreseen initiatives on Digital Twins is an opportunity to develop the needed assets for an evolution of industrial products life cycle that is relevant for Space but, we think, equally useful to other strategic industrial sectors.
The themes proposed for this Spoke on Digital Twin also foreseen to be carried out in synergy with other NRRP initiatives addressing the use of Digital Twin and production and in particular:

- PNRR M4C2 - Innovation Ecosystem of Lazio, Abruzzo-Molise-Umbria (MUR), Abruzzo (Ag.Co.)
- M1C2 - Space Factory
- ESA and ASI space calls involving Italian space industries (large like TAS-I, Leonardo, and SMEs, like Ingeniars)

Potential research areas of interest
Heterogeneous co-simulation and Digital Twin frameworks, for example:

- Co-simulation orchestration engine
- Open standards for models and interface and trade-off of available SW platforms/frameworks
- Scalable orchestrators, master algorithms and transparent middleware
- Semantics and languages (e.g. based on Process Algebra)

Data lake architectures and solutions, for example:

- Distributed / federated and IP regulated data lake
- Lifecycle oriented data and model management

Computing and Networking architectures and protocols, for example:

- In cloud co-simulation orchestration architectures and solutions
- Distributed co-simulation protocols (e.g. DCP)
- Real-time and HW/SW/Twin-in-the-loop
- Distributed ledger based M2M for automated supply chain
- DevOps concepts for distributed modelling and co-simulation (CI/CD)

Virtual testing and certification, for example:

- Certification and validation workflow based on digital models and digital test facilities replica
- Anechoic, Vibration and TVAC simulator and digital model's interfaces

Artificial Intelligence, for example:

- computer enabled models design and checking
- ML/AI on-board satellite data processing

Themes of interest as Proof-of-Concept (PoC) demonstrators
The themes of interest proposed in the following will be developed through Proof-of-Concept (PoC) demonstrators to achieve valuable building blocks and enabling technologies for applications in the frame of Earth Observations, Navigation, Satcom, Exploration Systems
Proposed projects will leverage on expertise and know-how in the field of ML/AI and HPC technologies for the development of highly reliable algorithms, advanced numerical models and software platforms use of accelerators like Soft-GPU and FPGA, in response to demands for more affordable and reliable Systems suitable for complex life cycle conditions. Deployment of the industrial research is expected to be in synergy between space and non-

space domains, to reach a common national research strategy and shared AI-based and HPC technological platforms

Synergy is expected with other NRRP opportunities:

- M4C2, CN HPC Spoke 1 Future HPC and Spoke 10 Quantum Computing
- M4C2, Expanded Partnerships – Artificial Intelligence: foundational aspects
- M1C2: Earth Observation
- M1C2: In-Orbit Servicing

Proposed partnership with other entities in the Spoke 6 (not exhaustive)

- University/Research Centre: UniBO, PoliTO, Sapienza, PoliMI, UniPI, CNR, Tor Vergata
- Industry: Leonardo Labs

List of possible Proof-of-Concept (PoC) demonstrators

- Cognitive radar: adaptive SAR sensing, dynamic waveform generation, real-time closed-loop radar, compressive sensing (sub-Nyquist sampling), autonomous tracking, efficient data compression techniques, on board SAR focusing algorithms
- Spacecraft health management: AI/ML-based on-board algorithms/solutions aimed at exploiting large dataset generated on board for supporting space operations, FDIR (Failure Detection, Isolation and Recovery) and satellite guidance; applications such as anomaly detection and prediction
- Edge computing, enhanced data inference for EO missions, through scalable AI/ML-based on board data processing for space systems; ML-based architecture optimization for on-board applications
- On board dynamic configuration for automatic constellation scheduling, autonomous navigation
- Distributed optimization control and learning for formation flying and space robotics
- AI-based data processing for rover vision, docking-grasping operations. Security monitoring on EO systems, aiming at analysing, detecting, and predicting attacks and intrusions in constellations of satellites
- Digital-twin modelling, data-driven simulation tools for space missions, advanced fluid-structure interaction and industrial production plant
- Virtual Reality environment for enhancing engineering applications and digital twins interrogation

Position from *LEONARDO*

- High fidelity/high computational efficiency
- Software certification
- Data lake, data maintenance, data standard, data protection, data sharing: GAIA-X
- Living labs
- Outer loops: optimization, low fidelity modelling, UQ,
- LBM (Lattice Boltzmann Methods)
- Computational electronics/electromagnetics for radar
- Computational fluid dynamics for helicopters, aerial vehicles and drones

Position from *FINCANTIERI*

- Digital twin of marine vehicles: operation, maintenance, Life Cycle for CO2 reduction
- Digital twin of infrastructures (bridges, ports, ...)
- Predictive maintenance and Health monitoring using data from distributed sensors
- Smart Building: Design, Progress estimation based on observables
- Hybrid diesel-electric propulsion: smart grid for optimal power balancing in variable absorption conditions
- Cryogenic hydrogen propulsion
- Manoeuvring optimization and automation in port
- Optimal port design
- Radar signature calibration (to maximize visibility)
- Applied Electromagnetics and Metamaterials
- Noise reduction
- 4D drones
- Unmanned and green ships for regional transport, Autonomous driving

Position from *ENI*

Topics of possible ENI interest in relation to the Centre's work plan are listed in bullet points below:

- Development of computational models in the geo-sciences:
  - development/improvement of specific modeling aspects of proprietary codes;
  - optimization of the same on Hardware accelerators (GPU, FPGA, quantum computing)
- Development of subsurface models
- Optimization of complex systems
- Exploitation of HPC applications in Plant Engineering
- Training of young engineers with skills in HPC and Big Data

Position from *TERNA*

TERNA S.p.A., is the owner of the Italian national transmission grid (NTG) for high and extra-high voltage power and is the largest independent electricity transmission system operator (TSO) in Europe. TERNA conducts grid planning, development and maintenance activities, bringing together expertise, technology and innovation to optimise high voltage electricity transmission. In operating the electricity system, TERNA ensures the balancing of power supply and demand 24 hours a day throughout Italy, also known as "dispatching" that is the managing of energy flows through the grid.

In this context, R&D & Innovation projects will be more and more crucial in order to accomplish the mission to guarantee the fulfilment of key system needs (RES integration, adequacy, reliability) at the lowest possible cost for the system with high quality standards.
Through the collaboration with Spoke 7 "Multiscale Modeling and Engineering Application", in particular, Terna could better address the items and projects related to the main innovation streams "Planning the power system at the minimal cost" and "Maintain grid stability and security" - part of the innovation roadmap for the system operator identified by Terna to face the system challenges expected in the short-medium term - with particular attention for:

- simulation (both for planning and operational purpose) of new hybrid AC / DC systems and support to related evaluation of replacement benefits
- development of simulation methods and tools for simplified evaluations in dynamic AC power system analysis
- explicit and equivalent modeling of HV grid components by use of open-source approach (i.e. Modelica language)
- new system, procedures and tools for infrastructure planning with sector-coupling
- new approaches to increase efficiency in optimal power flow and security assessment analysis, performed on large nodal grid models

Terna will also actively contribute, providing operational experience and data, input for use cases implementation and support for validation and testing, with main (but not exclusive) interest on WPs 2, 3, 4 & 5.

Spoke 7

The current industrial revolution is characterised by an increasingly pervasive use of data, digital technologies, new materials, and interconnected components. The knowledge, know-how, methodologies, technologies, and prototypes developed by spoke 8 will be made accessible to companies for the creation of new products, processes, and services. To this end, and from the very beginning, the spoke program includes specific actions for promoting industrial involvement, to which a spoke budget of ~ 1.5M € has been reserved. These resources will be mostly used to:

- Develop specific software and HPC applications (including web-based) dedicated to solving problems of industrial R&D in the field of molecular and materials simulation, with the ambition of customising the capabilities of the flagship computing codes to industrial requirements and of facilitating their operation from non-academic users. These new technologies will be defined on-demand on the basis of needs presented by companies interested in developing partnerships with the spoke. Wherever applicable, the new HPC tools will be made publicly available;
- Carry out research projects in collaboration with interested companies to address specific scientific problems relevant to the industrial application of advanced materials. Companies interested in developing

such partnerships with the spoke will be assisted by the spoke staff who will assess the technical feasibility of the proposed project and identify the most suitable competences available in the spoke;
- Support industrial users in the application of molecular and materials simulation through dedicated training programs and the activation of web channels for the provision of specific advice and support.

The involvement of industrial partners will be promoted throughout the project duration from the very beginning with the following industrial seed actions and will continue throughout the whole the spoke program:
- The organisation of an event to promote the spoke's opportunities and scientific offer dedicated to companies. The event will be in the form of a conference with open participation and it will include a round table with invited representatives from the private sector. Milestone M2.
- The organisation of a series of peer-to-peer meetings dedicated to companies interested in developing collaborative projects as described above, with priority being given to companies that are part of the Hub.
- The set up of an "industrial user contact point" for coordinating the industrial collaborations and requests.

At the moment, the spoke has started the discussion of possible forms of cooperation with some of the companies participating in the Hub. According to these preliminary discussions, ENI is interested in developing the partnership along the lines described above both in the area of HPC software and in its application to materials discovery and characterisation starting from the fields of nuclear fusion and solar energy. Similar discussions with Leonardo are in progress and are oriented to transistor technology.

Spoke 8

Humanitas

The AI centre of Humanitas Research Hospital are now stepping up the extent of the reach of interoperable data to the already massive DWH it maintains. Opposite to diagnostic images, which relies on an internationally and historically established protocol, DICOM; digital pathology and digital endoscopy offer a novel technical challenge due to the inherit complexity of the data they represent spatial and temporal high-resolution, non-standardized atomic samples.

To this end, Humanitas is developing IT and AI solutions to allow for the integration of massive sets of endoscopy videos and histological samples while reducing the burden on server and preserving clinical value. In Humanitas Research Hospital, the review of the information system has led to the creation of an Electronic Health Record platform that provides access to patient information with due reliability, granularity and security to a large number of users (doctors, nurses, health personnel), thus making data available immediately. Also, it ensures greater transparency and security of data.

The EHR is integrated with medical devices and telemetry and uses international standards IHE/XDS and HL7 communication protocols with data producers.

Our role in this application therefore aims to build up a data lake and to harmonize electronic clinical and -omics data, test-drive infrastructure to enable nationwide life science research, ensure data interoperability of existing fragmented information systems, and to raise the standard of Italian health system. Moreover, in-depth molecular and cellular, phenotypic, and functional characterization (i.e., Omics, drug screening) of tumours as well as of non-neoplastic diseases will be performed. To this end, we will test-drive the application of new digital technologies and computational tools towards clinical decision support. We hypothesize that multidimensional disease profiling will predict molecular and cellular drug response relationships that can inform clinical decision making for patient's care.

ENG

ENG is a worldwide ICT Company with a consolidated presence in healthcare market and a long-year experience on developing advanced data-driven solutions for innovative care delivery. ENG has also a relevant role within the European software research community, with a portfolio of several national and international projects and initiatives, where it contributed to aspects related to data interoperability and the entire life cycle of data analysis, from its collection, integration, storage, statistical processing, algorithmic and advanced visualization. ENG is in particular involved in research projects concerning the development of enabling technologies for telemedicine and, more recently, also in the field of precision medicine.

ENG will collaborate with the National Centre "HPC, Big Data & Quantum Computing" on WP5 activities, by specifically leading the T5.1 task. it will design and develop a holistic EHR, able to manage in a structured way holistic data of patients, including both ordinary data relating to the patient's phenotype and clinical history and those relating to omics investigations, and also providing added-value services and functions for end-users that maximize the exploitability of the data produced and facilitate the understanding of most relevant data. In particular, ENG will develop advanced and easy-to-use methods for an interactive and dynamic data visualization and automatic decision support, based on algorithms produced in the context of the other WP5 activities, as integrated features of the next-generation EHR.

Spoke 9

The Spoke identified a range of organizations that offer best opportunities for consolidating and validating the research results in the application domains previously described (availability and accessibility of real data, territorial relevance, etc.) and for ensuring long term exploitation and sustainability for the spoke. In general, possible stakeholders directly benefiting from the research of the Spoke include large companies as well as large and mid- to small-size cities. Further potentially interested organizations include Italian ministries (e.g., Ministero dell'Innovazione, Ministero delle Infrastrutture, Ministero della Ricerca, Ministero salute), local and regional health departments, municipalities (ANCI), research centres and universities, foundations, national and regional research consortia, competence centres, and Digital Innovation Hubs. While additional beneficiaries will be selected during the initial phases of the project, also based on the local links of the involved universities and research centres, the Spoke already collected an expression of interest from Fondazione Innovazione Urbana. The organization will push their "Gemello Digitale" project, targeting the urban area of Bologna, and already identified a direct involvement in the Spoke activities covering computing platforms, mobility, infrastructures and utilities, as well as environment.

The Spoke also successfully probed the potential impact and interest in the industry. The list of companies that expressed a strong interest in the activities of the Spoke includes Leonardo, Intesa San Paolo, and Engineering, covering the full spectrum of topics addressed in the project: modelling (with emphasis on data governance and privacy), computing platforms (architectures and technologies for distributed IoT sensing, integration with cloud/edge, architectures for twin systems), software systems (definition of system software platforms, software specification and maintenance), health and lifestyle (HPC and IoT technologies for e-Health, digital twins for smart hospitals and personal health records, big data platforms for advanced management of territorial medicine and for monitoring of behavioural reactions and infodemic outcomes associated with COVID-19 pandemic), mobility (modelling of citizens' mobility patterns and large-scale crowd dynamics in order to design more efficient traffic management systems, sustainable multimodal public transportation), socio-economical analysis (identification and tracking of misinformation, disinformation and hate social media, monitoring of urban crime using multiple sources of data), infrastructures and utilities (models and techniques for smart grids, drone-based maintenance, real-time monitoring/forecast for interconnected infrastructures, cyber-attacks for public spaces and critical infrastructures), and environment (monitoring and forecast for environmental conditions, regeneration of city biodiversity and ecosystems, water cycle management).

Spoke 10

The Italian industrial market strives toward a roadmap for the industrialization of quantum solutions. The spoke activities will contribute to create an Italian industrial ecosystem able to compete in a fast-growing market that will surely see Europe playing a leading role.
The activities of the spoke will cover the entire software and hardware industrial landscape, in collaboration with European industries and start-ups, from both supply (Pasqal, AQT, IQM, Atos, etc.), software (HQS, Cambridge Quantum Computing, ParityQC, Qruise, etc.) and demand (Thales Alenia, Intesa Sanpaolo, etc.).
The collaboration with industrial partners will be conducted from two perspectives. Some partners will supply hardware and infrastructure technologies to provide researchers with the required quantum computing capabilities, either in terms of physical devices, cloud resources or high-performance simulators. Other partners will provide industrial and challenging use cases that can be enabled by using quantum computing acceleration. One first example is from the space industry, by addressing the optimization problems when planning space missions (Thales Alenia). A second example is from the finance industry, with the design of AI-powered advisors for the automatic portfolio optimization (Intesa Sanpaolo).

Performance indicators

As detailed in Section C, Key Performance Indicators (KPIs, Section C) have been defined to quantitatively measure the impact of the project results, as defined by the CN Key Exploitable Results (KERs, Section C). In all cases, numeric targets are available and defined there. Without repeating Section C, part 6 here, a subset of the Proposed KPIs include:

- Resources available to users from scientific domains and beyond (measured by node hours and PB)
- Number of active users on the CN infrastructure (measured as number of unique users)
- Power efficiency of the renewed / new infrastructures (measured by Power Usage Efficiency – PUE)
- Number of centres connected with high bandwidth connections (with > 100 Gbit/s and > 1 Tbps)
- Number of PhD grants funded by the CN
- Number of researchers / technologists / technicians hired by the CN
- Number of use cases using advanced solutions like accelerators and quantum computers / emulators
- Number of industry-related testbeds and pilots
- Number of open calls involving industry
- Number of services or technologies getting medium-high TRLs (for "close to production" TRLs 6-9)
- Fraction of women in charge of tasks and work packages
- Fraction of women hired by the CN
- Fraction of women in the overall CN workforce
- Fraction of the investment in Southern Regions
- Fraction of demonstrators in Southern Regions
- Fraction of Open Calls in Southern Regions
- Fraction of PhD grants in Southern Regions
- Fraction of the overall CN workforce in Southern Regions

In addition to the KPIs related to the project impact, the execution of the program will be monitored through specific indicators, such as:

- The amount of recruited personnel
- The number of tenders
- The fraction of equipment procured
- The fraction of budget spent

The **monitoring of the project execution will be done via industry-standard project management techniques**; in particular, the Hub will monitor the advancement of the KPIs at least on an annual basis, by requesting reports and documentation to the Spokes. The progress and monitoring of the project will be a standard discussion item at the recurrent Scientific Board meetings.

C. Program impact

How the National Centre addresses the expected impacts of the Call

Overall impact of the National Centre

The National Centre (CN) on HPC, Big Data and Quantum Computing provides a pivotal opportunity for the national scientific, industrial, and economical system to address current and upcoming scientific and societal challenges, strengthening and expanding existing competences and infrastructural resources.

Digital infrastructures, high performance computing architectures, hyper scale cluster architectures and multi-terabit bandwidth network infrastructures will have a tremendous innovation in the next few years. The **global competition towards exascale computing** is imparting a strong acceleration to the deployment of supercomputing systems with a performance more than many hundreds of petaflops. The target for the availability of exascale computing systems is getting closer and closer. As is well-known, the US, China, and Japan are investing many billion dollars for supporting R&D with the target of having exascale systems. Along similar lines, the European Union created EuroHPC[1], a joint undertaking with the purpose of mobilizing a budget of billions of Euros for R&D for the development of European HPC technologies and for achieving towards pre-exascale and exascale systems to open production in Europe in the same time frame. This initiative closely connects to other European initiatives aimed at advancing European Competitiveness in Open Science Cloud Infrastructures[2] (EOSC), in the Data Strategy[3], in Microprocessor Technologies[4] (EPI), in sovereign data infrastructures[5] (GAIA-X), in multiple ESFRI initiatives[6], in a comprehensive recovery plan[7] (NextGenerationEU).

Italy is willing to significantly contribute to this European effort through the National Recovery and Resilience Plan[8] (PNRR in Italian): the proposal offers an innovative concept for a new CN based on a **supercomputing cloud distributed infrastructure, which is a pre-requisite for Italian competitiveness**, since data-driven R&Ds are key aspects in all relevant societal and industrial domains.

As per provision, the CN is structured into hub and spokes, where the hub aggregates and integrates the governance structure and coordinates transversal activities such as project management, education and training, dissemination and outreach. The major national industries and public entities directly affiliated to the Center participate in the hub. The spokes are structured to cover the strategic domains of the Centre's activity and the persistent development of the national high-performance digital infrastructure for technical-scientific computing and (big) data processing. The conceptualization of the Center intends to pursue the following objectives:

- **Infrastructure:** to deploy by means of an interconnection at multiple-terabits scale a federated national supercomputing and big data infrastructure providing access and services for the research and scientific community, industry (including SMEs) and the public sector. And to persistently maintain such an infrastructure in highly flexible configurations, adaptable to a wide range of applications and user needs. This state-of-the-art infrastructure will integrate the most advanced data and computing systems: petascale, pre-exascale and post-exascale class supercomputers,

[1] https://eurohpc-ju.europa.eu
[2] https://eosc.eu
[3] https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en
[4] https://www.european-processor-initiative.eu
[5] https://www.gaia-x.eu
[6] https://www.esfri.eu
[7] https://ec.europa.eu/info/strategy/recovery-plan-europe_en
[8] https://italiadomani.gov.it/en/home.html

highly distributed big data facilities, as well as new technologies such as quantum simulators, or the future quantum computers and multi-terabit transmission systems based on Open Line System technologies and orchestration, giving users direct access to the lower optical level for the data center interconnection.

- **Technologies and living labs:** to design future supercomputing and big data technologies at both the hardware and software levels, including the integration of new technology generations and computing paradigms (e.g. post exascale, quantum, neuromorphic, etc.). A major objective is the capability to maintain constant technological supervision for critical components, including microprocessors, system architectures, accelerators, and complex software stack and software tools developed to cover HPC and big data sectors, as well as broader markets and domains (such as extreme-scale, high-performance distributed big data, federated data spaces, embedded systems, edge computing applications).
- **Applications:** to promote and pursue a leading competence in the development of applications that are key for science, industry (including SMEs) and the public sector. Scientific and industrial codes utilizing HPC, Cloud and big data resources, applications and software packages in key areas will be co-designed, co-developed, ported, and optimised in strict collaboration with the high-level support team of DevOps staff provided by the infrastructure spoke. This will be done in connection with domain experts provided by the vertical domain spokes to fully exploit the performance of current and future computing systems. Excellence in research will be pursued, increasing the TRL levels as an outcome of the R&D activities of the domain spokes.
- **Enabling:** to enable Science, Large, Small and Medium Enterprises to benefit from the infrastructures and services provided by the CN. A closely coupled network of competence spokes, in partnership with academia and business enabling activities, will leverage knowledge and best practices to strengthen competitiveness of industry, science and public administrations.
- **Leadership:** to achieve and maintain leadership in the use of key supercomputing technologies and applications by widening the scientific and industrial use of HPC and big data applications, and by fostering knowledgeable leading scientific communities and a skilled workforce.

To this extent, a comprehensive composition of academia, research institutes and industry have joined forces in providing their experience and know-how to create a unique environment through which competitiveness and innovation can be enforced for the benefit of the entire National system.

The numerous challenges science, industry and society are facing are extremely diverse. The aim of the CN is to enable a national network of public and private research communities to face and tackle extreme challenges, such as contributions to the mitigation of climate change, to the phase out of combustion technology, with the consequent energy revolution that ecological transition imposes; develop scientific applications towards exascale in the fields of materials science, fundamental research, cosmology, astronomy and astrophysics, in space economy, multi scale engineering sciences, geophysics of the earthquakes, tsunami and volcano phenomena, so to face extreme natural events in a resistant and resilient way; develop methods and conceptualizations of the city digital twins for smart urban systems and optimized management of our cities, contributing to the safety of citizens; plan the future of mobility as an efficient and sustainable service, combined with the introduction of alternative energy sources and the introduction of autonomous driving; address the necessary developments for personalized precision medicine and new diagnostics and specific treatments, based on the processing of data deriving from the sequentialization of the human genome also in reference to the e progressive aging of our society.

These great scientific and technological challenges can be faced and solved only with the indispensable commitment of science that the CN will address with binding guidelines aimed at safeguarding ethical and moral values, and with the constraint of refraining from dangerous experiments. All this requires the availability of both start-of-the-art supercomputers and big data processing systems.

With these considerations at the foundation of the CN program, the strategy designed to effectively address the expected impacts mandated by the PNRR call revolves around three pillars:

- The Key Enabling Technologies (KET), as identified by the PNRR.
- A set of Key Exploitable Results (KER), serving as the general impact blueprint of the CN.
- A set of Expected Impacts and associated Key Performance Indicators (KPI), concretely implementing the KERs.

The general process of the project impact strategy is shown in the following picture:

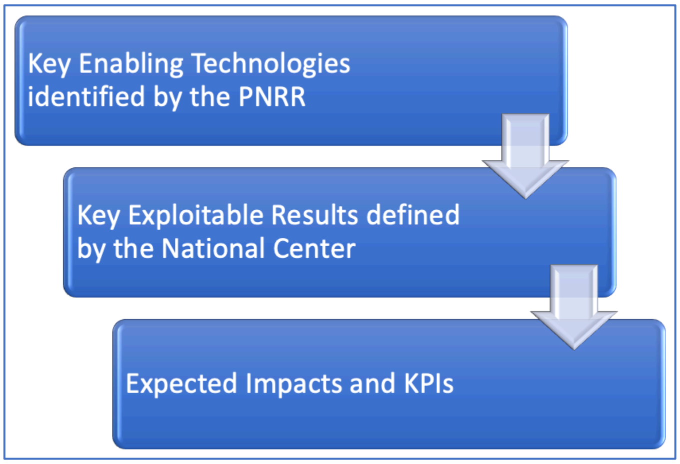


Figure 1 - General impact process

The CN will address several of the PNRR KETs through a two-fold architecture, centered on the strategic integration of a network of infrastructural resources on the one hand, and on an unprecedented aggregation of Thematic Scientific Spokes on the other.

The PNNR KETs specifically considered by the CN are listed in the following table:

PNRR KET ID	Description
PKET.1	**Advanced simulation, analysis and big data management on top of an open multi-terabit network**
PKET.2	**Technologies for climate, environmental and disaster protection**
PKET.3	**Quantum technologies**
PKET.4	**Material sciences**
PKET.5	**Health-related technologies**
PKET.6	**Smart cities and sustainable mobility**
PKET.7	**Industry 4.0**

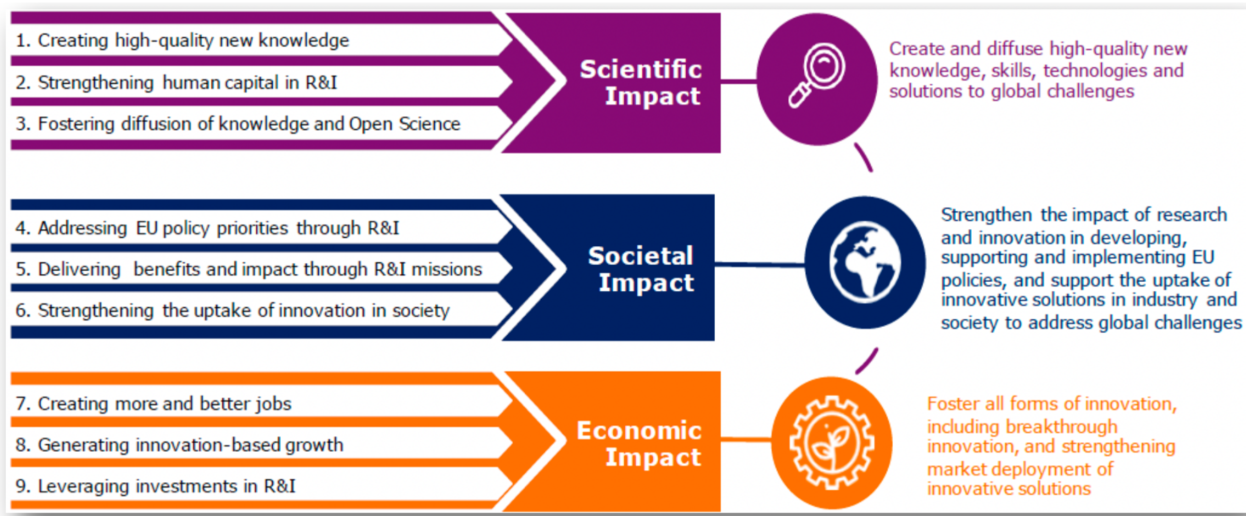
Table 2 – PNRR Key Enabling Technologies addressed by the CN

Impact on the Country's Economic, Social and Cultural System, and Positioning of the CN and Its International Image

As highlighted by the EU in the Horizon Europe Framework Program[9], the assessment of the impact of research and innovation projects has to consider the following key impact pathways:

- Scientific impact
- Societal impact
- Technological/Economic impact

which are depicted in detail below:



(Source: European Commission)

We can adopt this general view also regarding the overall impact of the CN.

EU policy trajectories are based on **green transition, digital transition, and justice and social inclusion[10].** In this context, **R&D models and the job market show the pervasive need for high performance simulations, computation and data analysis and the enhancement and development of HPC facilities, technologies, and skills (professional data scientists and various experts trained in data science)**. This trend is necessary to favour a holistic and integrated multidisciplinary and interdisciplinary problem-solving approach to support research infrastructures and industries in properly handling and gaining insight from big data. OECD confirms this as a global trend[11].

The CN will play a central role in fostering Italy's capacity of promoting socially inclusive growth and people-centred process of sustainable human development. Big data analysis and sophisticated socially-oriented management can improve citizens' quality of life and their access to fundamental private and public goods and services. This is the case first of all of citizens' access to health, education and knowledge. More in general big-data and advanced computing capacity will enable efficient and effective decision making offering bold solutions and responsive actions to central social issues. In this perspective, example of important fields of activity - where the CN will play a role with a strong and evident social impact - are: forecast of weather and natural disasters management; environment protection, m ammortamento onitoring

---

[9] https://ec.europa.eu/info/research-and-innovation/strategy/support-policy-making/shaping-eu-research-and-innovation-policy/evaluation-impact-assessment-and-monitoring/horizon-europe_en

[10] https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf

[11] https://www.oecd.org/sti/ieconomy/data-driven-innovation.htm

and management; energy production and consumption; innovative teaching and learning models; cultural heritage management and access; response to epidemic and pandemic; personalized healthcare and communities' healthcare provision; fraud and corruption detention, contrast to criminal organization; mobility and traffic management; urban, regional and community planning for public utilities services.

A significant indicator showing the social and economic impact of big data is how much industrial players are investing in data science professionals. According to the results of the 2020 European Data Market study[12], "*Enterprises will add 3.2 million data professionals' positions between 2019 and 2025, bringing the total to 9.3 million jobs. However, this will increase the potential data professionals' skills gap, which may become a bottleneck for some enterprises or regions, creating competition between enterprises for the most skilled professionals*". On the other hand, we continue "*to register an imbalance between the* ***demand and the supply of data skills*** *in Europe as the estimated gap reached approximately 459,000 unfilled positions in the EU27 plus the U.K., corresponding to 5.7% of total demand*". Similar trends are in the USA, where a report from the US Bureau of Labour Statistics suggested that the rise of Data Science needs is forecast to create roughly 11.5 million job openings by 2026.

The global trend also applies to Italy. For the sake of illustration, this is what emerges in Emilia Romagna from a study on the use of big data carried out in 2018 by Nomisma and commissioned by the Emilia Romagna region[13]. By exploiting big data, three out of four companies (71.4% of responses) can increase productivity or turnover and develop new processes and products, but to achieve these results it is necessary to wait at least two or three years. The research highlights that production and service companies use big data mainly to carry out market and customer analyses (83.3% of responses), to develop new products / services and to improve production (75% each) and to manage after-sales activities (41.7%). On the other hand, ICT companies use big data to improve production (64.7%), to develop products and services (58.8%), but also to analyse markets and customer behaviour (41.2 %) and to manage after-sales activities (47.1%). The research highlighted the obstacles that companies encounter in achieving business objectives. These are cultural factors for 60% of ICT companies and 81.8% of production and services companies: operators often do not know how to use the analysis procedures and do not know how to understand their functions or purposes. There are several strategic objectives that the company interviewed in the study would like to achieve using advanced data analytics techniques. Production and service companies, in order not to disclose data that could contain sensitive information, prefer (82%) to do the first analysis internally and mainly aim to obtain an increase in turnover and develop new processes and products (78.6% for each of the options). Equally important are the development of customer relations and marketing (71.4%), the efficiency of existing processes (50%), the increase in productivity and the identification of new markets (42.9% each). The objectives of ICT companies are instead mainly oriented towards increasing turnover (90%), increasing productivity, and improving the efficiency of processes (70% each). The business areas that most benefit from the use of big data in production and service companies are "business" (54.6%), IT (36.4%), "marketing" and research and development (27.3% each), and the production division (18.2%). Similar conclusions apply to Italy (see, again, the above-mentioned 2020 European Data Market study).

Dimensions of Impact

Specific patterns and dimensions of impact are recommended for the analysis of research centres and infrastructures (see European Commission, *Supporting the Transformative Impact of Research Infrastructures on European Research*)[14], recommendations for patterns and dimensions that we elaborate and rephrase for this CN.

---

[12] https://datalandscape.eu/sites/default/files/report/D2.9_EDM_Final_study_report_16.06.2020_IDC_pdf.pdf
[13] https://www.regione.emilia-romagna.it/datavalley/approfondimenti
[14] https://ec.europa.eu/info/publications/supporting-transformative-impact-research-infrastructures-european-research_en

Ensuring Scientific Excellence

The CN will be strongly engaged in high-quality research. Its scientific objectives are ensured by the predominantly international scientific communities and benchmarked at the global level. The usage of the CN resources and solutions will be competitive, with the standard criteria of scientific quality expected to be high and monitored by Key-Performance Indicators. Excellence is also guaranteed by the network model. Benchmarking of international labs and centres will be adopted. The training of young researchers through the access to the CN will be part of the monitoring activities. The adoption of proper and possibly evolving KPIs may help in evaluating the most appropriate configurations for the provisioning of research infrastructure services.

Attracting, Educating and Training Managers and Users

The CN will have a management structure that allows the selection and the progressive enrichment of the skills of its scientific, administrative, managerial, and human resources competences. Managing the activities of the CN requires skills that are often not only in the background of the scientists who, having been principal investigators of several international projects, must cope with the complexity of tasks that are not always rewarding from a purely scientific point of view level. An important issue of the CN is to attract added value from users, by sharing feedbacks for designing upgrades and building parts of the infrastructure with selected user groups.

Unlocking the Innovation Potential of Research Infrastructures

Monitoring the innovation potential of access to the research infrastructures is not obvious and only very partially reflected in patents and exploitation, e.g., by industry. The CN can maximise the potential of direct innovation by offering tailored access modes for industry (particularly SMEs), possibly including dedicated training and, whenever appropriate and legally possible, data exploitation support. The innovation potential will also be supported with FAIR data, with data services becoming progressively open to the economic sector, and with the development of an entrepreneurship mindset through the possible creation of start-ups or spin-off.

Measuring Socio-Economic Impact

The projects and specific activities of the CN specify KPIs that reflect impact on local or general economy, society at large and the local community. Indeed, "*few examples of studies on the social cost-benefit analysis of the infrastructures exist that adapt cost-benefit models to the specific case of RIs. No uniform methodology is in place and the RIs can only address this issue in a non- standardised manner which makes it hard to measure and benchmark impact. This point has been addressed in different reports (OECD, RI-PATHS)*"[15].

Creating Value from Data Generated by Research through the Open Data Paradigm

The CN will effectively create a National Big Data Value Network. A structural effort will be devoted to make data sets FAIR and to developing interoperable services. It will adhere to best practices in data archiving and retrieval, utilizing standards to develop appropriate data management plans in all projects. Through these standards, supercomputing, big data resources and services as well as external data networks will be integrated to maximize the exploitation of research data, creating new value from data.

Framework Conditions for Effective Governance and Long-Term Sustainability

Ensuring sustainability of the investment will concern all stages in the lifecycle of the CN. Accordingly, "the development of robust, and flexible, business plans are key to define the proper governance and identification of the stakeholders beyond the reference research community". A clear strategy will be

[15] EC, *Supporting the Transformative Impact of Research Infrastructures on European Research*, p. 50.

developed for the overall alignment in time of the national and European roadmaps and funding opportunities.

Structuring the International Outreach of the CN

Given the uniqueness of its expected Research Infrastructure, the CN has the ambition to attract international scientists and competences. All the institutes and their exposed staff are already part of the global scientific ecosystem, as part of international collaborations as typical in today's top science initiatives.

Overall selection of the Impact indicators

To determine the impact of its research infrastructure, the CN will consider, among others, the OECD *Framework for assessing the scientific and socio-economic impact of Research Infrastructures*[16]. As suggested there, the CN will pay attention that indicators "*should be linked with strategic objectives*", they "*have to provide information relevant to operational issues*" and "*should be time-bounded, so that they can be used to assess evolution over set periods*". Also, "*they should […] be stable over time to ensure consistency*".

The National Center Key Exploitable Results (KER)

According to the impact strategy shown in Figure 1, the PNRR KETs are modeled in the CN through the following strategic **Key Exploitable Results** (KER):

Key Exploitable Results ID	Description
KER.1	**An integrated HPC, Quantum and Big Data Platform**. The CN will deliver an unprecedented strongly interconnected set of turnkey resources and solutions, made available to the whole academic and research community via an integrated unified platform. This will effectively overcome the fragmentation deriving from independent funding, provisioning, and support measures, facilitating resource and knowledge sharing and efficient utilization.
KER.2	**Consultancy, technology transfer and training courses for effective exploitation of scientific applications**. The deployment, development and full adoption of modern applications exploiting state-of-the-art HPC and Big Data resources requires a paradigm shift and extensive training and consultancy services. The CN will therefore deliver a complete set of initiatives to stimulate adoption of the available solutions by science and society at large, ranging from dissemination events to PhD grants, to Technology Transfer initiatives.
KER.3	**A federated service management system.** A national HPC and Big Data infrastructure such as that made available by the CN must manage the key elements of its operation, monitoring, security, resilience, availability, and productivity. The CN will therefore strive to provide an integrated and federated service management system that allows the persistent development of network systems, computing resources and assets such as data, applications, software,

[16] https://read.oecd-ilibrary.org/science-and-technology/reference-framework-for-assessing-the-scientific-and-socio-economic-impact-of-research-infrastructures_3ffee43b-en#page1
https://community.oecd.org/docs/DOC-156270

	forming the basis for a sustainable and complete management framework even beyond the borders and constraints of PNRR calls.
KER.4	**An interoperable open multi-tier data repository**. The CN has the ambition to serve as provider and implementation conduit not only for the Spokes identified by the Center itself, but act as a persistent national infrastructure. This will be realized through adherence to *de jure* and *de facto* standards, by the adoption of open-source solutions, by the connection with key international initiatives and through extensive support to Open Science Cloud and FAIR data principles.
KER.5	**Strategy and recommendations**. The outcomes and results of the CN will be captured not only in the expected key scientific and technological results, but also in public whitepapers and policy recommendations. These strategic insights deriving from the CN experience have the ambition to help shaping and evolving national directions connected to leadership in science and in societal changes.
KER.6	**Domain-specific Scientific Outputs**. Through the integration of the 10 Spokes closely connected to the PNRR KETs, the CN will provide multi-disciplinary outputs related to the delivery of results making use of state-of-the-art high-TRL HPC and Big Data solutions (Spokes 2 to 9), and to the creation and evolution of top-level laboratories exploiting development and use of new, intermediate-TRL technologies (Spokes 1 and 10).
KER.7	**Dynamic co-creation of results through Open Calls**. By reserving substantial funds to each of the Spokes for the implementation of open calls through the project lifetime, the CN aims to the *dynamic extension* of its results and of the exploitation of its technological solutions. The open calls will effectively act as a *co-creation* conduit to reflect on the experience gained *en-route*, and will be targeted to attract new partners, technologies and strategic ideas.
KER.8	**Growth and Innovation**. Through the engagement of SME, industries and scientific partners, the CN will stimulate growth and innovation in the identified PNRR areas both through dissemination and exploitation activities, such as those identified by KER2, through the increase of the Technology Readiness Levels (TRL) identified by all the Spokes, and through the support to Open Science.
KER.9	**Open Multi-Terabit Network.**

Table 2 – Key Exploitable Results

The KERs mentioned in the table above are then delivered and measured through the Impacts described in the following sections.

Impact of the National Center as an Open Science enabler for innovation and sustainability

A horizontal impact spanning through all the Key Exploitable Results of the CN is represented by its methodology centered on Open Science, intended as a key vehicle to foster innovation and sustainability.

Open Science is intimately linked to the promotion of FAIRness in data management and solutions. Concretely, the neutral ground provided by the advanced CN infrastructure, its reliance on open-source solutions, the long-standing and recognized international experience of the partners delivering and managing the infrastructure, coupled with the strong connection between the infrastructure and the Thematic Scientific Spokes, facilitates the following **Open Science Enablers**:

Open Science Enablers	Contribution to innovation and sustainability
Common data sharing	Definition of public and common interoperability frameworks or toolkits. This will allow open and FAIR-based sharing of data whenever possible and appropriate and will be fostered by the creation of open catalogues of resources and services connected to analogous European initiatives, such as the EOSC Catalogue[17]. In turn, this will allow reuse and extension of methodologies and results.
Optimized scientific workflow platforms	An efficient platform of scientific workflows across the resources enabled by the CN open science infrastructure fosters innovation thanks to effective utilization of the available resources, and therefore to increased speed in the delivery of scientific outcomes. In turn, this contributes to minimize energy and waste, in accordance with sustainability objectives identified by the *Do No Significant Harm* criteria.
Integrated knowledge sharing	The ambitious program of the CN involving distributed training, support, and consultancy, especially with the engagement of young talents, leads to innovation through sharing of knowledge at both the national and international levels and through the creation of innovative bespoke solutions.
Economies of scale	The open science paradigm and the related scientific and technological integrations promoted by the CN allows to avoid lock-ins and closed circles. In turn, this leads to network building capabilities that promote efficient procurement of equipment and know-how, thus contributing to cost reductions and sustainability.
Open pilots, testbeds, and demonstrators	The creation of open pilots, testbed, and demonstrators, to be realized also via the open calls foreseen by the CN, creates innovation and sustainability especially in relation to the competitive creation of industrial opportunities.

Table 3 – Open Science Enablers

How the National Center connects to and implements the PNRR objectives

A set of objectives, described in detail in the implementation plan, has been defined for each of the 11 Spokes of the CN. All the objectives link to the PNRR KETs mentioned in a previous section, addressing foundational areas related to major social concerns, preservation of life and biodiversity, international cooperation, advancements in technology and science, promotion of cultural advancements and heritage, and global challenges such as well-being, clean energy, climate control.

The connection between the objectives and the PNRR KETs is realized through leading HPC, Quantum and Big Data federated resources, a dynamic portfolio of composable and reusable services, thematic data

[17] https://eosc-portal.eu

spaces and data analytics solutions. The following table summarizes the Key Spoke High-Level Objectives, mapping them to PNRR-related areas.

Key Spoke High-Level Objectives	PNRR-related Areas
Spoke 0, Supercomputing Cloud Infrastructure: upgrade of the current computing and service national infrastructure involving terabit-speed networking, HPC, HTC and Big Data systems distributed across the country. Deployment and operation of open-source middleware supporting composable federation services compatible with EuroHPC, EOSC and GAIA-X principles.	• All PNRR KETs.
Spoke 1, Future HPC & Big Data: creation of new laboratories as an integral part of a world-class national federated centre of competence for the advanced co-design of high-performance and high-throughput hardware and software systems, with the goal of strengthening Italy's leadership in the EuroHPC Joint Undertaking and in the data infrastructure ecosystem.	• PKET.1 ○ Microprocessor and system architecture ○ Software tools, mathematical libraries, application and simulation code
Spoke 2, Fundamental Research & Space Economy: creation and the optimization of algorithms and, in general, of computing solutions capable of maximizing the potential physics output from data and simulations.	• PKET.1 ○ Data insight ○ Data reduction ○ Machine learning
Spoke 3, Astrophysics & Cosmos Observation: development of HPC solutions supporting the new generation of observatories and instruments, such as the Square Kilometer Array, the new MeerKAT+ and LOFAR2.0 observatories, the Cherenkov Telescope Array, the European Extremely Large Telescope, the Vera Rubin Observatory, the low-frequency space-born Laser Interferometer Space Antenna, the EUCLID and WFIRST satellite missions.	• PKET.1 ○ Observation and monitoring data systems ○ Interferometry, management of spatial data management of satellite systems
Spoke 4, Earth & Climate: development and implementation of Earth System Models (ESMs) to address the complex challenges facing society in a changing climate.	• PKET.2 ○ Forecast and prediction of the extreme meteorological events
Spoke 5, Environment & Natural Disasters: modelling, simulation and management of natural and anthropic disasters and their consequences, to increase the resilience of local	• PKET.2 ○ Geophysics of earthquake, tsunami, vulcanological phenomena

systems by devising, developing, and implementing proper risk mitigation measures	○ Understanding of anthropogenic disaster phenomena such as pollution, habitat deterioration, spills into groundwater, sea, water basins, fires, loss of biodiversity, collapse of the ecosystem
Spoke 6, Multiscale Modelling & Engineering Applications: complex multiscale mesoscale modelling with quantum accuracy, development of ML and high-performance data analytics methods to extend high-precision quantum simulations to mesoscopic temporal and spatial scales. Simulation of a wide range of devices/systems. Development of advanced numerical methods and applications, open software tools and workflows, to integrate the computing simulation, collection, and analysis of data of interest for engineering applications.	• PKET.7 ○ Energy efficiency of products and production systems. ○ Overcoming the production of disposable and non-recyclable items. ○ Foreshadowing the functionalities of the devices and their life cycle
Spoke 7, Materials & Molecular Sciences: design, implementation, maintenance, and dissemination of enabling technologies for the high-performance numerical simulation of materials and molecular systems.	• PKET.4 ○ Overcoming the use of metallurgical materials and progressive design and use of innovative materials ○ Devices to produce energy from renewable and alternative sources, devices for energy storage.
Spoke 8, In-Silico Medicine & Omics Data: design and development of models and simulation platforms for in silico trials. Production and analysis of omics data through the development of next-generation bioinformatics pipelines and clinical machine learning codes.	• PKET.5 ○ Precision medicine ○ Personalization of diagnosis and treatment ○ studying complex diseases resulting from higher-order interactions of an entire network of genes ○ Study of the interaction between the effect of drugs in relation to therapy ○ Study and treatment of rare diseases
Spoke 9, Digital Society & Smart Cities: investigation of novel approaches that build upon and extend the concept of “digital twins”, digital representation of social and organizational structures of cities and communities and of their citizens, by exploiting available “big data” digital tracks, powerful data analysis and AI techniques and advanced simulation opportunities unlocked by HPC infrastructures.	• PKET.6 ○ Computer processing of digital twins of a complex system such as an urban system. A behavioral study, or simulation, of the twin aims to predict the behavior of the real-world counterpart. That is, the real-world data creates the twin (data driven modeling) and the simulated model information allows you to understand, improve or make predictions for the real-world counterpart.

	○ Large-scale data assimilation has become possible with almost ubiquitous fast internet connections. In this way, observations in the form of data can be continuously incorporated into a simulation and continuous corrections can be made to forecasts or processes. ○ Scalable AI frameworks for Deep Learning, Transfer Learning, Learning-to-Learn-Systems, Generative Adversarial Networks or Continual Learning.
Spoke 10, Quantum Computing: identification of quantum algorithms characterized by a speedup, even limited, if compared with the corresponding classical algorithm. Use of the quantum approach to perform calculations on state-of-the-art classical computers (quantum-inspired algorithms, emulators, tensor network computations) to gain significant benefits in terms of algorithms efficiency without using a quantum hardware.	• PKET.3 ○ integration of quantum computing architectures systems to realize the vision of a holistic future hybrid supercomputer ○ open new possibilities for demanding computing tasks in science and industry ○ industrial optimization and quantum chemistry ○ Industrial pilot users of quantum computing

Table 4 – Key Spoke High-Level Objectives related to PNRR areas

Synergy with Other Programs in PNRR for Mission 4, Component 2

Data-driven R&D is acknowledged as crucial in the PNRR since research and innovation are a booster for enhancing the digital transition of the country. Sharing strategies and trajectories with other PNRR centres, consortia or projects is a goal of this CN, captured especially by KER.1 and KER.4. This CN is thus expected to offer enabling technologies, projects, and expertise also to several other programs of the PNRR. The table below outlines perspectives specifically for synergies with other initiatives in Mission 4, Component 2. Possible joint activities will be considered according to PNRR rules and guidelines.

PNRR Relevant Programs[18]	**Synergy of CN Pillars on PNRR[19]**			**References in PNRR and additional comments[20] (relevant keywords in bold)**
	Enhancing and developing a world-	*Researc h and knowled ge-*	*Training for data-driven profession*	

[18] Names of the specific topics from NRRP programs are in Italian and quoted from governmental official documents, in particular the *NRRP MUR Linee Guida per le iniziative di sistema Missione 4* (https://www.mur.gov.it/sites/default/files/2021-10/Decreto%20Ministeriale%20n.1141%20del%2007-10-2021%20-%20Linee%20Guida_MUR_NRRP_M4C2.pdf).

[19] Table key: ✓ (low synergy impact), ✓✓ (medium synergy impact), ✓✓✓ (high synergy impact). The evaluation of impact is qualitative and based on the description of programs presented in the mentioned official documents of NRRP.

[20] Texts illustrating potential synergies are in Italian since they are excerpts from governmental official documents, in particular from the *NRRP MUR Linee Guida per le iniziative di sistema Missione 4* (see above for the full reference).

		class supercomputing cloud infrastructure	*transfer projects*	*als, scientists, and innovators*	
Other CNs "Campioni nazionali di R&S"	**Specific Topics**				"I Centri Nazionali (CN) sono dedicati alla ricerca di frontiera relativa ad **ambiti tecnologici** coerenti con le priorità dell'agenda della ricerca europea e con i contenuti del PNR 2021-27." The digital transition is at the core of PNRR and data-driven R&D is pervasive in all topics of CNs: all CNs are oriented to data-driven methods for R&D.
	Tecnologie dell'Agricoltura (Agritech)	✓✓✓	✓	✓✓	"Il Centro svolge ricerca e promuove lo sviluppo di **tecnologie innovative** nel settore agricolo […] anche attraverso la prevenzione, la resistenza e la resilienza rispetto ai rischi […]. Il Centro sfrutta le **tecnologie abilitanti come l'intelligenza artificiale** […] Un ulteriore tema del centro riguarda l'applicazione di sistemi di **intelligenza artificiale** in agricoltura e nell'indotto per implementare la sicurezza, la tracciabilità e la tipicità della filiera e dei prodotti agricoli."
	Sviluppo di terapia genica e farmaci con tecnologia a RNA	✓✓✓	✓✓	✓✓	"[…] Il Centro focalizza le sue attività in ambiti ad alto valore innovativo […] come la terapia genica, utilizzando tecnologie basate su RNA, competenze di **biocomputing avanzato e nanomateriali intelligenti**. […] Il Centro contribuisce a raggiungere gli obiettivi del PNRR in relazione al **digitale**."

	Mobilità sostenibile	✓✓✓	✓✓	✓✓	"Il Centro svolge ricerca e promuove l'innovazione di livello nazionale e internazionale sull'insieme delle **tecnologie che contribuiscono ai sistemi e alle infrastrutture di trasporto** […]Il Centro sviluppa una **valutazione sulle condizioni, la fattibilità e il grado di diffusione delle soluzioni tecnologiche possibili**, avendo anche riguardo agli aspetti comportamentali della mobilità. Ad esempio, il Centro sviluppa ricerca e tecnologie per […] i **modelli di simulazione** etc."
	Bio-diversità	✓✓✓	✓	✓✓	"Il Centro svolge ricerca e promuove lo sviluppo di soluzioni per monitorare, preservare e ripristinare la biodiversità funzionale […] L'elemento chiave del Centro di biodiversità sono le **Key Enabling Technologies**, come le **biotecnologie, l'intelligenza artificiale**, le tecnologie per le scienze della vita […] anche sostenendo e sviluppando **biobanche** […]"
Enlarged partnerships extended to universities, research centres, enterprises and funding basic research projects	**Specific Topics**				
	Intelligenza artificiale: aspetti fondazionali	✓✓✓	✓✓✓	✓✓✓	Il Partenariato si concentra sugli aspetti fondazionali del **machine learning e dell'intelligenza artificiale.** In particolare, **sulle grandi**

					sfide teoriche e computazionali aperte […]”.
	Scenari energetici del futuro	✓✓✓	✓✓	✓✓	“La ricerca dovrà svilupparsi in progetti […] ad alta innovatività, per la filiera dell’energia da fonti rinnovabili […] Questo grazie a **processi efficienti, scalabili, sostenibili** e che siano in grado di migliorare l’affidabilità, l’efficienza, la flessibilità e la resilienza del sistema energetico nazionale.”
	Rischi ambientali, naturali e antropici	✓✓✓	✓✓	✓✓	Il Partenariato si concentra sulla necessità di una migliore comprensione dei rischi ambientali, naturali e antropici […] al fine di garantire un miglioramento delle tecniche di **previsione dei rischi** […] e di **prevenzione e mitigazione dei loro effetti** sull’ambiente, nonché la capacità di adattamento dei sistemi. […] Il Partenariato valorizza anche **l'uso di dati** […]”
	Scienze e tecnologie quantistiche	✓✓✓	✓✓✓	✓✓✓	“Il Partenariato si concentra sulla ricerca […] nel campo delle scienze, delle tecnologie quantistiche per applicazioni radicalmente innovative […] **Le scienze e le tecnologie quantistiche sono abilitanti in molteplici campi**, e hanno un naturale approccio problem solving, multidisciplinare e olistico.”
	Cultura umanistica e patrimonio culturale come laboratori di innovazione e creatività	✓	✓✓	✓✓	Il Partenariato […] contribuisce a far progredire la ricerca e il dialogo interdisciplinare e a proiettare la cultura umanistica italiana nel mondo e ad **integrarla con le punte avanzate**

					dell'apporto scientifico e tecnologico."
	Diagnostica e terapie innovative nella medicina di precisione	✓✓✓	✓✓	✓✓	"[…] Tra le metodologie impiegate ci sono: le **piattaforme omiche** (es. genomica, metabolomica, proteomica, radiomica) […] **i metodi e le piattaforme di bioinformatica per l'analisi e l'integrazione di dati** genetici, molecolari, di imaging, clinici e relative agli stili di vita […] **i metodi e tools di machine learning** per la classificazione automatica dei fenotipi e per l'estrazione di signatures di genotipo/fenotipo con valore diagnostico e prognostico […]"
	Cybersecurity, nuove tecnologie e tutela dei diritti	✓✓✓	✓✓✓	✓✓	"[…] **l'utilizzo delle nuove tecnologie - ad es. software di analisi di dati tramite sistemi di intelligenza artificiale, anche basati su sistemi di "High Performance Computing"** […]"
	Conseguenze e sfide dell'invecchia mento	✓✓	✓	✓	"[…] Il programma permette di avere a disposizione un portafoglio di metodologie, e **soluzioni tecnologiche** e organizzative per **il monitoraggio e la valutazione**, nonché per la cura **delle patologie associate all'invecchiamento, in modo integrato con gli aspetti sociali, psicologici, giuridici, ed economici**."
	Sostenibilità economico-finanziaria dei sistemi e dei territori	✓✓✓	✓✓	✓✓	"Il Partenariato riguarda lo sviluppo di **un insieme integrato di basi di dati eterogenei geo referenziati per lo studio delle diverse dimensioni rilevanti per l'analisi dello stato** e

					dell'evoluzione delle condizioni economico sociali dei territori italiani e del sistema economico nel suo complesso. […] Il Partenariato **concorre alla costruzione di repository integrati e di piattaforme di analisi di dati da fonti eterogenee**. […]"
	Modelli per un'alimentazione sostenibile	✓✓	✓	✓	Il Partenariato affronta il problema di coniugare sostenibilità e salubrità alimentare intervenendo sul segmento della distribuzione e del consumo della filiera alimentare. […] Le azioni si rivolgono: allo sviluppo di **sistemi logistici intelligenti** e alla definizione di modelli comportamentali sostenibili per l'approvvigionamento alimentare; all'innovazione tecnologica e digitale per la qualità e la sicurezza dei prodotti alimentari. […]"
	Made-in-Italy circolare e sostenibile	✓✓	✓✓	✓	"[…] Il Partenariato […] riguarda **l'innovazione in chiave 4.0** nei settori nei quali l'Italia è leader e che rappresentano uno snodo per la transizione digitale e sostenibile. […]"
	Neuroscienze e neurofarmacologia	✓✓	✓	✓	"[…] è essenziale la collaborazione […] tra ricerca di base (come la conoscenza approfondita di pathways alterati nella patogenesi) e ricerca applicata (come lo sviluppo di terapie innovative) nelle seguenti attività: […] **l'analisi combinata e multiscala** di pathways molecolari e dei determinanti genetici della fisiologia neuronale; **dati genetici** e l'interazione cervello-corpo e cervello-

					ambiente; […] validazione di nuovi biomarcatori precoci e **modelli predittivi di malattia**. […]”
	Malattie infettive emergenti	✓✓✓	✓✓	✓✓	“[…] La necessità di implementare un sistema di controllo attivo della circolazione delle malattie emergenti in Italia […] Il Partenariato ha l’obiettivo di integrare gli attuali sistemi di monitoraggio per: […] **applicare sistemi di intelligenza artificiale e analisi di big data per costruire modelli previsionali e progettare azioni di intervento e contenimento**. […]”
	Telecomunicazioni del futuro	✓✓	✓✓	✓	“Nei prossimi anni lo sviluppo di interi settori dipenderà dagli investimenti che i Paesi riusciranno a realizzare in ricerca e sviluppo **su infrastrutture e servizi di telecomunicazioni** […].” HPC capabilities will be more and more needed for telecommunications to fully unlock the value of the data generated by devices.
	Attività spaziali	✓✓✓	✓✓	✓✓	“[…] **prevenzione** dei disastri e dello space weather; […] **modellizzazione dei processi complessi** indotti da eventi geologici estremi […] **previsione delle condizioni meteorologiche e climatiche** estreme […] capacità di **previsione delle diverse componenti del ciclo del carbonio**; dell’agricoltura sostenibile; […] **integrazione dei dati e della gestione di policies urbane e suburbane** […]”

Research and innovation infrastructures[21]	✓✓✓	✓✓	✓	Potential strong synergies with all projects involving research data infrastructures.
Innovation ecosystems and territorial champions of R&D	✓✓	✓✓	✓✓	"Gli ecosistemi dell'innovazione svolgono un ruolo cruciale **nell'attuazione delle attività di ricerca e innovazione nel campo della sostenibilità ambientale e sociale**. A tal fine, viene attuato un approccio orientato alle grandi sfide, favorendo la creazione di innovazione di impatto e l'imprenditorialità."
PhDs and researchers green and Innovation (REACT-EU)	✓✓✓	✓	✓✓	All doctoral projects based on data-driven studies.

Table 5 – Synergy with Other Programs in PNRR

How the National Centre connects and contributes to the NextGeneration EU objectives

NextGenerationEU is the temporary instrument designed to boost European recovery; it will be the largest stimulus package ever financed in Europe. It is cantered around supporting modernization through a few high-level objectives.

The CN, thanks also to the experience and key involvement of its partners in some of the largest European projects and frameworks, such as EuroHPC, the EOSC, the European Data Strategy, GAIA-X , the European Processor Initiative and in multiple ESFRI projects, will connect and contribute to the NextGenerationEU objectives with its Key Exploitable Results, according to the following table:

NextGenerationEU Objectives	CN Key Exploitable Results
Research and innovation, via Horizon Europe	• KER.1, KER.2, KER.3, KER.4, KER.5, KER.6, KER.7, KER.8, KER.9
Fair climate and digital transitions, via the Just Transition Fund and the Digital Europe Programme	• KER.1, KER.6, KER.7
Preparedness, recovery and resilience via the Recovery and Resilience Facility, rescEU and a new health programme, EU4Health	• KER.1, KER.2, KER.3, KER.4, KER.5, KER.6, KER.7, KER.8, KER.9
Modernization of traditional policies such as cohesion and the common agricultural policy, to maximise their contribution to the Union's priorities	• KER.2, KER.5, KER.7

[21] See, in particular, https://www.mur.gov.it/it/atti-e-normativa/avviso-n-3264-del-28-12-2021

Table 6 –NexGenerationEU Objectives related to the CN KERs

Expected impacts, KPI, TRLs and reference to the Key Exploitable Results

The following table lists the expected impacts of the CN, associated to the KERs and measured through **Key Performance Indicators** (KPI). A set of projected future metrics is given for each impact.

Impact 1: Support the whole research lifecycle through integrated and federated HPC and Big Data resources	
Project Contribution: KER.1, KER.2, KER.3	
Indicators (target)	Projected Future Impact
Conventional node hours available per year (equivalent to 4000 node): 30 Ml Accelerated node hours available per year (equivalent to 2000 node: 15 Ml Peta byte processed per month: 10 PB Integrated services per year: 100	10% annual increase of HPC adoption 20% annual increase of Cloud adoption 10% annual increase of multi-service selection
Impact 2: More scientific communities have access to state-of-the-art HPC and Big Data services	
Project Contribution: KER.2, KER.3, KER.4	
Indicators (target)	Projected Future Impact
Active user per year: 5000 New involved communities: 10 # of stakeholders trained (1K) # of handbook downloads (1K)	10% annual increase of users 10% annual increase of organized data spaces
Impact 3: Enforcement and delivery of best practices for Green Computing	
Project Contribution: KER.1, KER.5	
Indicators (target)	Projected Future Impact
# of data centres reducing their Power Usage Efficiency, or PUE (5) # of trainings delivered on Green Computing (50)	PUE lower than 1.2 for new system installations Improved application run time energy efficiency as a result of code optimizations
Impact 4: High-speed connectivity	
Project Contribution: KER.1, KER.9	
Indicators (target)	Projected Future Impact
# of cities or data centres connected with > 100 Gbit/s speed (20) # of cities or data centres connected with Terabit/s speed (5)	10% annual increase in overall network bandwidth utilization Research network capacity in the South and at national level: x10
Impact 5: Increase of available state-of-the-art facilities and know-how throughout the country	

Project Contribution: KER.1, KER.2, KER.9	
Indicators (target)	Projected Future Impact
Increase in the overall storage capacity (20%) # of new network point of presence / access with >100Gbps (20) # of PhD or grants funded by the CN (200) # of personnel hired by the CN (200) % of the budget dedicated to open calls (10%)	10% annual increasing porting of applications 10% annual increasing of data repository capacity Persistent increasing number of involved PhD students and structured researchers More than 20 new network points of presence with > 100 Gbit/s speed access; more than 4 data centres with > 1 Tbit/s speed access
Impact 6: Advancement in Quantum Computing	
Project Contribution: KER.1, KER.6, KER.8	
Indicators (target)	Projected Future Impact
# of use cases demonstrated on quantum computers, real and/or emulators (3) # of qubits implemented on a classical HPC emulator (100) # of qubits implemented on at least 2 technological platforms (5)	100% 100% 100%
Impact 7: Increase of integration with industry	
Project Contribution: KER.2, KER.7, KER.8	
Indicators (target)	Projected Future Impact
# of industry-related testbeds and pilots (15) # of analysed industrial use cases (15) # of open calls involving SME (3)	10 % increasing per year of industry Proof of Concept 10% increasing per year of test before investment and assessment of industry use cases
Impact 8: Increase of Technology Readiness Levels	
Project Contribution: KER.8	
Indicators (target)	Projected Future Impact
# of Spokes or solutions reaching TRL6 (30) # of Spokes or solutions reaching TRL7 (30) # of Spokes or solutions reaching TRL8 or above (20)	The domain PoC workflows designed by the spokes must have a TRL of no less than 6. The application solutions developed by the spokes must have a TRL of no less than 7. The enabling specialist support activated by the spoke infrastructure will have to increase the TRL level above 8 for at least 5 workflows and for at least 5 application solutions per year.
Impact 9: Reduction of geographical and gender-related digital divides	
Project Contribution: KER.1, KER.2, KER.5, KER.9	
Indicators (target)	Projected Future Impact

% of women leading Tasks or Work Packages % of women hired by the CN (40%) % of investments in the South (40%) % of demonstrators in the South (40%) % of open calls in the South (40%) % of PhD in the South (40%) % of personnel of the CN in the South (40%)	Investment, Demonstrators, Open Calls, PhD and personnel staff involved in the South regions: compliant with the target prescriptions. Gender equal opportunity: Elaboration, preparation and approval of the Plan for equal gender opportunities Personnel recruitment procedures: The procedures will be prepared with reference to the merit and competence criteria without any precautions regarding the gender of the candidates. All appointments defined through elective procedures will provide for the obligation to express two preferences, each for a candidate of a different gender

Table 7 –Expected Impacts

Risk Assessment

Foreword

Proper identification of events that may have a negative impact on the objectives of the ICSC project is essential to come to mitigation procedures aimed to minimize that impact. Beyond their identification, it is also very important to associate these events to a certain likelihood of their occurrence. In this section, we describe the **Risk Assessment Process** that we have followed in the ICSC project. The process is designed to eventually come to a proper, quantitative evaluation of these possible events.

In general, negative impact events can be described as **barriers or obstacles** to the full realization of the stated objectives of the project. These barriers and obstacles we categorize as possible **risks**; for easier categorization, they can be logically subdivided into several domains, listed below. Each risk associated to a certain event or scenario is assigned a likelihood, as well as a possible impact. Mitigation actions or strategies are then associated to each risk.

However, as already outlined in the preliminary submission of the proposal, it is important to note that proper Risk Assessment, especially for a project of this size, cannot be statically defined at any one point in time. For this reason, the input provided in this section will be complemented in ICSC throughout the course of the project by a **Continuous Risk Monitoring and Assessment (CRMA)** methodology, which is described later.

Definition of Risks

Risks have been categorized in a scale **from 0 (very low risk) to 8 (very high risk)**. A risk with a certain value is associated on the one hand to the negative impact that that risk is estimated to have and, on the other, to the likelihood for a certain scenario to occur.

There is naturally some uncertainty in categorizing risks and their related impact and likelihood. However, we have strived to minimize it in multiple ways: first, by resorting to best practices for risk assessment techniques, as described for instance in ISO/IEC 31010:2019[1], in particular taking into account the risk evaluation techniques described in pp. 31-37 of that document. Then, we have also specifically drawn on the extensive, multi-year international experience of the ICSC partners in proposing, implementing, and running state-of-the-art distributed scientific infrastructures, experiments, and collaborations. This full body of experience will also be brought forward throughout the project with direct involvement of the ICSC management structure, through the already mentioned CRMA methodology.

In general, using the 0-8 scale mentioned above, we have defined:

- **Low** risks: 0-2
- **Medium** risks: 3-5
- **High** risks: 6-8

The following table, adapted from the ENISA Cloud Computing Security Risk Assessment[2], shows the risk level as a function of the negative impact and of the likelihood of a certain scenario.

	Likelihood of incident scenario	Very Low (Very Unlikely)	Low (Unlikely)	Medium (Possible)	High (Likely)	Very High (Frequent)
Impact	Very Low	0	1	2	3	4
	Low	1	2	3	4	5
	Medium	2	3	4	5	6
	High	3	4	5	6	7
	Very High	4	5	6	7	8

Risk Domains

The following **logical risk domains** have been identified:

Scientific and Technological Domain

Risks in Scientific and Technological domain pertain to events that have an impact on the scientific results of the National Centre. This impact may be directly related to scientific outcomes or may affect technological installations or developments that are functional to the scientific outcomes.

Regulatory and Systemic Domain

Risks in the Regulatory and Systemic domain are connected to processes and procedures that are intimately linked to how the National Centre operates, to the efficiency of its governance structure, and to boundary conditions traceable to regulatory stipulations and agreements.

Economic Domain

Risks in the Economic domain are linked to events that may affect the ability of the National Centre to fulfil its commitments related to funding, running costs, tenders, or any other activity that has an impact at the financial level for the good operation of the Centre.

By categorizing the risk domains, ICSC intends to **better evaluate the relative severity of the identified risks**, which is a key factor in the definition of the priority of the mitigation measures. As it will become apparent when discussing the distribution of risk probabilities and impacts, for the ICSC project the risks in the Scientific and Technological Domain are typically clustered around low-medium levels, the risks in the Regulatory and Systemic Domain are clustered around medium levels, and the risks in the Economic Domain are clustered around medium-high levels. This is a useful information, not only from the prioritization point of view, but also because it allows to partition roles and responsibilities for certain risk domains, so that the best strategies and competences are activated to tackle each risk class. However, this is a general observation and notable exceptions may occur (see for instance R.11 in the table below). This points to the fact that Risk Assessment must always be carried on holistically, and that the right prioritization of the mitigation measures must always consider the actual details of each potentially negative event. This reinforces the need to adopt the CRMA process that was introduced in the Foreword above, and that will be described below.

Risks

The following table describes the **risks currently identified by the project**. Each risk is individually labelled (*Risk ID*), has a specific name (*Name*), a description of its potential consequences (*Consequence*) and belongs to a certain *Domain*. To each risk a *Probability* and an *Impact* are attributed, which jointly define the overall *Risk Level*, determined according to the previous table. Finally, the table below also contains the *Proposed Mitigation Measures* defined for each risk.

Risk ID	Description of Risk	Proposed Mitigation Measures
R.1	**Name**: Withdrawal of key partners, experts, or developers. **Consequence**: Key know-how in the project may be lost. The tasks assigned to the withdrawing partner might not be fully delivered. **Domain**: Regulatory and Systemic **Probability**: LOW **Impact**: MEDIUM	If a key partner, expert, or developer abandons the project, its tasks will be taken over by other partners involved in the same activities. A further possibility would be the involvement of external partners with similar expertise. Withdrawal of a key partner will cause delay and might result in incomplete works, but the main goals will be completed.
	Risk Level:	MEDIUM
R.2	**Name:** The needed manpower effort is underestimated. **Consequence**: Not all the goals of the projects might be met. **Domain**: Regulatory and Systemic **Probability:** LOW **Impact**: HIGH	The expertise gained by the ICSC partners in other national and international projects ensure a proper estimation of the efforts needed. Potential problems will be monitored at an early stage and can results in rescheduling efforts.
	Risk Level:	MEDIUM
R.3	**Name:** Insufficient development of the exploitation and sustainability strategy. **Consequence:** Exploitation or sustainability goals may not be reached. **Domain**: Regulatory and Systemic **Probability**: MEDIUM **Impact**: HIGH	ICSC will prepare an exploitation plan that will be evaluated and updated by the end of the project, based on the continuous observation of the ICSC needs and on feedback received from stakeholders.
	Risk Level:	MEDIUM
R.4	**Name:** Failure to timely recruit personnel. **Consequence:** There will be a delay of some project outcomes.	Recruitment of personnel will be closely monitored and publicized through multiple channels. If necessary, the federated nature of the ICSC project

	Domain: Regulatory and Systemic **Probability:** MEDIUM **Impact**: HIGH	will allow to shift resources or efforts to other locations or partners to meet the project milestones.
	Risk Level:	MEDIUM
R.5	**Name**: Delivery of services or of Spoke outputs is threatened due to COVID-19 (or other force majeure occurrences) limiting availability of experts, events, or resources. **Consequence**: Delay of project outcomes. **Domain**: Regulatory and Systemic **Probability**: LOW **Impact**: LOW	ICSC will closely monitor the situation and ensure Continuity plans are updated adequately. As contingency plan, the ICSC project will identify alternative sources of expertise or resources across the federation and will give a strong focus on a "Digital-first" attitude for events and outcomes to mitigate the risks of constrained physical availability.
	Risk Level:	LOW
R.6	**Name**: Difficulty in achieving sustainable collaboration between the funding agencies, the partners, or the associated communities beyond the project life. **Consequence**: Possible major disruption of the ICSC outcomes after the commitment of the project proponents ends. **Domain**: Regulatory and Systemic **Probability**: MEDIUM **Impact**: VERY HIGH	ICSC will closely monitor external risk factors related to sustainability. A Sustainability Roadmap will be delivered as a living document. A special focus will be given to collaborations with National and International initiatives, as well as with industry, to enforce sustainability.
	Risk Level:	HIGH
R.7	**Name**: Policies and recommendations issued by ICSC are not endorsed or adopted by high-level stakeholders. **Consequence**: Over time, the policy and strategic outcomes of the National Project might become out of sync with current practices. **Domain**: Regulatory and Systemic **Probability**: MEDIUM **Impact**: LOW	ICSC will organize periodic meetings with relevant high-level National and International stakeholders, to ensure its outcomes are properly taken into consideration and promoted.
	Risk Level:	MEDIUM
R.8	**Name**: Some of the scientific objectives of the vertical spokes turn out to be too complex to be tackled during the project. **Consequence**: Some expected key scientific results may not be realized. **Domain**: Scientific and Technological **Probability**: LOW **Impact**: LOW	The vertical spokes all work in a peer-controlled way, so that there is continuous feedback on the outcomes. Also, the proposed activities for each spoke have all been pre-validated by the ICSC partnership. In case any unforeseen negative event occurs that may prevent reaching some key result, the coordinating hub will be immediately informed to analyze the severity of the event.
	Risk Level:	LOW

R.9	**Name**: Some technologies expected to be used by some of the vertical spokes, such as for example advanced Machine-Learning models, turn out to yield poor predictions. **Consequence**: Proper scientific results are undermined. **Domain**: Scientific and Technological **Probability**: MEDIUM **Impact**: MEDIUM	As part of scientific exploration and innovation, it is possible that advanced technologies do not deliver immediate results. However, the know-how and experience of the partners that are part of the vertical spokes will either accept this as a data point to turn to different technologies or identify alternative paths. This may lead to some delays that is not possible to foresee in abstract terms, but it is not likely that this causes severe disruptions to the project.
	Risk Level:	MEDIUM
R.10	**Name**: Quantum Computing technologies turn out to be only useful for small-scale or small-impact problems or, in general, are not viable to complement traditional computing technologies. **Consequence**: The direct impact of results obtained thanks to Quantum Computing technologies is not apparent. **Domain**: Scientific and Technological **Probability**: MEDIUM **Impact**: LOW	Quantum Computing is still in its relative infancy. Therefore, it is possible that when applied to novel fields or problems it does not turn out to provide immediate results. However, even this case, if it occurs, will be very valuable in providing concrete and real-world experience within the project with Quantum Computing, whose strategic importance for Italy and beyond should not be underestimated.
	Risk Level:	MEDIUM
R.11	**Name**: The services provided by ICSC utilize various types of distributed hardware. Users will share hardware and software resources and, in several cases, will rely on logical isolation mechanisms to protect their data. These logical mechanisms may get compromised or exposed. **Consequence**: compromise of isolation mechanisms may lead to potential data breach. **Domain**: Scientific and Technological **Probability**: HIGH **Impact**: VERY HIGH	The project foresees the creation of an ICSC SOC (Security Operation Centre), specifically tasked to protect and react to security incidents. Beside this, the ICSC infrastructure is largely based on a homogeneous platform ran by few very experienced partners, where updates and patches will be rolled out much more easily than with a disjoint model.
	Risk Level:	HIGH
R.12	**Name**: In some cases, effective scientific outcomes may depend on fast and reliable connections between edge data producers and ICSC core resources. While distributed ICSC resources may have built-in redundancy, it may happen that edge locations get disconnected from the ICSC core or suffer degraded connections, due for instance to force of nature reasons, or to other technological or bureaucratic reasons. **Consequence**: This may lead to reductions in the scientific impact of ICSC.	Connection to edge sites will be realized through resilient network connections. ICSC foresees to also create content distribution networks to leverage automatic data replication. In extreme cases, the possibility to resort to temporary alternative network providers for connection to edge sites will be considered.

	Domain: Scientific and Technological **Probability**: LOW **Impact**: LOW	
	Risk Level:	LOW
R.13	**Name**: A participant to ICSC may find itself locked into the procedures, tools, services, or data formats offered by ICSC. **Consequence**: inability to port or expand some solutions to other providers. **Domain**: Scientific and Technological **Probability**: VERY LOW **Impact**: MEDIUM	ICSC is strategically focused and built on open solutions, interfaces, and full support of the FAIR paradigm for data, thus minimizing the risk of vendor lock-in.
	Risk Level:	LOW
R.14	**Name**: The different ICSC actors operating at the infrastructural level (INFN, CINECA, GARR) may need to adopt different strategies and procedures for what regards operational and security coordination, due to their user / funder needs. **Consequence**: disruption of service provisioning or in general potential disservice to users. **Domain**: Regulatory and Systemic **Probability**: MEDIUM **Impact**: MEDIUM	The actors working at the infrastructural level have decades of experience working together both at the national and international levels. They also participate to multiple initiatives where they have the best interest to coordinate efficiently. Besides, an ICSC Security Operation Center (SOC) will make sure that proper coordination of activities, especially related to security, is performed.
	Risk Level:	MEDIUM
R.15	**Name**: The participants to ICSC express regulatory requirements leading to the need of certain official certifications to use ICSC resources and solutions, and ICSC may not satisfy these requirements, or may only partially satisfy them. **Consequence**: the exploitation of the ICSC investment may be limited due to the inability to offer the required certifications. **Domain**: Regulatory and Systemic **Probability**: HIGH **Impact**: HIGH	Several of the ICSC partners already have multiple official certifications. Besides, the ICSC activity program foresees the achievement of additional certifications (such as ISO 27001, 27017 and 27018) at multiple locations. Should new certification requirements arise, ICSC will either redirect these requirements to solutions or locations already in possess of these certifications, or trigger activities to acquire them.
	Risk Level:	HIGH
R.16	**Name**: ICSC includes many participants. Some malicious insiders that are part of the ICSC technological infrastructure might cause significant disruption or even strategic damage to the project. **Consequence**: severe disruption of the ICSC project.	At the infrastructural levels, the ICSC partners will work with the rule of the least privilege. Besides, the Cloud infrastructure is designed with resiliency in mind. Finally, threat management will be also enforced through the definition of dedicated roles.

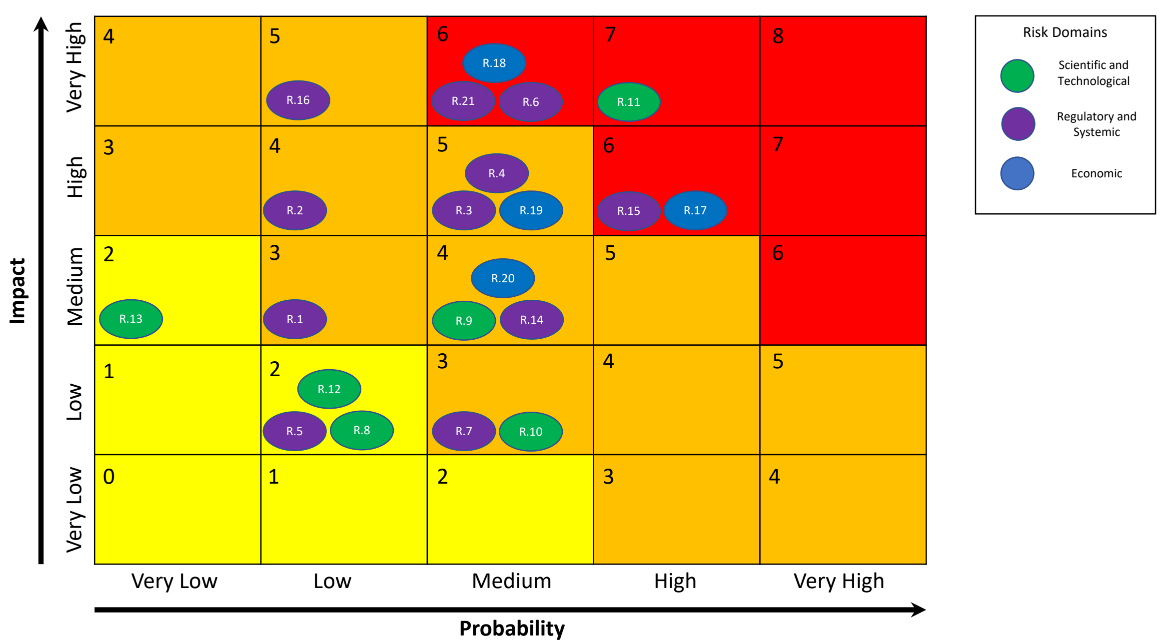
	Domain: Regulatory and Systemic **Probability**: LOW **Impact**: VERY HIGH **Risk Level:**	MEDIUM
R.17	**Name**: The cost of purchasing equipment will substantially raise beyond initial estimates. **Consequence**: It will be difficult or even impossible to come to the expected strategic and scientific results of the ICSC project. **Domain**: Economic **Probability**: HIGH **Impact**: HIGH **Risk Level:**	This a risk that is difficult to quantitively estimate in advance. However, given current world circumstances we foresee it has a high probability to occur. If it occurs, activities may need to be scaled down, unless mitigation support is externally provided in some form. Internally, wherever possible the tenders will be defined and aggregated so to profit from economies of scale and from the largest possible competition. HIGH
R.18	**Name**: The cost of running the ICSC infrastructure will substantially raise beyond initial estimates due, for example, to unforeseen increase of energy costs. **Consequence**: It will be difficult to maintain operational readiness of the ICSC services, thus degrading or potentially even prohibiting outcomes. **Domain**: Economic **Probability**: MEDIUM **Impact**: VERY HIGH **Risk Level:**	This a risk that is difficult to quantitively estimate in advance. However, given current world circumstances we foresee it has a high probability to occur. If it occurs, activities may need to be scaled down, unless mitigation support is externally provided in some form. Internally, all the infrastructural partners (e.g., INFN, CINECA) already have programs in place to move to greener infrastructures to reduce environmental impact and costs. HIGH
R.19	**Name**: The ICSC project is not funded according to the original requests. **Consequence**: scaling down of ambitions of the project and thus of its outcomes and impact is possible. **Domain**: Economic **Probability**: MEDIUM **Impact**: HIGH **Risk Level:**	In case ICSC is not funded according to the original requests, before scaling down activities the Hub will seek additional funding from either the public or private sectors. MEDIUM
R.20	**Name**: The acquisition of some equipment purchased by ICSC is delayed due to litigations connected to the tenders. **Consequence**: There may be a delay of some project outcomes. **Domain**: Economic **Probability**: MEDIUM **Impact**: MEDIUM **Risk Level:**	Perusing the extensive multi-decade know-how of its partners, ICSC will create transparent tenders that may foresee multi-year commitments (rather than spot acquisitions), provisioning of equipment through ready-made marketplaces wherever appropriate, and careful support provided by the experienced legal teams of the infrastructural partners (INFN, CINECA, GARR). MEDIUM

R.21	**Name**: The acquisition of some equipment purchased by ICSC is delayed due to shortcomings of the supply chain. **Consequence**: There may be a delay of some project outcomes. **Domain**: Regulatory and Systemic **Probability**: MEDIUM **Impact**: VERY HIGH	Providers of goods to be purchased will be given aggressive timelines, associated to possible penalties to reduce the duration of the procedures for ordering, delivery and installation. The market will also be informed as early as possible about acquisition procedures, to simplify and speed up pre-orders for procurement.
	Risk Level:	HIGH

Risk Distribution

The following picture shows the **distribution of the risk probabilities and impacts**. In the picture, risks are coloured according to their domain, using the following colour code:

- Risks belonging to the Scientific and Technological Domain are in GREEN.
- Risks belonging to the Regulatory and Systemic Domain are in VIOLET.
- Risks belonging to the Economic Domain are in BLUE.



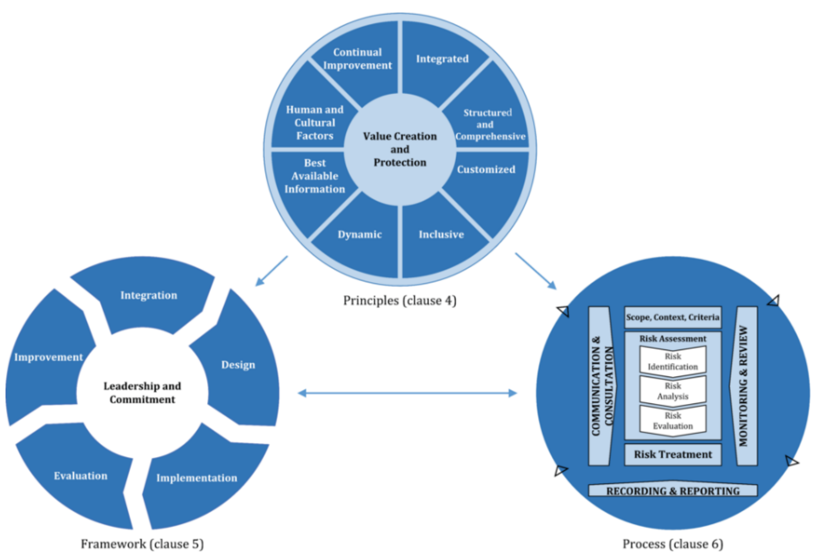
As the picture above shows, **no risks have been identified as "Extreme Risks"**, defined as risks with Very High Probability and Very Low Impact, or Very Low Probability and Very High Impact, or Very High Probability and Very High Impact. This in general points to an overall general stability and consistency of the project in terms of Risk Assessment, since extreme risks are often associated to little directly relevant data, while in terms of mitigation measures, they are often connected to the need to resort to expert opinions, rather than to quantitative best practices[22].

On the other hand, **the examination of the distribution of the Risk Domains** considered here shows that risks in the Scientific and Technological Domain are typically clustered around low-medium levels, risks in the Regulatory and Systemic Domain are clustered around medium levels, and risks in the Economic Domain are clustered around medium-high levels. This type of information, while general in nature, is useful for the project to start prioritizing actions for the different risk domains and, most importantly, to define roles and responsibilities for each risk domain.

Finally, the presence in the picture of some possible "outliers", such as R.11, points to the fact that it is important to **see Risk Assessment as part of a more general overall project strategy**, aiming to reach all project goals and taking into consideration more conditions than risks alone. This strategy is cantered around a Continuous Risk Monitoring and Assessment (CRMA) process, to which we now turn.

Continuous Risk Monitoring and Assessment (CRMA)

The **CRMA methodology** we have defined for ICSC heavily draws from guidelines related to risk management as specified by ISO 31000:2018[23]. In essence, the CRMA process requires to establish a virtuous circle between the three domains of **Main Principles, Overall Framework, and Process**. These links can be shown by the following picture, taken from the ISO 31000 guidelines:



We thus see that Risk Assessment is only a part of an overall vision that considers the success of the entire project as a whole.

---

[22] For some literature on Extreme Risks methodologies, see Franklin, J. (2008), "Operational risk under Basel II: A model for extreme risk evaluation" (https://web.maths.unsw.edu.au/~jim/oprisk.pdf) and Taleb, N.N. (2007), "The Black Swan: The Impact of the Highly Improbable" (ISBN: 978-1-4000-6351-2).

[23] https://www.iso.org/standard/65694.html

To concretely implement a CRMA in ICSC, it is therefore essential to include the CRMA itself in the running governance and management of the project. Only in this case will then be possible to gather feedback on the already identified risks, on their likelihood and proposed mitigation measures, and to extend and complete the Risk Assessment, taking into continuous account actual implementation constraints and opportunities. As the picture above clearly shows, what we simply define as "Risk Assessment" must be closely connected to the management commitment to evaluate, design, implement, integrate, and improve the proposed mitigation strategies.

Therefore, in the first months of the project, the ICSC Hub will publish a **CRMA Overall Strategy**, complemented by a **living Risk Monitoring and Assessment document** targeted to capture the evolving needs and results of the project for what regards risk management.

While the details of the CRMA will be published as a project deliverable, we foresee that the CRMA implemented by ICSC, based on both unstructured and structured data[24], will be **centred around four pillars**:

1. **Identification of the risks to be monitored**, complementing the ones that are described here.
2. **Development of relevant Key Risk Indicators** (KRI)[25]. The definition of the KRIs, often considered as useful *early warnings*, is essential to complement the KPIs, i.e., the metrics that help to assess progress toward the declared goals.
3. **Continuous assessment of the KRIs**. Since KRIs also depend on evolving conditions, strategic goals, and priorities, it is of the utmost importance that this assessment process is dynamic. It is also important to note the "K" of KRIs means that not all possible risks indicators applicable to ICSC will be tracked, but only those that, considering the "Principles" and "Framework" domains defined by the ISO/IEC 31000:2018 guidelines (see picture above), adapted to the ICSC context, can be considered key to the project results.
4. **Prioritization of the risk procedures to minimize risk exposure**. ICSC will strive to automatize the prioritization process as much as possible. In fact, since with a project this big the amount of information that needs to be processed to come to determining the right prioritization for a given risk is large, importance will be given to the definition of quantitative rather than qualitative methods to assess and prioritize appropriate risk procedures. Wherever this is not possible, for instance due to the lack of machine-actionable information, the ICSC Risk Monitoring and Assessment document referenced above will be used to capture information that will be fed back to the ICSC Governing bodies ("Framework" domain in the picture above).

Conclusions

The Risk Assessment strategy defined by ICSC is based on industry best practices tailored to an evaluation of possible Key Risks for the project that is as objective and as dynamic as possible.

We started with a general overview of the project Risk Assessment when submitting the Expression of Interest elicited by MUR.

We now complete that overview with the description of a fuller **two-stage strategy**. This strategy first focuses on **capturing likelihood, impact, and domain** for a substantial number of identified risks, to which we have associated proper mitigation strategies.

---

[24] For an overview of common CRMA methods, see https://www.linkedin.com/pulse/continuous-risk-assessment-leveraging-live-data-risks-regelbrugge/ and https://www.dhg.com/article/continuous-risk-management-conceptualizing-a-continuous-risk-management-framework-part-1.

[25] For an introduction to KRIs, see https://en.wikipedia.org/wiki/Key_risk_indicator.

This part will then be **complemented throughout the project lifetime with a Continuous Risk Monitoring and Assessment (CRMA) process**. The CRMA is designed to integrate the information available so far with dynamic risk assessment and with the definition of Key Risk Indicators, able to capture the evolving conditions and strategies of the project, thus contributing to the **goal of consistently minimizing the impact of negative events that may occur during the project**.

Measures to maximize Impact

Detailed dissemination and exploitation plans will be defined at the beginning of the project, with their outcomes captured through dedicated Deliverables. These activities will contribute to lower the barriers listed above and help in maximizing the Impacts described. They will be implemented and supported by **dedicated Communication and Exploitation teams**, whose full tasks will be detailed in the complete project proposal.

Moreover, the CN governance foresees an **Innovation Manager** and an **Education Manager**. Their duty is to evaluate innovations generated by the project and educational opportunities, connecting them to Communication and Exploitation, also assessing the potential for the generation of start-ups or spin-offs. Assistance to the maximization of Impact opportunities will also be provided by the establishment of an **Observatory on Supercomputing Trends and Applications**, as well as by the already mentioned connections to national and international activities. Then, to mitigate risks and at the same time maximize impact, the CN has decided for each Spoke to **nominate both a Spoke Leader and a Spoke Co-Leader**.

The general process for maximizing the Impacts of the CN is based on the **full integration of its resources and services**. Further, the proposed measures are targeted at making sure that they reach the largest relevant user and stakeholder groups, with the goal of facilitating and enhancing their research and mission. Consequently, the measures to maximize impact are all based on the **strong interconnection between communication, dissemination, and exploitation activities**. This process requires a dynamic attitude to these measures since the evolution in the achievement of the CN KERs will necessarily require adaptation of the communication and dissemination activities. In addition, it is important to note that the **close connections of the CN to other NRPP programs**, described in Table 6 above, will also shape communication, dissemination, and exploitation activities.

Dissemination of results

A previous paragraph already mentioned the important of Open Science for the CN as an enabler of innovation and sustainability. In the context of dissemination, Open Science naturally leads to the sharing of knowledge and results for others to use and will therefore drive all dissemination measures.

According to its general impact framework, the CN plans to connect the dissemination of results to its KERs, so that measures to maximize impact can be consistently mapped into the overall CN activities. The following table provides this mapping:

KERs involved	Dissemination message	Dissemination channel
KER.1 (An integrated HPC, Quantum and Big Data Platform), **KER.4** (An interoperable open multi-tier data repository)	Processing solutions on large e-Infrastructures, utilization of distributed data repositories.	Websites and news, scientific publications, social media platforms, conferences and events, newsletters.

KER.2 (Consultancy, technology transfer and training courses)	Programming on heterogeneous and quantum systems, green computing, analytics on Big Data, transfer to the productive system.	Trainings, MOOCs, webinars, knowledge bases, reports, white papers, press releases, social media, conferences.
KER.5 (Strategy and recommendations)	Policy recommendations, best practices, whitepapers, blueprints.	Mailing lists, website area and news, papers.
KER.7 (Dynamic co-creation of results through Open Calls)	Engagement opportunities, enhancement of existing results, exploitation.	Newsletters, social media platforms, collaboration with other projects, call for ideas.

Table 9 –Dissemination of the project KERs

Communication activities

Communication activities can be divided into measures aimed at communicating project activities, and measures aimed at communicating project results. The following table lists both measures, associating them to target KPIs. An important remark is that the communication tasks mentioned in the table will all cover multiple audiences, ranging from scientists, to citizens, to society at large, to media, to the CN stakeholders.

Objective	Measurable output	Activities	Results
Definition of the CN image (logo, branding, social media)	Communication kit and general infographics produced and downloaded at least 100x	X	
Promotion of the CN objectives	Online content (short videos, tweets) linked to social media and liked, viewed, or shared at least 1000x	X	
Promotion of internal face-to-face events	Barring force-majeure reasons, participation to face-to-face kick-off and to 2x flagship events with at least 90% partner representation	X	
Promotion of webinars and seminars organized by the CN	At least 90% of the estimated maximum number of participants registered		X
Promotion of coordination with other NRPP activities	Dedicated web pages visited at least 200x/month; 2x joint events organized		X
Promotion of open calls	Dedicated web pages visited at least 100x/month; material of the open calls material downloaded at least 50x		X
Promotion of recommendations	Recommendation material downloaded at least 100x		X

Table 10 –Communication actions

Exploitation activities

The CN exploitation activities are naturally centred around the already discussed Key Exploitable Results, which are the foundation of the CN implementation. In previous sections, the KERs have been mapped to the PNRR expected outputs, to the impacts, to possible risks, to dissemination, related to the other PNRR instrument, as well as to international activities. Exploitation is therefore an integral part of the CN.

To support long-term exploitation, the CN plans to dedicate a specific activity tasked to identify new opportunities and seek the means to exploit them during and after the project's execution, closely linking them to the generated innovations, through an Innovation Manager. These innovations will be continuously followed, to assess their possible use and ownerships. The detailed framework for exploitation activities will be delivered in the first months of the project; at the proposal stage, the CN plans to generally consider the following action plan:

- Continuous evaluation of the CN market opportunities
- Adaptation and refinement of selected CN outcomes
- Identification of IPR ownerships and flows
- Evaluation of the possibility to create ventures such as spin-offs or innovative start-ups, possibly through joint public and private funding and collaborations.
- Customer-centric communication campaigns focused on the promotion of the exploitable results.
- Integrated CN-wide collaboration strategy to search for opportunities and extensions in the use of the CN results. This point is especially important and amenable to promising developments, given the large base of experienced stakeholders forming the CN partnership.
- Connection with national and international collaborations and funding bodies to explore further possibilities of exploitation of the project results.

IPR management

While detailed IPR management will be described in detail at the beginning of the project, the CN will require that all partners agree on explicit rules concerning IP ownership, access rights to any Background and Results for the execution of the project, and to the protection of intellectual property rights (IPRs) and confidential information. This agreement will include details on the ownership of background knowledge, on the ownership of foreground knowledge, on the use of foreground knowledge, on the protection of foreground knowledge, and on provisions for the transfer of results.

For what regards IPR ownership, the detailed IPR management agreement will detail the actual provisions. One possibility that will be proposed is that results shall be owned by the project partner carrying out the work leading to them. On the other hand, if any results are created jointly by more than one partner, and if it is not possible to distinguish between the contributions of each of the project partners, such results, including inventions and all related patent applications and patents, will be jointly owned by the contributing project partners.