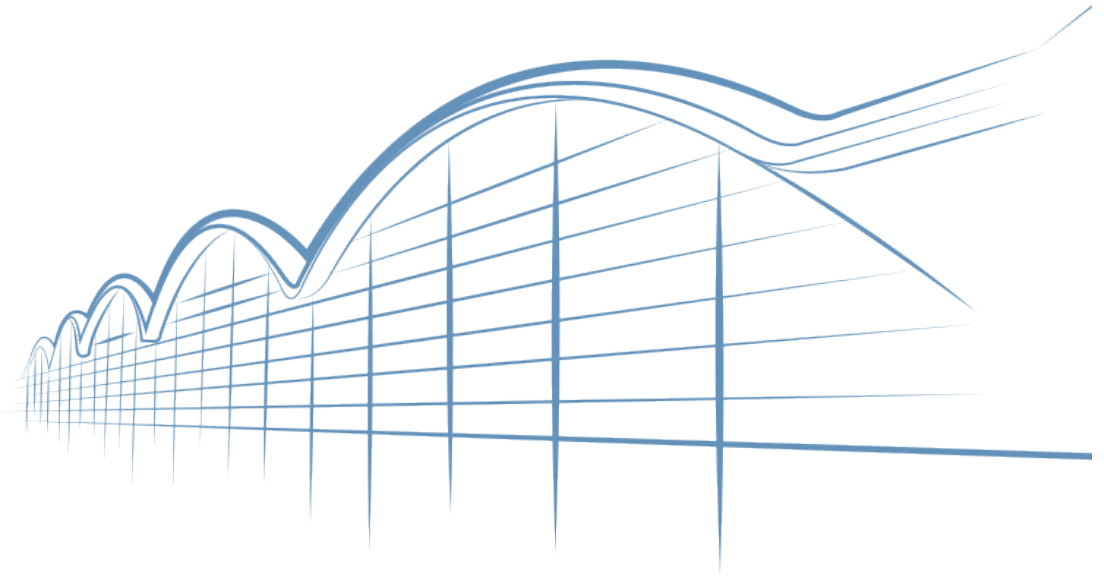
CINECA
HPC ANNUAL REPORT

2024/25



CINECA

CINECA 2024-2025

HPC REPORT

ISBN 978-88-86037-49-5

Cineca HPC Report 2024-2025
website: www.cineca.it
mail: info@cineca.it

Cineca Consorzio Interuniversitario
Via Magnanelli 6/3 40033 - Casalecchio di Reno (BO) - Italy

Presidente: Francesco Ubertini
Direttrice Generale: Alessandra Poggiani

Editorial Board
Sanzio Bassini, HPC Director
Gabriella Scipione
Antonella Guidazzoli
Maria Chiara Liguori

Design and layout project
Simona Caraceni
Daniele De Luca
Luca Govoni

Thanks to:
Angela Acocella, Fabio Affinito, Fauzia Albertin, Paola Alberigo, Federico Andrucci, Claudio Arlandini, Giovanni Baisi, Tiziana Bassi, Eleonora Bergamaschi, Carolina Berucci, Susana Bueno Minguez, Anita Camillini, Michele Carpene, Tommaso Ceccotti, Aldo Canfora, Patrizia Coluccia, Antonio Costantini, Marco De Masi, Mirko Degli Esposti, Chiara Dellacasa, Francesca Delli Ponti, Xhulio Dhori, Andrew Emerson, Matteo Falcioni, Francesco Falciano, Manuela Faustini Fustini, Gabriele Fatigati, Federico Ficarelli, Daniela Galetti, Thomas Gastaldo, Francesca Gebbia, Silvia Gioiosa, Tommaso Gorni, Alessandro Grottesi, Massimiliano Guarrasi, Silvano Imboden, Gianluca Iannuzzi, Mariella Ippolito, Antonino Amedeo La Mattina, Laura Lampariello, Piero Lanucara, Davide Giosuè, Lippolis, Gabriele Marconi, Luca Mattei, Giuseppe Melfi, Federica Mingarelli, Riccardo Mioli, Diego Molinari, Laura Morselli, Juan Mata Naranjo, Sara Oliviero, Daniele Ottaviani, Vito Palmisano, Rossella Pansini, Eric Pascolo, Giorgio Pedrazzi, Renata Pelosini, Diego Piacentini, Andrea Pilzer, Virginia Poli, Marco Petrone, Marco Puccini, Elisa Rossi, Giacomo Savelli, Donatella Sforzini, Luis O. Silva, Nitin Skhula, Saveria Teston, Translated, Roberta Turra, Michele Visciarelli, Marco Viceconti, Valentina Ventura, Cinzia Zannoni.

Dear Colleagues,

Distilling what has characterised the last 12 months of work into a few summary elements, given the number of achievements, is a truly challenging task. With regard to the computing infrastructure, the Leonardo system is now in full production. The EuroFusion Pitagora system, dedicated to the international nuclear fusion community, has also entered full production. The installation of the Gaia system for cloud computing services has been completed; it is the largest national system in terms of core count, integrated with a large-capacity multi-protocol data lake repository using flash memory technology. The Marcopolo system, designed for mission-critical services, such as operational weather forecasts for the Italian Meteo Agency and the processing of data from the INAF SKA and CTA projects, has been delivered and its installation is underway in preparation for production. The Leonardo machine room has been prepared for the installation of the Lisa upgrade. Finally, at Naples' data centre, the Megaride system is already at an advanced stage of installation. It is a system co-financed by the National Cybersecurity Agency and is one of the largest security operations centre systems in Europe.

With regard to production, the number of accredited users has grown by about a thousand units and the number of publications referring to Cineca as an enabling infrastructure for their research and innovation activities has grown to 24,000. At the same time, the H-index value relating to the set of 24,000 publications is equal to 16, which for the overall size on which the index was calculated represents a very high value, increased by one unit compared with the previous year.

With regard to research and development activities, we have not received any negative evaluations from project reviewers on any of the deliverables presented relating to participation in five PNRR project lines and over forty research and development projects financed by Horizon Europe. Finally, we have seen a very significant increase in collaborations involving private entities from both the services sector and the industry sector.

In this extremely positive picture, there is a key concern regarding the size of the department's staff, which continues to be inadequate despite the persistent continuous search for personnel and despite the fact that about 30 new colleagues have been brought on board. This difficulty in finding personnel with a STEM bakground is a widespread issue, particularly in the Emilia-Romagna region where Cineca has its primary premises. About 50% of the demand for STEM profiles requested by the private sector remains unmet, acording to data from the 'regional intelligent skill report'. This data on the difficulty in recruiting personnel with a technical scientific profile is a point of attention on which we need to work further in the near future by intensifying training activities and participation in all national actions relating to collaborations for education, doctorates, masters' degrees in HPC and AI and data scientists, including AI and data science experts as profiles increasingly required by the Italian and European systems. The speed of development and impact of artificial intelligence and machine learning is both impressive and revolutionary. On 30 November 2022, just one week after the President of the Republic had inaugurated Cineca's Leonardo system at Bologna's Technopole (which has since been renamed DAMA), OpenAI released the first version of ChatGPT. While coincidental, it had a significant practical impact on our work. First of all, the Leonardo system, configured with NVIDIA accelerators, which proved to be a very far-sighted choice, was the pre-exascale system that supported the first European response to the demand for computing

time on NVIDIA accelerators for training applications of large pre-trained generative AI models.

The Finnish Lumi system, slightly higher-performance in terms of floating point computing power, was ranked in third position in the TOP500 of November 2022, while the Leonardo system, ranked in fourth position, was configured with AMD processors, that proved to be less effective in responding to the demand for AI computing and the Spanish Marenostrum5 system was configured with a large scalar partition and a relatively small accelerated partition, albeit with NVIDIA processors. The German Jupiter system, the first European exascale system, the cycle following Leonardo's one, is scheduled to start production in the third quarter of this year 2025, while the French Alice Recoque system is still in the procurement phase, with completion scheduled yet to be defined. Therefore, the Leonardo system has greatly supported the European and Italian AI load related to the demand for computing time for ML/AI in scientific research, AI procedures for public administration and technological innovation in the private sector, with particular reference to small and medium-sised enterprises and start-ups.

A context as disruptive as that represented by the progressively growing availability of models and services on pre-trained LLMs is a technological foundation for creating tailored solutions and services for end users. These developments now influence both public and private digital infrastructures across all levels, national, European and international. could not fail to prompt a reaction, first of all from the European Commission – also an essential dimension for AI – as in its time for HPC and also from the national system. The concomitance has favored the possibility of successfully responding to the call by EuroHPC Joint Undertaking (EuroHPC JU) at the end of 2024 for expressions of interest regarding projects for the implementation of Factories for Artificial Intelligence. Cineca coordinates the national project IT4LIA (Italy for Artificial Intelligence) and will be the host of the HPC system optimised for AI that will join the Leonardo system from the beginning of next year 2026. The IT4LIA project, with a budget of €430 million, co- financed by the European Commission through EuroHPC JU and by Italy through the Ministry of University and Research and the National Cybersecurity Agency, joined by the Emilia-Romagna Region, the Italian Meteo Agency, INFN, Cineca, the AI for Industry Foundation and the Bruno Kessler Foundation, constitutes the largest European Artificial Intelligence Factory supported by a single Member State, the proposal of the consortium led by Finland (Lumi) has a slightly larger budget, but it is a consortium of all the countries of Northern Europe, while our project is by far the largest among those presented by a single Member State, such as France, Germany, Spain and others.

In conclusion, it has been a year of intense and stimulating work, which ended with renewed success and already heralds a new period of great commitment for the entire HPC Department, where the greatest asset, the true engine of the great results summarised in this brief introduction to the 2024/2025 report, is represented by the people who make up the department, their skills and their inexhaustible daily commitment.

Best wishes,

Director of HPC Department

Contents

Founded in 1969, Cineca is a non-profit public consortium that operates for the common good of the national academic and research community.
It includes 120 public institutions: 2 ministries, 70 Italian universities, and 48 national public institutions (14 research organizations, 10 university hospitals and IRCCS, 19 AFAM institutions, and 5 agencies and institutions).
Cineca is one of the largest computing centers in Italy and among the most advanced worldwide in the field of High Performance Computing (HPC). It is also one of the leading providers of solutions and services for universities, research centers, the Ministry of Education and Merit, and the Ministry of University and Research.
Cineca supports and promotes digital transformation through innovation and the development of cutting-edge technological solutions, the creation of integrated platforms designed for and with its member institutions, and the delivery of services to support public bodies and administrations. Its core mission is focused on serving its consortium members.
In the field of HPC, Cineca's activities range from infrastructure management to the development of frontier applications, co-design of technological solutions, and participation in research and innovation projects. On behalf of the Ministry of University and Research (MUR), Cineca represents Italy in PRACE (Partnership for Advanced Computing in Europe). It is also the hosting entity for the European High Performance Computing Joint Undertaking (EuroHPC JU) and a partner in the ITER project for nuclear fusion.

Evolution of the Cineca HPC infrastructure in 2024

Diego Molinari, Francesca Delli Ponti, Susana Bueno Minguez, Massimiliano Guarrasi
Cineca

In 2024, the Cineca HPC infrastructure underwent significant changes: the opening for production of the Data Centric/General Purpose (DCGP) partition of the Leonardo supercomputer and the start of Marconi's decommissioning, scheduled for early 2025. Moreover, in 2024, two smaller systems were shut down: DGX and WLCG-CNAF.

By late 2024, the Cineca HPC infrastructure included four clusters: the Tier-0 system Leonardo, the Tier-1 systems Marconi and Galileo100 and the cloud cluster ADA.

The Leonardo system is based on the BullSequana XH2000 architecture supplied by Atos and is particularly well-suited for high-intensity computing tasks such as data processing, High Performance Data Analytics (HPDA), artificial intelligence and machine learning. Leonardo has two main computing modules, named Booster and Data Centric. The Leonardo Booster partition includes 3.456 Atos 'Da Vinci' blades, each equipped with one Intel Ice Lake CPU with 32 cores and four customised versions of NVIDIA Ampere GPUs, called A100 SXM6 64GB. The Leonardo Data Centric partition is equipped with 1,536 conventional blades, each with two Intel Sapphire Rapids CPUs (56 cores each) and 512 GB of DDR5 RAM. The system also features more than 100 PB of rotative disk storage and more than 5 PB of full flash storage. While the Leonardo Booster module entered full production in summer 2023, the Data Centric module entered full production at the beginning of 2024, greatly increasing the number of standard (non-GPU) core-hours available to Cineca users.

The Marconi system is a X86 traditional cluster with about 3,000 nodes each equipped with 2 Intel Skylake (Xeon 8160) processors and 48 cores, connected with an Intel OmniPath network. It is essentially dedicated to the production running EUROfusion community. In summer 2024 it entered its decommissioning phase by by reducing the number of nodes in production by half. The cluster will be fully dismantled at the beginning of 2025 and will be replaced by a new system called Pitagora.

The Galileo100 Tier 1 cluster is devoted to smaller scientific and industrial workloads and is equipped with about 600 nodes each with 2 Intel CascadeLake 8260, 48 cores and at least 384 GB of RAM. It is also equipped with a 20 PB of rotative disks storage.

The ADA cloud infrastructure is based on OpenStack and hosts 71 interactive nodes each with 2 CPU Intel CascadeLake 8260, 48 cores, 768GB of RAM and 2TB SSD storage. The system is also equipped with 1 PB of dedicated CEPH storage.

In addition to the aforementioned systems, Cineca hosts other systems dedicated to specific communities:

- GATEWAY: 88 nodes (Skylake Intel Xeon 8160) with 48 cores each, connected with an Intel OmniPath network. It is a dedicated interactive cluster for EUROfusion;
- WLCG-CNAF: 517 Broadwell nodes (Intel XeonE5-2697) connected with a custom internal network and linked with a high-performance link with the Bologna INFN computing room. The cluster is dedicated to INFN users for the WLCG programme. IIt was shut down in November 2024 and will be fully decommissioned at the beginning of 2025.
- DGX system, an accelerated system for AI based on NVIDIA A100 GPUs decommissioned in September 2024.

The opening of Leonardo Data Centric module with its about 9 PFs of performance partially covered the decrease in global computational capacity due to the dismission of Marconi100. This transition year, marked by the decommissioning of Marconi and the opening of new machine rooms at the DAMA technopole, was necessary to free up space for new supercomputers. Their installation is expected over the next two years, bringing Cineca's total computational capacity to new heights.

	CPU	Total nodes/ cores	Memory/ node
LEONARDO "BOOSTER" (ATOS)	Intel IceLake, Intel Xeon Platinum 8358 (2.6 GHz, 32 cores) + NVIDIA Ampere GPUs	3,456 nodes 110,592 cores 13,824 GPUs	512 GB DDR4 3200 MHz
LEONARDO "DATA CENTRIC" (ATOS)	Intel Sapphire Rapids, 2x Intel Xeon Platinum 8480+ (2.0GHz, 56 cores)	1,536 nodes 172,032 cores	512 GB DDR4 4800 MHz
GALILEO100 (DELL)	Intel CascadeLake, 2x Intel Xeon Platinum 8260 (2.4 Ghz, 24 cores) + NVIDIA Volta 100	636 nodes 26,592 cores 72 GPUs	348 GB + 3.0 TB Optane on 180 "fat" nodes
MARCONI A3 – SKL (LENOVO)	Intel SkyLake 2x Intel Xeon 8160 (2.1GHz 24 cores)	3,124 nodes 149,952 cores	192 GB

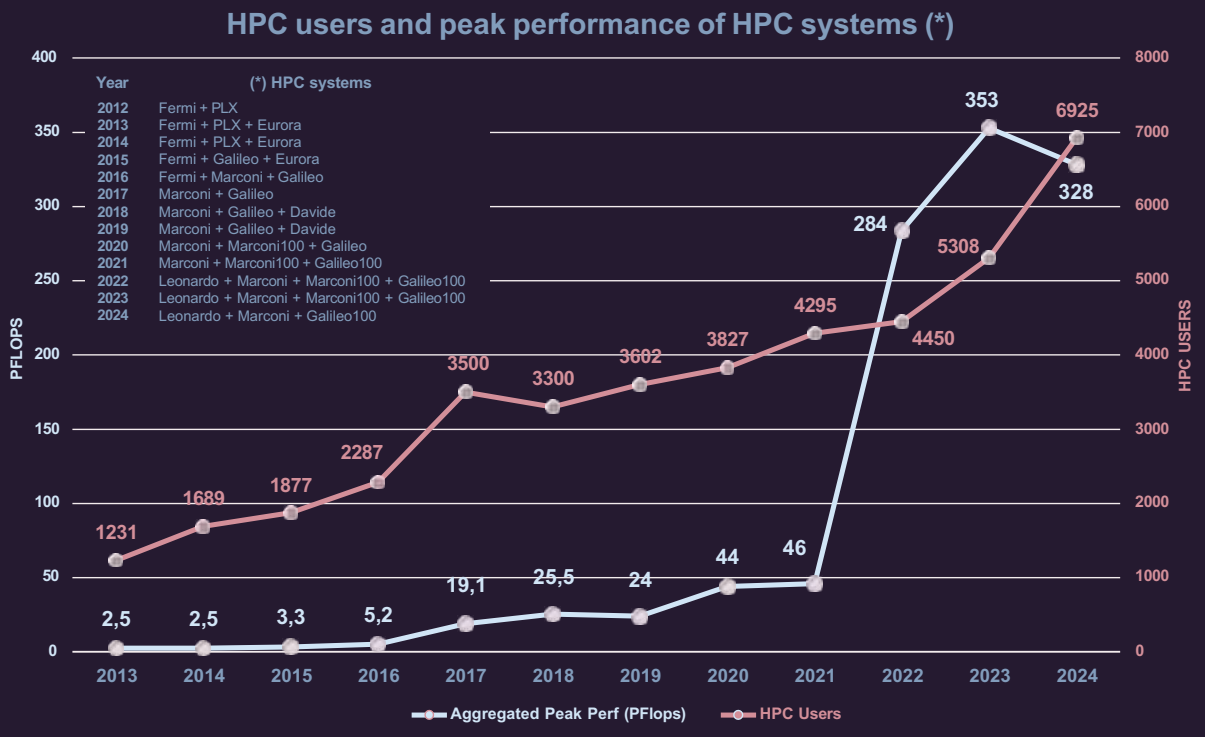


Figure 1: This chart represents the Peak Performance (PFlops) of all HPC Systems available for production since Fermi to Leonardo. In the same chart, we also present the number of active HPC users.

Supercomputer	Resources (Core Hours)	Performance capacity (Exaflops)
LEONARDO Booster	881,813,894	8,928,366
LEONARDO DCGP	1,246,095,213	192,255
Marconi A3	821,196,498	197,087
Galileo 100	231,756,803	61,358
Total	**3,180,862,408**	**9,379,065**

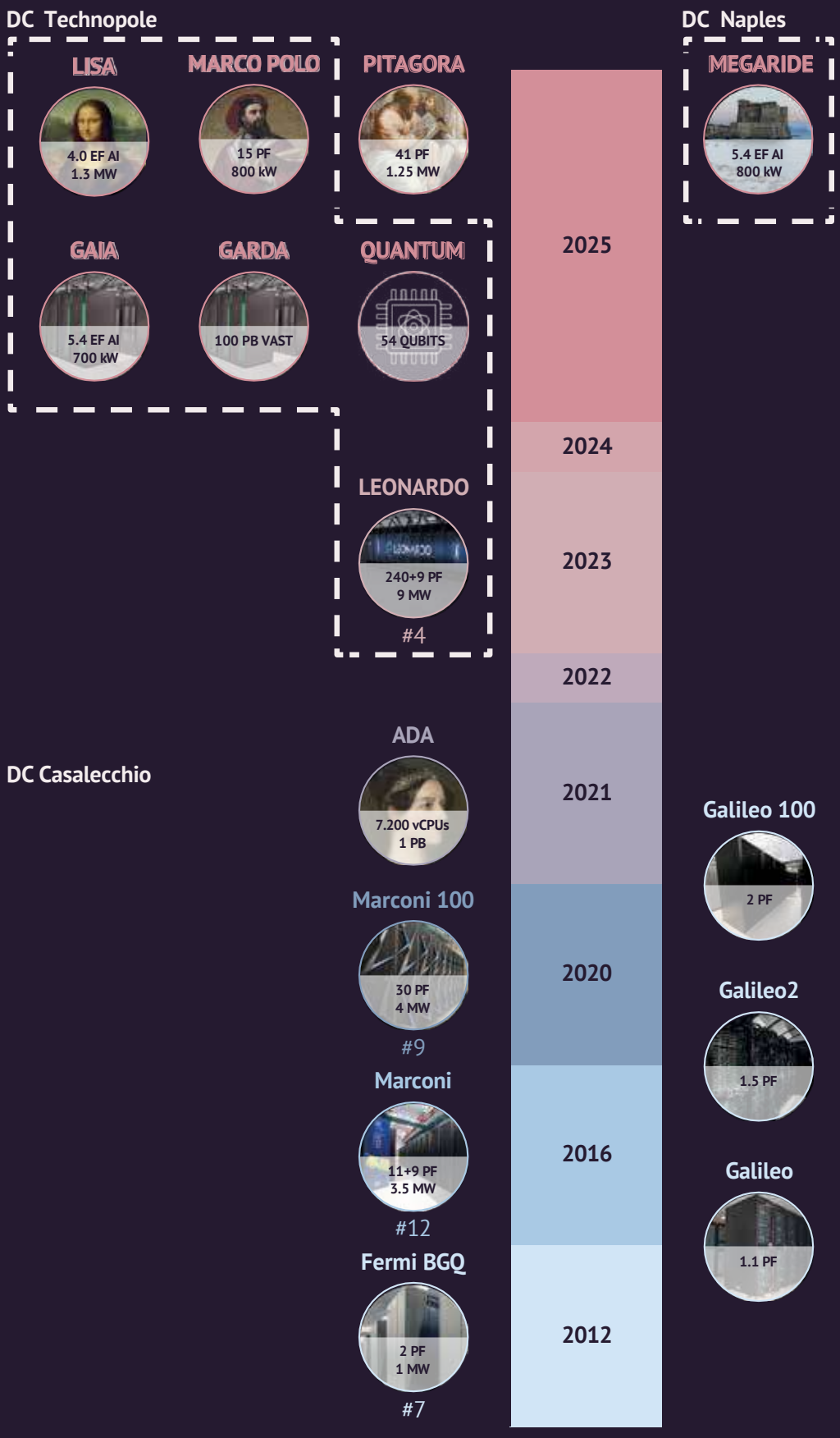


Figure 2: HPC timeline from 2012 to present, showing the Cineca Supercomputers Tier-0 and Tier-1 with the relative position in the TOP500 list.

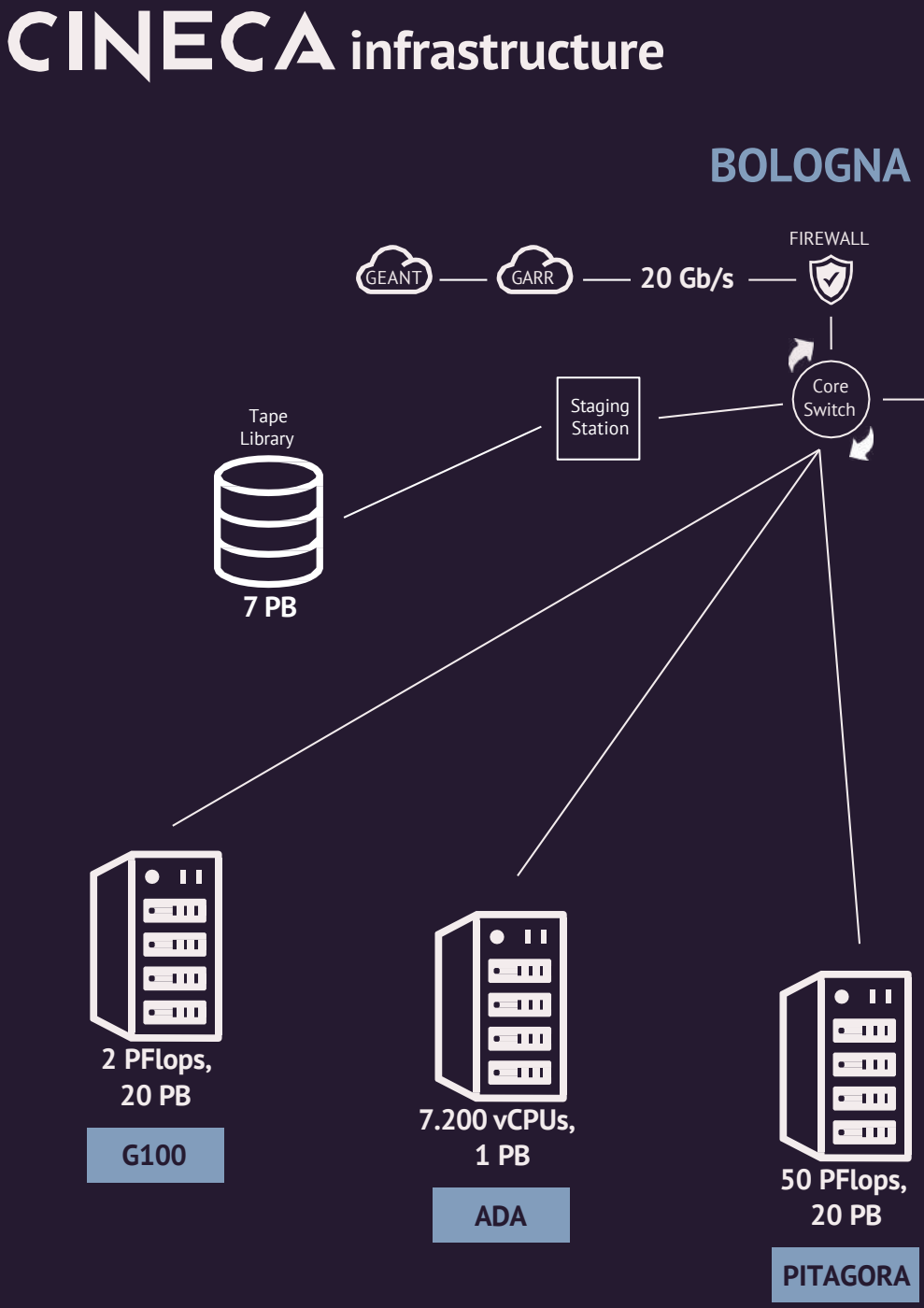
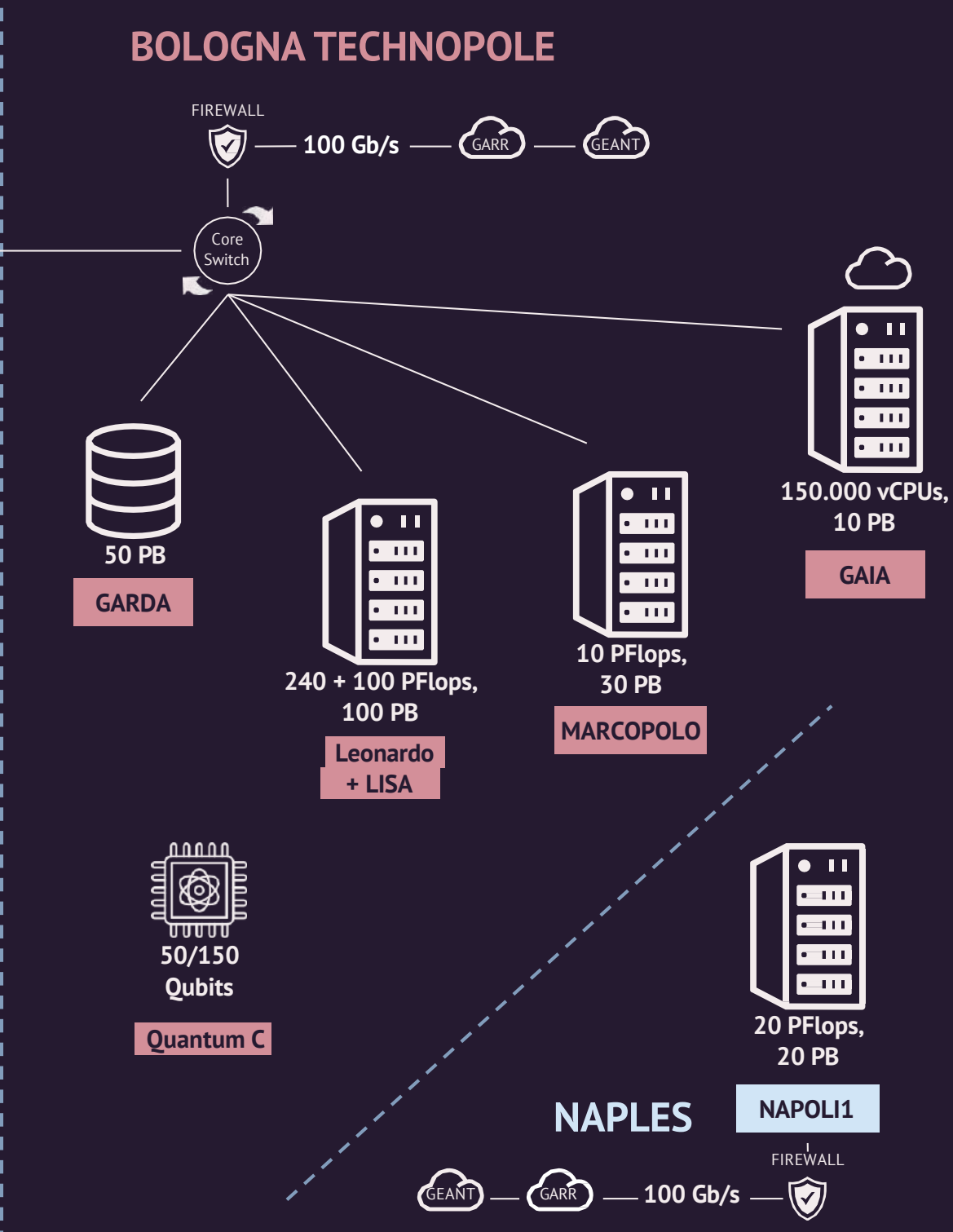


Figure 3: Cineca infrastructures.


BOLOGNA TECHNOPOLE
FIREWALL
100 Gb/s
GARR
GEANT
Core Switch
50 PB
GARDA
240 + 100 PFlops, 100 PB
Leonardo + LISA
10 PFlops, 30 PB
MARCOPOLO
150.000 vCPUs, 10 PB
GAIA
50/150 Qubits
Quantum C
20 PFlops, 20 PB
NAPOLI1
NAPLES
FIREWALL
GEANT
GARR
100 Gb/s

A Guinness World Record-breaking systems management group

Daniela Galetti
Cineca

The use of High-Performance Computing infrastructure in research and industry has become an essential cornerstone for advancing scientific exploration, modelling complex systems and conducting large-scale data analysis. In Italy, Cineca stands out as a national leader in HPC, facilitating significant contributions to academia and industry. Integral to the effectiveness of Cineca is its HPC division, particularly the Systems, storage and network management group, which plays a pivotal role in ensuring optimal performance and availability of computing resources.

The structure of the HPC management group at Cineca is distinguished by a collaborative architecture that promotes efficiency and effectiveness in its operations. The group is composed of senior and junior HPC experts and administrators.

Each of them are involved in the overall management of the HPC Cineca infrastructure. They focus on various operational aspects, including cluster network management, storage solutions and system performance optimisation.

Their responsibilities include monitoring systems availability and performance, troubleshooting issues and guaranteeing the security of computing clusters.

Equally important is the storage management that caters to the vast amounts of data generated and processed by HPC users. The team implements state-of-the-art storage technology, including parallel file systems, object storage and tiered storage solutions that balance performance, capacity and cost-effectiveness.

One of the key strengths of Cineca's HPC management group is its commitment to continuous improvement, applied daily, transforming repetitive actions into automated scripts and optimising all processes the team is involved in.

Self-training and professional development are prioritised to keep team members up to date with the latest technologies and methodologies in HPC management. This not only enhances individual expertise but also fosters a culture of collaboration and knowledge sharing within the group. Regular meetings are organised, providing opportunities for team members to discuss ongoing projects, share challenges and brainstorm solutions.

Furthermore, Cineca actively seeks collaboration with other research institutions and industry partners. Such partnerships enhance the capabilities of the HPC systems management group, allowing them to draw on external expertise, adopt best practices and implement cutting-edge technologies. By engaging with the broader HPC community, Cineca ensures that its infrastructure remains at the forefront of innovation.

While these activities are challenging and common across similar groups at other HPC centres, what distinguishes this particular team as a potential Guinness World Record holder?

The team manages 2 systems among the 10 most powerful HPC clusters in the world and this happened twice:

In June 2020, HPC5, ranked 6th and MARCONI100, ranked 9th[1].

In November 2024, HPC6, ranked 5th and LEONARDO, ranked 9th [2].

In fact, thanks to the competence and the wide experience described, the HPC systems, storage and network team

manages Cineca's HPC infrastructure but also helps manage clusters owned by other entities, like Leonardo, which is owned by EuroHPC JU and HPC5 and HPC6, owned by Eni, the Global Energy Tech Company.

In conclusion, the organisation of the HPC systems, storage and network management group at Cineca reflects a structured yet dynamic approach to managing advanced computational resources. Through skilled personnel, strategic planning and a commitment to collaboration, Cineca effectively addresses the challenges of high-performance computing. As the demand for innovative HPC solutions grows, Cineca's management group will continue to play a crucial role in driving scientific research and technological advancements, ensuring that Italy remains a key player in the global HPC landscape.

Figure 4,5: Top 500 screenshots

[1] https://top500.org/lists/top500/list/2020/06/

[2] https://top500.org/lists/top500/list/2024/11/

User statistics

Francesca Delli Ponti, Susana Bueno Minguez, Donatella Sforzini, Massimiliano Guarrasi
Cineca

Our HPC infrastructure has seen a significant increase in active users, reaching 6,925 by the end of 2024. This marks an increase of 1,617 users compared with the previous year, continuing a consistently positive trend over the past five years

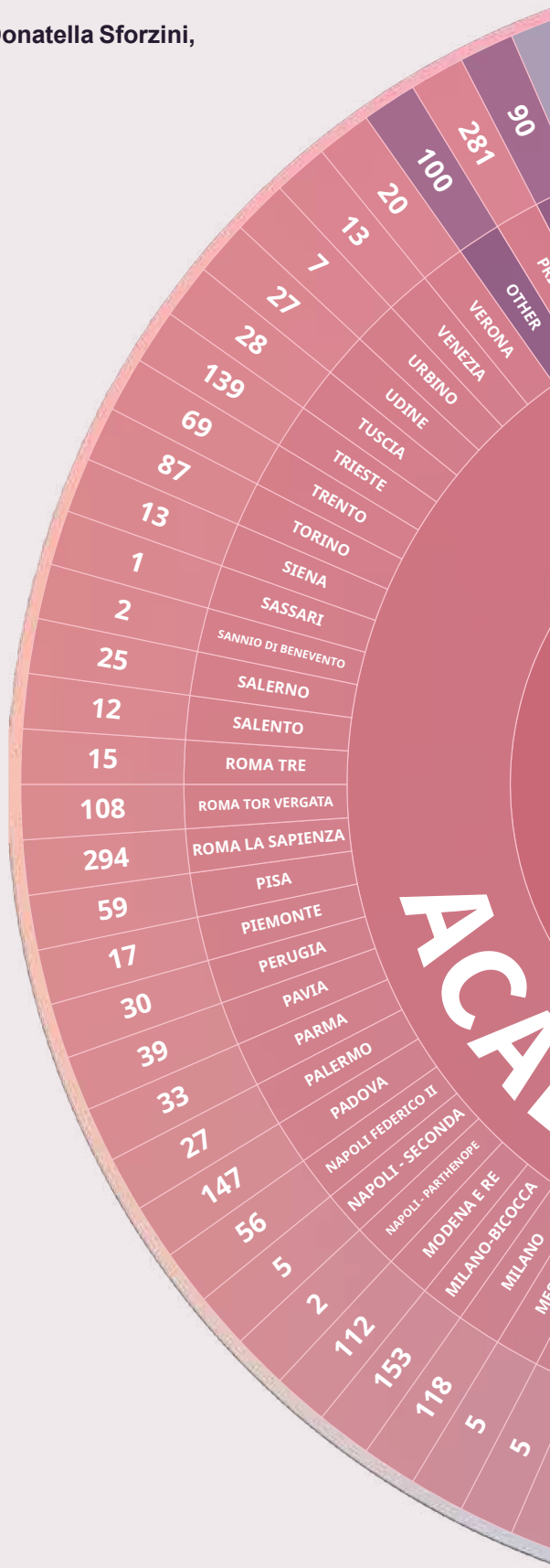
A large proportion of users are male (78%), with 41% under the age of 30 and 39% aged 31 to 45.

The 70% of users work for an Italian institution, concentrated in the Emilia-Romagna region (20%), Lombardia (17%) and Lazio (15%). At the city level, Milan, Rome, Bologna and Trieste each account for between 10% and 14% of users.

Most users are affiliated with European institutions: Italy (70%), Germany (5%), Spain (4%), France (3%) and United Kingdom, Switzerland and Austria (2% each). Most users work for universities and public or non-profit organisations.

The three best-represented disciplines among users' institutions are AI and Machine Learning (15%, double the previous year's percentage), Computational Chemistry (12%) and Condensed Matter Physics (9%). Computational Fluid Dynamics, Astrophysics and Plasma Physics each account for about 8%, followed by Nuclear Fusion, Computational Engineering and Earth and Climate Science, at 7% each.

The AI and Machine Learning scientific domain grew from 7% to 15% compared with 2023 and it is now in the top position. This demonstrates the growing interest and use of AI and Machine Learning and underscores the potential and future direction of HPC usage.



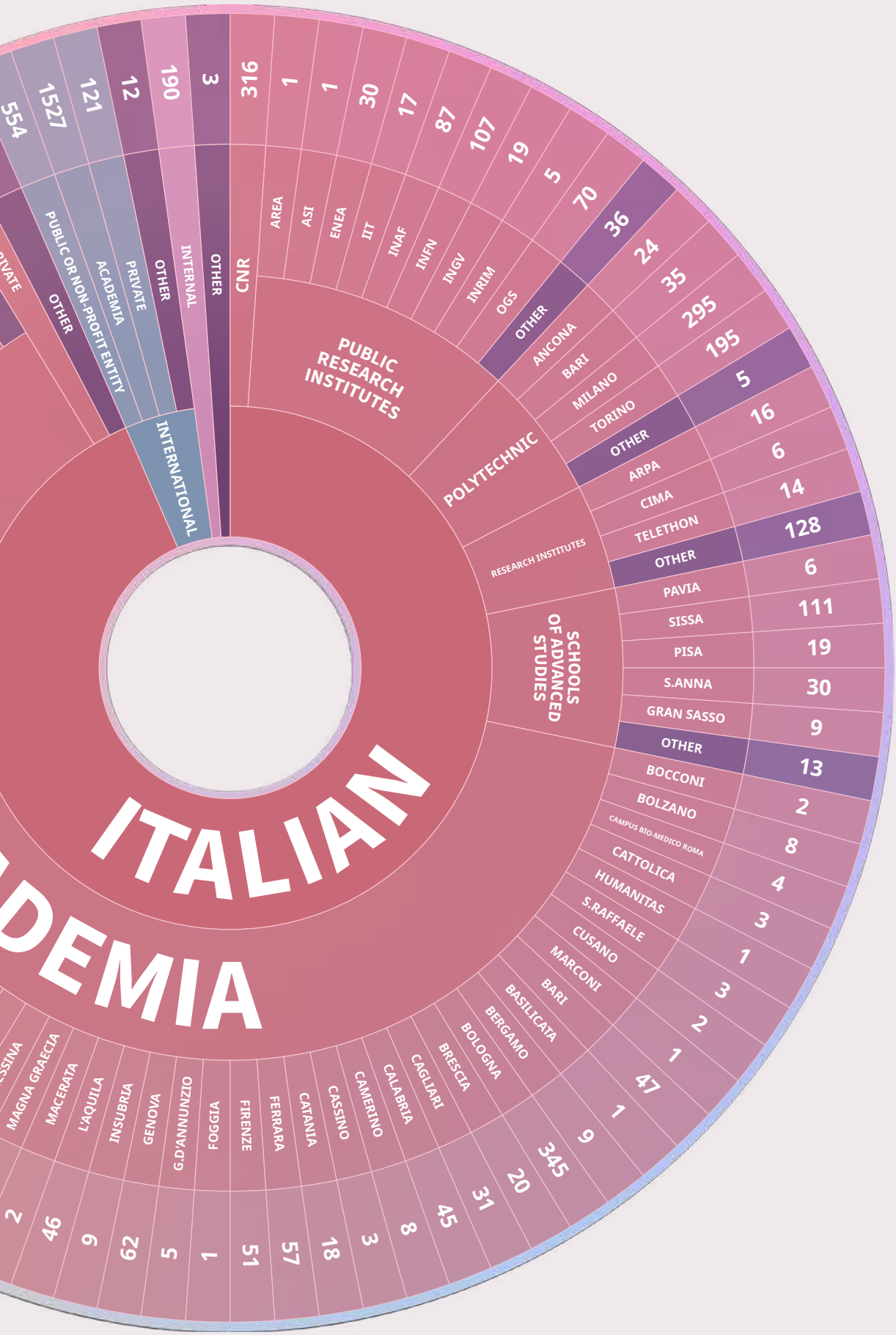


Figure 6: Classification of users' institutions.

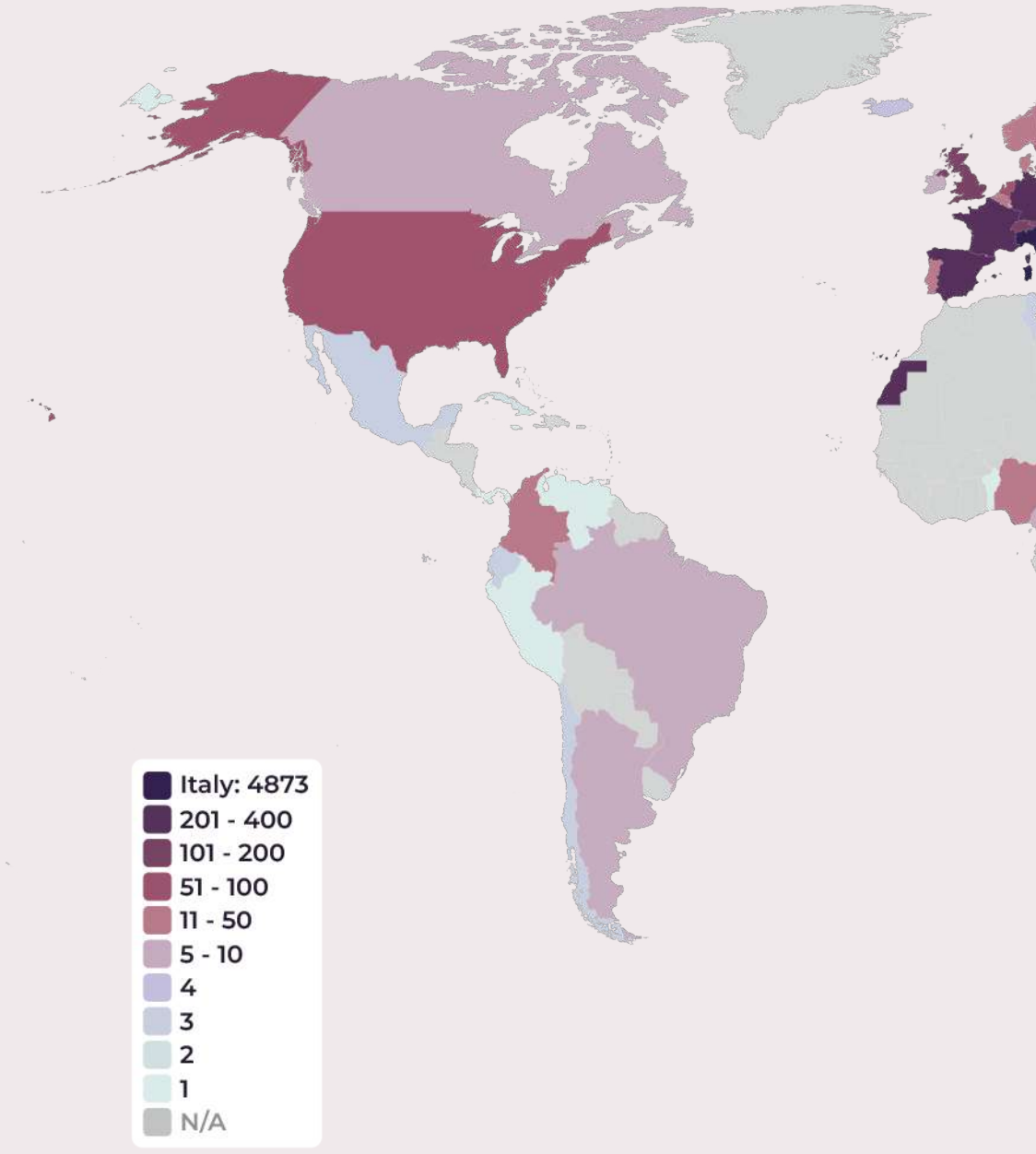
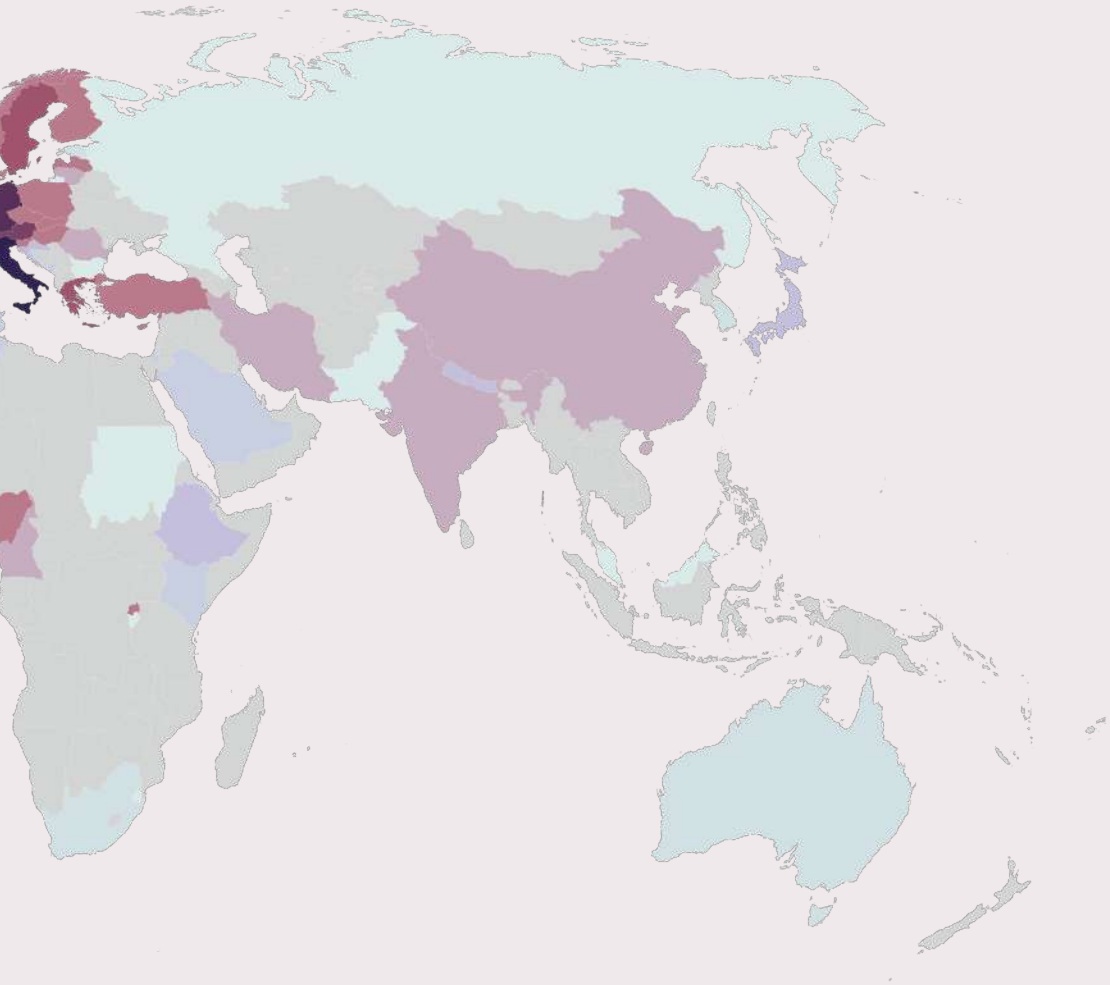


Figure 7: Geographic distribution of users' affiliations in the world.



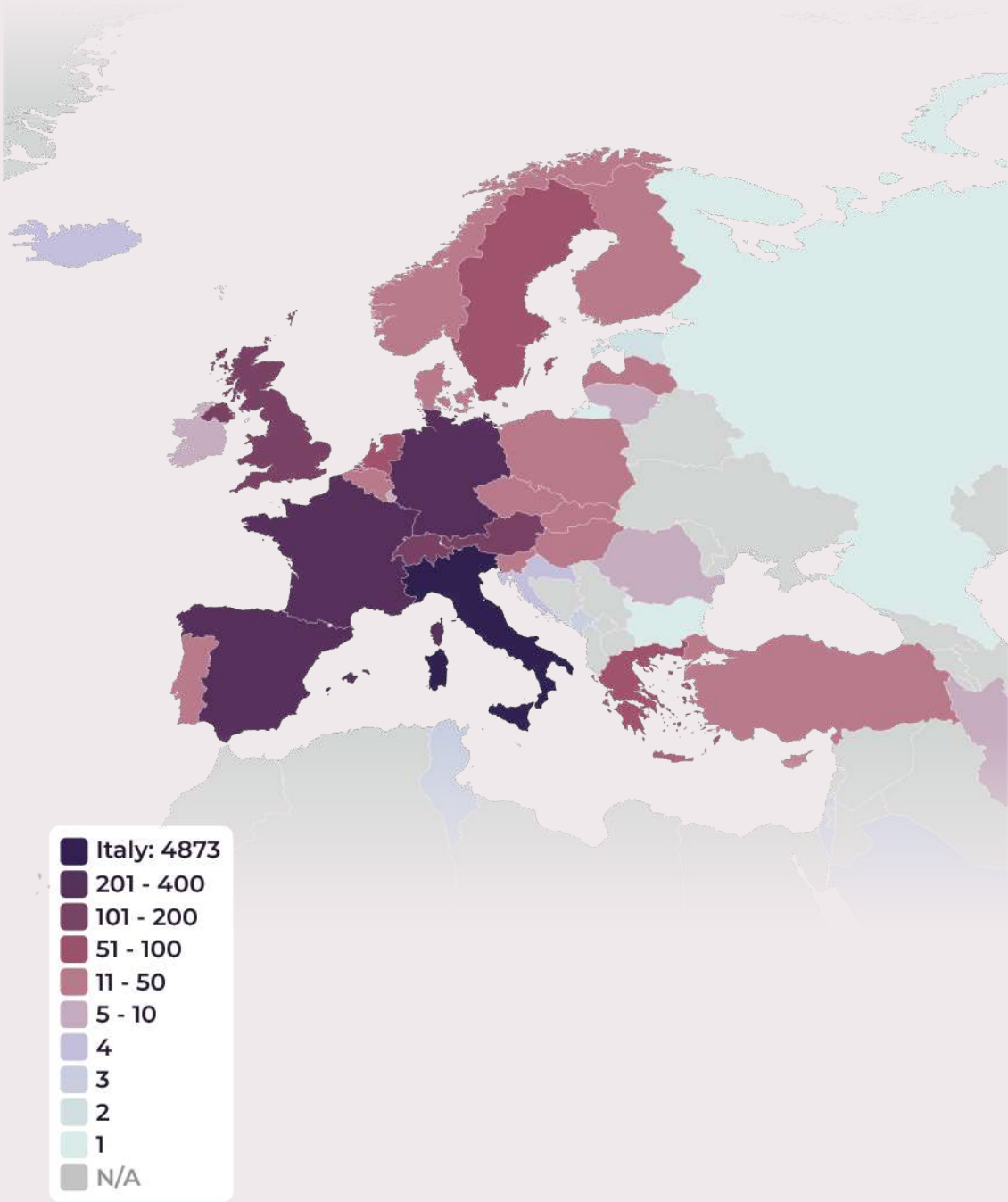


Figure 8: Geographic distribution of users' affiliations in the world with a focus on Europe.

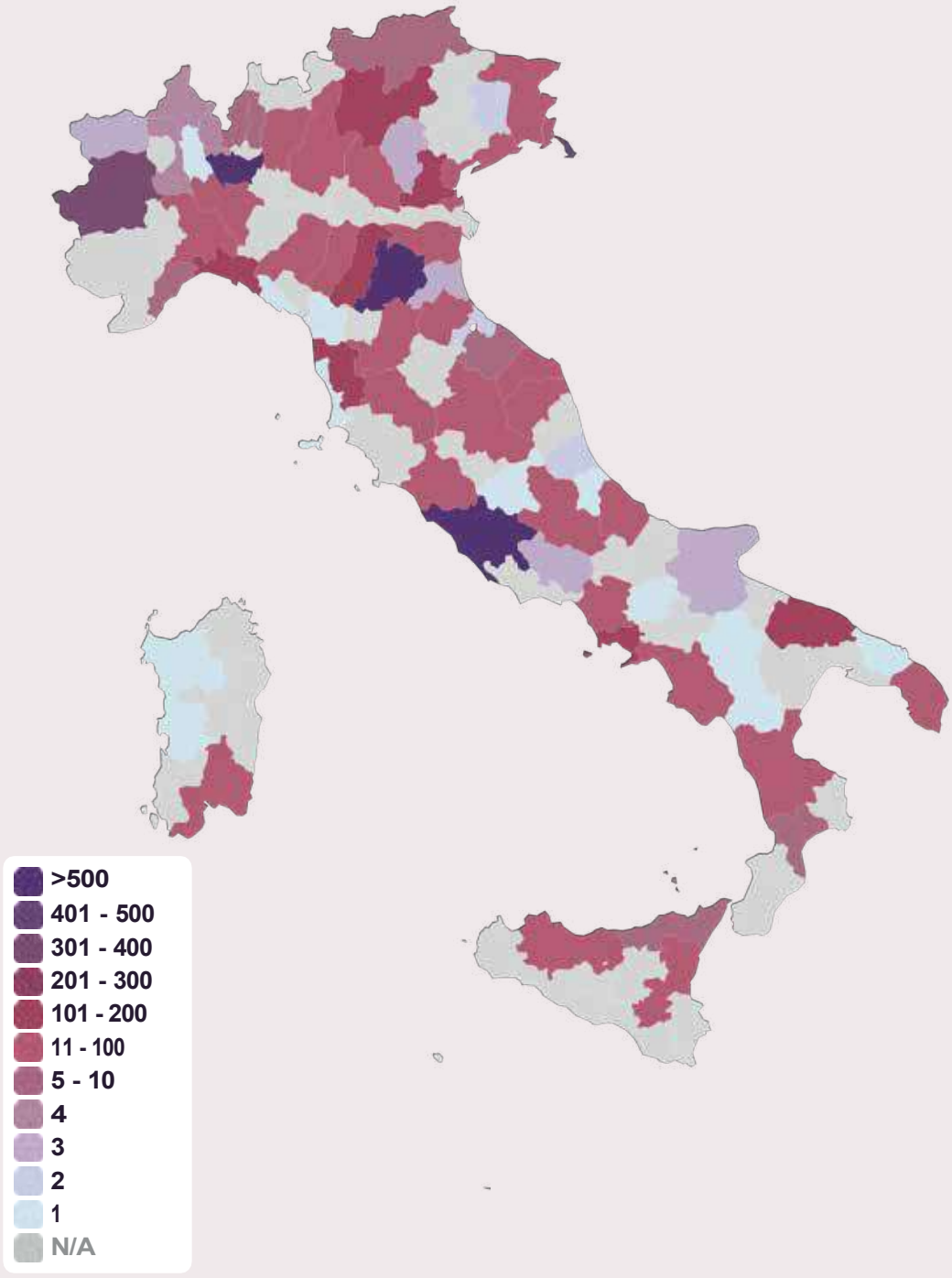


Figure 9: Geographic distribution of users' affiliations in Italy.

Resource allocation and usage: Projects, agreements, collaborations

Diego Molinari, Francesca Delli Ponti, Tiziana Bassi, Massimiliano Guarrasi
Cineca

The primary mission of the Cineca HPC department is to support scientific research. Cineca's supercomputers host Italian and European researchers who require computational resources for their projects. To access Cineca clusters, researchers submit proposals in response to calls that are opened almost every month, making Cineca's supercomputer resources available to the scientific community. A significant portion of Cineca's system allocations is offered through ISCRA and EuroHPC calls. The former is dedicated to researchers with Italian affiliation, while the latter is open to European researchers. In 2024, new ICSC calls were added, dedicated to the Italian High-Performance Computing, Big Data and Quantum Computing Research Centre. In all these cases, proposals are evaluated both technically and scientifically through a peer-review mechanism at European (EuroHPC) and national (ISCRA and ICSC) levels. Selected projects can use Cineca systems free of charge.

In addition to open calls, institutions interested in using Cineca's HPC infrastructure can sign a dedicated "Agreement" with Cineca. Resources obtained this way are then distributed by the institution among its researchers, following an internal peer-review process. It is also possible to participate to a national or European project in direct collaboration with Cineca; in this case, collaborators can access the HPC resources needed to achieve the common goal. Finally, in a few special cases, companies or private users can also sign an agreement with Cineca to use resources for R&D or proof-of-concept (PoC) projects.
In all of these cases, a user who is granted resources becomes the Principal Investigator (PI) of a dedicated project. Each project has a start date, an end date, a scientific domain and a budget of core-hours or GPU-hours (with GPU-hours converted to core-hours on the selected machine as needed). The PI can share the project's computational resources with collaborators.

In addition to these broad categories of projects sharing Cineca's HPC resources, small portions of Galileo100 and Leonardo are reserved for specific projects. On Galileo100, some nodes are dedicated to weather and biogeochemical forecasting activities. Reserved resources on Marconi A3 and the Leonardo Booster partition are dedicated to the EUROfusion consortium. For these portions, allocation is managed independently by the resource owners.

During 2024, including dedicated resources, a grand total of 1.249,5 million core hours was available on the overall Cineca HPC infrastructure. Of this amount, 1,243.5 million (99.5%) were allocated to projects and 1.074 million (86%) were actually used. The most significant portion of resources, in terms of core hours, was dedicated to peer-reviewed projects. The largest portion of allocated resources was distributed among EuroHPC (39.4%), ISCRA (20.8%), Agreements (14.2%) and EUROfusion (14.3%). About 5% of the resources were allocated to industrial projects.
When analysing the allocated resources by scientific domain, in 2024 the majority went to projects in Particle Physics (21.9%), Nuclear Fusion (14.3%), AI & Machine Learning (14.2%) and Condensed Matter Physics (10%).

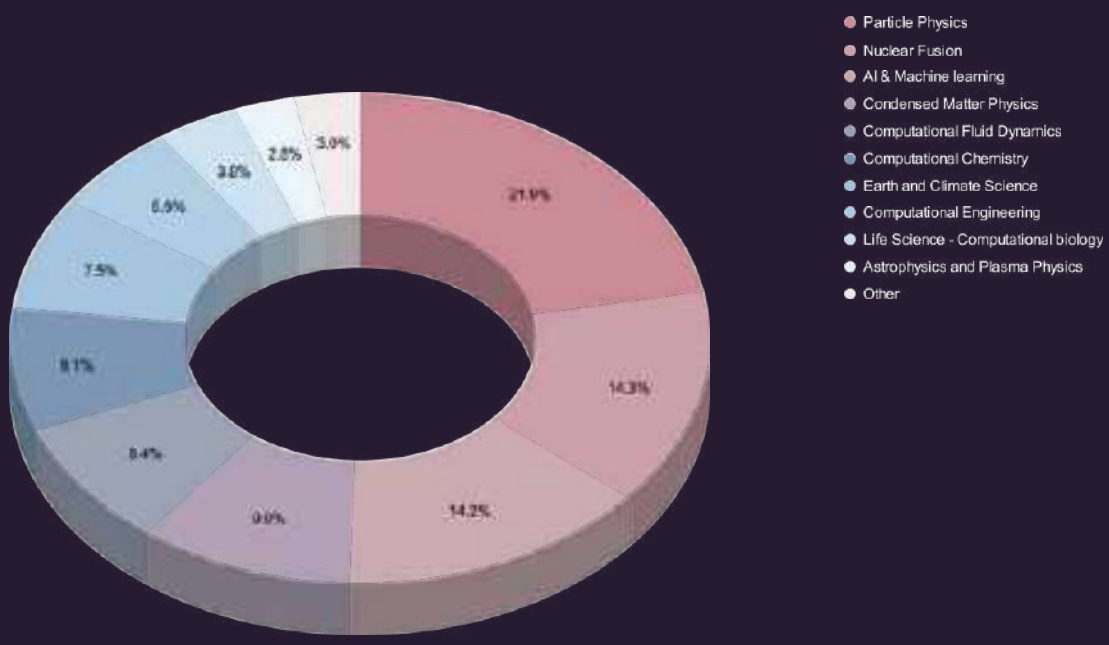
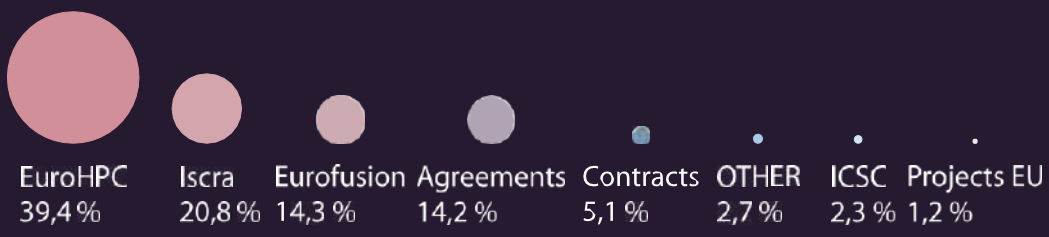


Figure 10: HPC allocation (in percentage) by project type.



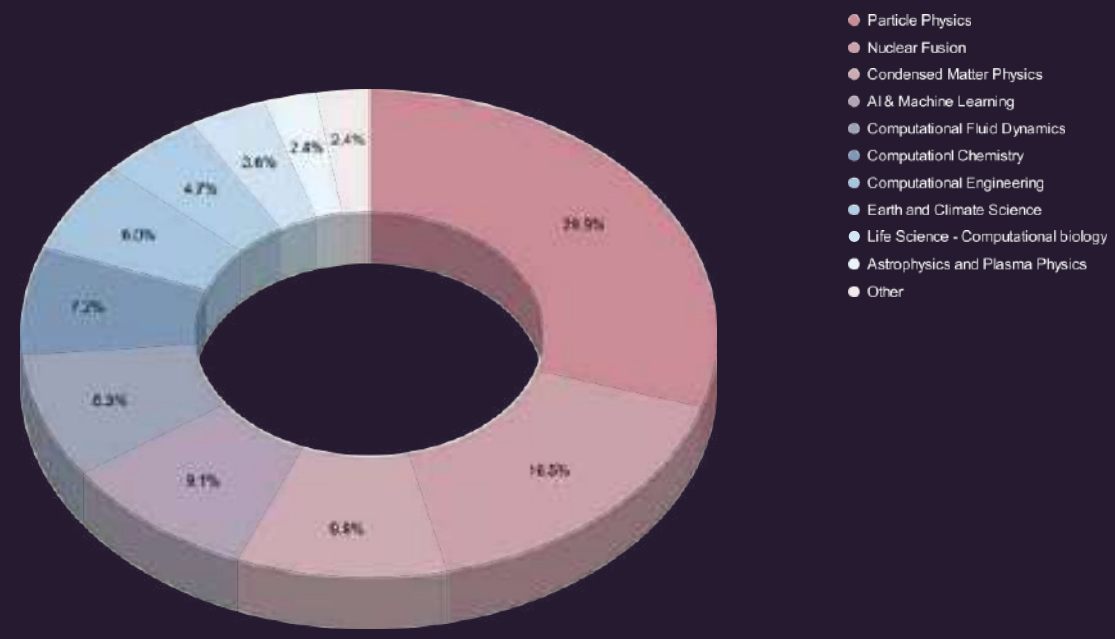
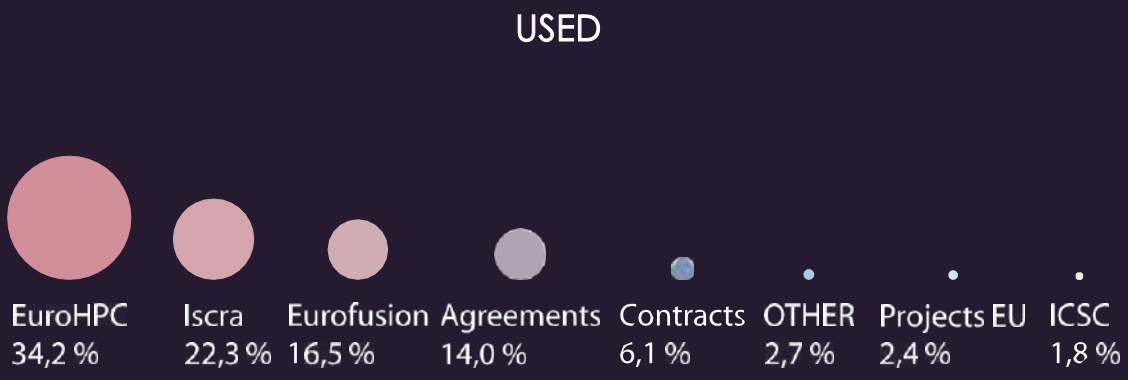


Figure 11: HPC resources used (in percentage) by scientific domain.



EuroHPC

Tiziana Bassi
Cineca

For the Extreme calls, a total of 10 projects were allocated on the Leonardo Booster partition and 3 on the Leonardo DCGP partition (including one project moved from Lumi to Leonardo DCGP), totaling 7,376,212 node hours. Of these, 5 projects were awarded to Principal Investigators (PIs) affiliated with Italian institutions.
For the Regular calls, 16 projects were allocated on Leonardo: 9 on the Booster partition, totaling 1,461,540 node hours and 7 on the DCGP partition (including 3 projects carried over from 2023). Notably, 9 of these projects were led by PIs affiliated with Italian institutions.
The Benchmark and Development calls aim to support researchers and HPC application developers in testing, benchmarking, developing and optimising their applications. In 2024, 158 projects were allocated on the Leonardo Booster partition, totaling 636,000 node hours, while 39 projects were allocated on the DCGP partition, with a total of 132,000 node hours.
The EuroHPC JU AI and Data-Intensive Applications Access call supports ethical artificial intelligence, machine learning and data-intensive applications, with a particular focus on foundation models and generative AI. Launched in April 2024, this call has granted 900,000 node hours across 18 projects.
Projects from the Extreme and Regular calls (with a cut-off in October 2024), as well as from the AI & Data-Intensive and Benchmark/Development calls (with a cut-off in December 2024), are currently undergoing peer review. The winning projects are expected to be allocated in the first quarter of 2025.
According to the EuroHPC agreement, 10% of the EuroHPC dedicated computational resources were awarded to strategic projects. Specifically:

- on the Leonardo Booster partition, the budget was split between Destination Earth (6 projects) and the winners of the AI Challenge (2 projects), amounting to 2,080,235 node hours;
- on the DCGP partition, 5% of the EuroHPC dedicated computational resources were awarded to Destination Earth (2 projects), corresponding to 424,447 node hours.

The details are provided in the table below.

		2023			2024											
Partition	Call	Oct	Nov	Dic	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dic	Total
LEONARDO Booster	AI & Data-Intensive Applications							8		6				4		Peer review	18
	Benchmark			2	2	1	7	2	7	9	7	2	6	5	4	Peer review	54
	Challenge AI-BOOST					2											2
	Destination Earth				4									2			6
	Development			5	3	6	7	10	12	3	9	5	15	18	11	Peer review	104
	Extreme		3					7						Peer review			10
	Regular		3					6						Peer review			9
	Training							2									2
	Core hours awarded	0 M	122 M	0,8 M	35,1 M	32,8 M	1,6 M	133 M	2,5 M	11 M	2,1 M	0,9 M	2,8 M	9,6 M	2 M	N.D.	364,2 M
LEONARDO Data Centric	AI & Data-Intensive Applications															Peer review	0
	Benchmark							2	3	3		2		2		Peer review	12
	Destination Earth				2												2
	Development							2	4	3	3	5	5	3	2	Peer review	27
	Extreme						1	2						Peer review			3
	LEAP				11												11
	Regular	2	1					4						Peer review			7
	Core hours awarded	6 M	15,7 M	0 M	48,6 M	0 M	53 M	75 M	2,5 M	2 M	1,3 M	2,7 M	2,2 M	1,8 M	0,9 M	N.D.	211,7 M
	Total Projects Allocated	2	7	7	22	9	15	45	26	24	19	14	26	34	17	N.D	267
	Total Node/h Allocated	0,05 M	4 M	0,02 M	1,5 M	1 M	0,5 M	4,8 M	0,1 M	0,4 M	0,08 M	0,05 M	0,1 M	0,3 M	0,1 M	N.D.	13 M

EPICURE – Unlocking European-level HPC support

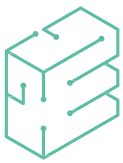
Andrew Emerson
Cineca

EPICURE (EuroHPC Applications Support Project) is a European initiative launched in February 2024, aiming to enhance user support services for HPC systems across Europe. Coordinated by CSC – IT Centre for Science, Finland, the project involves a consortium of 16 partners from 14 countries, including major HPC centres such as Barcelona Supercomputing Centre, Cineca, IT4I and Forschungszentrum Jülich.

EPICURE's primary goal is to establish and operate a distributed yet coordinated European HPC application support service. This service focuses on application porting, optimisation and execution support for projects that have been allocated computing time through EuroHPC's peer-review process. A key component of the project is the development of a European HPC Application Support Portal, designed to serve as a single point of contact for users from both public and private sectors, including SMEs. This portal will provide comprehensive information on EuroHPC systems, their architectures, access mechanisms and available support services.

The EPICURE project runs for 48 months with a total budget of €10 million, co-funded 50% by EuroHPC JU.

By the end of the first year, more than 50 projects had been helped to enable or optimise their applications on the EuroHPC infrastructure. Researchers who wish to know more about the services provided can contact the project via the EPICURE website: https://epicure-hpc.eu/



EPICURE

This project has received funding from the European High-Performance Computing Joint Undertaking under grant agreement No.101139786.

Co-funded by the European Union

EuroQHPC-I: Advancing European quantum-HPC integration

Sara Marzella
Cineca

The EuroQHPC-Integration (EuroQHPC-I) project, founded by EuroHPC JU, was launched in March 2025 and will run for four years. It aims to integrate quantum computing with high-performance computing (HPC). By combining these technologies, the project seeks to enhance computational capabilities for scientific and industrial applications.
EuroQCS-Italy will integrate quantum resources with the Leonardo supercomputer at DAMA technopole. A key milestone is the installation of Orion Beta, a 140-qubit neutral atoms quantum simulator from Pasqal, in Q2 2026. Initially analogue, it will be upgraded to support mixed analogue/digital computation, broadening its potential applications.
EuroQHPC-I also focuses on developing a cloud quantum computing service, allowing remote access to quantum resources.
The project will develop specialised software libraries and middleware to facilitate communication between quantum processing units (QPUs) and HPC architectures. Hybrid optimisation techniques will be implemented to enhance computational efficiency. Quantum resources will be integrated into SLURM, the widely used HPC job scheduler, simplifying hybrid computing adoption across Europe.

Strengthening Europe's quantum ecosystem

EuroQHPC-I involves six hosting entities, each contributing a different quantum computer. The project includes 30 partners from 17 countries, focusing on UX harmonisation, applications and benchmarks, standardisation and user support.
The hosting entities and their quantum computers are:

- EuroQCS-France: Photonic quantum computer by Quandela, integrated with the Joliot Curie supercomputer.
- Euro-Q-Exa: Superconducting quantum computer by IQM (54+ and 150 qubits), integrated with SuperMUC and BlueLion supercomputer.
- EuroQCS-Italy: Cold-atom mixed digital/analogue quantum computer by Pasqal (140 qubits), integrated with the Leonardo supercomputer.
- LUMI-Q: Superconducting quantum computer by IQM (24 qubits), integrated with the Karolina supercomputer.
- EuroQCS-Poland: Trapped-ion quantum computer by AQT (20+ qubits), integrated with the Altair-Proxima supercomputer.
- EuroQCS-Spain: Quantum Annealing computer by Qilimanjaro & HPCNow, integrated with the MareNostrum5 supercomputer.

In conclusion, EuroQHPC-I will advance scientific computing by integrating quantum technologies into HPC infrastructures, reinforcing Europe's technological leadership.

Co-funded by the European Union

EuroQCS-Italy partners

CINECA

JÜLICH Forschungszentrum

arnes povezujemo znanje

Jožef Stefan Institute Ljubljana, Slovenia





EUROQHPC-INTEGRATION

Towards a federated European HPC-QC infrastructure

Coord.: GENCI

30 partners

17 countries

6 quantum computers

Shared integration, development and support activities (LRZ)

- UX Harmonization
- Applications & benchmarks
- Standardization
- User support

EuroQCS-France

Photonic quantum computer by **Quandela**
Colocalization with the HPCQS system by **Pasqal**
Integration with the **Joliot Curie** supercomputer

Euro-Q-Exa

Superconducting quantum computer by **IQM** (54+ and 150 qubits)
Colocalization with German and Munich Quantum Valley systems by IQM, AQT and Planqc
Integration with SuperMUC and BlueLion supercomputers

EuroQCS-Italy

Cold-atom mixed digital/analog quantum computer by **Pasqal**
140+ qubits plus **partial-addressability** mode
Integration with the **Leonardo** supercomputer
Colocalization with Italian ICSC quantum computers

LUMI-Q

Superconducting quantum computer by **IQM**
24 qubits in star topology
Colocalization and integration with the Karolina supercomputer

EuroQCS-Poland

Trapped-ion quantum computer by **AQT**
20+ qubits in all-to-all topology
Integration with the **Altair-Proxima** supercomputer

EuroQCS-Spain

Quantum Annealing computer by **Qilimanjaro & HPCNow**
Colocalization with the Quantum Spain digital quantum computer by **Qilimanjaro & CMV**
Integration with the **MareNostrum5** supercomputer

EuroHPC

This project has received funding from the European High Performance Computing Joint Undertaking under grant agreement No 101159808

ISCRA 2024: Advancing scientific discovery through supercomputing resources

Paola Alberigo
Cineca

The Italian SuperComputing Resource Allocation (ISCRA) programme continues to be a cornerstone for advancing scientific research by granting access to HPC resources. Managed by Cineca, ISCRA supports researchers affiliated with Italian institutions by providing computational power and technical expertise necessary for complex simulations, data processing and emerging computing paradigms.
In 2024, ISCRA reaffirmed its commitment to excellence in computational research by launching ten ISCRA C calls and two ISCRA B and D calls, offering researchers access to a broad spectrum of computing resources including:

- HPC systems
- Cloud computing resources
- Quantum annealing and quantum simulation platforms
- Dedicated data storage for simulation-related data

Project Classifications and Evaluation

ISCRA proposals are evaluated under two main categories:

- Class C projects: Geared toward code testing, benchmarking and pre-production stages. These projects are assessed primarily on technical feasibility;
- Class B projects: Designed for full-scale production runs. These undergo rigorous evaluation based on both technical merit and scientific impact.

In 2024, the two ISCRA B calls garnered 168 applications, a testament to increasing demand for and the central role of advanced computational resources in modern research.

	call 29B 24 April 2024 to 24 April 2025	call 30B 20 January 2025 to 20 January 2026
Accepted	79	50
Rejected	12	27
Total number submitted projects	**168**	
Total accepted projects	**129**	

Assigned resources in local core hours per host and per call:

	Call 29B	Call 30B	Total
G100	46.169.253	30.385.524	76.554.777
LEONARDO_B	81.569.784	46.791.624	128.361.408
LEONARDO_DCGP	36.956.640	18.900.000	55.856.640
	164.695.677	**96.077.148**	**260.772.825**

New scientific panel for 2024

To guide the evaluation process and ensure scientific excellence, a new ISCRA Scientific Panel was appointed in 2024. The panel is chaired by Prof. Sergio Pirozzoli from the University of Rome La Sapienza, a leading expert in computational engineering. The panel is composed of domain experts across a wide range of scientific disciplines: this diverse team ensures that projects are evaluated with deep scientific insight and fairness across all disciplines.

Member	Scientific Area	Institution
Luciano Colombo	Condensed Matter Physics	University of Cagliari, Italy
Rossella Ferretti	Earth and Climate Science	University of L'Aquila, Italy
Alejandro Giorgetti	Life Sciences	University of Verona, Italy
Silvano Massaglia	Astrophysics and Plasma Physics	University of Torino, Italy
Michela Milano	Informatics – AI & Machine Learning	University of Bologna, Italy
Sergio Pirozzoli	Computational Engineering	University La Sapienza, Italy
Silvano Simula	Particle Physics	University of Roma Tre, Italy
Federico Toschi	Computational Fluid Dynamics	University of Eindhoven, Netherlands
Claudio Zannoni	Computational Chemistry	University of Bologna, Italy
Francesco Zerbetto	Computational Chemistry	University of Bologna, Italy

Looking Forward

The continued expansion of ISCRA's infrastructure and the integration of cutting-edge resources like quantum simulators mark Italy's commitment to maintaining a competitive edge in computational science. With growing participation and the support of top-tier experts, ISCRA is poised to foster transformative research across disciplines: from climate science and life sciences to AI, fluid dynamics and fundamental physics.
For more information about upcoming calls and how to apply, visit: https://www.hpc.Cineca.it

Agreements

Maurizio Cremonesi
Cineca

Researchers may face the need to use HPC resources for activities that do not have a definite objective but are no less important than others. Testing new algorithms, checking published results and training students are some of the activities that are difficult to justify when requesting resources in peer-review calls such as ISCRA or EuroHPC.

Some institutions are aware that their researchers need HPC computing resources and prefer to secure the necessary computing time without burdening them with applications that might be rejected. These institutions therefore gain a stable amount of computing hours that can be distributed to researchers according to internal criteria.

Whatever the motivation, also in 2024 several universities and research organisations have sought to reserve HPC resources by making agreements with Cineca. Moreover, Cineca has established collaborations with important organisations whenever it has been recognised that this provides mutual advantages in achieving common objectives. Thus organisations have therefore gained access to computing resources that would otherwise have been difficult for them to acquire.

On 2024, an important increase in the availability of HPC platforms managed by Cineca has led to a greater amount of computing time for agreements and collaborations than the previous year, with a total of 555 Mcore_hours, compared with the 239 Mcore_hours on previous year. Plus, the number of organisations that have been granted an important amount of computing time has increased from five to seven. The number of Universities with an agreement with Cineca is slightly lower than in the previous year but with a greater amount of booked resources. Even greater is the amount of computing resources granted to research centres.

A list of institutions with active agreements and collaborations in 2024 is provided on the following page.

Institution	Award resources (Mcore-h)
INFN (Istituto Nazionale di Fisica Nucleare)	406.97
Università degli Studi di Milano Bicocca	32.63
ICTP (International Centre for Theoretical Physics)	31.60
OGS (Istituto Nazionale di Oceanografia e di Geofisica Sperimentale)	24.97
SISSA (Scuola Internazionale Superiore di Studi Avanzati)	21.44
CMCC (Centro Euro-Mediterraneo sui Cambiamenti Climatici)	12.67
FAIR (Future Artificial Intelligence Research)	10.00
INAF (Istituto Nazionale Astrofisica)	3.95
AIRC (Fondazione AIRC per la Ricerca sul Cancro)	3.42
Università degli Studi di Trieste	1.48
Elixir (Distributed Infrastructure for Biological Data)	1.09
Politecnico di Milano	1.08
Università degli Studi di Genova	0.64
Università degli Studi di Milano	0.48
Università degli Studi di Brescia	0.44
Università degli studi del Piemonte Orientale	0.40
INRiM (Istituto Nazionale di Ricerca Metrologica)	0.39
Telethon	0.36
Ospedale Pediatrico Bambin Gesù	0.35
Università degli Studi di Brescia	0.30
INMI (Istituto Nazionale Malattie Infettive Spallanzani)	0.29
Istituto G. Ronzoni	0.24
IIT (Istituto Italiano Tecnologia)	0.16
Scuola Superiore Sant'Anna	0.05

Training HPC

Francesco Falciano
Cineca

In 2024, following the emergencies related to the COVID-19 pandemic, there was a strong emphasis on fully resuming in-person classes. However, due to classroom capacity limitations, the number of students allowed had to be restricted.

The policy for online courses has always been to limit the number of official students who could receive full assistance during exercises. However, all registered users were allowed to participate as auditors, meaning they could attend theoretical and practical lessons and ask questions, but did not receive direct support from instructors during exercises.

The in-person learning format undoubtedly fostered the social and collaborative aspects that had been lacking in online classes. For these reasons, the 2024 educational programme was structured to have the vast majority of lessons held in person, limiting online courses to locations where no suitable classrooms were available.

Additionally, the practice of recording online courses and making them publicly available has been consolidated. This has resulted in a rich collection of video courses in both English and Italian on the e-learning platform, accessible to people worldwide.

The reported student numbers refer to the number of attendance certificates issued – i.e. students who completed the course with the required number of hours. Those who dropped out early or had excessive absences and did not meet the minimum attendance requirement were excluded.

The HPC department, in partnership with international companies and universities, also participates in organising events and lectures such as Master's programmes, PhD courses, workshops and tutorials on HPC topics, digital humanities, and big data.

Since these are externally organised or outreach events, it was not possible to precisely calculate the number of participants and Cineca did not issue certificates for them. However, it is estimated that approximately 300 additional individuals were reached through Cineca HPC training events.

In addition to attendance certificates, course evaluation surveys were conducted to assess the perceived quality of instruction and gather suggestions for improving future editions. Student requests were incorporated into the next academic year's curriculum.

The student surveys from the Academy reported a high level of satisfaction. Specifically, in 2024, the analysis of the questionnaires revealed that, on average, the courses received an overall rating of 9.1 out of 10.

TRAINING NUMBERS

Detailed Data for 2024 HPC
Academy Courses and Schools at CINECA

Italian and European Students

Teaching Days

Instructors Involved

Courses

Schools

STAFF

The HPC department at Cineca employs 160 people distributed among four premises: 126 in Bologna, 11 in Milan, 22 in Rome, and 1 in Naples.
We wish the colleagues that were hired in 2024 a very fruitful and interesting career.

FAUZIA ALBERTIN

After graduating in Physics and obtaining a PhD in 2011, she worked in the academic field in various research groups, both in Italy and abroad, in the field of scientific analyses dedicated to Cultural Heritage. During her professional career, she has participated in several projects interfacing with museums, academic and archival institutions, acquiring expertise in the development of innovative scientific instruments, in data processing and management and in disseminating them to both a general and specialist audience.

FEDERICO ANDRUCCI

With a master's in Computer Engineering, he completed his thesis internship at the VisitLab of Cineca, working on the Bologna Digital Twin project. During this period, he developed a workflow to develop a prototype of a Digital Twin of Bologna, starting from raw data and arriving at the creation of a dynamic three- dimensional representation of the city.

MASSIMILIANO BONETTI

Graduated in Computer Engineering at Politecnico di Milano with a specialisation in artificial intelligence and a thesis on risk-averse reinforcement learning. In addition to reinforcement learning, he has a particular passion for parallel programming.

ANITA CAMILLINI

She holds a Master's degree in Theoretical Physics from the University La Sapienza of Rome. She is currently a PhD student at INL, a nanotechnology research centre located in the north of Portugal, where she is conducting studies on 'Architectures for scalable photonic quantum computing', with a deadline of 2025. During her academic career, she has participated in numerous international conferences, presenting the results of her research and cultivating collaborations with researchers and scientists from all over the world

MARCO CELORIA

Graduated in Physics at the University of Milan, he obtained a PhD in Astroparticle Physics at the Gran Sasso Science Institute and carried out a post-doc in Trieste at the Abdus Salam International Centre for Theoretical Physics (ICTP) dealing with cosmological inflation.
He subsequently completed the master in High Performance Computing at SISSA/ICTP, working on optimising of scientific codes for simulations of rarefied gases. Afterwards, he worked as a technologist at Area Science Park, where he dealt with the development of scientific software and artificial intelligence in the field of life sciences.

SIMONE CEPPI

After graduating in Physics, he devoted his PhD to the formation of planets and stars using numerical and analytical methods.
Using hydrodynamic simulations, he studied the dynamics of accretion disks orbiting multiple star systems. During his work, he contributed to the development of a state-of-the-art Smoothed Particle Hydrodynamics code (PHANTOM) popular in the stellar astrophysics community and wrote analysis tools used in scientific publications.
The need for computing power to run demanding simulations drove him towards HPC (MPI, OpenMP and CUDA).

DAVIDE CRISANTE

After graduating in Computer Science and Engineering at the University of Bologna, he is currently pursuing a Master's degree in the same field. In 2020, he followed the project "Scaling the DPD code, on large multi-GPGPUs architectures" in the United Kingdom, analysing the structure of the DL_MESO software to implement a GPU-compatible version, porting it from Fortran to CUDA.

ANTONIO DE NICOLA

After obtaining a Master's degree in Chemistry, he obtained the title of Doctor of Philosophy in Chemistry in 2014 with a thesis on molecular modelling on biological systems and composite materials.
He specialised in the computational study of "soft-matter" systems by developing specific Coarse-Graining models, contributing to the development of a Molecular-Dynamics code for special applications such as particle-field.
From 2018 to 2021, he held the position of Assistant Professor at Yamagata University in Japan, where he was the manager of a corporate project funded by Borealis AG. for optimising polymer production processes based on computational predictions.

MOHAMMAD ENAYATI

He received his PhD in Theoretical Physics (Quantum Cosmology) in 2017, specialising in group theoretical methods in physics. His research focused on de Sitter space-time symmetry, with applications in both classical and quantum physics. After several years of contributions in the field of theoretical physics, he moved to high-performance computing (HPC). He is attending the Master's in High-Performance Computing (MHPC) at the International Centre for Theoretical Physics (ICTP) as a student of the Scuola Internazionale Superiore di Studi Avanzati (SISSA)

FILIPPO FESTA

After obtaining a Bachelor's degree in Physics at the University of Padua, he completed a Master's degree in "Physics of Data". During his academic career, he developed skills in machine learning, data analysis and physics of complex systems. He carried out research internships at the CoMuNe Lab in Padua and the Universitat de Barcelona, studying complex networks and quantum information. He also gained experience as a teacher of mathematics and physics, during which he developed communication and organisational skills

ETTORE FADIGA

After his PhD in Computational Fluid Dynamics at the University of Ferrara, he held the position of HPC researcher in Leonardo SpA, where he specialised in numerical simulations, with a focus on aeroelasticity, GPU acceleration of finite volume codes and high-order solvers. Previously, he was a research fellow at the University of Ferrara, working on volumetric machines and open-source CFD. Expert in C++/Python development, simulation of industrial geometries with OpenFOAM/commercial software and management of research projects.

JONATHAN FRASSINETI

HPC Production and Support Specialist
He obtained a PhD in Condensed Matter Physics working on a project on the study of magnetic and charge interactions in innovative quantum materials using Nuclear Magnetic Resonance and ab-initio calculations (DFT).
He held the role of R&D designer in the simulations department with Digital Twin at Sacmi Imola for a few months.
He is passionate about computers, programming (Python and C++) and quantum computing.

ALEX GIANELLI

Master's degree in Computer Engineering from the University of Bologna, he carried out his thesis internship at the Cineca VisitLab working on a prototype for the Bologna Digital Twin.

GIUSEPPE GIAQUINTO

After graduating in Aerospace Engineering, presenting a study on the simulation of plasma synthetic jet actuators, he held the position of researcher in computational fluid dynamics at Leonardo labs.
He worked on software development of CFD codes and industrial simulations in the field of external aerodynamics. His professional aspiration remains that of expanding his knowledge of computational physics with particular regard to CFD and developing scientific codes on HPC platforms.

ALBERTO GUARNIERI

Master's degree in Computational Science & Engineering at the Technical University of Munich, with a thesis on the execution of particle simulations, based on the SPH method, on heterogeneous hardware using the SYCL parallel programming standard. In general, he has acquired skills in parallel programming for HPC, mainly in C++, with a focus on fluid dynamics and visualisation methods.

MATTEO IPPOLITI

During his PhD in Physics, he developed and implemented computational methods for quantitative Magnetic Resonance and medical imaging data analysis.
In 2020 he joined the Scientific Computing group of Elettra-Sincrotrone Trieste S.C.p.A., where he acquired skills in web applications aimed at data management and analysis. He participated in research collaborations with: the Charité University Hospital in Berlin, the Weill Cornell Medical College in New York and the National Institute of Metrology of the Federal Republic of Germany (PTB).

ALESSANDRO LAMBERTUCCI

Graduated in Molecular Biology with a specialisation in bioinformatics, he has developed transversal skills as a System Administrator in the HPC field, managing computational infrastructures to support the analysis of bioinformatics data.

FEDERICO LOMBARDI

Graduated with honours in Industrial Chemistry in 2021 at the "Toso Montanari" department of the University of Bologna with an experimental thesis focused on the study of the geometries of the transition states of the organo-catalysed asymmetric Michael reaction through DFT and Molecular Mechanics methods; he held the role of chemical laboratory technician in the field of environmental matrix analysis at a certified laboratory in Trentino Alto-Adige.
He subsequently joined the research and development division in the formulation of household detergents for hard surfaces, actively participating in finalising formulations contained in consumer products in Italy.
Since his early adolescence, he has cultivated a passion for programming and computer science, obtaining certifications of competence for the Python and C++ languages

SERENA LORENZINI

With a background in industrial biotechnology and a PhD in structural biology, she has taken an unconventional career path, moving from life sciences to the world of computer science.
In recent years she has gained extensive experience on Linux, HPC infrastructures and DevOps methodologies, expanding his skills and contributing to innovative projects in the field of bioinformatics.

GABRIELE MARCONI

While pursuing his Master's degree in Computer Engineering, he spent his internship at CNHI in Modena working on their Virtual Simulator for testing agricultural machinery.
During this period, he was able to develop an efficient method for creating virtual environments in Unity3D. He later participated as a programmer in an independent team in Bologna developing video games

RICCARDO GIUSEPPE MARGIOTTA

After obtaining a PhD in Applied Mathematics from King's College London (UK), he subsequently carried out research in data science and machine learning at SISSA (Trieste). He collaborated on national and international projects, developing specialist skills in deep learning, bioinformatics and data analysis with a focus on biomedical applications

MICHELE MASTROPIETRO

Graduated in Aerospace Engineering at the University of Pisa, he obtained a joint PhD in Astronomy from the Universities of Ghent (Belgium) and Groningen (The Netherlands) in the framework of the Marie Sklodowska-Curie Actions in the SUNDIAL project, performing hydrodynamic simulations of galaxies on HPC platforms. Afterwards, he held a position as technologist for INAF, at the Astronomical Observatory of Rome, where he worked on the software infrastructure as administrator of HPC and kubernetes systems for the analysis of Cherenkov telescope data and contributed to the parallelisation and optimisation of simulation codes.
His interests are cloud computing and HPC technology, with an emphasis on the containerisation aspect.

RICCARDO MIOLI

After graduating with a Bachelor's degree in Computer Science for Management, he continued his education with a Master's degree in Computer Science and a Master's degree in Data Science and Business Analytics.
During his studies, he carried out several projects in the field of Data Analytics, ranging from the automation of forensic artifact analysis to the analysis of transaction networks on blockchain. During the Master's degree, he carried out machine-learning projects in application domains ranging from sentiment analysis using supervised classification techniques to market analysis using clustering algorithms.
His passions include 3D printing, hiking, and more generally everything related to life in the open air.

ANDREA MONTERUBBIANO

He obtained a PhD from "La Sapienza" in Rome, focusing on programmable network devices and their applications, such as network flow monitoring. He collaborated with the Huawei Paris Research Centre on a project for the integration of a machine learning model in a programmable network card (smart NIC with FPGA). He knows low-level languages including C and Verilog. He worked with more user-friendly languages, such as Rust or Python. He is keen on the open-source world and is familiar with managing Linux servers.

ROSSELLA PANSINI

With a PhD in Archaeology, she was a research fellow and contract teacher at the University of Siena, Department of Historical Sciences and Cultural Heritage, where she collaborated on the projects SOS - Sotto Siena and ADABB - Archivio Digitale Aumentato Bianchi Bandinelli. Specialised in topographic survey, GIS, photogrammetry and 3D reconstruction for cultural heritage.

LUCA ROBBIANO

Having obtained a Master's degree in Computer Engineering in 2020, he is completing a PhD focused on deep learning and computer vision. In parallel with his research activity, between 2020 and 2024 he managed and maintained the IT infrastructure of his research laboratory, honing his skills in Linux and HPC systems

Unlocking excellence: Why Cineca summer school matters & our inspiring success story

Nitin Shukla
Cineca

In an era of rapid technological advancement, preparing the next generation of talent for both industry and academia is more critical than ever. With this goal in mind, every year Cineca offers a two-week summer school tailored to Italian and European undergraduates, PhD students and early-stage postgraduate students. In recent years, there has been an unprecedented surge in interest. In 2024 we received a record number of applications. Due to our competitive selection criteria, we were able to accept up to 25 highly motivated participants (see group picture). These participants came from a diverse range of fields, including computer science, aerospace engineering, chemistry and physics.

One of the main challenges was designing a course that would remain dynamic and motivating for two weeks. The course was designed to take participants from basic to advanced levels, exploring new ways of working with large HPC systems. In addition to hands-on sessions, we included additional activities to keep participants engaged and excited throughout the programme.
A new addition to the programme was the introduction of a project-based approach. We received encouraging feedback from the participants regarding this format. In fact, the Cineca Summer School received the highest evaluation rating of 9.3/10, reflecting the quality and excellence of the course and teaching. This success would not have been possible without the valuable support of all the staff involved in the school. Below are the testimonies from some of the students.

Figure 12, 13: Summer School

Manuel Casablanca
University of Padova – Department of Industrial Engineering

The Cineca HPC Summer School is a unique opportunity for personal and professional growth. The knowledge and skills acquired in HPC and parallel computing are immeasurable for my research, which is focused on computational fluid dynamics applied to aeronautical propulsion. These tools significantly enhance the performance of simulation codes, enabling the study of complex cases that would otherwise be impractical. Additionally, the Summer School offers an enriching international and interdisciplinary environment, allowing participants to collaborate with outstanding students and instructors while balancing productive work with moments of relaxation.

Khaled CHIKH
PhD & Research fellow at Hipert Lab | AI Engineer
University of Modena and Reggio Emilia

Attending the Cineca HPC Summer School was a game-changer for my research in deep learning and model optimisation. The hands-on experience with high-performance computing resources, parallel computing strategies and optimisation techniques allowed me to scale my object detection models efficiently. This training was particularly valuable as I transitioned from training on a single GPU to leveraging the Leonardo HPC system for large-scale distributed training. The school provided a solid foundation in HPC best practices, which directly impacted my ability to develop computationally efficient AI models for edge devices. I highly recommend it to researchers aiming to push the limits of their AI workloads.

Events

Cineca for artistic residencies

Antonella Guidazzoli, Maria Chiara Liguori
Cineca

Artistic residencies at Cineca, between AI and scientific data

The final exhibition of the GRIN project (Art-driven Innovation for Digital and Green Transition in European Regions), held at Serra Grande Kilowatt in September 2024, showcased the outcomes of three regional artistic residencies. The installations, created by Marco Barotti, Salomé Bazin and Calin Segal, integrated artistic practices with scientific data, leveraging AI and high-performance computing (HPC) for innovative experimentation. Following this initial engagement in artistic residencies, Cineca played a key role in the EU Digital Deal project, curated by Sineglossa and the artistic duo Dmstfctn. The project utilised Cineca's HPC resources to develop The Models, an AI-driven installation presented to the public at DAMA technopole in February 2025.

Figure 14: Destination Earth by Salomé Bazin.
Figure 15: Fungi – Symbiotic Harmonies by Marco Barotti.

Figure 16: Tales from the Receding Edge by Calin Segal.
Figure 17: Presenting The Models at DAMA technopole.
Figure 18: The Models by dmstfctn.

Events

Biannual workshops on AI, cultural heritage and art

Antonella Guidazzoli, Maria Chiara Liguori
Cineca

The most recent workshops organised by Cineca's VISIT Lab took place on 25 September 2024 and 21 February 2025. Continuing the tradition of previous editions, these events explored the intersections between artificial intelligence, cultural heritage and art. As discussions on these topics continue to expand, new themes were introduced, including:

- Authorship and AI: Examining the challenges and opportunities that AI presents in creative processes.
- The AI Factory Concept: Investigating its implications and the expectations it generates among stakeholders.
- AI and Environment: Analysing the relationship between AI and environmental sustainability from a broad perspective.

As in past editions, the workshops were streamed online and the proceedings are available for reference.

Figure 19: Invitation to the fourth workshop (25 September 2024).
Figure 20: The fifth workshop, hosted by PwC Italia, Bologna office.
Figure 21: Link to the streaming page for all events.
Figure 22: Supercomputing

Events

SC2024 - Atlanta

Daniela Galetti
Cineca

The Supercomputing Conference 2024 (SC24), which took place in Atlanta, offered Cineca, an excellent platform to engage with global leaders in the supercomputing arena.

Taking place from 17 to 22 November 2024, Supercomputing 2024 drew thousands of participants from around the world, including researchers, scientists, educators and industry professionals.

As in past years, the event focused on the latest advancements in supercomputing technology, applications and solutions that drive research and innovation across diverse fields, including healthcare, climate modelling and artificial intelligence.
Cineca showcased its latest achievements in HPC, relevant to both European and global audiences, at adjacent to the INFN one. By sharing expertise and insights with visitors to the booth, Cineca aimed to foster collaboration and knowledge sharing, essential for addressing complex scientific challenges.

Networking is also a key component of the Supercomputing Conference, allowing Cineca to connect with other leading supercomputing centres, their researchers and the principal technology providers. The information and new technologies seen at the event helped Cineca pave the way for future collaborations that will drive advancements in AI and the supercomputing environment.

HPC H-index

WebofScience

1084
Results found

2391
Number of citations

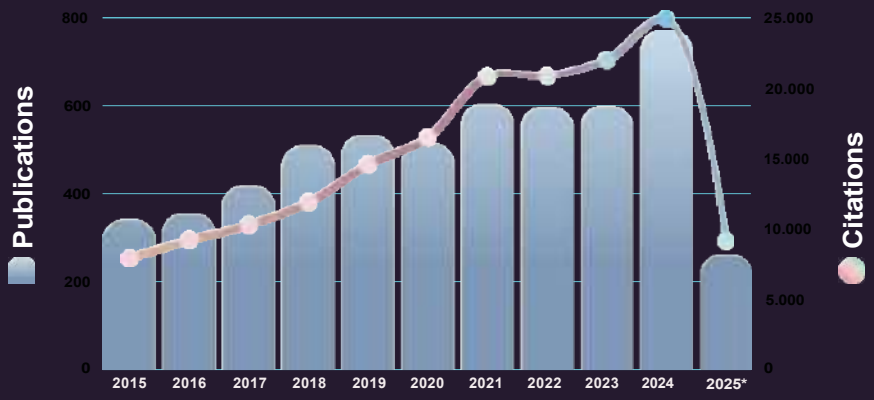
2,21
Avarage
Citations per item

16
H-index

www.webofscience.com - Database: Web of Science Core Collection
Search: (FT=Cineca or FO=Cineca) and (PY=2024-2025) - Date: 20.05.2025

Publications mentioning CINECA

* Updated to may 2025



Top affiliation of the authors

NATIONAL RECOVERY AND RESILIENCE PLAN

ECOSISTER ecosystem for Emilia-Romagna's sustainable tradition

Angela Acocella
Cineca

The National Recovery and Resilience Plan (NRRP), under Mission 4, "Education and Research," has promoted the creation of "Innovation Ecosystems", as part of Component 2 Investment, founded by the European Union. The primary goal of this initiative is to strengthen national research and industrial ecosystem, fostering economic growth, innovation and the digital transformation of production processes, with a particular emphasis on economics and environmental sustainability. In Emilia-Romagna, this initiative was activated through collaboration between several universities and public research in- stitutions, coordinated by ART-ER (Attrattività Ricerca Territorio), which has allowed to create a territory-oriented Innovation Ecosystem: "Ecosystem for the sustainable transition of Emilia-Romagna" (ECOSISTER). The Ecosystem is structured according to a Hub & Spoke model, to facilitate synergies between different institutions. HUB members include the University of Bologna, the University of Ferrara, the University of Modena and Reggio Emilia, the University of Parma, the National Research Council (CNR), the Catholic University of the Sacred Heart, the Polytechnic University of Milan, the National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), ART-ER and the National Institute for Nuclear Physics (INFN). Spokes promote activities focused on research, innovation and technology transfer, aimed at developing advanced materials, clean energy technologies, sustainable production processes, mobility/housing solutions and technologies for natural resource management.

In collaboration with INFN, Cineca participates in Spoke 6, titled "Ecological Transition based on HPC and Data Technology", playing a central role inenhancing Emilia-Romagna's position as the National Data Valley through the promotion of advanced computing and data technologies, enabling researchers to exploit HPC capabilities and encouraging sustainable research practices. The contribution of Cineca in supporting ECOSISTER partners involved multiple activities. A tailor-made resource allocation model has been developed and implemented via a platform that hosts quarterly allocation calls, optimizing resource usage in a targeted manner to address the specific needs of ECOSISTER researchers. In collaboration with INFN, Cineca has also contributed in drafting the reference document that defines paradigms supporting advanced computing activities, aiming to promote an efficient use of HPC infrastructures. As part of high-level support activities, Cineca is involved in a collaboration with the University of Modena and Reggio Emilia (UNIMORE) in optimising and debugging specific routines of the flagship Quantum ESPRESSO code, a reference software in the field of materials science. A workshop specifically designed for Ecosister researchers is also being prepared, focusing on introducing advanced parallel computing techniques to provide participants with tools and methods aimed at enhancing the performance of their scientific simulations. All these support practices promoted by Cineca contribute to improving the efficiency of leading codes, enabling more effective management of computational resources and reducing energy consumption.

Figure 23: Ecosister pillars

ecosister

D3 4Health

Elisa Rossi
Cineca

D3 4Health "Digital Driven Diagnostics, prognostics and therapeutics for Sustainable Health care" is a project funded by the National Plan for Investments Complementary to the NRP and intended for "Research Initiatives for Innovative Technologies and Pathways in Health and Care.''
The project is coordinated by La Sapienza University of Rome and composed of 28 partners and aims to enabling new technologies for data collection and analysis in order to provide personalised medicine. It promotes the development of innovative predictive, diagnostic and therapeutic models, making use of the most advanced digital technologies, represented by AI algorithms, wearable devices and sensors. The expression of scientific research advancement will be represented by the development of a Digital Twin and a Biological Twin.
D3 4Health consists of four spokes aiming to develop new solutions for the diagnosis, mo- nitoring and therapy of five different diseases: colon cancer, liver cancer, central nervous system cancer, type 1 diabetes and multiple sclerosis.
Cineca participates in Spoke 2, "Multilayer platform to support the generation of the Patients' Digital Twin", which integrates different multidisciplinary professional expertise. The aim is supporting the generation of innovative AI solutions for data collection and analysis in healthcare.
Cineca provides the necessary infrastructure (cloud and HPC) and resources to support computational activities. Given that the project is expected to handle an increasing volume of data and the development of additional microservices, Cineca is responsible for ensuring that the HPC infrastructure remains efficient in meeting all demands.
In addition to infrastructure and computational resources, Cineca is responsible for creating a research platform that will enable the use of AI models developed by project partners. The research platform will host a model registry, which will be connected and synchronised with SRACE and integrate an inference service to support data analysis, facilitating model distribution and use by healthcare professionals.

* Progetto PNC 0000001 D3 4Health – CUP B83C22006120001 – Piano nazionale per gli investimenti complementari al PNRR, finanziato dall'Unione europea – NextGenerationEU.

D3 4HEALTH
Digital Driven Diagnostics, prognostics and therapeutics for sustainable Health care

The computational platform of the National Biodiversity Future Centre (NBFC)

Antonio Costantini, Gabriella Scipione, Giuseppe Trotta, Marco Puccini, Giuseppe Melfi, Juan Mata Naranjo, Xhulio Dhori, Davide Crisante, Elena Saluzzi
Cineca

The National Biodiversity Future Centre (NBFC) aims to monitor, conserve, restore and valorise (MCRV) Italian and Mediterranean biodiversity. It has been created to address interdisciplinary and frontier research and innovation activities devoted to knowledge monitoring, conservation, restoration and valorisation of Italian biodiversity as a National Champion of R&D based on Key Enabling Technolo- gies for Biodiversity and Environmental Sustainability.

Cineca is a Partner of one of the eight Spokes of the Project. Initially, Cineca's role was to develop four different platforms, for each of the identified macro-areas: "Biodiversity Collections", "Molecular Biodiversity", "Bioresources" and "Ecosystem Function and Monitoring (BEF)". After discussions with the Platform's Spokespersons and the Researchers from the NBFC, the decision was made to aggregate these four platforms to create a single National Access Point to Biodiversity Data and Tools.

In other words, Cineca has a double role in NBFC: to build the NBFC Platform that includes Biodiversity Collections, Molecular Biodiversity, Bioresources, Ecosystem Function and Monitoring and to provide cloud and HPC resources with its infrastructure, including Galileo100 and the pre-exascale system Leonardo.

To develop the platform, Cineca decided to follow the containerised microservices architecture strategy. An ecosystem of services is under development, that can be divided into two categories. General-purpose services such as Identity and Access Manager, Object Storage, Graph Database, (asynchronous) Task Queue, Log Aggregation System, etc. are implemented to manage and maintain the platform itself. On the other hand, there are many domain-specific services, requested by the project's partners: External Data Retrieval from international repositories (e.g. GBIF, BOLDSystem, GenBank, Copernicus), Comparator of Primary Biological Sequence Information, Bayesian Analysis of Molecular Sequences, GPU-accelerated Genome Analysis Tools, Joint Species Distribution Models, Hydrology Models, etc. From this point of view, the main purpose of the platform is to provide FAIRness of data and tools. The development team is also working to build such an ecosystem on top of a Kubernetes cluster, to guarantee stability of the services.

Recently, a notable activity has been the implementation of AI services for NBFC researchers. One such service is Image Recognition, primarily aimed at supporting the Biodiversity Collections area by processing handwritten and typewritten labels from natural history museums. Another AI service will leverage large language models (LLMs) to facilitate easy retrieval of information from the platform's graph database.

ARTIFICIAL INTELLIGENCE

AI FACTORY

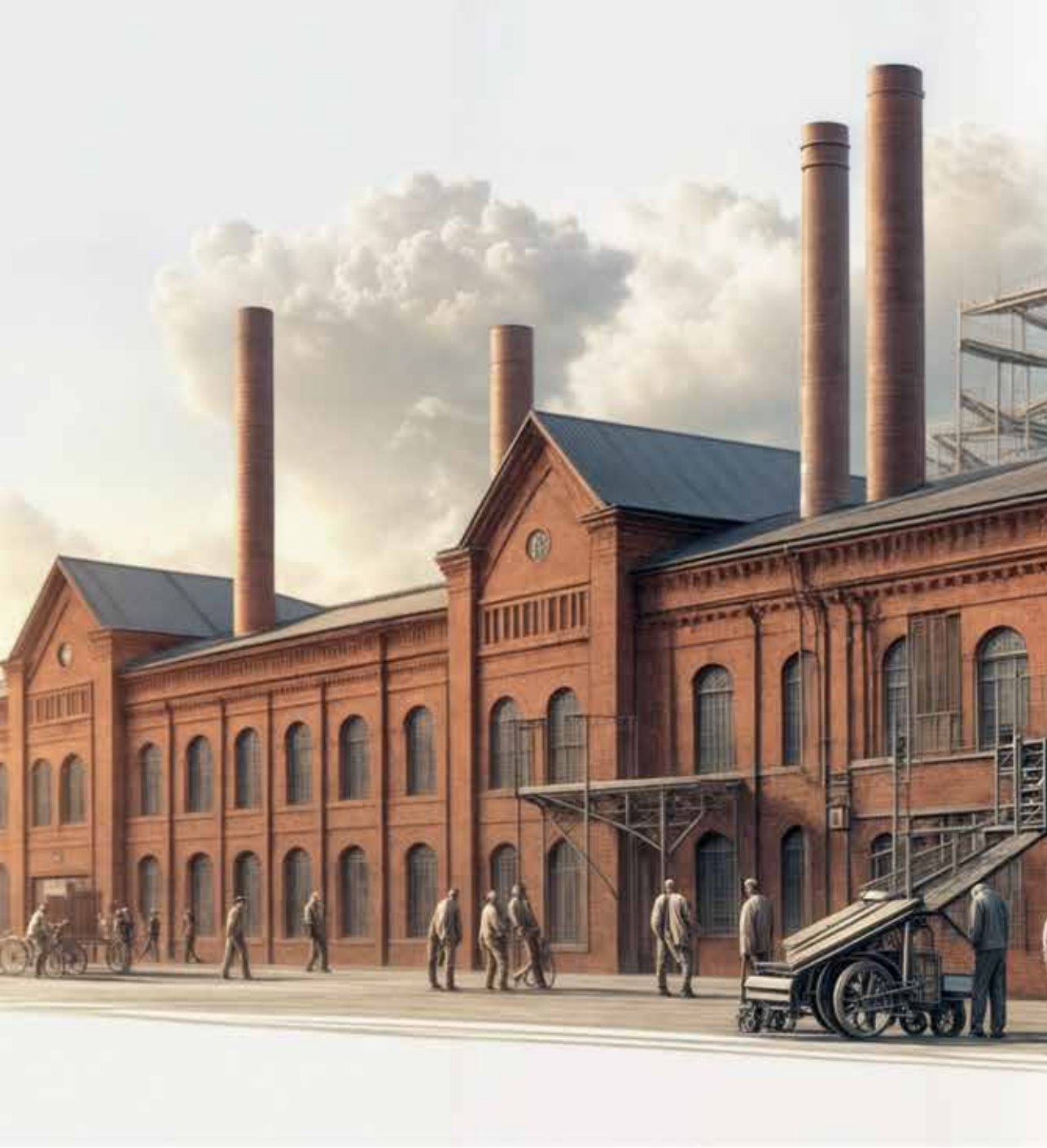


Figure 24: AI Factory, Maria Chiara Liguori

Empowering Italian and European AI ecosystems: the launch of IT4LIA

Laura Morselli
Cineca

The Italian AI Factory (IT4LIA), starting in April 2025, is one of the first AI Factories selected and founded by the European High-Performance Computing Joint Undertaking (EuroHPC JU). This marks a significant achievement for Italy and a major step toward building a powerful, coordinated European AI ecosystem.

Coordinated by Cineca, IT4LIA is a strategic initiative aimed at transforming the national and European AI landscape through advanced infrastructure and a comprehensive portfolio of services. IT4LIA benefits from EuroHPC Leonardo supercomputer, its AI-enhanced LISA upgrade, GAIA cloud and the upcoming AI-optimised system, all situated within the DAMA technopole data centre. This world-class infrastructure hub provides a complete computing continuum and will serve Italy's and Europe's rapidly expanding AI ecosystem, fostering Italian and European competitiveness worldwide. Overall, the infrastructure is anticipated to offer over 20.000 GPUs, establishing it as an exceptional facility to attract talent from Italy and Europe and enable them to effectively implement their most advanced and innovative ideas.

The AI Factory will offer specialised services in key verticals, including agri-tech & agri-food, cybersecurity, weather & climate and manufacturing, ensuring sector-specific demands are addressed with precision. Complementing these vertical offerings, the AI Factory will provide a full suite of horizontal services aimed at supporting all stakeholders within the AI ecosystem. A core feature of the AI Factory is its data-related services, anchored by a vast, co-located data facility with 24 petabytes of on-premise data (and growing). A strong emphasis is placed on supporting SMEs and start-ups through a set of customised services as well as via open calls that will provide free access to computational resources, expertise and end-to-end project support. This targeted support will be key in ensuring that smaller, emerging players can leverage advanced AI resources and bring innovative solutions to market, effectively bridging the gap in resources and technical expertise.

Additionally, the AI Factory will implement a training programme that includes a learning platform for self-paced learning, hackathons, internships and numerous in-person training events. This programme will ensure broad access to up-skilling opportunities while addressing the diverse needs of the AI ecosystem.

IT4LIA is designed to integrate and complement various national and European initiatives which will serve as AI gates, acting as vital connectors between the AI ecosystem and the AI Factory's resources and services. By fostering collaboration and leveraging synergies, the AI Factory will create a cohesive ecosystem where resources, expertise and innovation converge, ensuring that these efforts collectively advance AI development and adoption across Italy and Europe.

The selection of IT4LIA as one of the first AI Factories underscores Italy's commitment to fostering innovation and collaboration in the AI domain, positioning the country as a pivotal player in shaping the future of artificial intelligence in Europe

Large-scale AI on Leonardo: Powering Italy's foundational models

Laura Morselli
Cineca

In 2024, Leonardo supercomputer became a key enabler of Italy's large-scale AI development. Several Italian foundational AI models were trained on Leonardo, reinforcing its role as essential infrastructure for national and European AI innovation.

The open-source foundational Large Language Model (LLM) Italia, developed by iGenius in partnership with Cineca and released in June 2024, is designed specifically for the Italian language. With 9 billion parameters, it marks a significant step in building transparent, secure AI aligned with European values.

The Minerva models, developed by the Natural Language Processing group of the Sapienza University of Rome with support from the FAIR project and Cineca, are Italian-language LLMs trained from scratch. The 7-billion parameter Minerva model demonstrates strong performance and linguistic alignment with national needs.

Velvet, developed by Almawave, is a multilingual and multimodal LLM family (including Velvet 14B and 2B) supporting six major languages. Trained on Leonardo in 2024, it is designed for applications across sectors such as healthcare, finance and public services.

LLaVA-MORE, from AIMagelab, is a family of multimodal LLMs that integrate visual and textual data. Trained using Leonardo's computational power, these models advance capabilities in reasoning, generation and following instructions.

Leonardo also supported the Large AI Grand Challenge, an initiative by the EuroHPC JU and the European Commission to foster innovation and excellence in large-scale AI models. Two European SMEs were granted the opportunity to train large-scale LLMs on Leonardo by getting an allocation each of 2 million GPU hours: Lingua Custodia (from France) and Unbabel (from Portugal).

Translated partnered with Cineca in November 2024 to train a machine translation model on Leonardo. Built on a dataset collected over more than 15 years and enriched with human revisions, the model focuses on Italian–English translation and will be released with open weights. The project aims at a "universal translator" and shows how HPC can drive impact in multiilingual communication and AI for language.

Together, these initiatives demonstrate how Leonardo is shaping the foundation of Europe's AI future. By enabling the development of sovereign, trustworthy and multilingual AI models, Leonardo is cementing Italy's role as a strategic hub in the global AI race and ensuring that European values remain at the core of tomorrow's digital intelligence.

AI APPLICATIONS

Advancing AI: From cultural heritage to weather forecasting and beyond

Roberta Turra
Cineca

The growing demand for AI solutions across diverse domains provides valuable opportunities to test and refine rapidly evolving technologies. Over the past year, alongside its core activities (facilitating access to computational resources for AI projects and supporting users with training programmes) Cineca has actively participated in research projects and proofs of concept (PoCs). This involvement has enabled the assessment and comparison of some of the most advanced techniques, models and tools available in the AI field. In particular, an automated image captioning system was developed inside the MIC project, for the digitisation of cultural heritage collections, requiring the selection and adaptation of the best-performing object-detection, image-recognition and language-generation technologies. In the domain of image analysis, AI-driven approaches have also been applied to weather forecasting. Specifically, models based on U-Net architecture, spatio-temporal learning (STL) and transformers have been developed and validated, in collaboration with Illumia, for the AI General Circulation Model (AI-GCM), an Italian deep learning-based weather model capable of generating hourly forecasts. Additionally, for short-term weather prediction (nowcasting), AI models leveraging satellite and radar data have been developed in partnership with FBK, ARPAE and the Italian Meteo Agency. Large language models (Llama 3.1 8B and Qwen 2.5 32B) have been used to develop CinecaBOT, a prototype chatbot designed to assist both expert and non-expert users by providing answers based on Cineca's technical reports, user manuals and project documentation. Language technologies have also been used to identify and anonymise personal data in administrative acts and decrees issued by local authorities, ensuring compliance with data protection regulations in the Savia project. Given the increasing need for anonymisation across multiple projects, various techniques have been studied and tested also for structured data collections, notably within the GRINS project. Additionally, foundational models for time series, such as Amazon Cronos, IBM Granite and Google TimesFM (large quantitative models, or LQMs) are being studied and applied to air-quality sensors data, in the AIQS project to generate forecasts in the same way that LLMs generate text, with very promising results. The same techniques are being tested on Leonardo monitoring data to predict job behaviour and to anticipate which node is going to fail.

Cineca GRINS project summary - 2024/2025

Giorgio Pedrazzi
Cineca

For the PNRR project GRINS (Growing Resilient, Inclusive and Sustainable), in the second half of 2024, Cineca significantly expanded the AdaCloud tenant available for the GRINS AMELIA platform, doubling its previous capacity. This upgrade strengthened AMELIA's computational capabilities and laid the groundwork for a further planned expansion, which will integrate GPU resources into the cloud infrastructure. These developments complement the ongoing pilot projects on Leonardo, carried out in collaboration with Exprivia and other GRINS partners, aimed at leveraging Cineca's high-performance computing infrastructure for advanced AI model training.

Moreover, Cineca collaborated within the GRINS project, specifically focusing on Work Package 3, in Spoke 0, addressing critical challenges related to anonymising student data for research purposes. Cineca led the identification and application of advanced anonymisation techniques, selecting the sdcMicro tool to effectively manage sensitive information.

The case study involved integrating personal, enrolment and academic data into an unified analytical table, posing distinct privacy risks due to the potential for re-identification. Cineca developed a comprehensive three-step anonymisation strategy, tailored to the specific nature of the data:

- Cross-sectional data: Cineca applied recoding and suppression methods, setting a k-anonymity threshold of 3 to protect against re-identification while maintaining data utility. Continuous variables, such as grades and financial indicators (ISEE), were anonymised through binning and noise addition.
- Group-level data: Recognising the re-identification risk in courses with small enrolments, Cineca strategically merged smaller academic programmes into larger groups, balancing privacy protection with data integrity.
- Longitudinal data: To secure individual privacy across repeated observations (grades, programme participation), Cineca implemented innovative techniques like grade aggregation, microaggregation, noise addition and random swapping, effectively mitigating privacy risks associated with unique individual patterns.

Throughout the case study, Cineca emphasised balancing privacy with analytical utility, fostering continuous dialogue within the GRINS project team. This iterative approach facilitated customisation of anonymisation techniques, highlighting the importance of flexible strategies and tailored solutions to diverse research needs. The insights gained underline Cineca's commitment to refining methodologies for secure, privacy-conscious data sharing, setting a robust framework for future research collaborations.

SAVIA project: Normative drafting

Michele Visciarelli
Cineca

Legislative texts rely on clear formal structure and substantive intent, yet many laws suffer from issues like excessive chain references, vague provisions and disjointed amendments, which obscure meaning and undermine legal certainty. These flaws affect public administration, contributing to defensive bureaucracy and litigation. To address these challenges, a drafting manual was developed to enhance clarity by combining formal techniques with an assessment of legislative impacts, public consultation and modern information technologies.

Building on these principles, the SAVIA project stands out as a pioneering initiative that applies advanced natural language processing (NLP) techniques to systematically enhance legislative drafting. The project harnesses tools that extract quantifiable textual indicators such as word count, sentence complexity and the frequency of gerunds, passive constructions and chain references from digitised legislative texts. Using methods like lemmatisation, stemming and tokenisation, the project transforms legal texts into a standardised string format that enables automated analysis. This text is then processed to calculate a composite quality score based on the L1 norm of the vector representing the law, where lower scores denote clearer and more concise language. The scores are normalised on a scale from 0 to 1 and laws are ordered and segmented into quality tiers, each receiving a qualitative rating from 1 (best) to 5 (worst).

Moreover, SAVIA integrates this textual analysis with judicial data by linking legislative texts to administrative acts and subsequent judicial decisions. Advanced prompt engineering with large language models (LLMs) is employed to automatically extract references to administrative acts from judicial pronouncements, thereby bridging the gap between legislative drafting and its practical implications. In parallel, the project leverages zero-shot named-entity recognition (NER) techniques to classify judicial outcomes. The system encountered difficulties in distinguishing similar legal categories, such as "inadmissible" versus "improcedible" appeals. By consolidating these into two overarching classifications (outcomes unfavorable to the legislator versus those favorable), the approach achieved an accuracy of approximately 97%. This integration of sophisticated NLP methods with AI-driven analysis not only provides a robust quality measure for legislative texts but also offers a dynamic tool for continuous improvement in lawmaking.

While traditional drafting challenges persist in obscuring legislative intent and hindering administrative effectiveness, the SAVIA project demonstrates a promising convergence of legal expertise and technological innovation. By quantifying textual quality and linking it to judicial outcomes, the initiative offers a clear, data-driven framework for producing more coherent and effective legislation, in the hope of fostering transparency and reducing administrative disputes.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani

DIGITAL TWINS

From open data to energy simulation: Building UBEM for Bologna

Eleonora Bergamaschi, Riccardo Mioli
Cineca

Urban Building Energy Models (UBEMs) are key tools for analysing and optimising urban energy consumption, supporting sustainable city development. They simulate energy performance by aggregating building models, providing insights into consumption patterns and enabling different energy scenario evaluation [1][2]. Integrating UBEMs with Urban Digital Twins enhances urban planning strategies for energy efficiency and decarbonisation.

A UBEM has been developed as part of the Bologna Digital Twin project, coordinated by Fondazione Bruno Kessler (FBK) in collaboration with Cineca, the University of Bologna, and Fondazione Innovazione Urbana (FIU). The case study aims to simulate energy use for as many buildings as possible and analyse results by neighbourhood. EnergyPlus [3], a widely used open-source simulation engine, was chosen for the project. The simulation is based on some input files that specify the characteristics of a building. At the end of the simulation, output files describing the "energy behaviour" of the building are produced.

The accuracy of UBEMs depends on input data such as building geometries, materials, HVAC systems and occupancy patterns. Bologna's model uses Open Data from the Municipality to extract building geometries, which were converted into EnergyPlus input files. Two-dimensional floor plans were transformed into 3D models, incorporating window surfaces and shading effects. The thermal properties of each building were determined using TABULA [4], a European project that defines building archetypes based on construction year, ensuring realistic energy performance modelling.

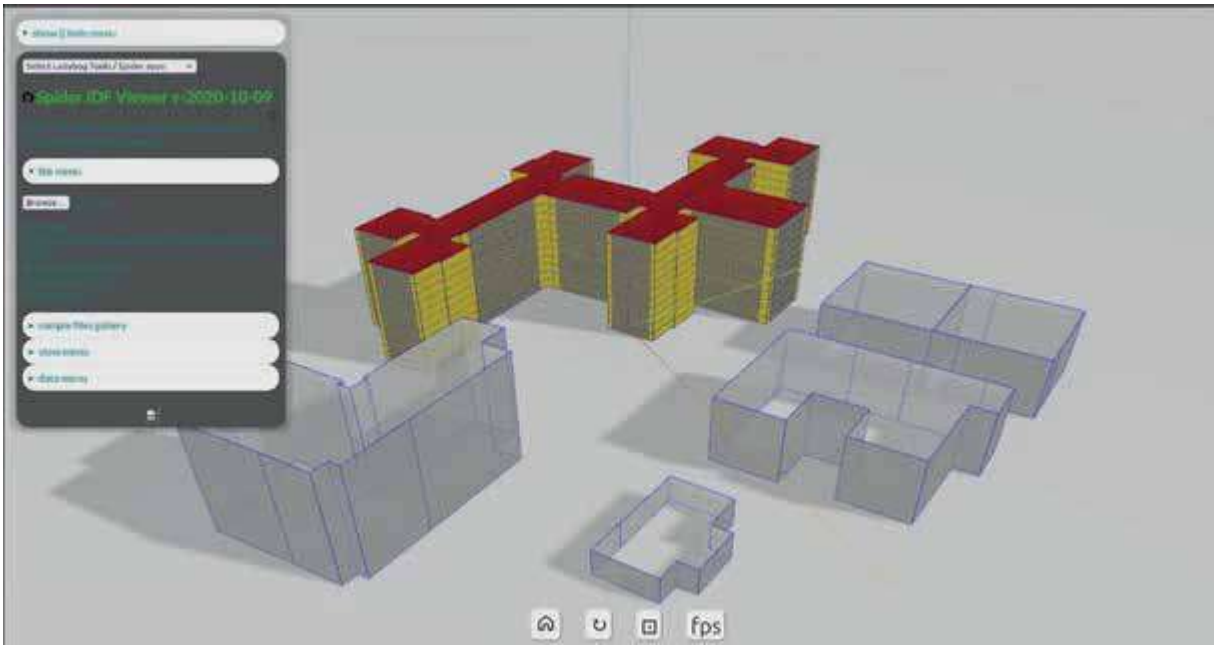


Fig. 27 Visualisation of an EnergyPlus input file using Spider IDF Viewer

Input data was collected for approximately 24.000 buildings. The simulation of that number of buildings requires high computational power. Workloads were distributed across 10 nodes of Leonardo's Data Centric General Purpose (DCGP) partition with 112 CPU cores each, using a task-parallel approach with Ray [5], an open-source framework designed for scaling Python and Machine Learning applications with massive distributed computing capabilities. This enabled thousands of simulations to run simultaneously, completing the full-scale run in 1.5 hours. Further optimisations aim to improve efficiency.

Next, the model will be validated by comparing simulated data with actual energy consumption. Bayesian optimisation will refine archetype parameters, enhancing accuracy. Once validated, the UBEM will support scenario analysis for urban planning. It can assess the impact of solar panels, insulation materials, energy-efficient retrofits and district heating systems. These insights will help policymakers and urban planners make data-driven decisions for energy efficiency and sustainability.

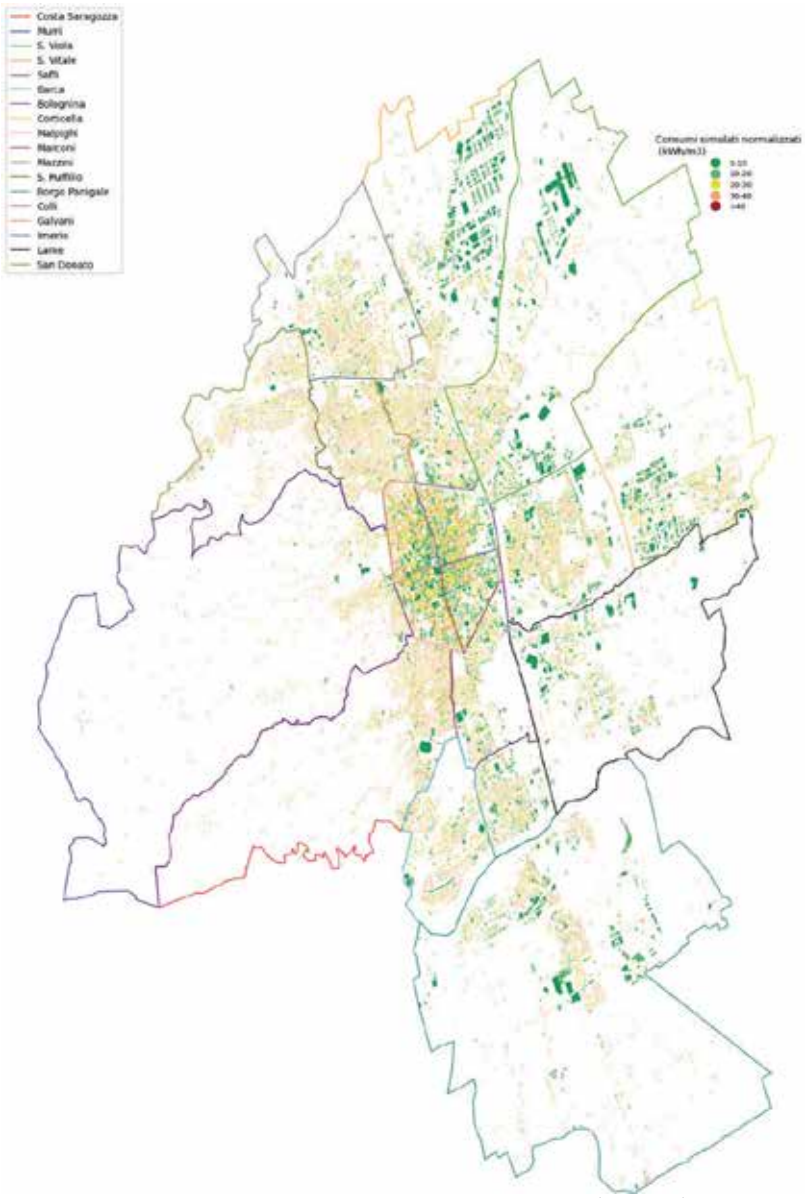


Fig.28: First outputs of initial data pipeline

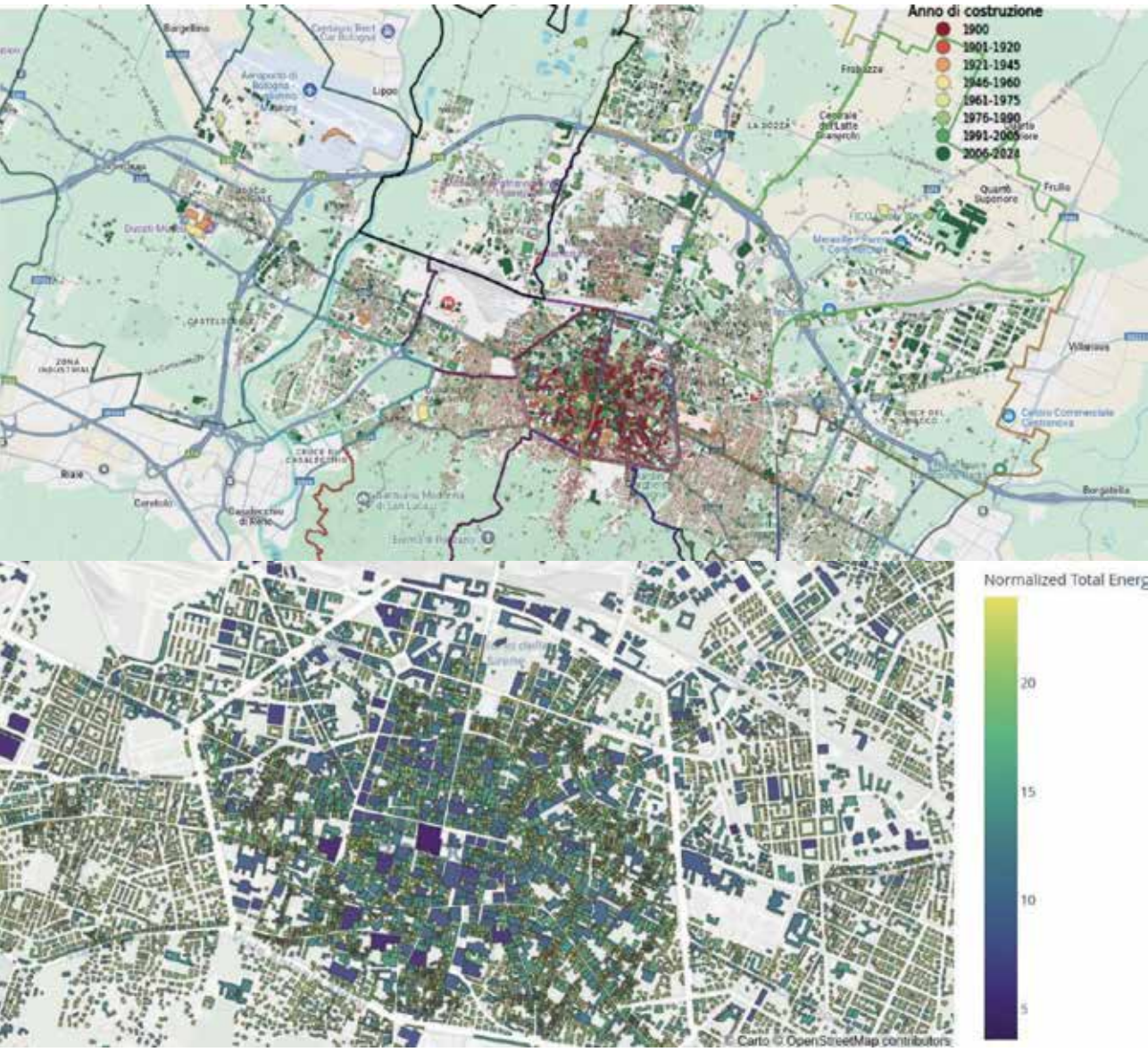


Fig.29: Distribution of construction years for the simulated buidlings".
The second is "Normalised energy consumption (simulated) KWh/m3"

[1] Ferrando, Martina & Causone, Francesco & Hong, Tianzhen & Chen, Yixing. (2021). Urban Building Energy Modelling (UBEM) Tools: A State-of-the-Art Review of bottom-up physics-based approaches. 10.48550/arXiv.2103.01761.

[2] Johari, Fatemeh & Peronato, Giuseppe & Sadeghian, Paria & Zhao, Xiaoyun & Widén, Joakim. (2020). Urban Building Energy Modelling: State of the Art and Future Prospects. Renewable and Sustainable Energy Reviews. 128. 109902. 10.1016/j.rser.2020.109902.

[3] EnergyPlus is funded by the U.S. Department of Energy's (DOE) Building Technologies Office (BTO) and managed by the National Renewable Energy Laboratory (NREL). EnergyPlus is developed in collaboration with NREL, various DOE National Laboratories, academic institutions and private firms. https://energyplus.net/

[4] TABULA Project Team. Typology approach for building stock energy assessment. Main results of the TABULA project: final project report. Darmstadt: GmbH; 2012. Available from: http://episcope.eu/building-typology/

[5] Ray Documentation: https://docs.ray.io/en/latest/index.html

Preserving heritage through digital twins: The Garisenda Tower project

Stefano De Miranda
University of Bologna
Manuela Faustini Fustini
Bologna Municipality
Andrea Pilzer
NVIDIA
Chiara Dellacasa
Cineca

The iconic skyline of medieval Bologna is shaped by its towers, remnants of a time when powerful families constructed these lofty structures as both status symbols and defensive bastions. The Garisenda Tower stands out among them, leaning by four degrees beside the taller Asinelli Tower. In 2023, unexpected shifts in its structure raised alarms, sparking urgent restoration efforts to prevent the risk of collapse. At the forefront of this initiative is a groundbreaking digital twin project that combines artificial intelligence, high-performance computing and real-time sensor data to protect this historic landmark.

This digital twin is the result of a collaborative effort between the Municipality of Bologna, the University of Bologna and Cineca, with NVIDIA providing the technological infrastructure. Central to the project is NVIDIA Omniverse, which serves as the data integration and visualisation platform. Using Universal Scene Description (USD) technology, it helps create a dynamic, photorealistic replica of the tower; approximately 100 sensors embedded in the tower provide data on vibrations, tilt, temperature, wind conditions and material strain directly into this virtual model.

While structural monitoring is not new, what sets this project apart is how the data is used. Leveraging the power of supercomputing and AI, the system will be able to run predictive simulations that assess the tower's integrity, to simulate its behaviour in different scenarios and test the outcomes of possible restoration strategies – virtually – before any physical intervention is made.

Advanced imaging techniques like high-resolution laser scanning, photogrammetry and detailed structural modelling further enhance the system. High-resolution digital meshes can help conservators monitoring tiny changes over time, such as cracks, deformation or erosion, that might otherwise go unnoticed.

The platform enables seamless real-time collaboration. Engineers, architects, AI experts and other specialists can work within the same shared digital space, updating models, running simulations and exchanging insights instantly. This interconnected approach speeds up decision-making and guarantees that every stakeholder works from the most accurate and up-to-date information.

As the project progresses, it represents more than just a method to save a leaning tower. It signals a broader shift toward predictive, data-driven conservation strategies – blending tradition with cutting-edge technology. Once proven successful, the Garisenda Tower's Digital Twin could serve as a template for safeguarding other culturally significant structures across Italy, helping preserve the nation's architectural heritage for future generations.

Figure 30: Rendering of the digital twin of the Garisenda Tower created using high-fidelity 3D meshes within NVIDIA Omniverse platform. The model integrates the historic structure into the urban digital environment of Bologna, enabling advanced simulations for structural monitoring, preservation planning and immersive visualisation.

Preserving heritage through digital twins: The Garisenda Tower project

Eleonora Bergamaschi, Chiara Dellacasa, Anna Nikishova, Michele Visciarelli
Cineca
Aldo Canfora, Mirko Degli Esposti, Matteo Falcioni, Tommaso Rondini
University of Bologna

As part of the project "Gemello Digitale: Governo e Valorizzazione del Patrimonio Dati", University of Bologna and Cineca are working together to build the classification of Bologna's LiDAR dataset and develop a pipeline for individual tree segmentation from airborne LiDAR data in Bologna.

The data, acquired with a Leica CityMapper-2 sensor by CGR, integrates high-resolution aerial imagery and dense LiDAR point clouds to support detailed 3D urban modelling. Leveraging the HPC capabilities of the Leonardo supercomputer, we trained a Random Forest classifier on 2.5 million annotated points to label each LiDAR point into six semantic classes (Grass, High Vegetation, Building, Railway, Road and Car) achieving 95.1% accuracy. The model utilised a rich set of geometric, radiometric and structural features. In parallel, we explored an experimental instance-level segmentation pipeline based on Meta AI's Segment Anything Model (SAM), combining 2D projection with 3D reprojection to generate preliminary segmented outputs.

These results provide a foundation for Bologna's digital twin, enabling applications such as vegetation mapping, 3D reconstruction and infrastructure monitoring to support urban planning and sustainability initiatives.

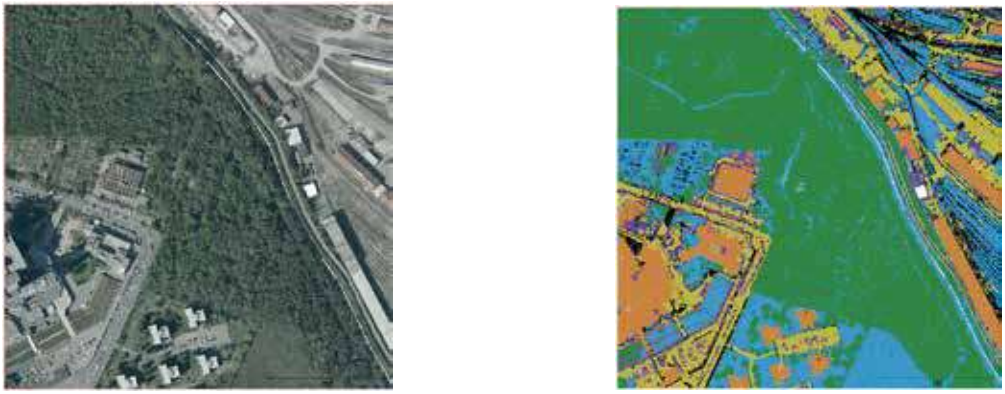


Figure 31: Example of Lidar point classification.

In particular, we focus here on tree segmentation – the process of identifying and isolating individual trees from remote sensing data, such as airborne LiDAR point clouds. This technique is fundamental for tree inventory, ecological monitoring, urban planning and biomass estimation, as it enables the extraction of key metrics like tree height, crown width and spatial distribution.

In our analysis, we explore and compare two popular methods for individual tree segmentation: watershed and region growing. Each method offers unique strength in handling different canopy structures and terrain complexities, which we assess in the

context of our dataset. The watershed method leverages a Canopy Height Model (CHM), derived by subtracting the Digital Terrain Model (DTM) from the vegetation-filtered Digital Surface Model (DSM). Local maxima within the CHM are used to detect treetops, from which watershed lines are propagated to delineate tree crowns. This process proved effective in densely vegetated urban environments.

The second method adopts a point-based, iterative technique where each high-vegetation point in the LiDAR dataset is assigned to a tree instance based on proximity and local geometric criteria. Starting from the highest point, the algorithm progressively assigns neighboring points to existing or new tree instances, considering canopy structure and spatial continuity.

This segmentation refines vegetation classification into grass, shrubs and trees, producing a detailed green cover map. Integration with Bologna's open dataset of approximaely 80.000 geolocated trees aims to enhance spatial precision, update attributes and map previously undocumented vegetation, supporting broader environmental analyses. Both methods produce segmented 3D tree instances from the classified LiDAR data.

Fig. 32: Example of tree segmentation using the watershed algorithm. The boundaries of individual tree crowns are highlighted in light green, treetops are in dark green and geopoints from Bologna Open Data are in orange.

Fig.33: Tree cloud points segmented.

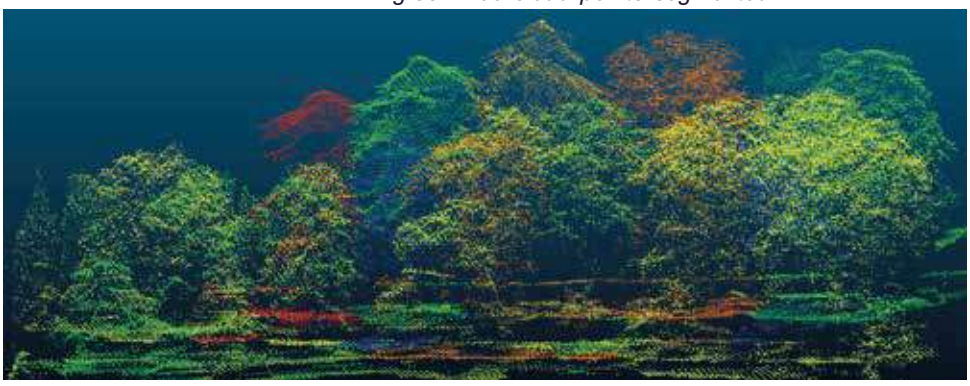


Fig 34: Example of three different cloud trees extracted.

METEO

From COSMO to ICON: Advancing weather modelling in Italy

Thomas Gastaldo, Virginia Poli
Agenzia nazionale per la meteorologia e climatologia "ItaliaMeteo"
Agenzia regionale per la prevenzione, l'ambiente e l'energia dell´Emilia-Romagna (Arpae)

Renata Pelosini, Gianluca Iannuzzi
Agenzia nazionale per la meteorologia e climatologia "ItaliaMeteo"

In the diverse landscape of Italian meteorology, Agenzia ItaliaMeteo plays a central role in coordinating and harmonising national numerical weather prediction (NWP) activities, as well as in carrying out related operational and research tasks. Specifically, the Agency is responsible for maintaining and developing an NWP system over Italy. This work is conducted in collaboration with Arpae Emilia-Romagna, which has operated a national NWP system on behalf of the National Civil Protection Department for many years and with Cineca, which provides the computational infrastructure and technical support necessary to run reliable and efficient advanced forecasting models.

As part of the COSMO consortium, Agenzia ItaliaMeteo contributes to a shared scientific framework that recently adopted ICON as the new reference model, replacing the long-standing COSMO system. This transition, the result of long-term research, international cooperation and careful technical planning, represents a significant advancement in Italy's operational forecasting capabilities.

ICON (ICOsahedral Nonhydrostatic), developed by the German Meteorological Service (DWD) and the Max Planck Institute, in collaboration with KIT, DKRZ, CSCS, the COSMO Consortium and the CLM Community, is a new generation model designed for both global and regional simulations. Its unstructured icosahedral grid and modular architecture make it particularly suitable for high-resolution applications over complex terrain such as the Italian Peninsula. Moreover, it is optimised for modern HPC environments. These features make ICON significantly more scalable and flexible than COSMO, which was a key reason for its adoption. Additionally, ICON provides more accurate forecasts due to its advanced physical parameterisations.

Following this transition, Agenzia ItaliaMeteo is now responsible for the national implementation of the ICON model in Italy. In collaboration with Arpae Emilia-Romagna, the Agency maintains an operational configuration known as ICON-2I. The model domain covers all of Italy and surrounding areas (Figure 1), with a horizontal resolution of 2.2 km. ICON-2I produces forecasts twice a day, initialised at 00 UTC and 12 UTC, extending up to 72 hours. These runs serve various purposes, including general weather forecasting, support for civil protection alerts and input to downstream environmental models.

To complement the main forecasts, a rapid update cycle configuration (ICON-2I-RUC) provides 24-hour forecasts every three hours, enabling close monitoring during high-impact weather events and offering more timely information to forecasters and emergency services. Additionally, an ensemble prediction system (ICON-2I-EPS), based on an ensemble of 20 members, is run daily to support both traditional forecasting and probabilistic guidance, helping assess forecast confidence and uncertainty. All these forecast runs rely on initial conditions generated through a continuous data assimilation system, which

incorporates conventional observations from surface stations, aircrafts, soundings and national radar data provided by the National Civil Protection Department.

The entire NWP chain runs operationally on the Galileo100 HPC system at Cineca, with

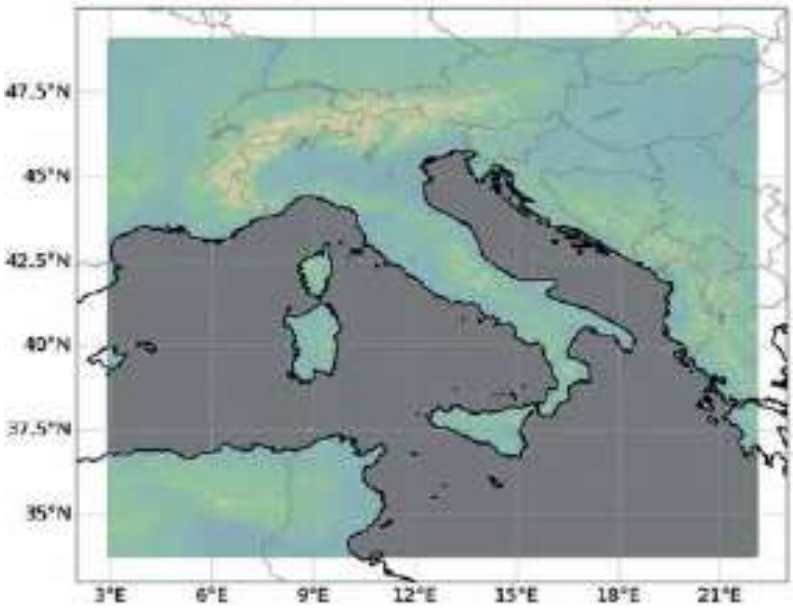


Figure 35: Model domain of ICON-2I

Leonardo supercomputer serving as a back-up infrastructure. The reliability and speed of execution provided by Cineca's HPC environment are essential to ensure that all forecasting chains are completed within strict operational timeframes, a fundamental requirement for operational NWP. Forecast outputs are made publicly available through the ItaliaMeteo website and the MeteoHub platform, jointly maintained by ItaliaMeteo and Cineca following its original development within the Mistral project.

Before its release, ICON-2I underwent an extensive testing phase across various periods and weather regimes. Comprehensive verification was carried out on several forecast parameters at the surface and different atmospheric levels, including precipitation, temperature, relative humidity and wind. As an illustrative example, Figure 2 shows the 24-hour accumulated precipitation on 18 September 2024 during a severe weather event that caused flooding in Emilia-Romagna. The panel on the left displays observations, while the central and right panels show forecasts from the COSMO-2I and ICON-2I models, respectively, both initialised at 00 UTC on the same day. In this case, ICON-2I captures both the intensity and spatial distribution of the precipitation much more accurately than COSMO-2I, highlighting the improved performance of the new forecasting system in high-impact weather situations.

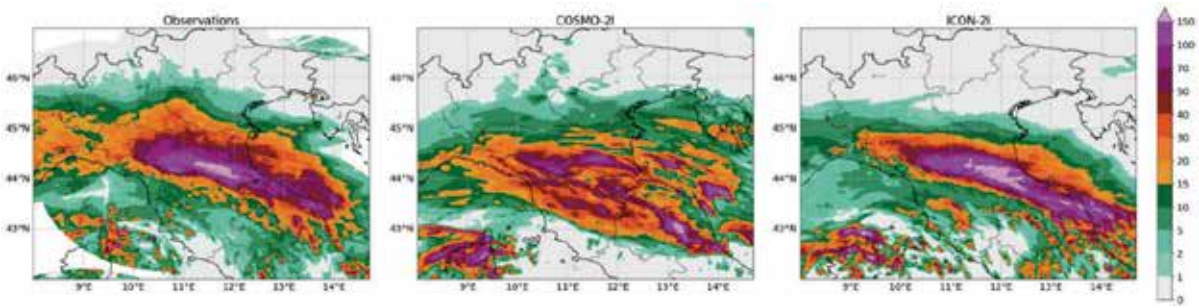


Figure 36: Observed daily accumulated precipitation on 18 September 2024 (left panel) and corresponding forecast of COSMO-2I (central panel) and ICON-2I (right panel.

Further evidence of ICON-2I's improved performance is provided in Figure 3, which summarises verification results for near-surface temperature, wind speed and dew-point temperature, based on nearly 300 forecasts covering different seasons. Across all variables, ICON consistently shows lower root mean square errors (RMSE) compared with COSMO, with improvements reaching up to 30%. In addition, a reduction in systematic bias is observed, particularly for dew point temperature, further confirming the enhanced accuracy of the new model across a variety of meteorological conditions.

Looking ahead, ICON's modular design and multi-scale capabilities provide a strong foundation for future developments, such as increasing the model's horizontal resolution, the frequency of runs and the assimilation of new types of observational data. Achieving these goals will rely heavily on the use of Cineca's cutting-edge HPC infrastructure and its extensive expertise. Furthermore, the active involvement of ItaliaMeteo, Arpae Emilia-Romagna and other Italian members in the COSMO Consortium will continue to promote international collaboration and innovation, ensuring that national advancements remain closely connected with the broader scientific community.

By combining the advanced capabilities of ICON, the computational power of Cineca and the institutional role of ItaliaMeteo – including the coordination of distributed expertise on NWP – Italy is well equipped to face the evolving challenges of weather forecasting, especially given the growing frequency and intensity of high-impact weather events.

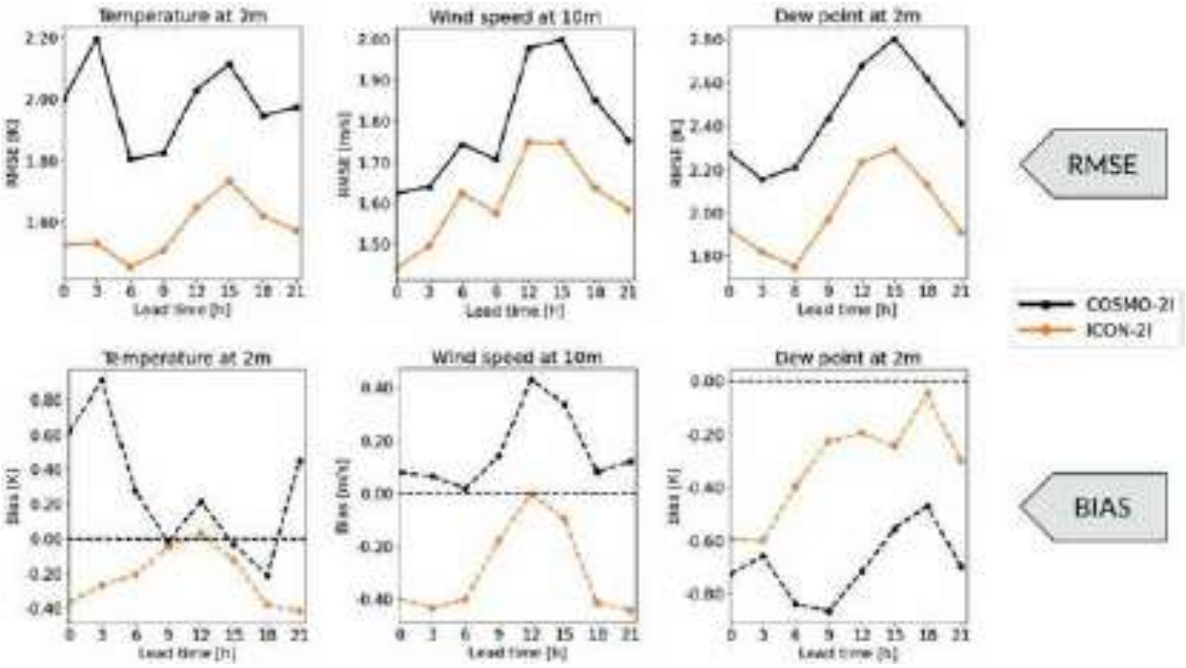


Figure 37: Verification of 2-metre temperature (left panel), 10-metre wind speed (central panel) and 2-metre dew-point temperature (right panel) in terms of root mean square error (top) and bias (bottom) for ICON-2I and COSMO-2I forecasts.

Cineca's path to exascale: GPU optimisation and portability for ECMWF models

Piero Lanucara, Gian Franco Marras, Fabio Di Sante, Massimo Gisonni, Giorgio Amati, Luca Ferraro, Nitin Shukla
Cineca
Thomas Geenen
ECMWF

The transition to exascale supercomputing requires a fundamental rethinking of scientific models, especially in fields such as weather and climate, where simulation codes need to scale efficiently on heterogeneous, GPU-accelerated architectures. Within the Destination Earth (DestinE) initiative, Cineca has contributed to this transition by analysing and optimising two critical components of ECMWF's modelling infrastructure: the ecWAM ocean wave model and the IFS-based RAPS system used for model profiling and development. Alongside performance engineering, Cineca also focuses on containerisation, a key enabler to simplify deployment and ensure the reproducibility of complex applications across diverse HPC platforms.

Profiling RAPS to guide GPU migration

RAPS, the reduced version of the ECMWF IFS model, is used for code development and performance studies. Using a variety of profiling tools and metrics, such as Score-P and Darshan, Cineca engineers identified key bottlenecks that impact scalability, including communication overheads in MPI routines and load imbalances in I/O workloads. This analysis enables more effective planning of the upcoming GPU porting of the full IFS model.

Porting ecWAM to GPU with OpenACC

Cineca also worked on the ecWAM ocean wave model, successfully porting it to GPU using OpenACC, targeting the core routines propag_wam and propags2. These routines are among the most computationally demanding and were carefully parallelised and offloaded. The result was a speed-up of more than five times compared with the CPU-only version, marking a significant milestone in preparing ECMWF components for deployment on EuroHPC systems like Leonardo. These improvements have been successfully added to the main branch of IFS.

Ensuring portability with containerised models

To support reproducibility and simplify deployment across heterogeneous systems, Cineca developed a containerised workflow for the RAPS and ecWAM models using Singularity/Apptainer. The same container images are used across multiple EuroHPC platforms, reducing configuration effort and ensuring consistency between environments. Benchmarking confirms that the containerised models perform on par with their native counterparts, validating this approach as both practical and efficient for scientific HPC

Cineca's role: From profiling to development

Cineca provided both technical profiling support and active development, contributing directly to the GPU acceleration of ecWAM. This dual role is crucial for enabling the long-term goal of running full Earth system models efficiently on European exascale infrastructures.

This work marks a key step in the broader roadmap of ECMWF's integration within the Destination Earth initiative, a major European project aimed at building a digital twin of the planet. The performance gains obtained so far are promising and lay the foundation for a full transition to next-generation computing platforms.

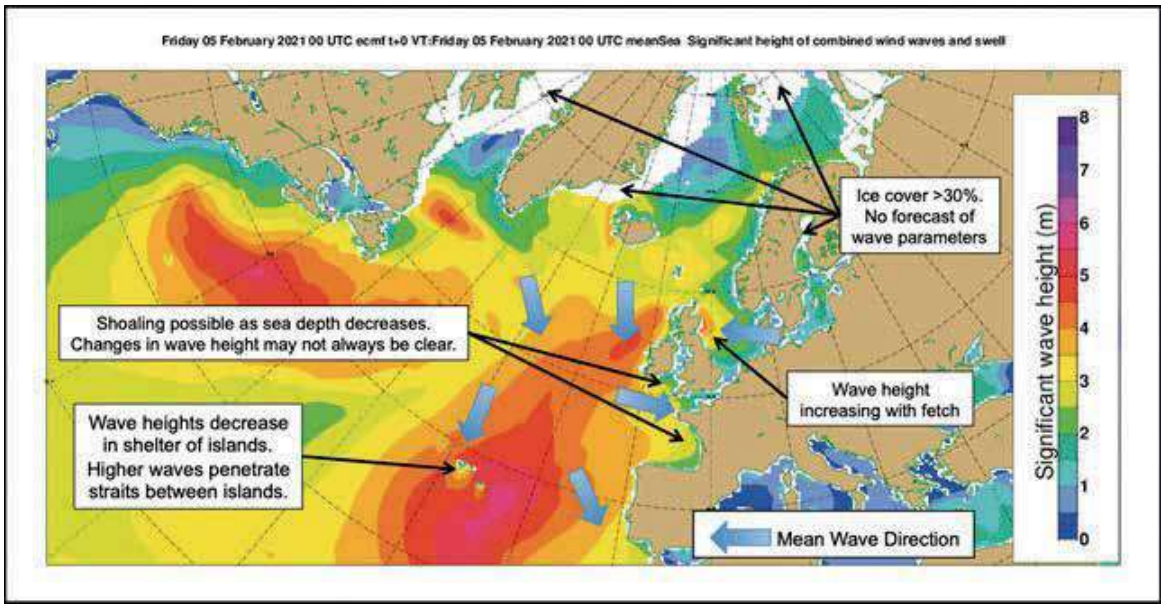


Figure 38: An example chart of wind waves and swell, displaying significant wave height forecast.

HPC FOR SCIENCE

BIOINFORMATICS AND LIFE SCIENCE

ELIXIR-ITA@CINECA collaboration

Silvia Gioiosa
Cineca

As the infrastructure partner of ELIXIR-ITA node, Cineca, reaffirms its commitment to innovation and excellence in computational life sciences, supporting bioinformatics researchers by providing access to advanced computational resources. While Cineca already offers various agreements for its consortium members, this dedicated framework is designed exclusively for life sciences researchers, addressing the unique computational challenges arising from biological and biomedical data.

Importantly, the agreement ensures not only access to high-performance computing but also a tailored support structure for scientists who may not have extensive experience with HPC environments – such as those in clinical and translational research fields – helping them leverage these resources effectively.

The recently renewed agreement will offer up to three million core hours and 300 TB of cumulative storage, allocated based on scientific merit. Each selected project will initially receive 50.000 core hours and 1 TB of storage on an HPC cluster, with the possibility of extension upon motivated request. Applications are accepted year- round on a first-come, first-served basis. Proposals undergo a peer-review process jointly managed by a technical and scientific committee to assess feasibility, ensuring that awarded projects receive allocated resources within a maximum of 14 days after submission.

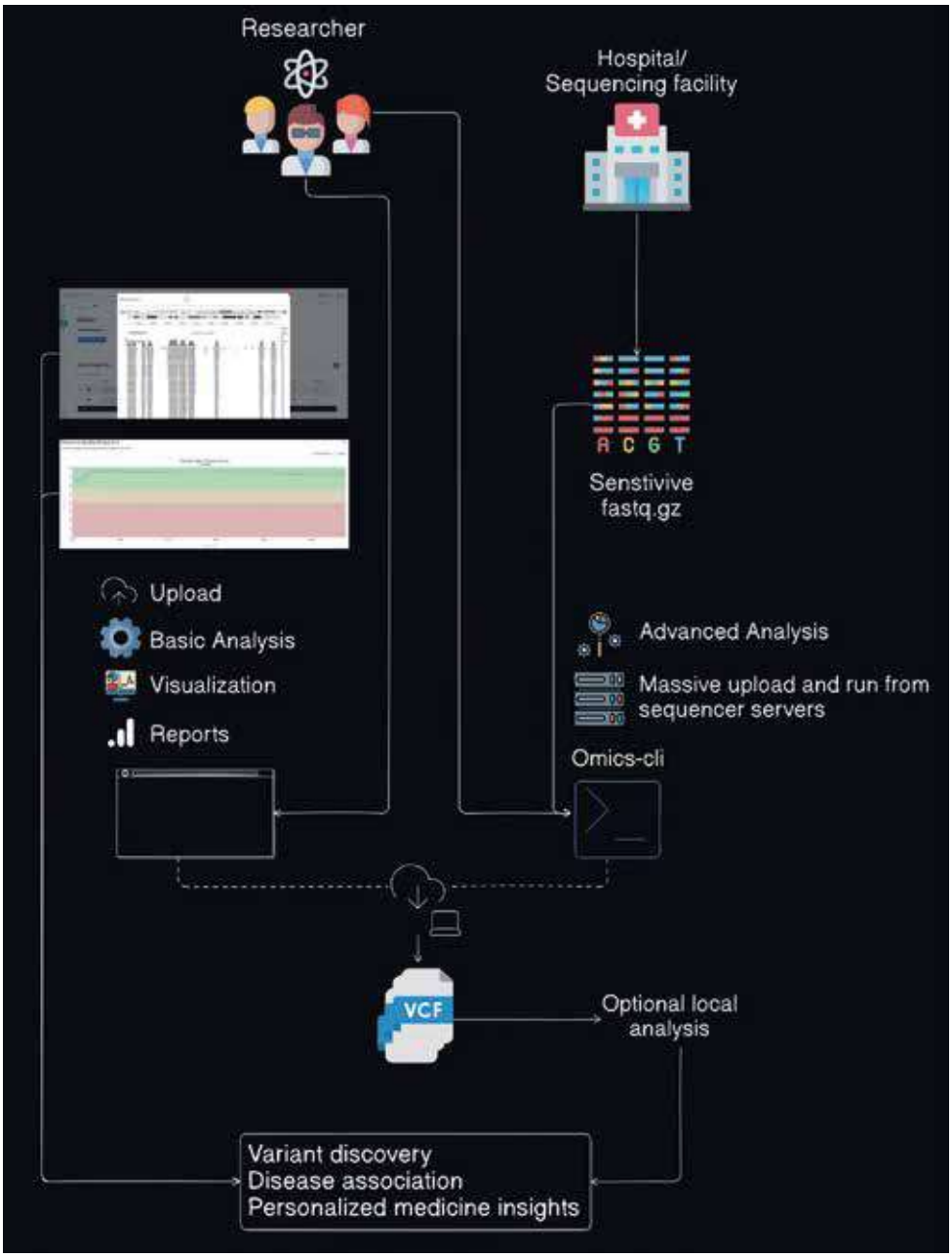
A significant milestone in this renewed agreement is the launch of a dedicated pilot initiative for the analysis of sensitive human NGS data, fully compliant with the 2018 EU GDPR regulations. This marks a major step forward in extending HPC services to projects involving protected genetic information – a crucial requirement for many clinical and precision medicine studies. To meet these needs, researchers will be granted access to encrypted virtual machines equipped with 96 vCPUs, 2 TB of disk space and a preconfigured software environment. These virtual machines will be managed by researchers themselves and will have a maximum lifespan of six months to ensure efficient project turnover in human genomics. The selection process for GDPR-compliant projects will take place twice a year, enabling a structured and secure allocation of resources.

With this initiative, Cineca and ELIXIR-ITA reinforce their role as key enablers of next-generation biomedical research, providing both specialised computational power and domain-specific support to a growing community that increasingly relies on large-scale data analysis. Principal investigators of awarded projects are encouraged to acknowledge Cineca and ELIXIR-ITA's support in scientific publications reporting results obtained using these resources. The continuation of this agreement represents a strategic step forward, further strengthening Cineca's integration into ELIXIR's digital infrastructure and enhancing its role in supporting the Italian and European bioinformatics ecosystems.

Omics platform: Advancing secure and scalable genomic analysis

Xhulio Dhori, Giuseppe Melfi, Juan Mata Naranjo, Silvia Gioiosa, Alessandro Grottesi, Elisa Rossi
Cineca

Genomic research and personalised medicine require efficient tools for handling whole genome sequencing (WGS) data. However, hospitals often face challenges related to storage and computing power. To address these issues, Cineca developed OMICS, a secure, high-performance genomic data analysis platform with both a web interface and a Command-Line Interface (CLI) for seamless data upload and processing. OMICS leverages NVIDIA Clara Parabricks, a GPU-accelerated suite that reduces whole genome



Figures 39, 40: Omics abstract

sequencing (WGS) analysis time by about 30 times, from approximately 30 hours to 45 minutes. To enhance accessibility and automation, OMICS features a command-line interface that interacts with the Omics API, facilitating data uploads and advanced job submissions. Users can choose from three containerised pipelines: Germline, Somatic and DeepGermline, each optimised for specific variant detection and annotation needs. To ensure data privacy and regulatory compliance, the platform is hosted on a secure, encrypted cloud virtual machine (VM) equipped with EU GDPR-compliant data protection protocols.

The infrastructure currently includes: 2 Cascade Lake GPUs for accelerated processing; 96 vCPUs and 2 NVIDIA V100 Tesla GPUs to handle large-scale computations; 336 GB RAM and 1 TB per-user storage for managing extensive genomic datasets. To further strengthen security, ISO 9001 and ISO/IEC 27001 certifications guarantee compliance with industry best practices for cloud computing and data management in life sciences. Future enhancements will incorporate additional authentication and encryption protocols to fortify data integrity and access control. The Omics Platform provides a robust back-end infrastructure designed with a microservices architecture. The platform integrates containerisation, distributed computing and stringent authentication protocols. Users can explore genomic variants through interactive, open-source visualisation tools such as IGV (Integrative Genomics Viewer). Additionally, OMICS supports downstream annotation and machine learning-driven predictive analytics, offering advanced insights into disease mechanisms and potential therapeutic targets.

OMICS is poised for continuous innovation, with a roadmap that includes expanding support for additional omics data types, including transcriptomics and epigenomics and integrating new AI-powered bioinformatics tools to enhance variant classification and enhancing interoperability with clinical systems, allowing seamless integration with hospital electronic health records (EHRs) and research databases. Future infrastructure upgrades will incorporate next-generation GPUs to further accelerate genomic analysis, enabling real-time insights for precision medicine applications.

Cineca's OMICS Platform is a cutting-edge, GDPR-compliant solution that empowers hospitals, researchers and clinicians with high-performance genomic data analysis capabilities. By combining advanced computational power, robust security measures and scalable infrastructure, OMICS is shaping the future of precision medicine and genomic research, driving faster and more accurate discoveries in healthcare and life sciences.

BIOENGENEERING

HPC-enabled in silico trials of new interventions to reduce the risk of hip fracture in frail elderly people

Antonino Amedeo La Mattina
Laboratorio di Tecnologia Medica, IRCCS Istituto Ortopedico Rizzoli
Giacomo Savelli, Marco Viceconti and Sara Oliviero
Dipartimento di Ingegneria Industriale, University of Bologna

One thousand years after Ali Ibn Sina's Canon of Medicine first described how to test the efficacy of a new medical treatment, the fundamental experimental process remains the same: a group of patients with similar characteristics is randomly divided into two groups – one receives the treatment being tested and the other a placebo. A primary outcome is used to determine whether the health of the treatment group is significantly better than that of the placebo group. This approach is reliable but also very expensive and, in some cases, ethically challenging. If we are able to build a patient-specific model that can accurately predict the primary outcome depending on the treatment, in principle, we could run a clinical trial in silico using a cohort of virtual patients to test the efficacy of a new intervention. Our team has developed and extensively validated a digital twin in healthcare (BBCT-Hip), a computer model that can predict the risk of hip fracture upon falling for a given person, based on their age, height, weight and detailed femoral morpho-densitometry as provided by a CT scan (Schileo et al., 2008; Bhattacharya et al., 2019). Using this modelling technology, we could run in silico trials on new interventions to reduce the risk of hip fracture after addressing some specific points: first, we developed a virtual cohort generation procedure (La Mattina et al., 2023). We then had to model the effect of a specific intervention; as a first example, we selected hip protectors, a class of medical devices aimed at reducing fracture risk. Lastly, we had to make the solution computationally efficient. The digital twin requires about 1.5 core hours for each simulated fall. For a 1,000 virtual patients, each requiring 28 falls per year over ten years, the workflow presents a computational cost of approximately 500,000 core-hours per trial arm. In the in silico trial, called BoneStrength, because the fall biomechanics model is stochastic in nature, we used a Markov-chain approach to reduce the computational cost to 200,000 core hours per arm (Oliviero et al., 2024). Once the BoneStrength simulator was operational, we validated it by reproducing the placebo arms of completed large-scale clinical trials in silico. The number of hip fractures observed in the trials fell within the range predicted by the stochastic model (Savelli et al., 2024). Once validated, we used BoneStrength to confirm that the key factor in the efficacy of hip protectors is patient compliance (Oliviero et al., 2024).

References

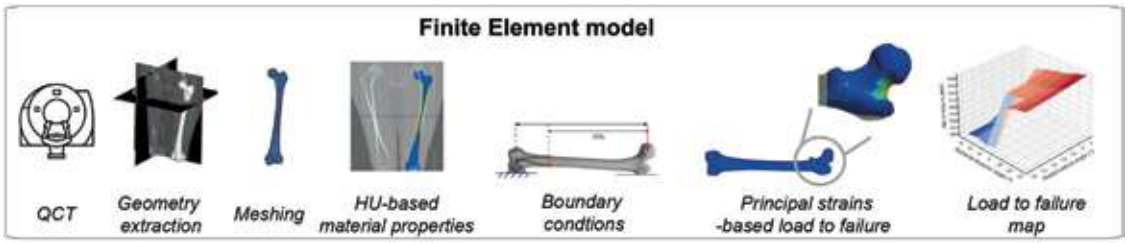
Bhattacharya, P., Altai, Z., Qasim, M., Viceconti, M., 2019. A multiscale model to predict current absolute risk of femoral fracture in a postmenopausal population. Biomech Model Mechanobiol 18, 301–318. https://doi.org/10.1007/s10237-018-1081-0

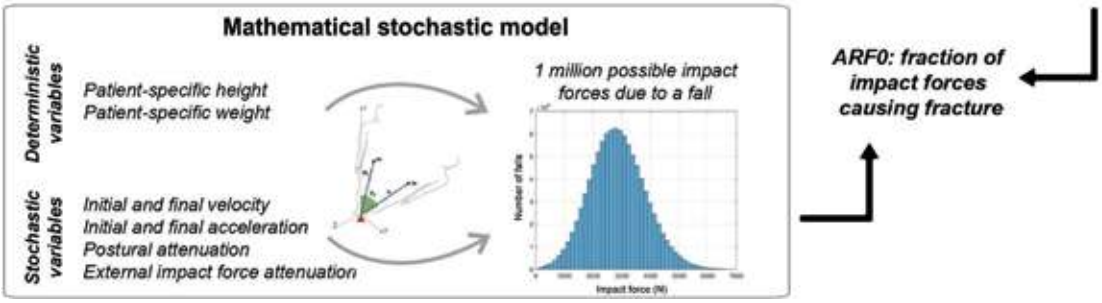
La Mattina, A.A., Baruffaldi, F., Taylor, M., Viceconti, M., 2023. Statistical Properties of a Virtual Cohort for In Silico Trials Generated with a Statistical Anatomy Atlas. Ann Biomed Eng 51, 117–124. https://doi.org/10.1007/s10439-022-03050-8

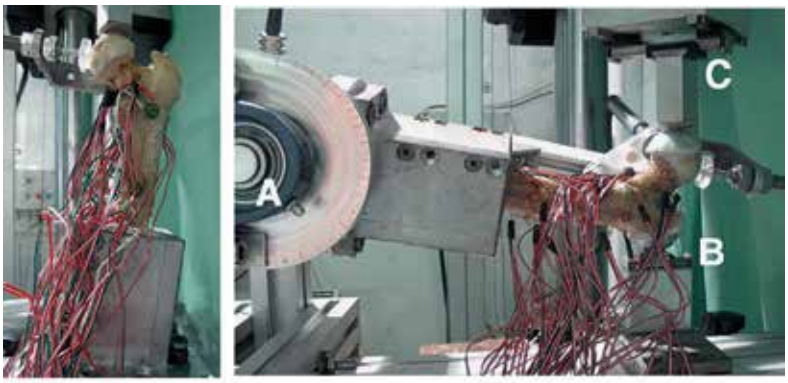
Oliviero, S., La Mattina, A.A., Savelli, G., Viceconti, M., 2024. In Silico clinical trial to predict the efficacy of hip protectors for preventing hip fractures. J Biomech 176, 112335. https://doi.org/10.1016/j.jbiomech.2024.112335

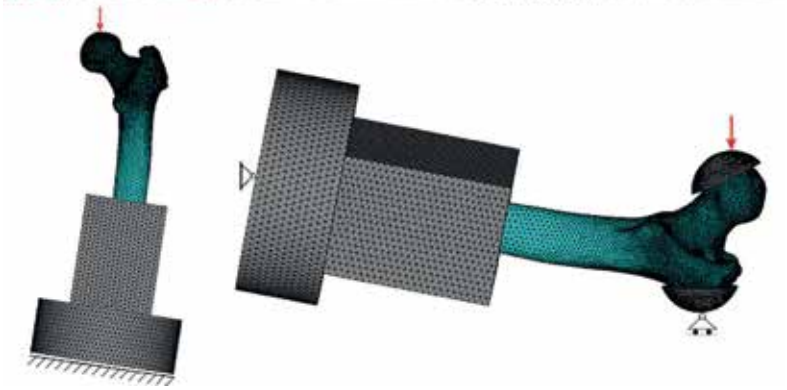
Savelli, G., Oliviero, S., La Mattina, A.A., Viceconti, M., 2024. In Silico Clinical Trial for Osteoporosis Treatments to Prevent Hip Fractures: Simulation of the Placebo Arm. Ann Biomed Eng. https://doi.org/10.1007/s10439-024-03636-4

Schileo, E., Dall'ara, E., Taddei, F., Malandrino, A., Schotkamp, T., Baleani, M., Viceconti, M., 2008. An accurate estimation of bone density improves the accuracy of subject-specific finite element models. J Biomech 41, 2483–2491. https://doi.org/10.1016/j.jbiomech.2008.05.017









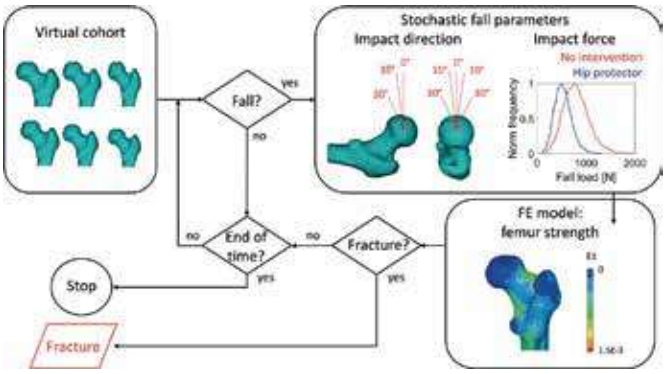


Figure 41. Logical structure of the BBCT-Hip Digital Twin in Healthcare.
Figure 42. Experimental validation of the Digital Twin using cadaver femurs.
Figure 43. Conceptual framework of the BoneStrength in silico trial.

Towards 3-pillar Cineca support for the Solid Earth community

Piero Lanucara, Michele Carpene, Davide Crisante, Laura Lampariello, Giuseppe Trotta
Cineca

In 2024, Cineca's support for Solid Earth activities was driven by an innovative and cross-disciplinary vision, designed to have a significant impact on both computational and non-computational efforts. This vision is structured around three key pillars:

1. Flagship Code – Supporting and improving a community code.
2. HPC + Cloud Infrastructure – Leveraging high-performance computing and cloud resources efficiently.
3. Simulation Data Lake – Generating valuable data for scientific research and community support (e.g. dissemination, publications).

All activities related to various projects and work packages (WPs) have been categorised within these three logical frameworks.

Flagship code

In 2024, the 11 flagship codes of ChEESE-2P underwent significant development and optimisation. As in previous years, Cineca played a leading role, managing WP2 of ChEESE-2P and supporting both scientific and production activities within the project. A notable milestone was the finalisation of Deliverable D2.2, which introduced the innovative concept of performance portability. This concept was applied to both the 11 flagship codes and the so-called mini-apps (computational tools closely related to the flagship codes but more lightweight and manageable).

This work resulted in quantitative measurements of performance portability, a key metric that, along with traditional performance indicators, will be crucial in the next phases of the project. Given the continuous evolution of codes, programming models and hardware architectures, these assessments are essential for meeting the project's computational objectives.

Once a flagship code reached a stable, efficient and multi-platform state, it was integrated into various computational workflows. One example is its application in curiosity-driven scientific research within the Geo-INQUIRE project. Through Transnational Access (TA) calls, flagship codes from ChEESE-2P were deployed on Cineca's high-performance computing infrastructure (Leonardo and G100) in collaborative scientific projects with academic partners. These calls focus on specific research topics and computing installations.

One remarkable example is the Probabilistic Volcanic Hazard Assessment, a collaborative effort involving INGV Bologna, CSIC (Spain) and Cineca. In this project, the FALL3D flagship code was successfully optimised for CPU execution on G100, enabling advanced simulations of volcanic hazards.

HPC + cloud infrastructure

Cineca's federated infrastructure, based on FENIX services (https://fenix-ri.eu/), has been configured and implemented to provide DT-GEO users with the necessary resources and technologies for executing scientific workflows within individual experiments. These workflows, treated as executable software, were installed, executed and monitored within the DT-GEO infrastructure provided by Cineca (as a FENIX partner) on the G100 and Leonardo clusters. This setup ensures users can efficiently utilise hardware resources, services and libraries available in the computing environment.

Moreover, the FENIX infrastructure allows seamless access to federated resources and services. Users can authenticate from any federated site using their existing credentials, eliminating the need for multiple registrations and ensuring a smooth user experience.

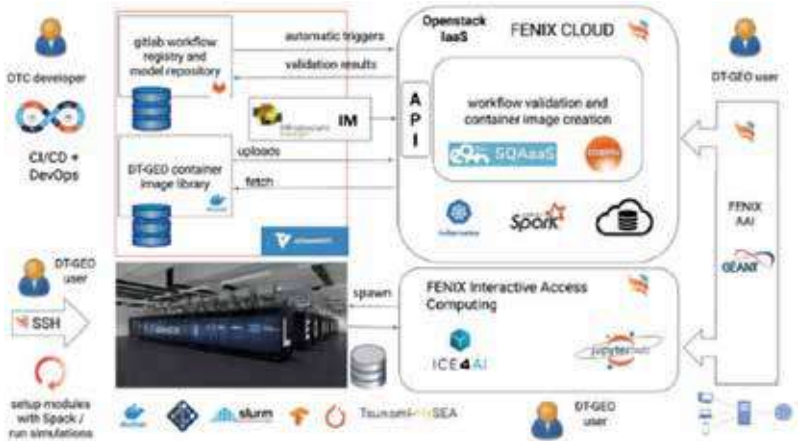


Figure 44: Cineca/FENIX Infrastructure at the Service of the DT-GEO Developer Community

The main active work packages (WP) have been WP5 (Volcanoes), WP6 (Tsunamis) and WP7 (Earthquakes). Individual users and researchers have been assisted in installing and running their applications on the machines by leveraging the HPC and Cloud infrastructure. For each DTC (Digital Twin Component), the relevant applications and main frameworks have been installed on the Cineca machines and cloud to execute the applications. Some examples include FALL3D, GALES, HySEA and SeisSol (with FALL3D, HySEA and SeisSol being flagship codes of ChEESE-2P).

Regarding the Volcanoes group (WP5), a specific activity was carried out to support the GPU porting of the GALES code for solving partial differential equations using finite element methods (FEM).

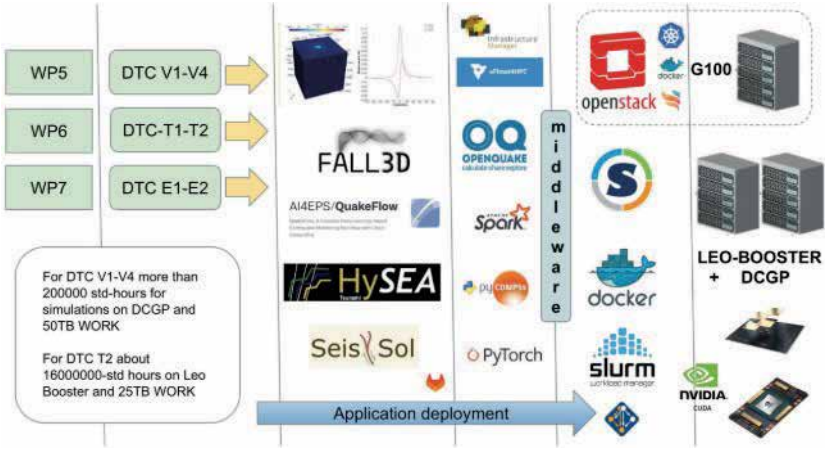
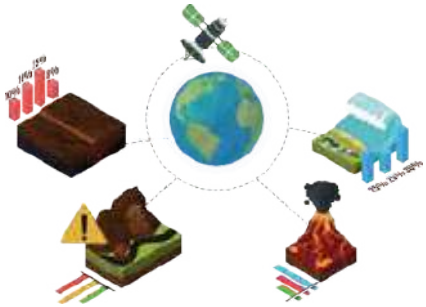


Figure 45: Deployment of DT-GEO Applications for various experiments and DTCs

In the near future, the infrastructure is expected to expand with new resources and further explore topics such as artificial intelligence and the use of deep-learning and machine-learning algorithms. These technologies will be used to analyse simulation-generated data and compare it with real-world data recorded by sensors in specific application contexts.

The simulation data lake - The "vision"

The "Simulation Data Lake" (SDL) is a scalable repository designed to store and manage large amounts of data produced by scientific simulations. It accepts data of various types and formats, including simulation inputs and outputs, as well as pre- or post-processed datasets and provides flexible access to the data, allowing researchers to retrieve it on demand according to specific scientific needs.



SDL's vision is to provide an infrastructure that ensures data storage and management in accordance with the FAIR principles, ensuring that data is easily 'Findable, Accessible, Interoperable and Reusable'. This approach promotes scientific collaboration and innovation, enabling users to integrate and analyse data from different sources efficiently.

The SDL promotes interoperability with other geospatial research platforms and tools through integration with external workflows via APIs. This functionality is essential for the creation of digital twins of geophysical events, permitting the connection of simulation big data with real-world observations. To this end, the link between the SDL and HPC becomes fundamental, allowing the management and analysis of large amounts of pre-/post-processed data and therefore leveraging the computing power of supercomputers to improve the accuracy of simulations and their integration with observational data.

The Simulation Data Lake (SDL) originated as a fundamental component within the Geo- INQUIRE project, aimed at offering a scalable cloud-storage solution for managing large volumes of geospatial data. Its main objective is to facilitate the access, analysis and visualisation of geospatial data, supporting applications in the fields of environmental monitoring, urban planning and disaster management.

Within the project and in synergy with related initiatives such as ChEESE-2P and DT-GEO, the SDL serves as a "virtual access" for all data generated by the TA (Transnational Access) activities, progressively archived here.

Technical Overview

The architectural design follows a microservices-based approach, ensuring modularity and flexibility. The back-end services are responsible for managing business logic, user authentication, data storage and external integrations.

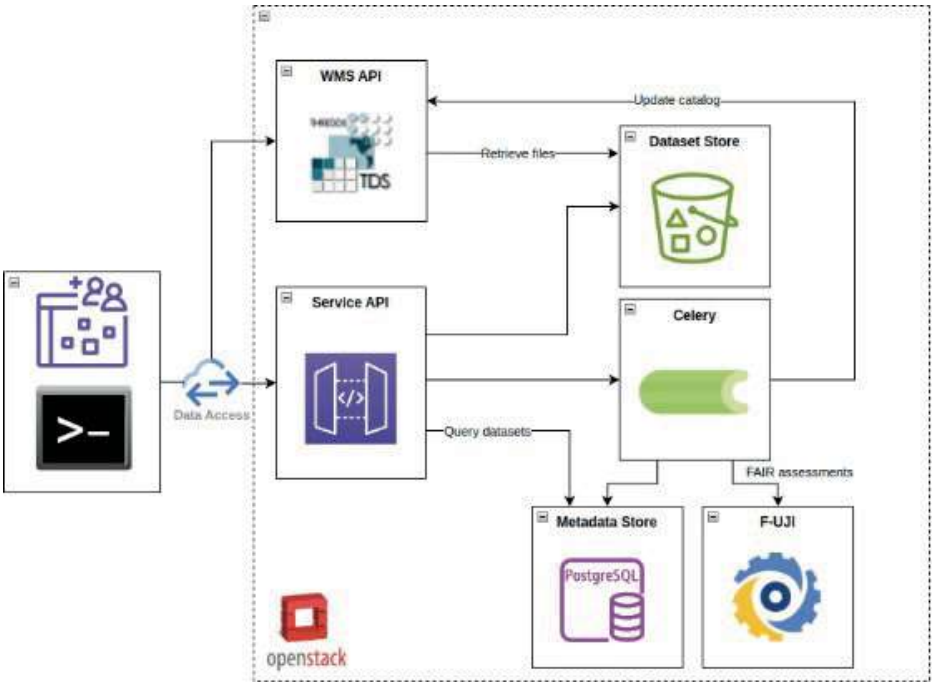
The system uses a centralised database and communication between components takes place primarily via RESTful APIs. All operations on files and metadata within the system are strictly regulated by the user's authentication status, ensuring consistent enforcement of access control policies.

The following figure presents a high-level view of the entire service and its macro-components.

2023 ChEESE-2P | Funded by the European Union. This work has received funding from the European High Performance Computing Joint Undertaking (JU) and Spain, Italy, Iceland, Germany, Norway, France, Finland and Croatia under grant agreement No 101093038

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement n° 101058129.

Geo-INQUIRE is funded by the European Commission under project number 101058518 within the HORIZON-INFRA-2021-SERV-01 call.



Below is a description of the functionalities of the Simulation Data Lake:

- Advanced search functionality, including searches by spatial and temporal coverage: This feature allows users to perform more targeted and precise searches, using spatial parameters (such as geographic coordinates) and temporal parameters (such as specific time intervals) to quickly find data that meet specific criteria.
- Publication of experiments: Users can manage access permissions for experiment data, with the option to make them public (open access), restrict access (limited access), or place the data under embargo (access deferred to a future date).
- Issuance and management of DOIs: The system supports the generation of DOIs (Digital Object Identifier) for experiments, ensuring efficient management and academic recognition of datasets through a persistent identifier.
- Integration with the EPOS Data Portal: The platform is integrated with the EPOS (European Plate Observing System) portal, enabling the sharing and access of scientific data related to geoscience.
- OGC Services: The system supports OGC (Open Geospatial Consortium) services for spatial data interoperability, allowing the upload and download of data and metadata in standardised formats such as RO-Crate, which are used for managing metadata in scientific contexts.
- Workflow management: The platform allows the definition of workflows associated with experiments, supporting the CWL (Common Workflow Language) format for process description and offering diagram visualisations for more intuitive management.
- KPI Dashboard: A dashboard enables the monitoring and evaluation of data performance through KPIs (Key Performance Indicators). The system automatically evaluates data compliance with the FAIR principles (Findable, Accessible, Interoperable, Reusable) using the F-UJI tool.
- Data proximity and integration with HPC systems: The data in the Simulation Data Lake is integrated with the high-performance computing (HPC) systems of Cineca, enhancing the efficiency and speed of computational analyses thanks to the physical proximity of the data.
- High-performance storage system: The data is stored on a infrastructure based on state-of-the-art SSDs (Solid State Drives) that guarantee high performance in terms of access speed and reliability in data management.

MATERIAL SCIENCE

Supporting material science codes towards exascale

Mariella Ippolito
Cineca

Cineca is a partner of the MAX HPC Centres of Excellence (CoE), currently in its third stage. For many years, MAX has played a crucial role in advancing the research and development of high-performance materials through the use of Density Functional Theory codes (i.e. QuantumESPRESSO, YAMBO, SIESTA, FLEUR and BigDFT). Leveraging the powerful computational resources of exascale, the centre aims to accelerate the discovery and optimisation of materials for a wide range of applications, from energy technologies and advanced electronics to biological and bio-mimetic materials.

The overall objective of the MAX Centre of Excellence is to provide the European materials simulation community – both developers and end users – with the capabilities needed to make the best use of computational resources to address scientific challenges hitherto considered prohibitive. To achieve this, in addition to running individual MAX flagship codes on thousands of accelerated nodes, MAX also aims to enable these codes to work cooperatively within tightly bound exascale workflows, which have so far been considered simply unfeasible. Cineca takes care of the continuous assessment and analysis of the parallel performance of the MAX flagship codes, by improving the uniformity and consistency of the data produced for each code and establishing common metrics for performance evaluation.

For the benchmarking activity we adopted JUBE, which is a tool developed at Jülich Supercomputing Centre that provides a script-based framework to easily create benchmark sets, run them on different computer systems and evaluate the results. For each benchmark application, the benchmark data is written in a certain format that allows JUBE to extract the desired information. We organised the MAX benchmark repository so that the data can be easily accessed by users and developers. We also developed a tool allowing automatic plotting of the data available in the repository. All MAX codes have been ported and benchmarked on Leonardo. Moreover, all of them support other EuroHPC architectures; in some cases, they have also been deployed on those systems and their benchmarks are available.

MAX released its first Performance and Scalability brochure on December 2024, gathering the latest results of the MAX lighthouse codes demonstrating outstanding performance and scalability on various EuroHPC architectures.

One other objective of the MAX CoE is to ensure that data produced and stored within the project complies with the FAIR principles (making data Findable, Accessible, Interoperable and Reusable).

To this end, it was agreed that EU HPC centres would host mirrors of the Materials Cloud Archive, one of the two official repositories for materials data recommended by the EU Commission in Open Research Europe*, which is currently hosted at the Swiss National Supercomputing Centre (CSCS) in Lugano.

MaX - Materials design at the Exascale has received funding from the European High Performance Computing Joint Undertaking and Participating Countries in Project (Czechia, France, Germany, Italy, Slovenia and Spain) under grant agreement no. 101093374.

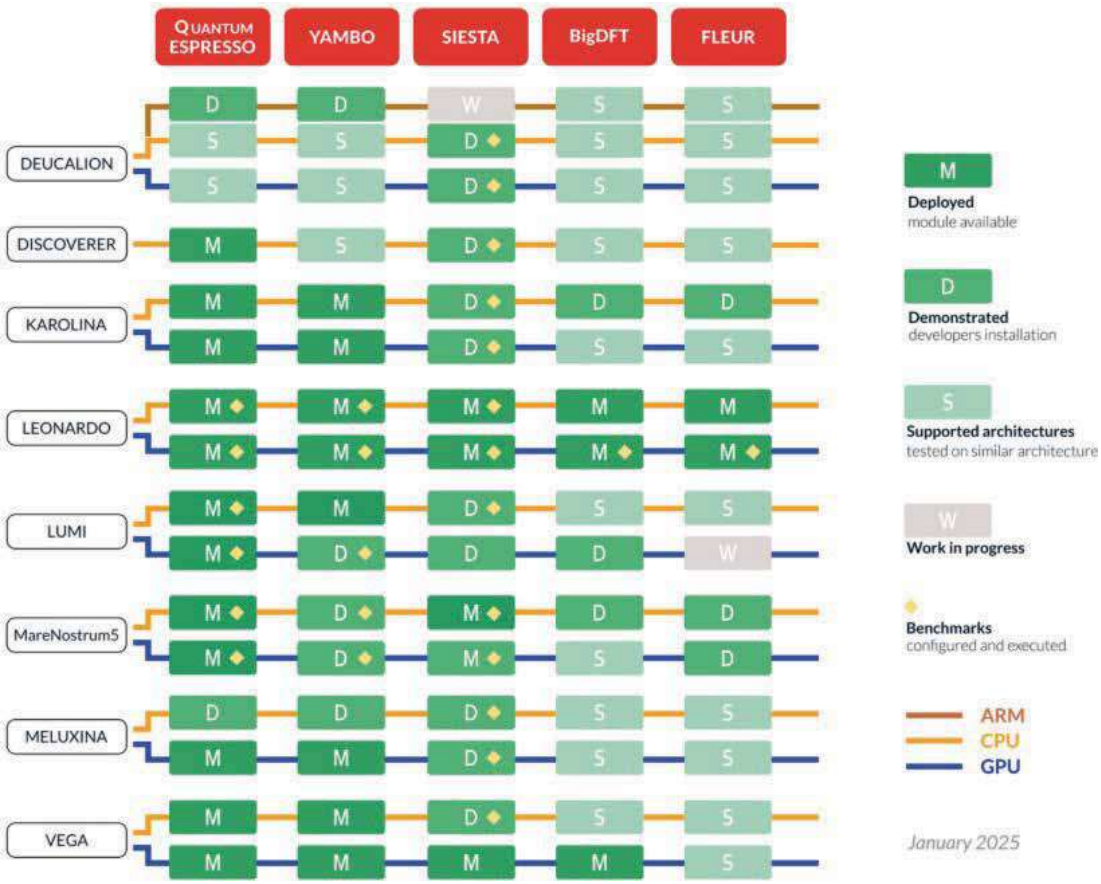


Figure 48: MaX codes deployment status on EuroHPC architectures

Thanks to collaboration between Cineca and CSCS staff, a nightly back-up of the Materials Cloud Archive has been set up from the CSCS servers to ADA Cloud, the Cineca HPC cloud infrastructure. The back-up data is currently available on the mirror web site**, where it is already freely accessible and downloadable. Such a deployment can also act as a failover for the main Materials Cloud Archive, which can redirect its traffic towards the Cineca mirror if the main CSCS repository is unreachable.

* https://www.fz-juelich.de/en/ias/jsc/services/user-support/software-tools/jube?expand=translations,fzjsettings,nearest-institut

** https://open-research-europe.ec.europa.eu/for-authors/data-guidelines#selectarepository

Interfacing QE to the MiMiC multiscale modelling framework

Angela Acocella
Cineca

Within a collaboration with the Technical University of Denmark (DTU), the École Polytechnique Fédérale de Lausanne (EPFL) and Forschungszentrum Jülich, we are working on interfacing QuantumESPRESSO (QE) to MiMiC, an integrative, high-performance framework for multiscale modelling. MiMiC is designed to efficiently perform Quantum Mechanics/Molecular Mechanics (QM/MM) simulations, by loosely coupling different programmes, each one specialised in calculating molecular properties for individual subsystems at different levels of theory and resolution.

The framework, based on a client/server approach and a Multiple Programme Multiple Data (MPMD) paradigm, enables data transfer while concurrently calculating interactions between subsystems. Data exchange is facilitated through a message-passing, interface-based (MPI) mode of communication, implemented in the MiMiC Communication Library (MCL), specifically developed to link external programmes without altering their underlying parallelisation structures. By employing a command-based approach, MCL allows MiMiC to exploit the efficiency of different codes, preserving their individual optimised performance on HPC architectures.

The currently released version of MiMiC interfaces CPMD, which works as both the MD driver and QM engine, but also as a server for communication, with GROMACS, acting as the MM client code. Ongoing developments are focused on a major refactoring of the MiMiC framework aimed at increasing its flexibility to integrate additional client programmes, supporting large-scale QM/MM molecular dynamics (MD) simulations on (pre-) exascale supercomputers, leveraging the functionalities of different high-performance codes. This update includes coupling MiMiC with GPU-enabled QM client programmes, such as QE, to enhance its computational efficiency and scalability.

Due to the high modularity of QE, a few modifications allow the code to act as a QM client, performing QM/MM simulations under the electrostatic embedding scheme developed by Laio et al. The developed MiMiC interface in QE mainly consists of newly added routines across two distinct layers of the code. In the Modules directory, MCL wrapper routines allow for initialising the MiMiC communicator and managing data exchange. A communicator module handles communication with the server through type-bound procedures. Within the Within the PWscf package, a newly added routine facilitates client operations, ensuring data exchange with the MD driver during the MiMiC multiscale simulation workflow.

The effect of the external electrostatic potential generated by the MM point charges on the QM wave function has also been included by properly incorporating it into the QE routine responsible for computing local potentials. Validation tests on the interface developed demonstrate that the QE-MiMiC interface is stable across multiple intra- and inter-node MPI processes, showing numerical consistency with trajectories obtained implementing other QM clients. Currently, we are focusing on testing and optimising simulation performance on the Leonardo supercomputer, comparing CPU and GPU implementations. The next goal will be to integrate the developed MiMiC interface within the official release of QE.

Automated tribology workflows for real materials

Tommaso Gorni
Cineca

Cineca participates in the development of the TribChem code, developed by Prof. Clelia Righi's group (University of Bologna)

TribChem is a workflow manager for the ab initio calculation of the tribological properties of materials such as adhesion, shear strength and charge redistribution. The typical workflow automatically retrieves the crystal structures from an online database (materialsproject.org), manages the GPU-accelerated atomistic simulations via dedicated python packages and stores the results on a mongoDB server.

Starting from late 2023, the TribChem group asked initially for assistance in the deployment of TribChem on the Leonardo+ADACloud infrastructure (with ADA hosting the mongodb server) and later for Epicure support aimed to optimise some Python bottlenecks in the data processing.

These first two interventions quickly evolved into a stable collaboration between the TribChem developers and Cineca's high-level support team (HLST), which is now actively contributing to the TribChem code base. Assistance has been provided both at the DevOps (CI pipelines, documentation, repository management) and HPC level (benchmarking, profiling and optimisation), resulting in two new code releases (v0.1.0 and v1.0.0) during 2024. Cineca HLST is also involved in the development decisions regarding the technical infrastructure of the code, such as the choice of libraries and dependencies.

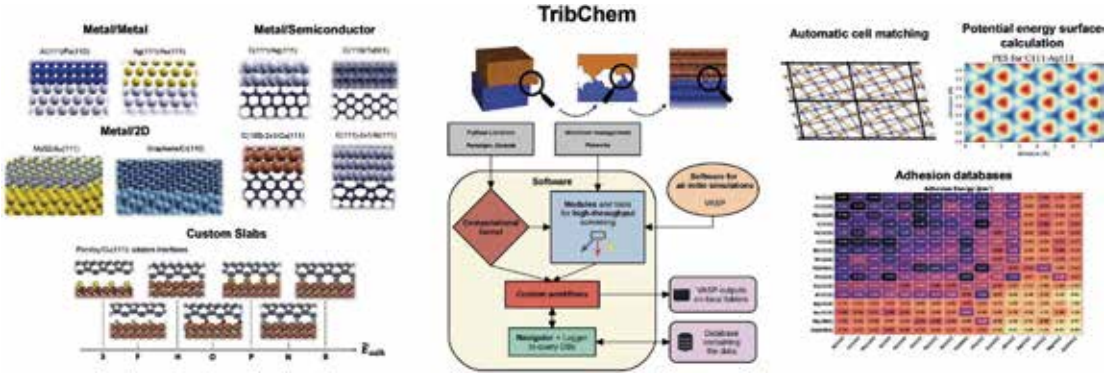


Figure 49: Tribchem structure and some of its applications (Refs.: J. Chem. Theory Comput 19. 5231; ACS Appl. Mater. Interfaces 15, 19624; Carbon 230, 119555)

ASTROPHYSICS

Redesigning astrophysical codes and advancing the frontiers of exascale computing

Nitin Shukla

Cineca

The European High-Performance Computing Joint Undertaking (EuroHPC JU) has played a pivotal role in developing world-class supercomputers and pre-exascale systems across European countries over the past decade, such as Leonardo, Lumi and MareNostrum 5. It will soon launch Europe's first exascale system named Jupiter, currently being installed in Jülich. These state-of-the-art exascale machines can perform 10^18 (1 followed by 18 zeros) calculations per second, harnessing the full power of accelerators like Graphics Processing Units (GPUs), Field-Programmable Gate Arrays (FPGAs) and more. The EU is supporting several Centres of Excellence (CoEs) across various research and development fields to ensure that scientific and industrial applications fully benefit from exascale computing capabilities.

The SPACE CoE is dedicated to optimising key computational astrophysical and cosmological codes for the demands of exascale computing. By integrating innovative programming methodologies and software solutions, SPACE empowers scientists to investigate intricate astrophysical phenomena. A core aspect of its mission is fostering collaboration between researchers, software engineers, HPC specialists, hardware manufacturers and developers, ensuring the efficient adaptation of astrophysical and cosmological applications to both exascale and post-exascale computing environments.

In addition, SPACE is committed to high-performance data analysis, addressing the immense data volumes generated by both astrophysical observations and exascale simulations. By employing machine learning and advanced visualisation tools, it enhances the interpretation and management of complex datasets. Scientific visualisation tools play a crucial role in analysing extensive simulation and observational data, helping to overcome challenges related to data accessibility, storage and processing on high-performance computing platforms. A fundamental goal of SPACE is to develop these tools and methodologies, facilitating deeper exploration and analysis of vast and intricate datasets (see first figure, credit to Guillermo Marin).

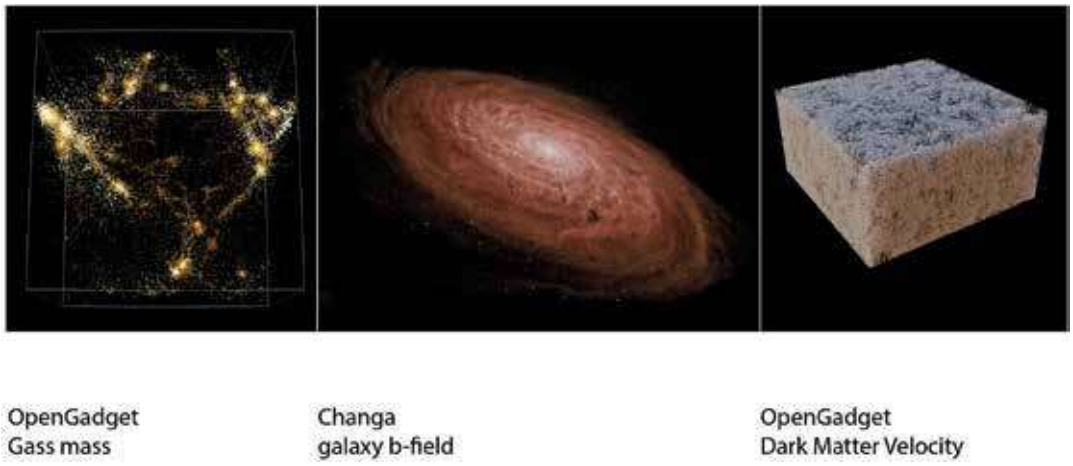


Figure 50. From visible to invisible: The left and middle images showcase stunning galaxies, while the right reveals the hidden structure of dark matter. (Credit: BSC, IT4I)

The SPACE CoE is now entering its third year. Since its inception, we have worked closely with code developers to achieve the project's objectives.

One key achievement has been our collaboration with researchers at University of Torino (the project coordinator) to modernise their SPACE application, gPluto. We helped rewrite the code in modern C++ with a GPU-friendly approach and accelerated it using OpenACC. As a result, gPluto can now efficiently scale across thousands of GPUs.

The accelerated version of gPluto has, for the first time, enabled highly detailed simulations of the tearing instability (see figure) a plasma instability believed to occur in various space and astrophysical environments, such as solar flares, black hole magnetospheres, pulsar magnetospheres, fast radio bursts and blazar jet flares. These simulations, which bridge spatial and temporal scales previously unattainable without GPU acceleration, allowed researchers at University of Torino to observe the onset of secondary instabilities at smaller scales relative to the tearing instability. These secondary instabilities could play a crucial role in explaining particle acceleration in the aforementioned astrophysical scenarios, an open question in astrophysics.

For studies of this kind, GPU acceleration is essential, dramatically reducing computation times and making large-scale simulations feasible.

Additionally, we collaborated with researchers from Istituto Nazionale di Astrofisica and Ludwig-Maximilians-Universität Munich to optimise their SPACE code, OpenGadget. Thanks to these efforts, OpenGadget can now efficiently utilise over 10.000 GPUs, significantly enhancing its performance.

Within SPACE, we offered training opportunities open to the entire science community. In July, we organised the hybrid workshop Maximising Performance in High-Performance Computing – Tools, Techniques and Hands-On Sessions. The event featured special sessions with NVIDIA and DDN scientists, providing valuable insights into the latest advancements in GPUs and storage technologies. The workshop was a great success, bringing together over 20 researchers from various CoEs at Cineca headquarters.

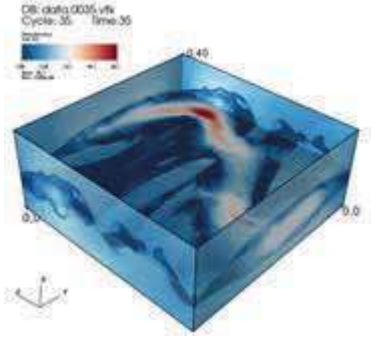


Figure 51: High-resolution reconnection simulations (768×384×384 grid) require approximately 13.392 minutes (~9 days, 7 hours) on OCCAM, whereas GPUs complete the task in just 742 minutes, demonstrating a significant speed-up

Scalable Parallel Astrophysical Codes for Exascale - SPACE CoE
Funded by the European Union. This work has received funding from the European High Performance Computing Joint Undertaking (JU) and Belgium, Czech Republic, France, Germany, Greece, Italy, Norway and Spain under grant agreement No 101093441.

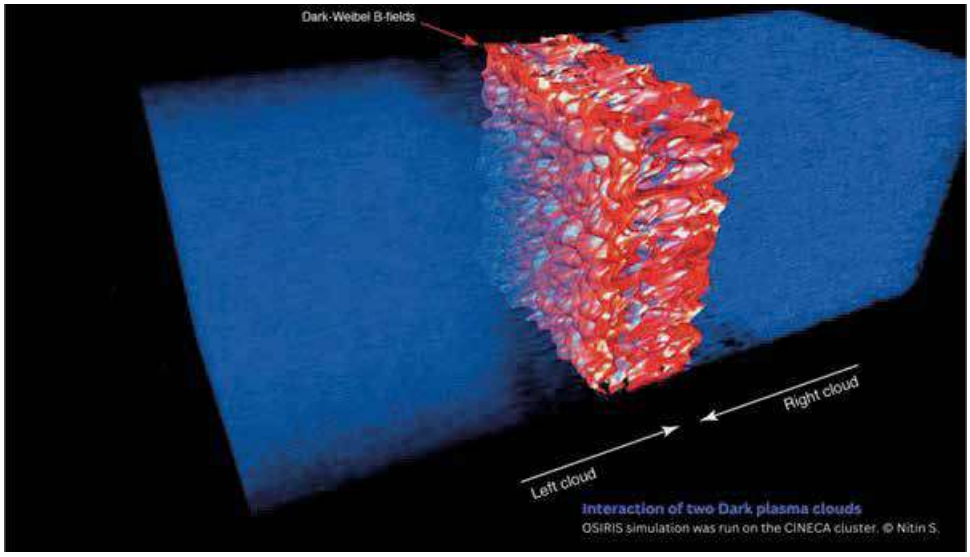
Strong bounds on dark-matter interactions established with simulations at Cineca

Luís O. Silva
Distinguished Professor, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

A team of scientists from Ruhr University Bochum, in Germany, Cineca in Bologna (Italy) and Instituto Superior Técnico in Lisbon (Portugal), recently established stringent constraints on the electromagnetic equivalent of dark-matter self-interaction, also known as dark electromagnetism, based on results from large-scale kinetic simulations performed at Cineca, with additional support from EuroHPC JU.

Hailed as one of the central questions in fundamental physics, the nature of dark matter has eluded scientists for decades. It is known that dark matter only interacts with regular matter via gravitational force. Scientists have hypothesised that dark matter could interact with itself, with the simplest interaction being dark matter's equivalent of electromagnetism.

In a recent work published in Physical Review D (Letters), highlighted as an Editor's Suggestion and in Physics, the online magazine of the American Physical Society, the team demonstrated that the most straightforward form of dark electromagnetism is highly unlikely to be present in dark-matter self-interactions. The scientists used kinetic plasma simulations with the state-of-the-art particle-in-cell code OSIRIS, which, with appropriate rescaling, can model dark electromagnetism and dark-plasma-like dynamics (see figure).



They compared the predicted slowing-down of dark-matter clouds from simulations performed at Cineca with astronomical observations, concluding that, if dark electromagnetism were present, the dark matter would slow down much more significantly than observed. This allowed the scientists to restrict the range of possible dark electromagnetic parameters (such as charge and mass) to a very narrow region of parameter space. Further studies are expected to narrow this region, potentially allowing us to fully discard either the simplest or additional variants of dark electromagnetism as a viable mechanism for dark-matter self-interaction.

QUANTUM COMPUTING

Two new quantum computers for the DAMA Technopole: IQM Radiance and Pasqal

Daniele Ottaviani
Cineca

The DAMA technopole is set to make a major leap in scientific research and high-performance computing: by the end of the year, it will install IQM Radiance, a 54-qubit quantum computer based on superconducting technology. By the middle of next year, a Pasqal system featuring 140 qubits based on neutral atoms will follow. These two machines will be the most powerful quantum computers ever installed in Italy and will be integrated with Leonardo, Cineca's supercomputer.

IQM Radiance: Superconductors for quantum computing

IQM Radiance is a quantum computer developed by the Finnish company IQM Quantum Computers, a European leader in the field. The system uses superconducting qubits, a well-established technology that employs low-temperature electrical circuits to achieve highly controllable quantum states.

With its 54 qubits, IQM Radiance is optimised for combinatorial optimisation, quantum simulations and machine-learning applications. Its architecture is designed for seamless integration with HPC systems, enabling the exploration of hybrid quantum-classical algorithms with Leonardo.

Pasqal: Neutral atoms for quantum computation

Pasqal, on the other hand, brings a completely different technology: its 140-qubit quantum computer leverages neutral atoms manipulated by laser beams to create highly scalable and reconfigurable qubit networks. This approach, inspired by Feynman's quantum simulators, is particularly well suited to quantum material simulations, quantum dynamics problems and combinatorial optimisation.

The installation of the Pasqal system at the DAMA technopole marks a significant advancement for quantum research in Italy, allowing scientists to explore interacting qubit networks, which are difficult to achieve with other technologies.

Towards a European quantum ecosystem

The integration of IQM Radiance and Pasqal with the Leonardo supercomputer is a crucial step in building an advanced quantum ecosystem in Europe. These systems will provide researchers and enterprises with a unique infrastructure to develop and test quantum algorithms, accelerating progress towards practical quantum advantage.

With these installations, the DAMA technopole positions itself as a centre of excellence for quantum computing, playing a key role in strengthening Europe's autonomy and competitiveness in the field.

Quantum Suite: Empowering quantum research on HPC

Vito Palmisano
Cineca

As we navigate the Noisy Intermediate-Scale Quantum (NISQ) era, having an up-to-date Quantum Suite is vital for advancing research and to access reliable and high-performance quantum circuit emulators. Quantum hardware is still limited, expensive and not always accessible, making emulators indispensable for testing, benchmarking and developing quantum algorithms. The Quantum Suite aims to address this need by providing a robust, ready-to-use environment optimised for HPC platforms like Leonardo, where setting up such tools can be complex and time-consuming. The goal is to streamline workflows, enabling researchers to focus on science rather than set-up.

What are the changes inside the quantum suite?

In recent months, we carefully selected a set of libraries to update inside the Quantum Suite, choosing the ones designed to maximise the potential of HPC systems.

- cuQuantum SDK, (including cuStateVec and cuTensorNet) enables GPU-accelerated simulations with high performance and scalability. it has been updated from version 23.06.1.8 to 24.11.0.21.
- cuTensor: Enhances tensor operations, which are critical in quantum simulations and tensor network algorithms. It was updated from version 1.5.0.3 to 2.0.2.5.
- Cirq: Allows the emulation of quantum circuits leveraging GPU and multi-threading capabilities. The version has been updated from version 1.2.0 to 1.4.1.
- PennyLane: (Chosen for its versatility in hybrid quantum-classical workflows) Supports multiple back-ends, integrates seamlessly with classical ML libraries, and leverages GPU and multi-threading capabilities. it has been updated from version 0.32.0 to 0.40.0.

Together, these libraries provide flexibility and power, allowing researchers to efficiently simulate complex quantum systems.

Work in progress

We are currently updating or installing the following libraries:

- Qiskit – IBM's flagship framework with a strong community and support for both emulation and hardware execution.
- QuEST – a C-based, MPI-compatible library for large-scale state vector simulation.
- QMatchaTEA – a tensor network-based emulator tailored for simulating highly entangled quantum systems efficiently on multi-node HPC set-ups.

Cineca and the quantum leap: Powering the future with QEC4QEA

Anita Camillini
Cineca

As quantum computing steadily moves from promise to practice and with new systems taking shape at DAMA technopole, Cineca is working to ensure these technologies become powerful and accessible tools for researchers and innovators across Europe. This is the aim of the Quantum Excellence Centre for Quantum-Enhanced Applications (QEC4QEA), a major initiative funded by the EuroHPC JU, placing Cineca at the core of Europe's emerging quantum ecosystem.

Bringing together 18 leading institutions from 10 countries, QEC4QEA is building a pan-European one-stop shop for quantum-enhanced applications. The project is designed to support users from academia and industry in developing, accessing and running applications that combine quantum and classical computing – with seamless integration into HPC systems such as Leonardo.

Cineca leads the Italian cluster of QEC4QEA and is working alongside national partners E4 Computer Engineering, LINKS Foundation and the University of Padova. As the hosting entity for EuroQCS-Italy, Cineca provides the infrastructure for real quantum workloads on systems like IQM Radiance and Pasqal and also the technical expertise to integrate these platforms into hybrid HPC-QC workflows. This includes user-friendly software interfaces, API development and support for quantum-classical co-processing.

Beyond infrastructure, Cineca plays a key role in developing quantum-ready applications across a wide range of domains. These include drug discovery and molecular simulations, using hybrid quantum-classical optimisation techniques; scheduling and combinatorial optimisation, explored through quantum annealers and neutral atom platforms; post-quantum cryptography and integer factorisation, contributing to the future of secure digital infrastructure; simulation of partial-differential equations, with applications in physics, finance and engineering. All applications within QEC4QEA are designed to be hardware-agnostic, meaning they can run on different quantum architectures or emulators. Cineca contributes to a shared European repository of these applications, together with associated compilers, execution APIs and training resources. The project also emphasises long-term sustainability: Cineca supports the benchmarking and validation of quantum software, the development of robust middleware tools and the federated access to quantum resources across Europe. In doing so, it ensures the QEC4QEA platform will remain adaptable, scalable and useful for years to come.

By participating in QEC4QEA, Cineca is not only helping users access quantum technologies – it is actively shaping how these technologies will be used, fostering a new generation of quantum-enhanced, high-performance applications. Combined with the deployment of advanced quantum hardware at DAMA, QEC4QEA marks a decisive step in positioning Italy and Europe at the forefront of the quantum revolution.

HUMANITIES AND SOCIAL SCIENCES, DIGITAL ECOSYSTEMS

Enabling a data space for Italian culture

Antonella Guidazzoli, Silvano Imboden, Davide Giosuè Lippolis, Luca Mattei, Donatella Sforzini, Rossella Pansini, Fauzia Albertin, Giorgia Cardano, Simona Caraceni, Maria Chiara Liguori, Fede- rico Andrucci, Alex Gianelli, Gabriele Marconi, Gabriele Fatigati, Giovanni Baisi, Tommaso Cec- cotti, Diego Piacentini, Marco Petrone, Roberto De Nicolò, Antonio Russo, Cinzia Zannoni

Cineca

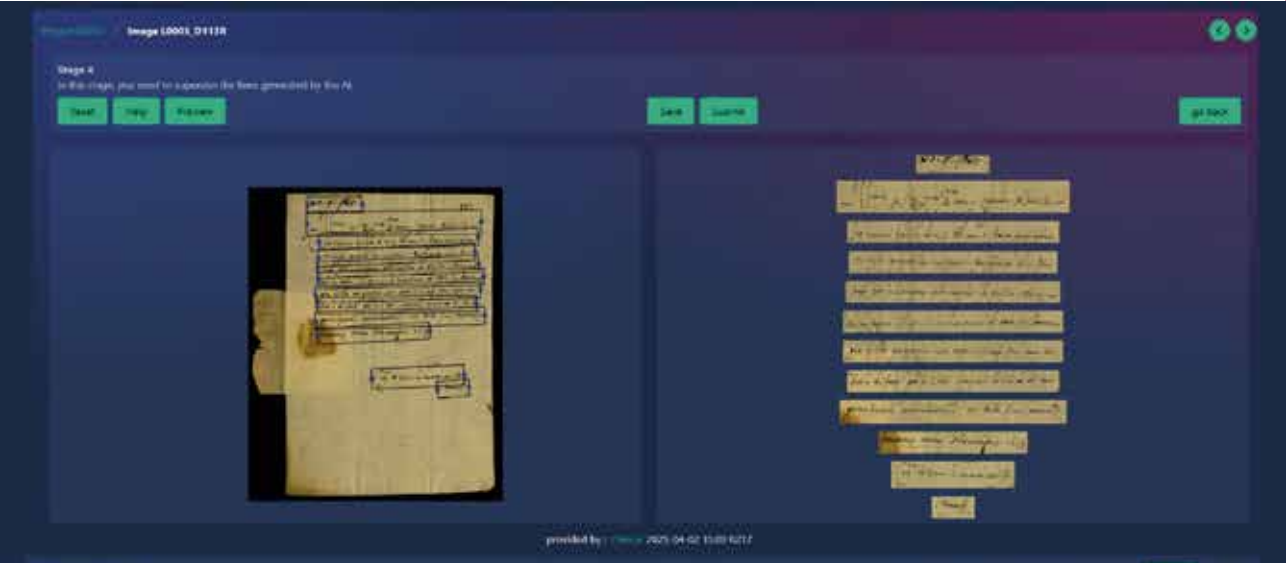
The Central Institute for the Digitisation of Cultural Heritage – Digital Library coordinates and promotes cultural heritage digitisation within the Italian Ministry of Culture. It develops and implements the National Plan for the Digitisation of Cultural Heritage and leads projects under the National Recovery and Resilience Plan, particularly investment M1C3 1.1, which focuses on digital strategies and platforms for cultural heritage.

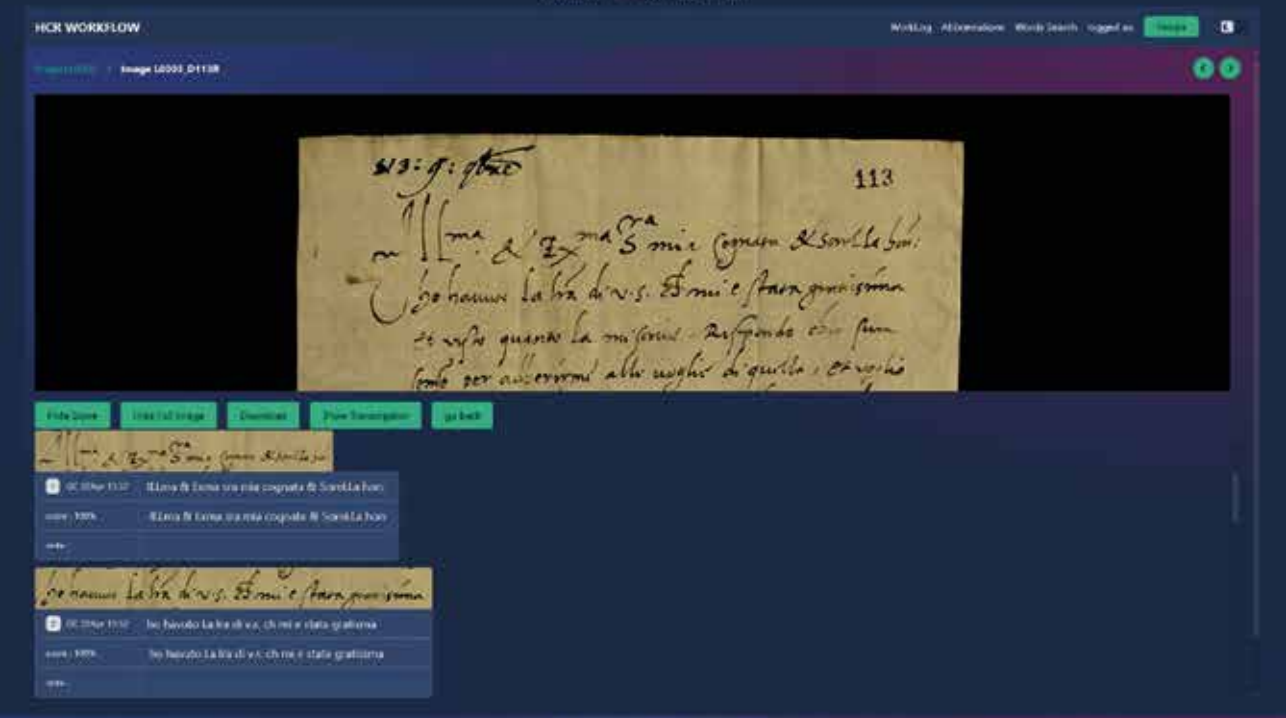
I.PaC (Infrastructure and Digital Services for Cultural Heritage) is a key initiative, a national data space designed to integrate, preserve and enrich digital cultural heritage. I.PaC addresses the fragmentation of digital systems by providing flexible and secure models for managing diverse data formats, domains and protection policies. It is a core part of the ministry's digital transformation strategy (2022-2026).

Built on cloud-oriented technologies, I.PaC offers advanced digital services, including AI tools for managing, processing and presenting digital assets.

AI models are designed to enhance the accuracy and depth of cultural data, promoting continuous evolution in how cultural information is managed and shared. By leveraging AI, I.PaC aims to facilitate more efficient data processing and uncover hidden connections between cultural entities. In this context, Cineca delivers advanced AI tools within an open science environment, supporting the I.PaC platform with interoperable, transparent and reusable technologies that foster collaboration and innovation in the cultural heritage domain. Moreover, Cineca enables new workflows for research environments, providing a framework to support the transcription of manuscripts from different historical periods, such as humanist and Mecantesque texts. New synergies will be explored to foster the use of AI in the digital humanities, in connection with the AI Factory ecosystem and the MIC Dataspace.

Figures 51, 52, 53, in the following pages: HCR Workflow is a platform designed to support handwritten transcription using AI tools. Users can train models on specific manuscripts or use pre-trained templates. The web app serves scholars in archival studies, paleography, codicology and more.
It also guides non-experts through each step of pre-processing and transcription. Designed for large-scale manuscript work, it balances accessibility and rigor. AI assists with character recognition, speeding the process up significantly.
Human reviewers correct errors, ambiguities and unique historical variants. This collaboration ensures high accuracy and supports scholarly standards. The system evolves with each use, improving its recognition capabilities.
HCR blends automation and philological precision in an iterative workflow.





AI Services Demo

Available services

Gliner *Zero shot NER*	**Keras** *Keras OCR for text detection and recognition*	**Kosmos2** *Image caption generator*	**LayoutParser** *Layout recognition for magazines and newspaper*	**Llama31** *Llama 3.1 8B LLM*	**NERs** *NERs - named entity recognition*	**Oemer** *Optical music recognition*
P2PaLA *Handwritten lines recognition*	**Qwen2** *Multimodal LLM*	**RealESRGan** *General purpose image upscaler*	**StableDiffusion** *Text to image model*	**Tagging** *Text tagging*	**Tesseract** *Multilingual OCR for printed text*	**TrOCR** *Handwritten text recognition (single line)*
		Whisper *Italian speech to text*	**Yolo** *Object detection and recognition*			

CHEMISTRY

LIGATE: In silico drug design with Ligate LIgand Generator and portable drug discovery platform AT Exascale

Andrew Emerson
Cineca

The LIGATE project [1] is a European initiative aimed at enhancing computer-aided drug design (CADD) by leveraging high-performance computing (HPC) systems, including future exascale architectures. Coordinated by Dompé Farmaceutici in Italy, LIGATE seeks to integrate and co-design top-tier European open-source components with proprietary intellectual properties to maintain global leadership in CADD solutions. This integration aims to significantly reduce the execution time of virtual screening campaigns and improve drug development effectiveness.

The project enhances the Exscalate platform by incorporating LiGen (molecular docking), GROMACS (molecular dynamics), HyperGen(flexible scheduling) and a new AI engine for efficient molecule screening and simulation. A key focus is on automation, which minimises human effort in managing massive datasets and optimising parameters.

Building upon the successes of previous initiatives like Exscalate4CoV, the LIGATE project demonstrated the efficacy of its platform by identifying Raloxifene as a potential COVID-19 treatment, which subsequently passed a phase II clinical trial.

In LIGATE, Cineca plays a crucial role in deploying cutting-edge HPC systems, including the Leonardo pre-exascale system. These advanced platforms serve as testbeds for LIGATE technologies, facilitating the evolution of the platform toward exascale capabilities throughout the project's duration.

Additionally, Cineca contributes its expertise by developing tools and scripts for high-throughput molecular dynamics simulations. This includes Python-based launchers designed to efficiently manage extensive GROMACS molecular dynamics jobs, optimising computational workflows. This year, Cineca has played a key role in enhancing the launching and analysis of molecular dynamics (MD) simulations for binding affinity prediction. These scripts handle large datasets, verifying equilibration and computing key metrics like complex stability, ligand and ion dissociation energies and free energies. [2]

By the end of next year, approximately 40 TB of scripts, simulations and analyses will be freely accessible, thanks to the Molecular Dynamics Data Bank (MDDB) initiative. MDDB is a European initiative providing a unified repository for molecular dynamics simulation data. It aims to make biosimulation data accessible to researchersinbiochemistry,pharmacology and personalised medicine. [3]

[1] https://www.ligateproject.eu/

[2] https://gitlab-hpc.Cineca.it/aemerson/ligate

[3] https://mddbr.eu/

LIGATE has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956137. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Italy, Sweden, Austria, Czech Republic, Switzerland.

COMPUTING PROVISIONING SERVICES

INNOVATION AND TECHNOLOGY TRANSFERT

INDUSTRIES

Empowering European SMEs: Overcoming innovation barriers with HPC and AI Factories

Claudio Arlandini

Cineca

European SMEs face major innovation barriers, including financial constraints, skills shortages, digitalisation challenges and market knowledge gaps. Access to advanced computing is critical, especially in AI and data-intensive fields. EuroHPC JU HPC centres and the upcoming AI Factories are key infrastructure projects aimed at democratising high-performance computing (HPC) for SMEs, reducing reliance on large tech firms and fostering independent innovation.

Funding challenges particularly impact micro-enterprises, limiting R&D and tech investments. The shortage of skilled talent in AI and data science further hinders innovation, while digitalisation remains a major obstacle, especially in eastern and southern Europe. Regulatory complexities add further burdens to SMEs. To address these challenges, EuroHPC JU has launched initiatives supporting SMEs in HPC-driven product development, process optimisation and AI applications. These shared computing resources lower costs and provide essential technical expertise.

AI Factories aim to break down entry barriers for AI development by offering computing infrastructure, AI model training, data storage and technical support. Acting as centralised innovation hubs, they help SMEs develop and validate AI solutions without prohibitive costs. Together, HPC centres and AI Factories provide SMEs with the tools needed to drive digital transformation and strengthen European innovation competitiveness. Strategic deployment across Europe can also bridge regional digital divides, ensuring more equitable access to cutting-edge technologies.

In 2024, the FF4EuroHPC project transitioned into FFplus, introducing a four-year framework with six open calls: three for HPC and AI business experiments and three for GenAI innovation studies. The initiative saw overwhelming SME participation, with Italian companies achieving notable success in early calls. AI-BOOST also launched its first challenge, awarding two million GPU hours on Leonardo to large-scale GenAI projects: Lingua Custodia (France), specialising in AI-driven financial language processing and Unbabel (Portugal), combining AI and human translation for multilingual support in all 24 EU languages. Meanwhile, the EuroCC project continues its mission of HPC training and literacy in Italy, developing new service prototypes to support SMEs in high-performance computing adoption.

FF4EuroHPC project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 951745.

A digital twin for airflow and inhaled drug delivery in human airways

Francesca Gebbia

Cineca

The DiTAiD project was developed under the FF4EuroHPC initiative.

Background and challenges

Respiratory diseases such as asthma and COPD are caused by airway narrowing, often requiring drug delivery via inhalers. However, the deposition of drug particles can be uneven, reducing treatment effectiveness and increasing side effects. Traditional assessments rely on radio-labelled drug inhalation, exposing patients to radiation, or Computational Fluid Dynamics (CFD) simulations, which require weeks of work and HPC resources due to manual 3D model creation from CT scans.

Proposed solution

The DiTAiD project introduced a Digital Twin (DT) to predict inhaled drug deposition in human airways. Based on a Reduced Order Model (ROM), the DT uses mesh morphing of a baseline airway model to simulate 1.000 geometries via CFD. The ROM compresses the simulation data, enabling near instantaneous results with high accuracy.

The software extracts airway parameters from a patient's CT scan and uses the ROM to deliver CFD-equivalent results without rerunning costly simulations. This significantly reduces costs and processing time, making it viable for clinical use.

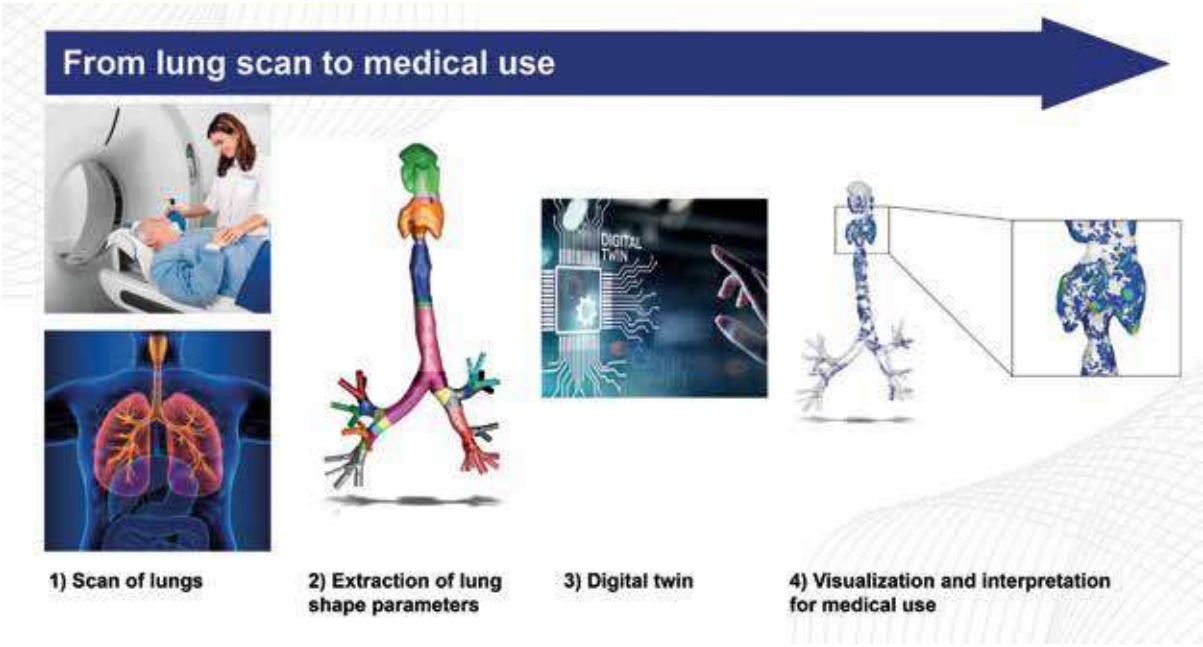


Figure 54: From lung scan to medical use

Impact and benefits

This solution enables hospitals to avoid radiative testing and costly simulations, either via direct use or licensing. Pharmaceutical companies can use it to improve drug and inhaler design, while healthcare consultants can integrate it into their services.

From a business perspective, this innovation supports the expansion of Digital Twins into healthcare. It is expected to create 5 high-level jobs and generate €1 million in revenue within five years.

Recent developments

The project led to the creation of two new companies: the Italian LivGemini (born from the EU project Meditate, in which Cineca is a partner) and the Dutch CFD Solutions, founded by the DiTAiD coordinator and a collaborator after the acquisition of One Simulations by Sweco.

RBF Morph, now scaling up, is gaining clients in the US healthcare sector thanks to the project demonstrators. Together with LivGemini and two French partners, they recently launched the "PANDORA: Pre-operative AssistaNt based on Data-driven apprOaches for vasculaR grAfts surgery" experiment under FFplus, the new FF4EuroHPC edition.

Vectoria private industrial RAG application: A success story in technology transfer

Carolina Berucci, Eric Pascolo

Cineca

Vectoria is a remarkable example of successful technology transfer, showcasing the collaborative efforts between Cineca and Leonardo Spa. This innovative project led to the development of a Retrieval-Augmented Generation (RAG) application tailored to meet an internal company need: the creation of "BMS Compass," a navigator for company procedures on top of Leonardo Business Management System (BMS). This tool equips new hires with a modern instrument to learn essential procedures accurately and swiftly, enhancing their onboarding experience.

Vectoria is the open-source version of the workflow underlying BMS Compass and has been released on GitHub on the EUROCC Italy page, a European project in which Leonardo Spa and Cineca are partners.

The Vectoria project was initiated through the funding provided by two scholarships from Leonardo Spa at Cineca, underscoring the significance of investing in technological innovation and education. This collaboration between the two entities combined their expertise and resources to develop an advanced solution tailored to meet specific business needs.

The application leverages RAG technology to offer precise and contextually relevant information, ensuring that users can quickly grasp the necessary knowledge to perform their tasks effectively.

The key strengths of Vectoria

Vectoria centralises diverse documentation from multiple repositories and formats into a unified, easily accessible system, enabling efficient and accurate information retrieval while minimising errors such as incorrect or fabricated responses (hallucinations). It employs advanced vector database retrieval techniques to provide precise, timely answers, significantly enhancing the overall user experience. Operating entirely within the company's private infrastructure, Vectoria ensures strict data privacy and security, eliminating the need for external generative AI services or subscriptions. Its modular architecture offers extensive customisation through configuration files, allowing adaptation to various organisational needs, document types, underlying models and retrieval parameters. As an HPC-ready application, Vectoria leverages high-performance computing resources to support scalability and rapid processing of large data volumes without requiring code modifications. The system is designed for straightforward installation on private infrastructures such as local servers or workstations, using Python virtual environments and Ansible playbooks to ensure a secure, reliable setup that maintains data confidentiality. This combination of flexibility, security and performance makes Vectoria an effective solution for organisations seeking advanced, customizable knowledge management powered by retrieval-augmented generation technology.

Vectoria exemplifies the power of collaboration and investment in technology, demonstrating how innovative solutions can address specific business needs while ensuring privacy and efficiency. The success of this project underscores the potential of RAG technology in transforming the way companies manage and access information.

Innovation and mobility: A data ecosystem for traffic management in Ravenna

Saveria Teston

University of Bologna

Chiara Dellacasa

Cineca

CTE COBO (Casa delle Tecnologie Emergenti del Comune di Bologna e Ravenna) is a distributed technological infrastructure across the Emilia-Romagna region, designed to foster innovation in strategic sectors such as Industry 4.0, Cultural and Creative Industries and Innovative Urban Services. Acting as an ecosystem enabler, CTE COBO promotes cross-disciplinary collaboration and positive cross-fertilisation among key innovation stakeholders, encouraging both technological advancement and business development.

In this context, the Municipality of Ravenna and Cineca are coordinating the implementation of a proof of concept (PoC) dedicated to mobility and traffic management. The initiative aims to support the Municipality and the Port System Authority in more effectively managing and coordinating traffic flows along two key transport axes: the road connecting the city of Ravenna to Porto Corsini and the road and bike/pedestrian path connecting the city to Marina di Ravenna, located in the coastal area. These mobility corridors are impacted by multiple, potentially conflicting traffic flows: logistics transport, tourism-related flows, public and private transport used by residents, workers, cruise ship crews and cargo ship personnel. A comprehensive data ecosystem has been leveraged, including:

- The Ravenna instance of TIM's Urban Genius platform, enhanced with zoning services to improve data granularity across seven municipal districts affected by the targeted traffic corridors;
- Traffic camera systems installed by both the Municipality of Ravenna and the Port System Authority;
- Video cameras managed by the Municipal Police, upgraded through AI and software technologies provided by TrafficLab, a company selected via CTE COBO's second TechTransfer call;
- Temporary video cameras installed by TrafficLab in collaboration with the Municipality.

The ultimate goal is to generate data-driven scenarios to reproduce realistic mobility conditions while showcasing the specific strengths of each technology involved, as well as the added value of their integration.

To support this objective, a graphical mock-up will be developed to illustrate possible use cases for data visualisation and interaction. The outcomes of the PoC will be explored through three complementary formats: a dashboard, designed not only for operational monitoring and real-time interaction, but also to support long-term mobility planning; a digital twin, offering a spatial and dynamic representation of traffic and infrastructures; and, where applicable, digital art, which will serve as a creative tool to help citizens engage with data in a more accessible and meaningful way.

Figure 55: CTE COBO

ENI

Eni's HPC6 supercomputer: Powering the energy transition

ENI

Eni has achieved an important milestone with the launch of its new supercomputing system in November 2024, HPC6, which entered the TOP500 list in an outstanding fifth place. HPC6 is the first European system and the world's first industrial-use supercomputer by computational power. It is also the only non-US system among the global top five. The launch of HPC6 marks a major increase in Eni's peak computational power, scaling from the previous 70 PFlop/s of HPC4 and HPC5 combined, to over 600 PFlop/s peak performance representing an increase of nearly an order of magnitude.

Built on a cutting-edge architecture featuring HPE Cray EX4000 and HPE Cray ClusterStor E1000 systems, HPC6 comprises 3.472 computing nodes. Each node integrates a 64-core AMD EPYC CPU and four AMD Instinct MI250X GPUs, for a total of 13.888 GPUs. The system relies on an HPE Slingshot high-speed network to ensure low-latency, high-throughput communication, capable of managing exascale-class workloads. HPC6 is hosted at Eni's Green Data Centre and employs direct liquid cooling technology, dissipating 96% of the heat generated while optimising energy efficiency. Its maximum power consumption is 10.17 MW.

HPC6 is a strategic asset for Eni's decarbonisation pathway and the development of new businesses related to the energy transition. Innovation and technological evolution are central to maintaining and strengthening Eni's leadership. Supercomputing is fully integrated across Eni's value chain, representing a key enabler to achieving net-zero objectives and fostering industrial competitiveness. HPC6 will support the optimisation of plant operations, the advancement of geological and fluid dynamic studies for CO_2 storage, the development of next-generation batteries, the enhancement of biofuel production chains, the design of new materials for biogenic and circular chemistry and the simulation of plasma behaviour in magnetic confinement fusion.

This achievement builds on the longstanding collaboration between Eni and Cineca, which has been instrumental in managing Eni's HPC infrastructure. Cineca's expertise covers both system management and the engineering and optimisation of critical scientific applications. Thanks to its experience in addressing the unique challenges of HPC software development for the energy sector (scalability, efficiency, numerical stability, robustness, fault tolerance, usability and portability) Cineca has contributed to maximising system performance.

Under the existing framework agreement, Cineca will continue to support HPC6's management, software engineering and porting activities. This partnership exemplifies the synergy between industry and research institutions, driving technological innovation and advancing Eni's strategic goals in sustainability and energy transition. Cineca's continued support of HPC6 reinforces Eni's computational capabilities, consolidating a collaboration that plays a pivotal role in shaping the future of high-performance computing for industrial applications.

DRUG DESIGN

Dompé and Cineca: Extreme-scale drug discovery

Federico Ficarelli
Cineca

The long-standing collaboration between Dompé Farmaceutici and Cineca has played a pivotal role in advancing drug discovery through HPC. Spanning more than two decades, this partnership has harnessed state-of-the-art HPC technologies to drive pharmaceutical research, leading to major breakthroughs, particularly during global health emergencies. A prominent example is the Exscalate4CoV European project, which was instrumental in accelerating the response to the SARS-CoV-2 pandemic by enabling large-scale virtual screening and drug repurposing efforts [1] [2]. At the heart of this collaboration is the Exscalate[3] platform, which combines supercomputing and artificial intelligence to evaluate extremely large chemical spaces. Key components of the platform have been co-developed by Dompé and Cineca and, through the years, have been optimised for supercomputing environments and further enhanced through the LIGATE[4] project and Cineca's technical leadership.

A strategic challenge on which Cineca is focusing is the integration of classical chemical methods with AI models, both during training and in production. These models are embedded directly into scientific workflows to enhance key applications, such as improving the accuracy of force fields used for ligand generation and conformational sampling. The synergy between traditional algorithms and machine learning enables more predictive simulations and accelerated hypothesis testing. The adoption of MLOps practices further ensures scalable, reproducible deployment and continuous improvement of AI components across the pipeline. Key improvements have also been made in GPU acceleration of numerical algorithms, allowing the platform to scale out efficiently on exascale supercomputers.

Cineca is also expanding its HPC footprint in Naples, in collaboration with the University of Naples Federico II and the City of Naples. The upcoming facility will host a new supercomputing system, fully integrated with Leonardo, Cineca's flagship HPC infrastructure. This expansion will significantly boost computational capacity for joint research efforts with scientific and industrial partners and the new data centre will further strengthen the collaboration between Dompé and Cineca: The large-scale virtual screening and big data analytics infrastructure will be hosted in this new facility, creating a tightly integrated innovation hub. This strategic investment is not only a milestone for the partnership but also a major opportunity for Italy to lead in digital pharmaceutical innovation and exascale scientific computing[5].

[1] Gadioli, D.; Vitali, E.; Ficarelli, F.; Latini, C.; Manelfi, C.; Talarico, C.; Silvano, C.; Cavazzoni, C.; Palermo, G.; Beccari, A. R. EXSCALATE: An Extreme-Scale Virtual Screening Platform for Drug Discovery Targeting Polypharmacology to Fight SARS-CoV-2. IEEE Trans. Emerg. Topics Comput. 2022, 1–12. https://doi.org/10.1109/TETC.2022.3187134

[2] Giulio Vistoli, Candida Manelfi, Carmine Talarico, Anna Fava, Arieh Warshel, Igor V. Tetko, Rossen Apostolov, Yang Ye, Chiara Latini, Federico Ficarelli, Gianluca Palermo, Davide Gadioli, Emanuele Vitali, Gaetano Varriale, Vincenzo Pisapia, Marco Scaturro, Silvano Coletti, Daniele Gregori, Daniel Gruffat, Edgardo Leija, Sam Hessenauer, Alberto Delbianco, Marcello Allegretti & Andrea R. Beccari (2023) MEDIATE - Molecular DockIng at homE: Turning collaborative simulations into therapeutic solutions, Expert Opinion on Drug Discovery, DOI: 10.1080/17460441.2023.2221025

[3] https://www.exscalate.com/

[4] https://www.ligateproject.eu/

[5] https://leonardo-supercomputer.Cineca.eu/it/Cineca-nuova-sede-napoli

INTERNATIONAL PROJECTS

HPC collaborations beyond EU: HANAMI and GANANA

Fabio Affinito

Cineca

In the last two years, EuroHPC JU has launched two initiatives aimed at strengthening Europe's bilateral collaboration with India (the GANANA project) and Japan (the HANAMI project).

HANAMI is dedicated to supporting and enhancing collaboration between European and Japanese scientific teams to advance the performance and adaptability of HPC applications. By leveraging the expertise of the scientific communities in both regions, the initiative will focus on user-driven development in key research areas.

The ongoing collaboration between the European consortium and Japanese research institutions will consolidate the necessary expertise to enhance applications' performance. This includes designing innovative models, optimising and developing computational methods and integrating ambitious workflows to fully utilise pre-exascale, exascale, post-exascale and hybrid architectures that are either already operational or soon to be deployed in Europe and Japan.

HANAMI targets climate simulation, materials science and biomedicine, building upon and expanding the strong scientific partnerships already established between Europe and Japan.

HANAMI is coordinated by CEA (France) and Cineca's contributions will focus on enhancing collaboration to develop modules for the QuantumESPRESSO package.

The GANANA project, which will be launched in 2025, will focus on expertise exchange and partnership in software development with India. GANANA will be based on the domains of biomedical and life sciences, natural hazards and weather modelling.

GANANA is coordinated by KTH (Sweden) and Cineca will contribute to the design and optimisation of a simulation pipeline for understanding properties of biomolecules.

NEXT TRENDS

Trends for the future

Sanzio Bassini
Cineca

In our 2023/2024 report, we shared the vision that the Bologna Technopole, now renamed DAMA – Data Manufacture, was to become a Factory for Artificial Intelligence.

This vision has become reality and the IT4LIA project, aimed at creating a European Factory for Artificial Intelligence, has been successfully realised. The project's kick-off formally took place at the start of last April.

This achievement reinforces a forward-looking perspective regarding the evolution of the HPC system architecture, becoming increasingly multi-tier and data-centric and the evolving structure of skills and services within the HPC Department.

Regarding the computing system architecture, there is a high-speed broadband connection between Naples and Bologna, where two multi-protocol data lake repositories are installed. This network backbone, extended across data centres, connects the data lakes and HPC systems that share the data repositories. The tier hierarchy is as follows: Tier 0 systems include Leonardo/Lisa and the upcoming AI-optimised IT4LIA system; Tier 1 systems, characterised by application functionality, are Megaride, Marcopolo and Gaia; Tier 2 systems, based on geographic rationale, include Trieste SISSA, Milano Bicocca and Palermo University.

First prospective consideration: the IT4LIA system, currently in conceptual design, will be optimised for AI workloads and will not be the successor to the Leonardo system, which will have a successor dedicated to scientific computing expected at the end of Leonardo's lifecycle in 2028. The design shift is from co-design focused on extreme performance to a model emphasising high availability and quality of service. This strategic choice reflects the need to support both research excellence and the AI-driven digital transition in public administration and industrial innovation.

This positioning reinforces one key aspect of Cineca's institutional mission: technology transfer and collaboration with the private sector. Traditionally viewed as one of the most vital players in the European HPC ecosystem, Cineca now requires a leap in role, from interlocutor to promoter, given the innovation challenges posed by AI.

Second element of perspective: The HPC access model is converging towards the hyperscale access model widely adopted by big tech companies. This model is based primarily on Kubernetes orchestration of microservices, offering modes of (virtual) infrastructure as a service, as well as service platforms able to efficiently host and manage the production workflows essential for AI solutions. Cineca HPC has developed initial capabilities in this direction with the Galileo system, soon to be expanded with the new Gaia system. These systems use an OpenStack software environment with Kubernetes orchestration, sharing data from the data lake repository with Leonardo/Lisa and, prospectively, the AI-optimised IT4LIA system, thus guaranteeing HPC services for training and refining large language models (LLMs), as well as supporting reasoning and inferential production through a cloud access system.

Third element of perspective: From an organisational point of view, three development priorities emerge for the HPC department. First, increasing complexity in system management requires simultaneous attention to both classic HPC services and

cloud services. Second, the growing complexity of the architecture necessitates the establishment of a team dedicated to the continuous evolution of HPC systems architectures, alongside a specialised support team focused on AI. Third, recent initiatives by the European Commission under the strategic action Invest AI demand focused attention.

To foster competitive European alternatives to global big tech cloud and data services for AI production, the European Commission has launched a call for expressions of interest to establish about four or five AI Giga Factory-scale infrastructures in Europe. This entrepreneurial project is on the order of billions of euros in investment, with the Commission co-financing 35% and private entities covering the remaining 65% through public-private partnerships.

Hosting an AI Giga Factory in Italy is of undeniable strategic importance for the national system. Cineca's desirable role in this context is to act as the public system's representative, bolstered by support from Italy's leading blue-chip companies.

In this role, Cineca must also prepare for and be ready to integrate its AI Factory infrastructure and services, developed under the IT4LIA project, within the ecosystem of a prospective AI Giga Factory located in Italy. This would combine public investment in AI solutions with a sovereign, European alternative to big tech services for AI inference production.

The second half of 2025 will be dedicated to the development and activities guided by these strategic elements, building upon an established technological base and skillset that requires further strengthening both technologically and organisationally.

Cover

This year our Report cover focuses on the inspiring principles of IT4LIA AI Factory: Science, Innovation and the Common Goods.
The graphic proposal, which simulates a series of overlapping and intersecting planes, goes hand in hand with the idea of the multiplicity of fields of application. The style based on transparencies and translucencies allows this message to be conveyed.
The palette and gradient themes come from the desire of avoiding cold, flat colours and thus "humanise" the AI concept.
The cover design started with 2D AI mock-ups with Krita AI and experiments with 3D modelling in Blender. The cover was then finalised with Blender and The GIMP, rendered in high quality with Cycles render engine and composited within Blender.

Questo report è stato stampato su carta ecologica prodotta con fibre riciclate.