CINECA
HPC ANNUAL REPORT

2023/24

CINECA

2023 - 2024

HPC annual report

ISBN 9788886037471

Cineca HPC Report 2023-2024
website: www.hpc.cineca.it
mail: info-hpc@cineca.it

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

Editorial Board
Sanzio Bassini
Gabriella Scipione
Antonella Guidazzoli
Maria Chiara Liguori

Design and Layout project
Simona Caraceni

Print
Grafiche Zanini
Via Emilia 41/e - 40011 Anzola dell'Emilia BO

Thanks to:
Fabio Affinito, Paola Alberigo, Federico Andrucci, Matteo Angelinelli, Claudio Arlandini, Mauro Bigi, Giacomo Barigazzi, Sanzio Bassini, Tiziana Bassi, Giovanni Bellavia, Eleonora Bergamaschi, Neva Besker, Susana Nausica Bueno Minguez, Francesco Califano, Simona Caraceni, Mattia Carello, Balasubramanian Chandramouli, Marcello Crosta, Francesco Maria Cultrera, Mirko Cestari, Beatrice Chiavarini, Patrizia Coluccia, Maurizio Cremonesi, Davide Crisante, Antonio Costantini, Mirko Degli Esposti, Chiara Dellacasa, Daniele De Luca, Francesca Delli Ponti, Xhuilio Dhori, Andrew Emerson, ENI S.p.A, Francesco Falciano, Daniela Galetti, Federica Gebbia, Alex Gianelli, Silvia Gioiosa, Massimo Gisonni, Luca Govoni, Alessandro Grottesi, Massimiliano Guarrasi, Antonella Guidazzoli, Mariella Ippolito, Piero Lanucara, Maria Chiara Liguori, Davide Manzini, Giacomo Masdato, Luca Mattei, Juan Mata Naranjo, Giuseppe Melfi, Riccardo Mengoni, Andrea Mignone, Diego Molinari, Rebecca Montanari, Luisa Monti, Laura Morselli, Lucia Rodriguez Munoz, Juan Mata Narajo, Maurizio Ortali, Daniele Ottaviani, Stefania Paolazzi,Eric Pascolo, Marco Pistore, Giorgio Pedrazzi, Oreste Pezzi, Marco Puccini, Elisa Rossi, Elena Saluzzi, Gabriella Scipione, Donatella Sforzini, Nitin Skhula, Luca Tornatore, Giuseppe Trotta, Claudia Truini, Roberta Turra, Cinzia Zannoni, Paolo Zuzolo.

Cover
To create this issue's cover, AI tools were integrated with traditional digital art tools. An isometric prototype was created with DALL-E, showing buildings and structures surrounded by elements such as displays and circuits. These geometries were then modelled in 3D using Blender and rendered to produce various outputs such as Z-Depth, Cryptomatte and outlines using Freestyle.
These outputs were then integrated into a pipeline using ComfyUI and the desired image was generated through a series of steps.
Finally, GIMP and a graphics tablet were used to add details such as lettering on buildings and stylised figures, and the central building was modified to resemble the appearance of Bologna's Technopole. The colour palette taken from the cover helped to create the visual identity of this report.

Dear Colleagues,

I am pleased to confirm the successful evolution of the Leonardo development strategy. The system is now in full production, a strategy based on the continuous expansion of the HPC Department staff and the development of a national federated and distributed HPC infrastructure managed by CINECA.

Despite the strong demand for technical-scientific skills in the labor market, which has increased the need to improve and make the onboarding process of new staff more effective, the Department has remained attractive. The Department's staff size is now over 150, with more recruits on the way. Our objective remains to reach a critical mass of 200 staff members over the next 6-8 months, bringing on board an additional 35-40 staff members.

Regarding the development of the national HPC infrastructure of federated systems, all procurement tenders for Tier 1 systems have been activated. Lenovo has been assigned the provision of the Pitagora system for Eurofusion, with an investment of €28 million, and E4 Engineering will provide the Galileo100+ system, with an investment of €16 million. Procurement is underway for two Tier 1 systems, to be installed at the Tecnopolo in Bologna and the San Giovanni a Teduccio center in Naples, with investments of €19 million and €26 million respectively. Overall, approximately €90 million of investments have already been activated. Key elements have also been established to proceed with the acquisition of three Tier 2 systems, which will be installed in Milan, in collaboration with Bicocca University; in Trieste, in collaboration with SISSA; and in Sicily, in collaboration with the University of Palermo.

This infrastructural framework will soon be completed with the upgrade of the Leonardo high-end system, the Lisa project, and the acquisition of two quantum computing systems, to be installed at the CINECA data center at the Tecnopolo of Bologna, one based on neutral atoms and the other on superconductivity technology. Overall, infrastructure investments will exceed €150 million. The development strategy is therefore not in question and will indeed be accelerated, to ensure that the HPC Department of CINECA can continue to play a strategic role in supporting advanced scientific research in Italy and Europe and in driving innovation for industrial production and service companies in our country, despite the significant changes in the world.

Over the last twelve months, we have all witnessed an extraordinary revolution in knowledge production and scientific discovery, driven by the game-changing impact of large language models as the basis of generative

Artificial Intelligence. The demand for computing power for the automatic training of foundational LLMs has exploded, saturating all available resources and creating exceptional demand for infrastructural and organizational investments, especially from the private sector and big-tech industries, as well as from the public sector at both European and national levels.

This revolution has not only changed the development model paradigm but also the value chain paradigm in information technology. We have shifted from hardware to software. The largest stock market capitalizations, once dominated by large manufacturing and energy companies, are now dominated by software solution developers. Among the so-called trillion-dollar companies, apart from two energy companies, Aramco and PetroChina, the rest are all software industries: Microsoft, Apple, Nvidia, Alphabet, Amazon, Meta, and Tesla. Don't be fooled by the final products of Nvidia and Tesla, which are presented as floating-point accelerator microprocessors and electric cars. The real added value lies in their software. Nvidia is a fabless company designing microprocessor architecture to support LLM training workloads, where the greatest added value is the software ecosystem. Tesla assembles electric cars using off-the-shelf components, with added value coming from its vehicle software platform for autonomous driving, mobility, and connectivity functions.

At CINECA, we cannot remain indifferent to this epochal change. Four vertical projects have been activated for training foundational LLMs in Italian. One project is in collaboration with the national academic partnership for Artificial Intelligence, FAIR; three are in collaboration with private partners, aiming to provide open-source foundational LLMs, both text-based and multimodal, available for research and innovation in vertical domains.

This is just a glimpse of the commitment that will be at the center of our work over the next twelve months, with the ambition and objective of finalizing a coordinated action with the national stakeholder ecosystem to complete the vision project of the Tecnopolo of Bologna, the hub of the national big data valley. This aligns perfectly with the European Commission's request to create AI Factories, the factory of the future, based on data and automatic training of foundational LLMs, usable for domain inference in the most challenging areas of scientific research and technological innovation.

2023
IN NUMBERS

Cineca HPC is the main Center for scientific computing in Italy. It runs a large computing infrastructure and make it available to Italian and European Researchers, as well as to important Italian Companies within a programme for supporting national industrial competitiveness.

Here we present data about usage and users, projects, events and classes offered in the field of High Performance Computing.

At the end of this section we present our staff, the true resource that makes everything work.

Evolution of the Cineca HPC infrastructure in 2023

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

During the first half of 2023 the Cineca HPC infrastructure was composed by four clusters, the Tier-0 systems Marconi and Marconi100, the Tier-1 system Galileo100 and a Cloud cluster named ADA cloud.
The Marconi system is a X86 classical cluster with about 3000 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 runs of the EUROfusion community.
On the other hand, Marconi100 is an accelerated cluster equipped with 980 compute nodes each with 2 IBM Power 9 processors and 4 Nvidia V100 GPUs. It was devoted to large production runs of the Italian and European scientific community.
The Galileo100 Tier 1 cluster is devoted to smaller scientific and industrial workloads and it is equipped with about 600 nodes each with 2 Intel CascadeLake 8260, 48 cores and at least 384 GB of RAM.
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 the second half of 2023 the Cineca HPC infrastructure evolved further.
Particularly, on May 29th Leonardo (Booster partition), that was already inaugured on November 24th 2022, entered in its pre-production phase. During this phase it was tested by a selected number of users to verify its functionality.
LEONARDO's system is based on the BullSequana XH2000 architecture supplied by Atos and is specifically 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 General Purpose. Leonardo Booster partition includes 3456 Atos "Da Vinci" blades, each equipped with 1 Intel Ice Lake CPU with 32 cores and 4 customized version of NVIDIA Ampere GPUs, called A100 SXM6 64GB. The Leonardo General Purpose partition is equipped with 1536 conventional blades each equipped with 2 Intel Sapphire Rapids CPUs and 512 GB of DDR5 RAM. The Leonardo system is also equipped with more than 100 PB of rotative disks storage and more that 5 PB of full flash storage.

On July 19th 2023 our old accelerated Tier-0 cluster, Marconi100, has been shutdown for the final dismission after end of production. Login nodes and disks remained available until the first part of 2024 in order to allow users to copy their data on the other Cineca Clusters.
On August 3rd the Leonardo Booster partition started the production phase and the first production project has been allocated.

Finally, at the end of 2023 also de General-Purpose partition of Leonardo has been accepted for starting the production on February 12, 2024.
In addition to the aforementioned systems, other infrastructures are present and dedicated to special users or to specific communities.

These includes:

- 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 WLCG program.
- DGX system, an accelerated system for AI based on NVIDIA A100 GPUs.

Thanks to the start of production of Leonardo the computational capacity of the Flagship Tier-0 system hosted by Cineca has been increased by more than a factor 7: from 32 PFs of Marconi100 to more than 240 PFs of Leonardo Booster. At the same time also the total storage capacity of Cineca increased by more than a factor |5|.
The evolution of the clusters in production is still ongoing and will further increase the computational capacity of Cineca HPC infrastructure.

	CPU	Total nodes/ cores	Memory/ node
LEONARDO "BOOSTER" (ATOS) Opened to production on August 3rd, 2023	Intel IceLake, Intel Xeon Platinum 8358 (2.6 GHz, 32 cores) + NVIDIA Ampere GPUs	3456 nodes 110.592 cores 13.824 GPUs	512 GB DDR4 3200 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
MARCONI100 (IBM) dismissed on July 19th, 2023	IBM Power AC922 + NVIDIA Volta 100	980 nodes 31.360 cores 3.920 GPUs	256 GB
MARCONI A3 – SKL (LENOVO)	Intel SkyLake 2x Intel Xeon 8160 (2.1GHz 24 cores)	3.124 nodes 149.952 cores	192 GB

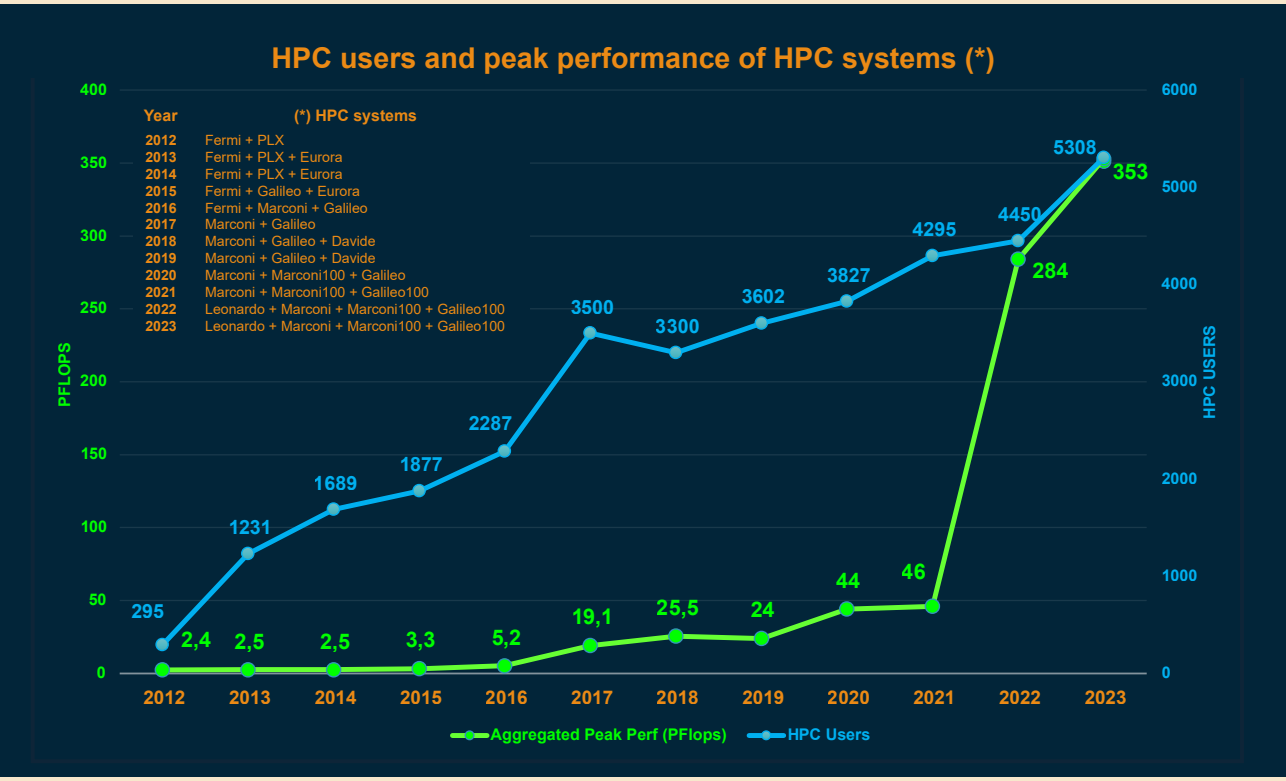


Figure 1: in this figure it is represented the Peak Performance (PFlops) of all HPC Systems available for production since Fermi to Leonardo. With the arrival of Leonardo, the peak performance of our HPC systems increases significantly by reaching a peak performance of 353 PFlops. In the same figure we also present the number of HPC users that shows a trend that grows together with the aggregated performance of our supercomputers.

Supercomputer	Resources (Core Hours)	Performance capacity (Exaflops)
LEONARDO	261.542.144	2.648.114
Marconi -A3	1.298.317.824	311.596
MARCONI100	159.559.680	574.415
Galileo 100	238.075.776	63.031
Total	**1.957.495.424**	**3.597.156**

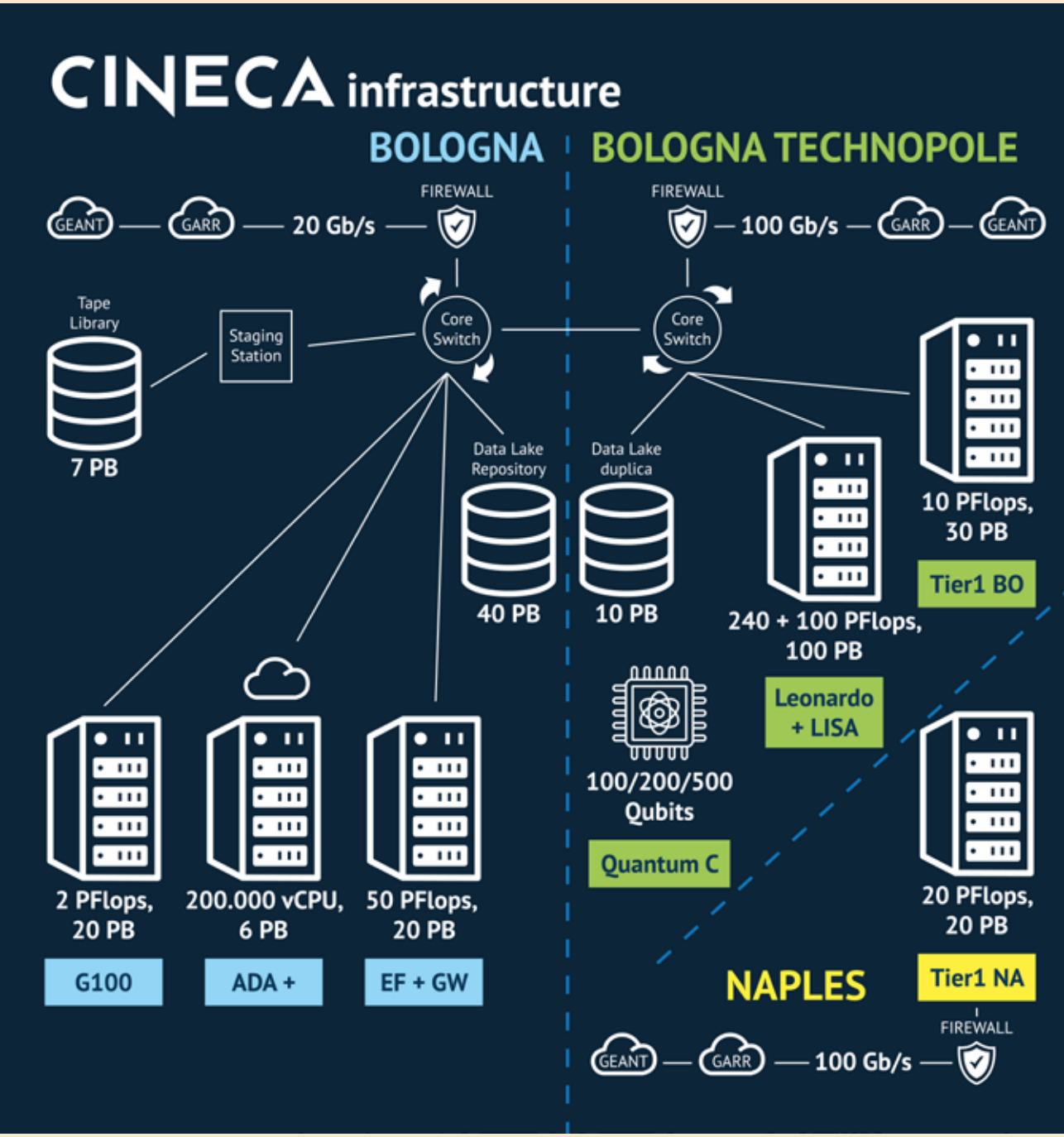


Figure 2: Cineca infrastructures.

Figure 3: HPC timeline from 2012 to nowadays the Cineca Supercomputers Tier-0 with the relative position in the TOP500 list. Below the Supercomputers Tier-1, the different configurations of Galileo.

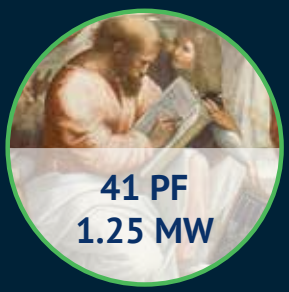
DC Technopole

Leonardo

240+9 PF
9 MW

#4

Pitagora



41 PF
1.25 MW

2022 2023 2024

Galileo 100+

Users Statistics

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

Our HPC infrastructure has seen a significant increase in active users, reaching 5308 by the end of 2023. This marks a growth of 858 users compared to the previous year, a trend that has been consistently positive over the past five years. A large part of the users are males (78%), with 40% aged between 31 and 45 and 37% under 30 years old.

67% of the total number of users work for an Italian institution, clustered in the Emilia Romagna region (19%), Lombardia region (18%), and Lazio (16%). At the city level, Milan, Rome, Bologna, and Trieste have a user percentage ranging from 15% to 11%.

The higher concentration of users is in European institutions, Germany, with 5%, and the United Kingdom, France, and Spain, with 3% each. Users mainly work for universities and public or non-profit organizations.

The three most represented disciplines are Computational Chemistry, with 11%, Condensed Matter Physics, and Computational Fluid Dynamics, with about 10% each, followed by Nuclear Fusion and Astro and Plasma Physics (9% each), Computational Engineering with 8%. Earth and Climate Science, as well as AI & Machine Learning, are represented at 7 % each. The AI & Machine Learning scientific domain grew from 4% respect in 2022. This upward trend reflects the increasing interest and utilization of AI & Machine Learning but also underscores the potential and future direction of HPC usage.

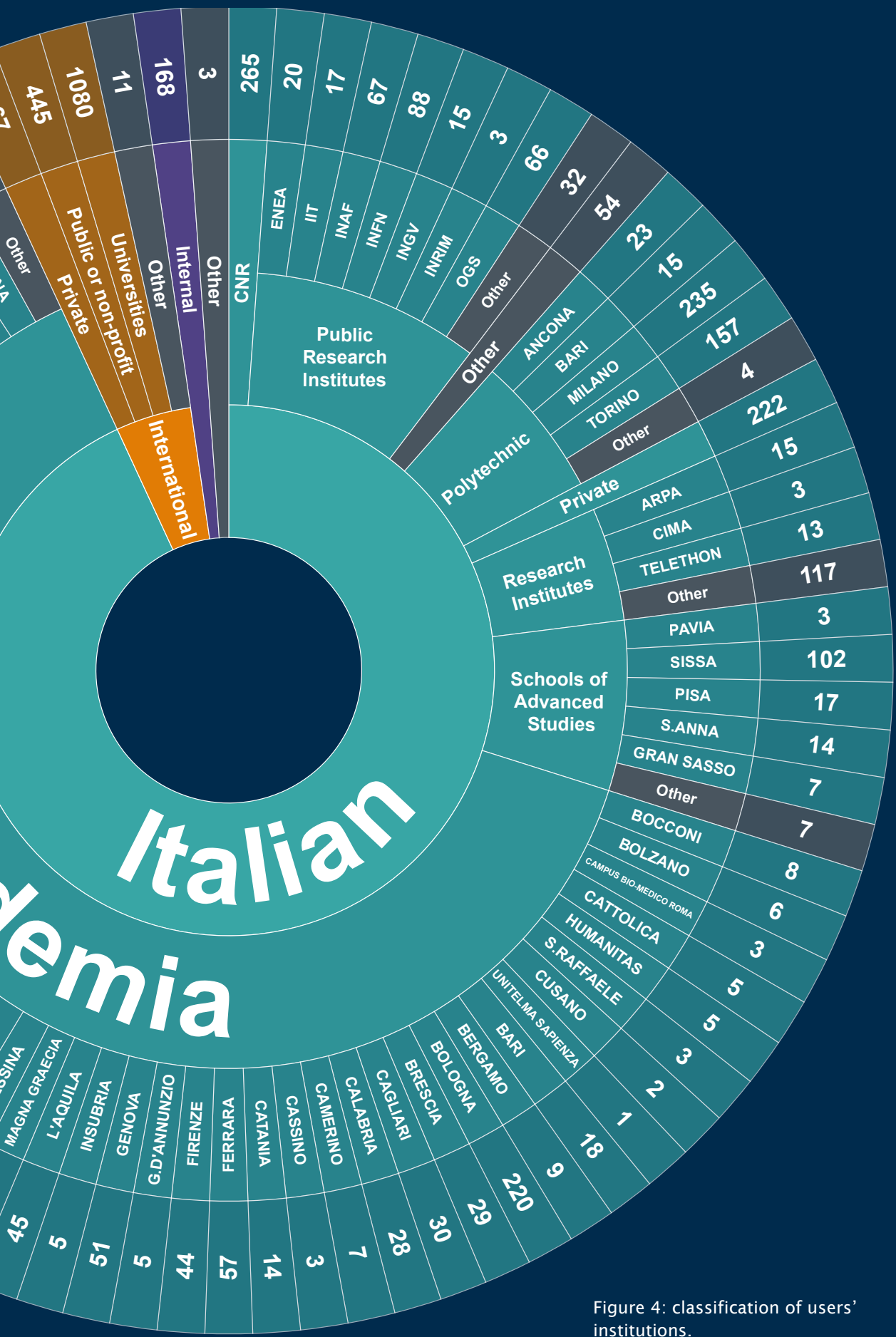


Figure 4: classification of users' institutions.

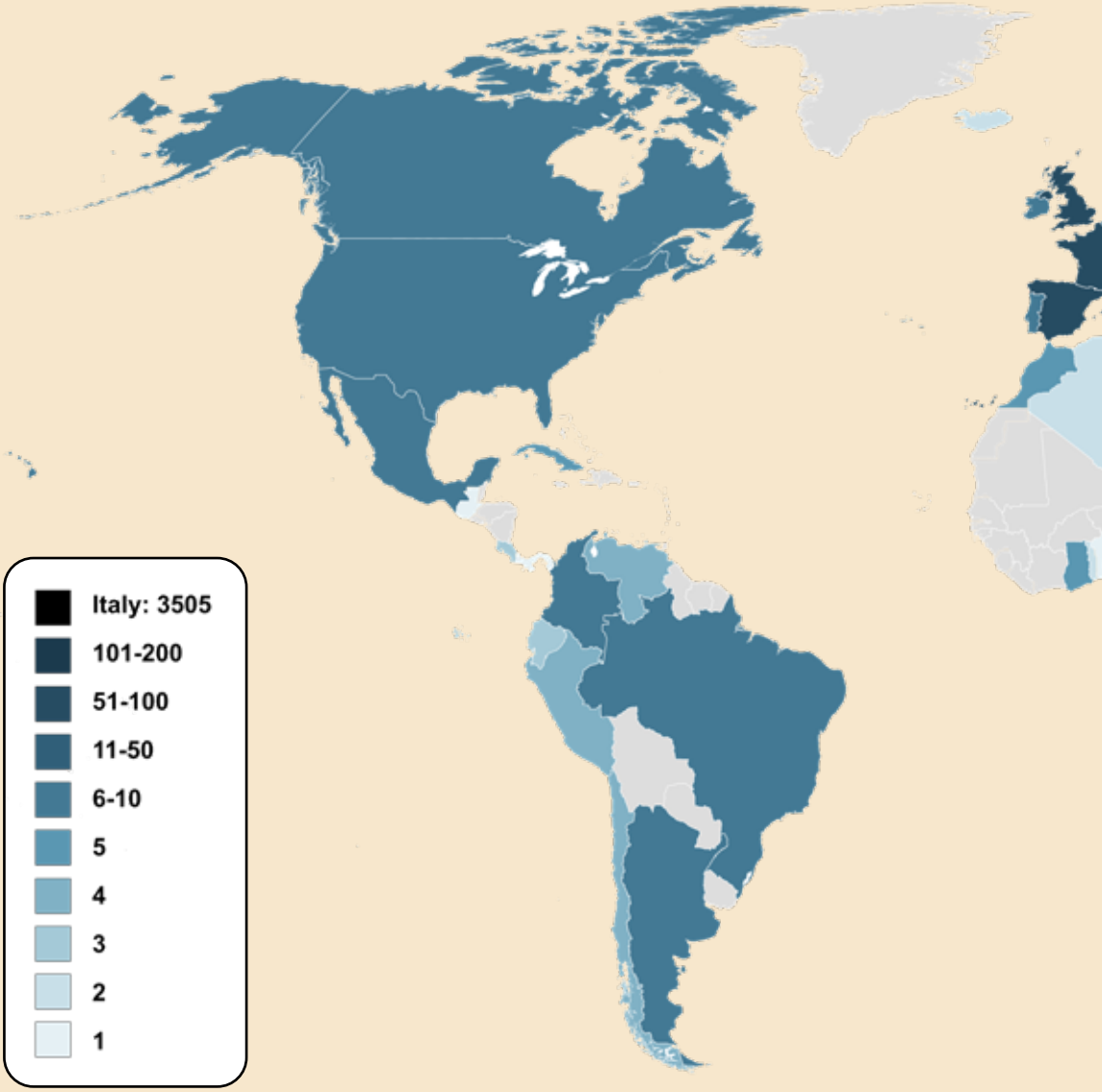
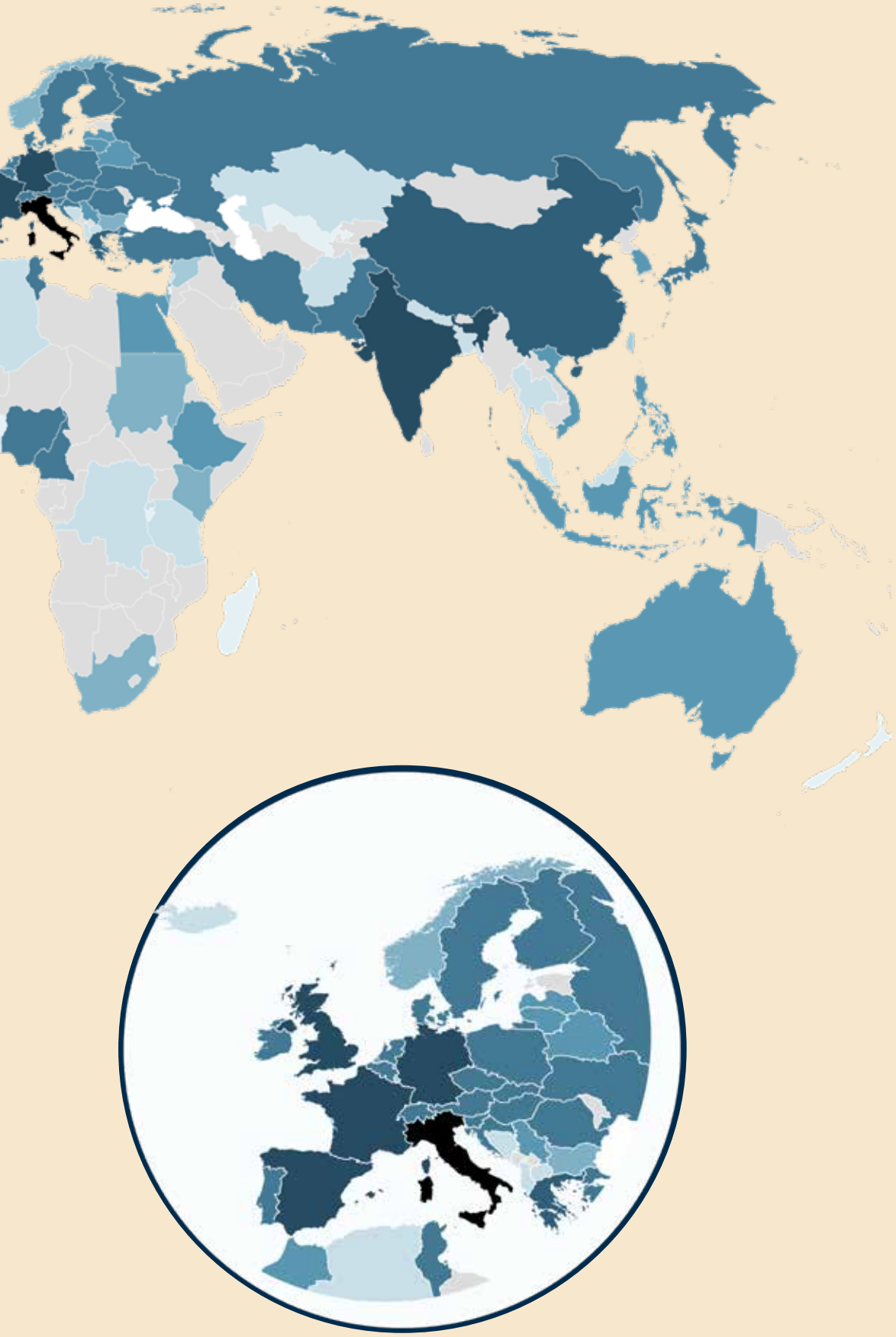


Figure 5: geographical distribution of world and Europe users' institutions.



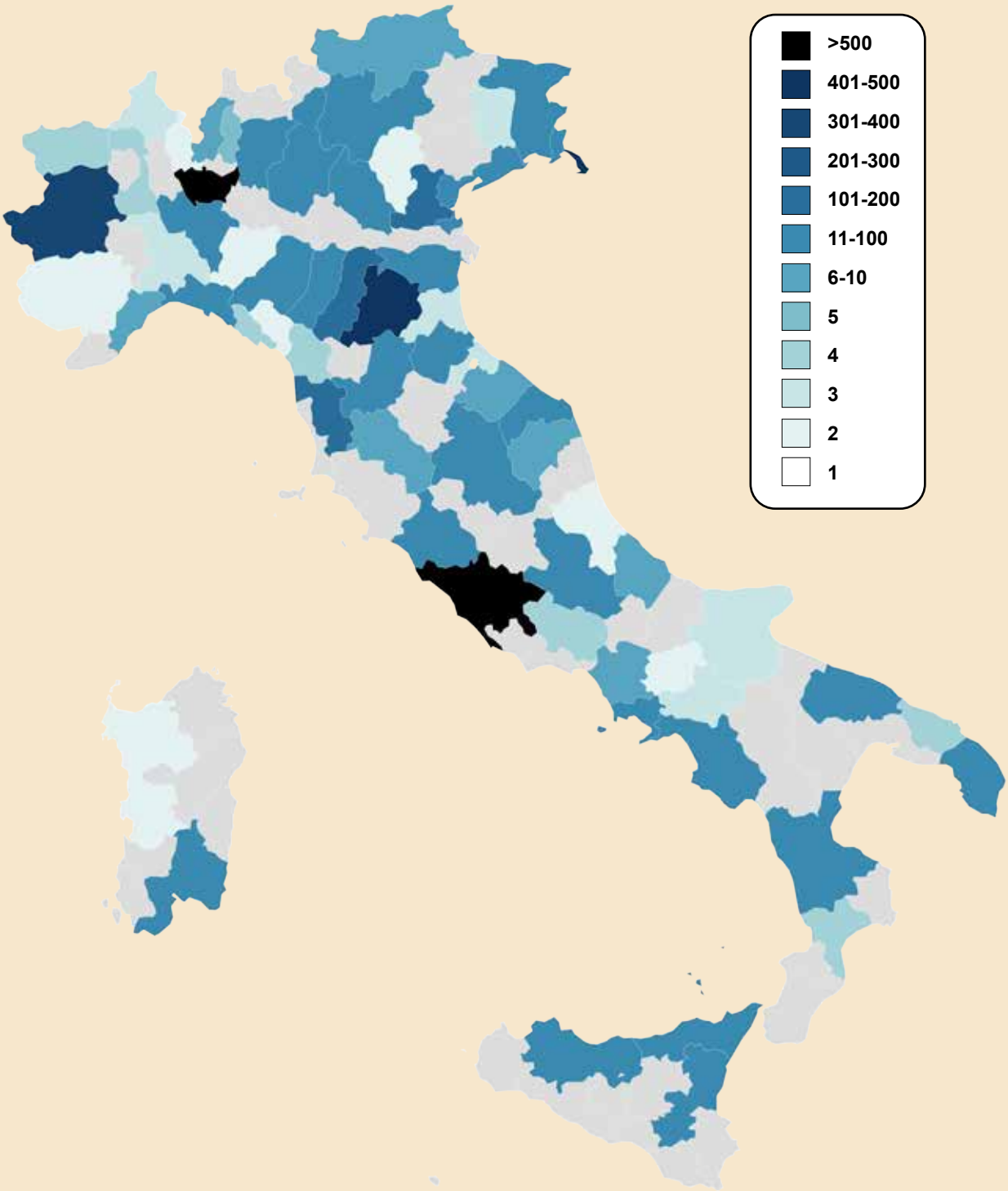


Figure 6: geographical distribution of Italian users' institutions.

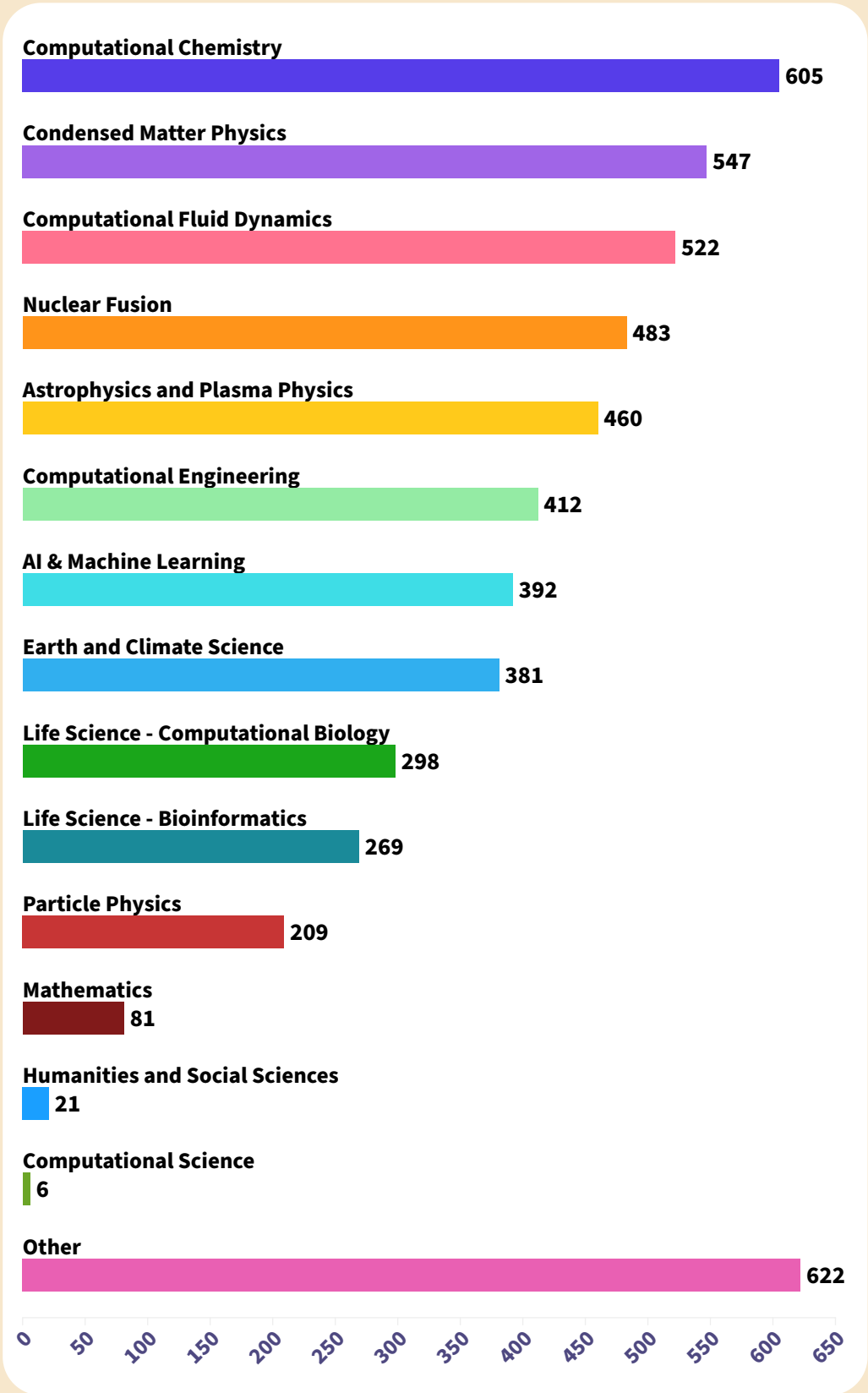


Figure 7: main disciplines for users' projects.

Resource allocation: projects, agreements, collaborations

Diego Molinari, Neva Besker, Massimiliano Guarrasi
Cineca

Currently Italian and European researchers are taking advantage of the HPC clusters in Cineca to perform their computational research. A significant part of the allocations on Cineca's systems are based on a peer-review process to ensure the highest scientific value of the selected projects.

EuroHPC and ISCRA calls allocate most of the computational resources on Leonardo and Galileo100. Both calls are based on the peer-review mechanism at the European and national level, respectively.

On the other hand, Institutions interested in using our HPC infrastructure can sign dedicated "Agreements" with Cineca. Generally speaking, the resources of the agreements are then allocated by the corresponding institutions always after an internal peer review process. In the same framework, it is also possible to allocate Cineca HPC resources on the basis of a collaboration to reach some common results. Furthermore, in some special cases, Industrial or private users can also sign an agreement with Cineca to use our resources for R&D or PoC projects.

For running on our HPC machines, users must have a dedicated project. In general, a project is characterized by a starting date, an end date, a science domain, a budget (in terms of core-h/GPU hours) and a PI (Principal Investigator). It is also possible to share the computational resources of a project with several collaborators.

Along with these wide classes of projects, there are a dedicated partitions Galileo100 and some resources on Leonardo-General Purpose for weather and biogeochemical forecast activities. Galileo100 also host some industrial projects while special partitions of Marconi A3 and some resources on Leonardo-Booster are dedicated to the EUROfusion consortium. For these partitions, the allocation is made autonomously by the owners of the resources.

During 2023 were available on overall Cineca HPC infrastructure a total of 1,957 million core-h whilst the total allocation was of 1,894 million core-h.

Excluding the resources of Eurofusion, the most significant part of the resources in terms of core hours has been dedicated to peer reviewed projects as in the following: 29% ISCRA, 20% to EuroHPC, 37% agreements. Less than 1.5% of the total resources are allocated to industrial projects.

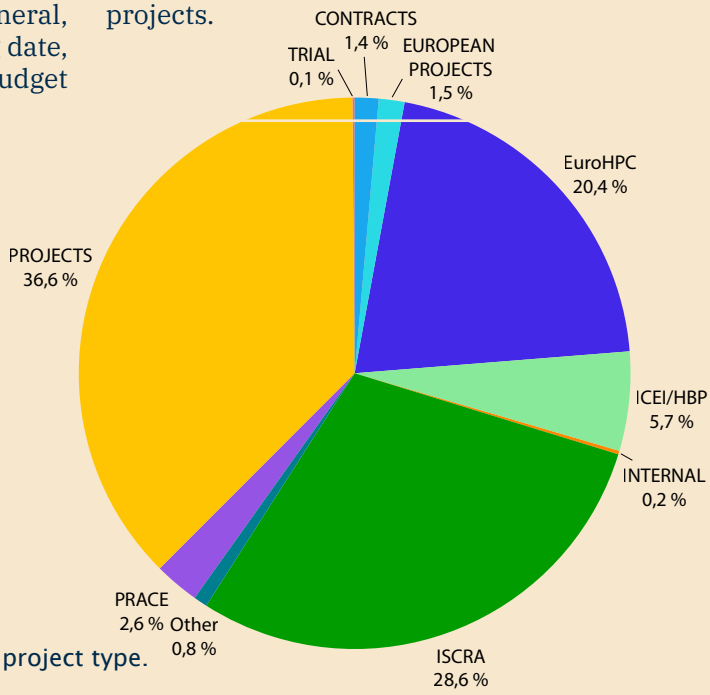


Figure 8: HPC allocation (Core hours) by project type.

EuroHPC

The EuroHPC Joint Undertaking (JU)

Tiziana Bassi
Cineca

The EuroHPC Joint Undertaking (JU) facilitates cooperation among European Union member states and participating countries to establish Europe as a global leader in supercomputing. This initiative aims to advance Europe's scientific and industrial capabilities, drive digital transformation, and safeguard technological sovereignty. The EuroHPC JU focuses on developing, deploying, and maintaining a cutting-edge, secure, and interconnected supercomputing and quantum computing infrastructure within the EU. It supports the creation and adoption of innovative and competitive supercomputing systems tailored to user needs, while also promoting a resilient supply chain to minimize disruptions. Additionally, the EuroHPC JU aims to encourage the development of applications optimized for these systems and broaden access to supercomputing resources for both public and private users. It also prioritizes the enhancement of high-performance computing skills across European science and industry.

At present, the JU consists of 35 participating states. CINECA serves as Italy's representative in the EuroHPC JU and hosts Leonardo, one of the three pre-exascale EuroHPC supercomputers funded by the initiative. Installation of Leonardo commenced in 2022 at the new data center located in the Technopole of Bologna, Italy.

During the pre-production phase of Leonardo, a total of 21 projects Leonardo Early Access Projects (LEAP), were granted access to operate on the Leonardo Booster partition. These projects had the opportunity to test their code on the cluster and explore its capabilities. Cumulatively, these projects consumed 133.014.662 Core Hours over the span of 8 months (April – December 2023).

The production phase of Leonardo officially began on August 3rd.

Since August 2023, Leonardo Booster partition has been awarded 10 EuroHPC calls: 2 Extreme, 1 Regular and 7 Benchmark/Development calls. EuroHPC allocated a total of 377.580.928 Core Hours (equivalent to 11.799.404 Node Hours), as detailed in the following table.

For the Extreme calls, a total of 12 projects have been allocated on Leonardo Booster for a total of 379.659.008 Core hours . Of these, 6 projects were led by Italian-affiliated Principal Investigators (PIs).

Regarding Regular calls, 6 projects have been allocated on Leonardo: 4 on the Booster partition for a total of 9.497.920 Core hours and 2 on the DCGP partition (not in production yet, thus deferred to 2024). Notably, 3 projects on the Booster partition have been spearheaded by Italian-affiliated PIs.

The Benchmark and Development calls are aimed at supporting researchers and HPC users in testing, benchmarking, developing, and optimizing their applications. During the 5 production months of 2023, 62 projects were allocated on Booster partition, totaling 6.944.000 Core Hours.

Projects from the Extreme and Regular calls with a cut-off in November 2023 underwent peer-review processes, and the awarded projects are anticipated to be allocated in March 2024.

According to the EuroHPC agreement, 10% of the EuroHPC dedicated computational resources have been awarded to Destination Earth, corresponding to 18.048.000 Core Hours.

Accepted projects	call 11/2022	call 04/2023	call 06/2023	call 07/2023	call 08/2023	call 09/2023	call 10/2023	call 11/2023	call 12/2023	Destination Earth	Total
Extreme call (3 calls: 11/2022 - 04/2023 - 11/2023)	7	5						Peer review process			12
Regular call (2 calls: 07/2023 - 11/2023)				4				Peer review process			4
Benchmark call (1 per month from June 2023)			1	4	5	3	3	1	7		24
Development call (1 per month from June 2023)			3	5	6	4	11	2	7		38
Destination Earth										4	4
Total accepted projects	7	5	4	13	11	7	14	3	14	4	82
Local core Hours	2.000.000	144.659.008	448.000	10.505.920	1.232.000	784.000	1.568.000	336.000	1.568.000	18.048.000	377.580.928

EuroHPC projects awarded and corresponding assigned resources, in local core hours, per call.

ISCRA

Italian SuperComputing Resource Allocation

Paola Alberigo
Cineca

ISCRA (Italian SuperComputing Resource Allocation) grants computational resources and technical support to PIs affiliated to Italian research institutions. It provides HPC resources through competitive calls: Class C projects (for code testing, pre-production and benchmarking) are evaluated on a technical basis and Class B projects (full production) both on technical basis and on scientific merit.
In 2023 a new class of ISCRA projects has been launched, where PI's can apply for storage space.
The space offered should be used to store data related the HPC simulations and requests could be up to 50 TB of data storage on File System and/or 100 TB of Archive.
ISCRA D calls will be launched twice a year with the same schedule of the B calls.
Since August 2023 the BOOSTER partition of the new LEONARDO machine is available to Italian Researchers trough ISCRA calls, the General-Purpose partition has been made available in ISCRA from spring 2024.
In 2023 CINECA issued ten ISCRA C calls, two ISCRAB and D calls.
Two IscraB calls received a total of 289 application reported in the table below:

	call 27B Spring 2023	call 28B Autumn 2023
Accepted	64	73
Rejected	47	33
Total number submitted projects	**289**	
Total accepted projects	**152**	

The assigned resources in local core hours are the following:

	Call 27B	Call 28B	Total
G100	25.819.722	29.997.918	55.817.640
LEONARDO_B	77.822.904	68.416.712	146.239.616
LEONARDO_DCGP		47.743.638	47.743.638

We would like to thank the Scientific Panel who served CINECA in last four years and the over one thousand independent referees that contribute with their experience to the peer-review process.

New Supercomputing Facility for Fusion Energy Research

Agreement signed between Eurofusion, ENEA and Cineca to install in Italy EU supercomputer dedicated to nuclear fusion

Paola Alberigo, Patrizia Coluccia
Cineca

On June 8, 2023, CINECA, EUROfusion, and ENEA signed an agreement for a €50 million investment over five years to establish a new high-performance computing (HPC) infrastructure dedicated to fusion energy research. This project aims to enhance the scientific community's ability to perform complex simulations and analyses, essential for developing fusion as a sustainable energy source.

Fusion energy, often described as the "energy of the stars," is pursued for its potential to provide a nearly limitless, clean energy source. European fusion researchers aim to replicate the processes that power the sun to produce energy safely and sustainably on Earth. Fusion research began in the 1950s, and today, EUROfusion is a member of ITER, which represents the largest international effort in this field, involving multiple global partners.

Expected to be operational by the end of 2024 at CINECA's headquarters in Casalecchio di Reno, Bologna, PITAGORA the new supercomputer will have a performance capacity of approximately 47 petaflops. This facility will support numerical simulations of plasma physics and structural analyses of advanced materials, crucial in the research and development of nuclear fusion technology.

The new HPC facility will bridge the EUROfusion scientific community to the Italian ecosystem of international relevance and strategic importance for Europe in the Bologna Technopole, sitting alongside the computing facilities for weather and climate forecasts of the European center ECMWF and the European supercomputer Leonardo.

The project includes a 47 petaflops supercomputer divided into a general-purpose partition (13.6 petaflops) and an accelerated partition (33.7 petaflops), along with an additional smaller computing system. These resources will support the European fusion scientific community, enabling researchers to conduct more advanced and comprehensive studies.

The collaboration between CINECA, EUROfusion, and ENEA highlights the importance of coordinated efforts in scientific research. It begun in 2016 by providing a partition of the Marconi supercomputer to the nuclear fusion science community. In 2020, the Marconi system was ranked among the top 10 most powerful in the world, paving the way for the realization of the Leonardo project, now ranked 7th in the world, and the creation of this new supercomputer dedicated exclusively to fusion research.

ITALY'S CONTRIBUTIONS TO FUSION RESEARCH

Italy has been actively involved in the global fusion research community, notably through projects like the Divertor Tokamak Test (DTT) machine at the ENEA Research Center in Frascati. The new supercomputing facility at CINECA will further bolster Italy's contributions to fusion research, supporting both national and international projects aimed at making fusion a viable energy source.

PARTNERS AND THEIR ROLES

CINECA, Italy's largest computing center, founded in 1969, provides extensive support to the scientific community through its powerful technological infrastructure. ENEA coordinates the national fusion research program and partners with the European Agency Fusion for Energy (F4E) and EUROfusion. EUROfusion, established in 2014, coordinates fusion research activities among 31 research organizations from 29 European countries, facilitating collaboration and innovation in the field of fusion energy.

LOOKING AHEAD

The establishment of this new HPC facility marks a significant step forward in fusion energy research. By providing the necessary computational power and fostering collaboration among key research institutions, this initiative aims to accelerate the development of fusion technology, bringing us closer to a sustainable energy future.

Figure 9: EUROfusion consortium.

This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grant Agreement No 101052200 — EUROfusion)

Agreements

Maurizio Cremonesi
Cineca

Iscra, EuroHPC and other peer-review calls are great means to allow valuable HPC resources to Italian and European researchers, but may not be as suitable if large quantities of stable resources are desired or if researchers and teachers are willing to carry on small experimental activities, student training and other jobs that lack a defined goal.
Whatever the reason, also in 2023 several Universities and Research Institutions decided to make agreements to acquire some HPC resources from Cineca. These agreements grant access to computing resources from one to a few years, without the need for researchers of submitting project proposals.

In total almost 240 Mcore-hours were granted for agreements on Cineca computing platforms. The largest budget corresponds to INFN, again at top position this year. ICTP follows with roughly half resources, then Università degli Studi di Milano Bicocca, SISSA are a little far behind; INAF also holds important agreements.
Just to summarize, non academic Research Institutions sum up a total 199,5 Mcore-hours, Universities an other 33,8 Mcore-hours and Life Science Institutions hold a remaining 6,4 Mcore-hours grant.
A list of the Institutions with active agreements in 2023 are listed below.

Institution	Award resources (Mcore-h)
INFN (Istituto Nazionale di Fisica Nucleare)	106,00
ICTP (International Centre for Theoretical Physics)	45,46
Università degli Studi di Milano Bicocca	24,43
SISSA (Scuola Internazionale Superiore di Studi Avanzati)	20,06
INAF (Istituto Nazionale Astrofisica)	13,20
Centro Euro-Mediterraneo sui Cambiamenti Climatici	8,15
Istituto Nazionale di Oceanografia e di Geofisica Sperimentale	5,76
AIRC	2,73
Elixir (Distributed Infrastructure for Biological Data)	2,73
Università degli Studi di Trieste	1,46
Università degli Studi dell'Aquila	1,36
Politecnico di Milano	1,27
Università degli Studi di Milano	1,74
Università di Bologna	1,21
IIT (Istituto Italiano Tecnologia)	0,82
Università degli Studi di Genova	0,75
Politecnico di Torino	0,56
Università degli Studi di Brescia	0,52
INMI (Istituto Nazionale Malattie Infettive Spallanzani)	0,40
Università degli Studi di Bergamo	0,30
Ospedale Pediatrico Bambin Gesù	0,28
Telethon	0,26
Scuola Superiore Sant'Anna	0,16

Training HPC

Francesco Falciano
Cineca

Training has always been a significant commitment and distinguishing feature of the activities carried out by Cineca for the Italian research community and users. Since 2012, it has been extended to European researchers, as Cineca has been recognized as a PRACE Advance Training Center (PATC) in HPC.

In 2023, Cineca's contribution to the PRACE project came to an end, but the commitment to ensuring internationality for some courses remained. This was achieved not only by collaborating in other international projects but also by simply delivering the most distinctive courses in English so that international students could participate.

The years of the COVID-19 pandemic, which forced online teaching, had the positive effect of developing distance learning technologies and also accustomed students and teachers to this type of teaching. For this reason, in 2023, alongside the return to in-person classroom teaching, courses were delivered in three modalities: online, in-person, and hybrid (in-person and streamed). The educational platform adopted by Cineca is Microsoft Teams. In the course of 2023, 15 courses and 4 schools were held, in addition to other teaching activities and/or participation in the organization of events by other institutions. 584 Italian and international students actively participated in the courses and schools, meeting the minimum attendance requirements, although online survey data revealed that several hundred more attended just one course lecture. In addition to these, there were participants in uncertified Cineca events for which attendance was not recorded but (based on registration data) amounted to more than 300.

The level of satisfaction measured in the questionnaires administered to participants revealed an average rating of 8.9/10 in the overall evaluation of the educational event, with individual scores for each teaching consistently above the standard set at 7/10.

The HPC Department staff was also involved in partnerships with Research Institutes and Universities, providing support in organizing and teaching Master's and PhD events, workshops, tutorials on HPC, Big Data, and Artificial Intelligence themes, and organizing scientific and training events in collaboration with external companies and international projects.

These are the "numbers" of our training activities:

- 584 certified Italian and European Students
- More than 300 not certified who attended at least one course lecture
- 78 Days of lectures
- 50 Teachers
- 15 classes
- 4 Schools
- Collaboration in events
- Guided tours

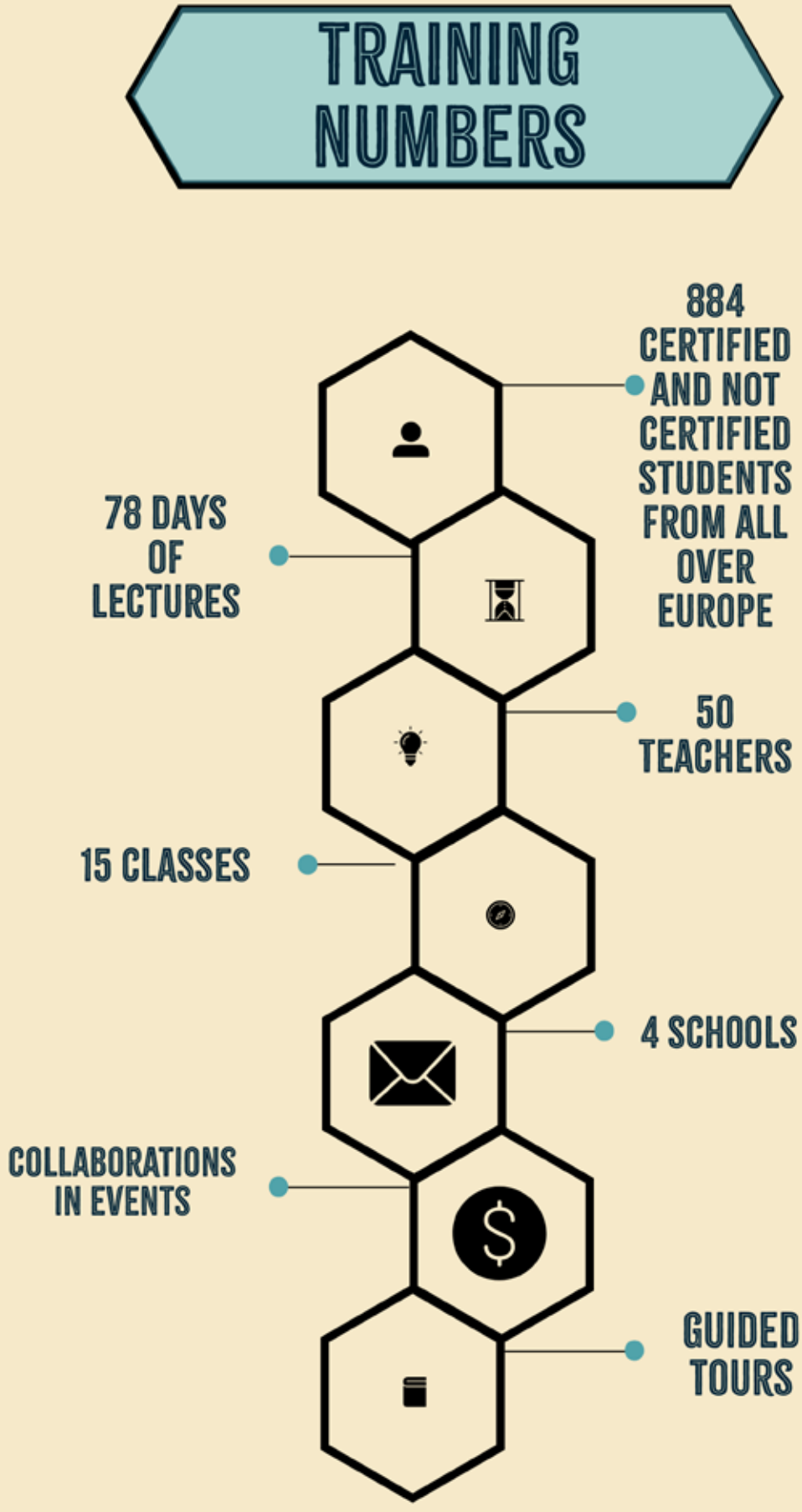


Figure 10: numbers in training.

2024 PLAN

In the second part of 2023 we tried to take back all of our courses in classroom fashion. Nevertheless pandemic experience forced all teaching institutes and teaching platforms to improve their technology to enrich remote courses of the advantages of an in classroom event. This is not always reachable. On the other hand in classroom events can give lessons just to a number of people limited to the phisical seats of the classrooms. For this reason in 2024 our objective is to get back all the courses to in classroom , expecially where interactions between people were the main goals, but for many courses we will try to do also online connections with advantages to make lessons available for remote students who can also interact live with teachers and trying to develop a sort of “hybrid” online-in presence way of teaching.

Of course we will improve the way of interactions between students and trying to minimize the disvantages of a remote connection.

Our effort is constantly focused in increasing the number of courses and schools developed by Cineca HPC Department, teaching the most new technologies available and trying to admit the largest number of students worldwide. For doing this we will try to admit as much students as possible, giving them access to tutors’ support and, for each course, a reasonable numbers of auditors giving them access at least to an email support service. For some courses we will open online forums so the students will be able to discuss and support each others with the tutoring of our experts. We will try to record and put online large part of the lessons, tentatively in English language, making them feasible.

Furthermore we are experimenting new ways of interaction between people remotely connected like closed rooms of discussions with few students.

STAFF

In the HPC department of Cineca work 160 people, distributed among the three premises: 126 in Bologna, 11 in Milan and 22 in Rome and 1 in Naples.
Several new colleagues were enrolled in 2023: we wish them a very fruitful and interesting job.

LUCA BABETTO

Luca Babetto graduated in Chemistry at the University of Padova, and subsequently obtained a Ph.D. in Molecular Sciences at the same university, with a thesis on theoretical investigations of lanthanide-based luminescent molecular thermometers. He worked on redox flow battery research at Green Energy Storage within a project funded by IPCEI, in which he developed a high-throughput workflow for the discovery of new organic electrolytes. He is currently working as HPC Specialist in the POC business unit.

ELISABETTA D'AIUTOLO

Elisabetta got a degree in Computer Science at "La Sapienza" University in Rome.
During her activity in academics and research she worked on system and network security at National Research Council.
Main expertise: automation and scalability of a hybrid infrastructure, IT automation and DevOps operations, management of academic and industrial systems.
She is currently member of Sys Admin group in HPC department at CINECA.

FRANCESCA GEBBIA

Francesca Gebbia obtained the Bachelor degree in Physics at the University of Palermo and the Master degree in Theoretical Physics at the University of Bologna. Later on, she pursued her theoretical physics studies with a focus on open quantum systems and dynamics at the University of Trieste, where she achieved her Ph.D. in 2022. Her research production centered on completely positive maps, quantum entanglement, quantum metrology, and quantum control protocols.
Since 2023, she has been working at Cineca as HPC specialist and project manager in the PoC solutions unit.

MICHELE VISCIARELLI

Michele Visciarelli has obtained a BSc/MSc in Electronic Engineering at the University of Bologna, and a PhD in Molecular Modelling at the Scuola Normale Superiore in Pisa. After 6 years of post-doc, at University of Bologna and KTH Stockholm, working on simulations on nanodevices and magnetic materials, he worked for Ericsson in Stockholm for 3 years as a data scientist, developing machine/deep learning models for demand/supply planning. Since 2023, he joined Cineca as Data Scientist, focusing on projects involving large language models.

LEONARDO DE CARLO

I have a background in mathematical physics and graduated from the Department of Physics at “La Sapienza”. I got a Ph.D. in Mathematics from GSSI/SISSA, on stochastic processes and discrete mathematics. I worked as Math Researcher at IST in Lisbon, SNS in Pisa, and LUISS in Rome.
In the past I also taught to high school students.

FEDERICO VENTA

Software Developer with a strong passion for Artificial Intelligence, Machine Learning, and Data Science. Experienced in architecting robust software and crafting high-quality applications. Holds a Master’s degree in Theoretical Physics from the University of Bologna. Currently contributing as a HPC Data Engineer at CINECA, following a previous role at Accenture as a full-stack Developer and Functional Analyst.

LUCA MATTEI

Luca is a Data Scientist and Machine Learning Engineer who got a BSc in Statistics and a MSc in Data Science, both from “La Sapienza”.
In order to write his Master Thesis, titled “AWE-T, a new Adaptive Weights Ensemble using Transformers: an application on Twitter data”, he collaborated with Cineca as an intern. Later in 2023, he joined as HPC Data Engineer, focusing on Machine Learning and Deep Learning on a cultural heritage project.

MANDANA SAFARI

After finishing M.Sc. in Nanophysics (Iran), I pursued a Ph.D. at SISSA, exploring charge transfer mechanisms at surfaces using DFT, DFPT, genetic algorithms, and NNIP in phonon calculations. Completing a master in High-Performance Computing at SISSA-ICTP, excelled me in applying neural networks to predict Quantum Espresso time-per-call on supercomputers. As an HPC Specialist at CINECA, I contribute to the MAX project, leveraging high-performance computing for scientific advancements.

MARCO PUCCINI

Marco got a degree in Physics at La Sapienza University of Rome designing and developing a FEM model optimization procedure for historical buildings and the remoting of experimental tests on earthquake shaking tables. He got a Scholarship at CNR INM (Institute of Marine Engineering) working on data analysis from experimental fluid dynamics. He taught at High School Math, Physics and Computer Science meanwhile collaborating at Molecular Medicine Department of La Sapienza with data analysis for novel PET scintillator crystals testing. Marco got a Research Fellowship in the HPC Laboratory at ENEA working on Big Data management and DevOps in a wide range of projects. Currently he is an HPC Data Engineer at the HPC Data Managment group at CINECA.

LORENZO SERAFINI

Lorenzo, born in 1997, has studied Astrophysics at “Università di Firenze”,
getting his master degree in February 2023 with a thesis involving turbulence in relativistic environments (GRMHD). During the writing of his thesis he managed to accelerate using GPUs the ECHO code for GRMHD (originally written for CPUs), reaching an acceleration of 16 times. Developing interest in the HPC architecture and environment, he works for CINECA since February 2023 in the industrial support department.

FLAVIA ZANON

Flavia is a project and grant manager specialized in the management of R&D and innovation projects. She has a long experience in the management of national and European grants and, in CINECA, she is devoted to enhancing the impact of HPC technologies on the research and innovation capabilities of the private sector and SMEs. Flavia holds a PhD in International Relations from the University of Trento and specialisation on EU public policies.

SARA MARZELLA

I graduated in applied mathematics, bachelor's degree at La Sapienza in Rome, master's degree Bicocca in Milan. I specialized in Machine Learning, then moved on to Quantum Computing and started an adventure in science dissemination. All this led me to land in the Quantum Computing Lab of Cineca, where I work on integration, application and dissemination of hybrid HPC-QC systems.

MASSIMO GISONNI

Massimo holds a PhD in Geometry and Mathematical Physics from SISSA, specializing in the connection between random matrices and combinatorics. He is enrolled in the MHPC program at the International Center for Theoretical Physics in Trieste, nearing his thesis' completion. In 2023, Massimo joined Cineca as an HPC Scientific Application Engineer, contributing to the development of containerized applications and workflows within the field of weather forecasting.

NANDHANA SAKTHIVEL

After completing bachelor's degree in computer science and engineering and working as a software development engineer in India, Nandhana moved to Trieste to pursue a master's degree in computational science and engineering with a specialization in quantum computing from University of Trieste, ICTP, SISSA and University of Udine. Then, she was working as a Research engineer at Barcelona Supercomputing Center before moving to CINECA to work as a HPC scientific application engineer since November 2023.

Events

The European Researchers' Night of Society project

Simona Caraceni
Cineca

The well-lo SOCIETY Consortium organised the 2023 edition of the European Researchers' Night (ERN) within the framework of the project SOCIETY riPENSAci. The project title, and main theme, is an invitation to think twice, i.e. to revisit our own knowledge of the world, opening ourselves to different points of view and different forms of understanding. The project sets the ERN as the venue where visitors can explore the world through the eyes of researchers from different disciplines, experiencing different perspectives and challenging common misconceptions.

As in the previous edition, the events organised for the ERN took place in various cities in the Emilia Romagna region, including Bologna, Cesena, Faenza, Ferrara, Forlì, Ravenna, Rimini. ERN activities were proposed by researchers from the SOCIETY partners, including the University of Bologna (UNIBO), or from the local branches of National Research Institutes, INAF (Istituto Nazionale di Astrofisica, OAS-IRA), INFN (Istituto Nazionale di Fisica Nucleare, Sezioni di Bologna e Ferrara, CNAF), INGV (Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna), CINECA (Consorzio Interuniversitario – Dipartimento di Supercalcolo), under the coordination of CNR (Consiglio Nazionale delle Ricerche – Area Territoriale di Ricerca di Bologna). All activities were organised and promoted with the substantial contribution of the two Ltd companies ComunicaMente and Naxta that participate in the consortium. The collaboration with the University of Ferrara granted once again the setup of the ERN in Ferrara.

Several exhibits were proposed by prestigious research institutions, such as ISTAT, ISPRA, ENEA and many more, that hold long-lasting collaboration with the project even though they do not participate in the consortium.

Considering all venues, hundreds of activities were proposed to the public by more than 600 researchers. The effort was rewarded by a great turnout, with almost 10000 people attending.
Cineca was present with 3 activities.
SUPERCOMPUTING, THE FRIEND OF RESEARCH. Researchers proposed a journey to discover Supercomputing and the cutting-edge research carried out with modern computing power. Visitors could see what it means to compute millions of billions of operations per second, and the state of this fascinating technological challenge, with an eye on the new super-friend Leonardo.
QUANTUM COMPUTING: AN EMERGING REALITY. Team quiz, online quantum game, Quantum Computing-themed video playback.
VISUALISING AND EXPLORING COMPLEXITY. From science to cultural heritage, from the micro to the macro. Computer graphics, computer applications and artificial intelligence to facilitate and enhance research results.

Events
“Festival della Scienza” in Genoa

Neva Besker
Cineca

Cineca participated in the Festival of Science, one of the most prominent events for promoting scientific culture, in Genoa from October 26th to November 5th. Over the years, it has evolved into a globally recognized cornerstone for science enthusiasts. For eleven days, the city of Genoa opened the doors of its wonderful cultural venues to host a rich programme of conferences, exhibitions, workshops, shows and other events. Total attendance was 200,000 people, with the public coming from 10 regions.

This year, CINECA introduced the interactive workshop titled ‘If I were a Supercomputer?’ based on a game where visitors become integral components of a computer to solve complex problems collaboratively. It was a funny and interactive way to learn and understand more about supercomputers, especially Leonardo, Europe’s second most powerful computer. It was an opportunity to introduce to the general public to the importance of high performance computing, some of Cineca’s core projects and future challenges in research and industry.

The workshop was hosted at the Palazzo della Borsa, together with the other workshops dedicated to Artificial Intelligence and algorithms.

The CINECA workshop had over 1,500 attendees, including schools and the general public, and always with a lot of curiosity and positive feedback.

Figure 11, 12: ERN in Bologna.
Figure 13: Festival in Genoa.

Events
The European Researchers' Night in Rome

Neva Besker
Cineca

Cineca with all the major local (Lazio) Research Institutes and Universities, such as CNR, Ispra, INAF, INFN, INGV, ENEA, University of Rome Sapienza, University Tor Vergata, University of Tuscia and Telematic University UniNettuno forms a partnership NET- scieNcE Together to spread scientific culture and knowledge of the research professions in an informal and stimulating context.

The partnership NET, with the CNR as a coordinator, organized, in 2023, the third edition of the European Researchers' Night in Viterbo and Rome at Città dell'Altra Economia (CAE). The Europen Researcher's Night (ERN) is an initiative promoted by the European Commission since 2005 that involves thousands of researchers and research institutions every year in all European countries. The goal is to create opportunities for citizens to meet the scientific culture and knowledge of the research professions and for researchers to disseminate their works and scientific expertise in an informal context.

The activities carried out for the NET 2023 Night included a program dedicated to students of Italian primary and high schools mainly from Rome. CINECA has carried out interacting activities in primary and secondary schools on supercomputing ("If I was a supercomputer", "Bit wars" and "Supercomputer, researcher's best friend"). Moreover, CINECA has collaborated with INGV to realize the Children's drawings calendar with the theme "A future for the planet" involving about 200 Italian primary schools.
NET Consortium organized the final event principally in the centre of Rome (Testaccio) inside the "The City of the Other Economy - CAE" which is one of the first spaces in Europe entirely dedicated to the promotion of the other economy. CAE is a big outdoor place where thousands of people could be hosted even according to the anti-covid rules. The final event was organised for two nights September 30th and October 1st (Friday and Saturday) from 18:00 to 24:00. CINECA was present with its stand, with videos of hardware facilities and HPC applications in research and industry, as well as virtual reality and interactive games to explain parallelism and binary calculations. Researchers involved in the final event inside CAE welcomed about 5150 visitors with a wide offer of activities and about 30 stands. Many activities were designed for families and general audiences, offering hands-on activities, games, science demonstrations, and informal conversations with researchers and science entertainment.

Figures 14, 15: European Researchers' Night in Rome.

Events
Supercomputing 2023: confirmations and novelties

Daniela Galetti
Cineca

The location was a confirmation: Denver. For most of the Cineca SC23 team was a return, finding the well-known Convention Center, with the blue bear standing at the glass wall, the restaurants, the hotels, where meet vendors, users, customers and all the interesting HPC software and hardware innovations.
The booth was a confirmation for the structure, but as every year, new panels confirmed the importance of Leonardo and his (for us is not a thing...) rank at the top of the top500 list (only a shift of two positions, from 4 to 6!). For the first time since many years, all the team where there for the booth construction and we realized that there's an optimal number of people that can collaborate to complete the booth...It was also the first time that Cineca and INFN had noncontiguous booths, compensate by the fact that next year we will have contiguous booths without doing any change.
The gadgets were also a novelty, not only why we change them every year, but because for the first time we have decided to have two levels of gadgets: one for all and one for selected people.
The idea to share these details not strictly High-Performance Computing consists in the intent to show the big work of designing and the sequence of many small and big decisions that occurred during the year before the SC show and that drive the Cineca team to a successful experience in the USA exposition.
As usual the booth was also the occasion to do presentations about our supercomputers, users and applications success stories, sharing experiences with colleagues from the worldwide HPC sites.
From the conference side, AI was everywhere.
All the vendors, no matter if they produce and or sell hardware, software, product support, spoke about AI.
They have many talks about AI, how they do AI, how they use AI.
Recycling for the AI a Dan Ariely quote about Big Data, we could say:
AI is like teenage sex; everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it.
Another hot topic at SC was Quantum Computing, like recently past years with a visible increase of quantum start-ups.
As usual, all the standard HPC topics were treated too and we came back home with a lot of interesting information about vendors' roadmaps, new releases (or absence of it, that it is a news, too).

Figure 16: Leonardo delegation at Supercomputing.

Events

Workshop AI, Cultural Heritage and Art – A topic for endless thoughts

Antonella Guidazzoli, Maria Chiara Liguori
Cineca

Since the first edition, held in January 2023, VisitLab Cineca has organized a bi-annual workshop dedicated to reflecting on artificial intelligence applied to cultural heritage and art. The wide range of guests invited to speak on the topic made it possible to approach the discussion from different perspectives, which was the starting point for the following meeting.
Both the streaming - in Italian - and the proceedings - in English - of each event are available. The proceedings of the two-day workshop held in February 2024 will be available for the opening of the next workshop edition which is scheduled in September 2024.

Figure 17, 18, 19: posters of the workshops with their QR for streamings and proceedings. The first QR is alwais for streaming, the second for the proceedings of the Workshops.


AI, BENI CULTURALI
E ARTE
TRA RICERCA
E CREATIVITÀ
SECONDA EDIZIONE
22 SETTEMBRE 2023
CINECA, BOLOGNA - SALA FIBONACCI
CINECA
S·T·ARTS
SCIENCE · TECHNOLOGY · ARTS
European Commission
Society
NOTTE EUROPEA DEI RICERCATORI

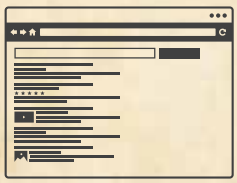

AI e Beni Culturali. Tra ricerca e creatività
20 gennaio 2023
Cineca, Bologna, Sala Fibonacci
CINECA

HPC H-index

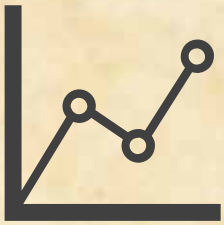
WebofScience https://www.webofscience.com/
Database: Web of Science Core Collection
Search: (FT=Cineca or FO=Cineca) and (PY=2023-2024)
Date: 28.05.2024

Results found: 841

Number of Citations: 1680



Average Citations per Item: 2



H-index: 14

Information obtained by analyzing the publications in
“Web of Science Core Collection”, in all the science and engineering fields, citing Cineca, from year 2009 to 2023.
Results found: 6.696
Number of Citations: 177.044
Average citations per item: 26,44
H-index: 139

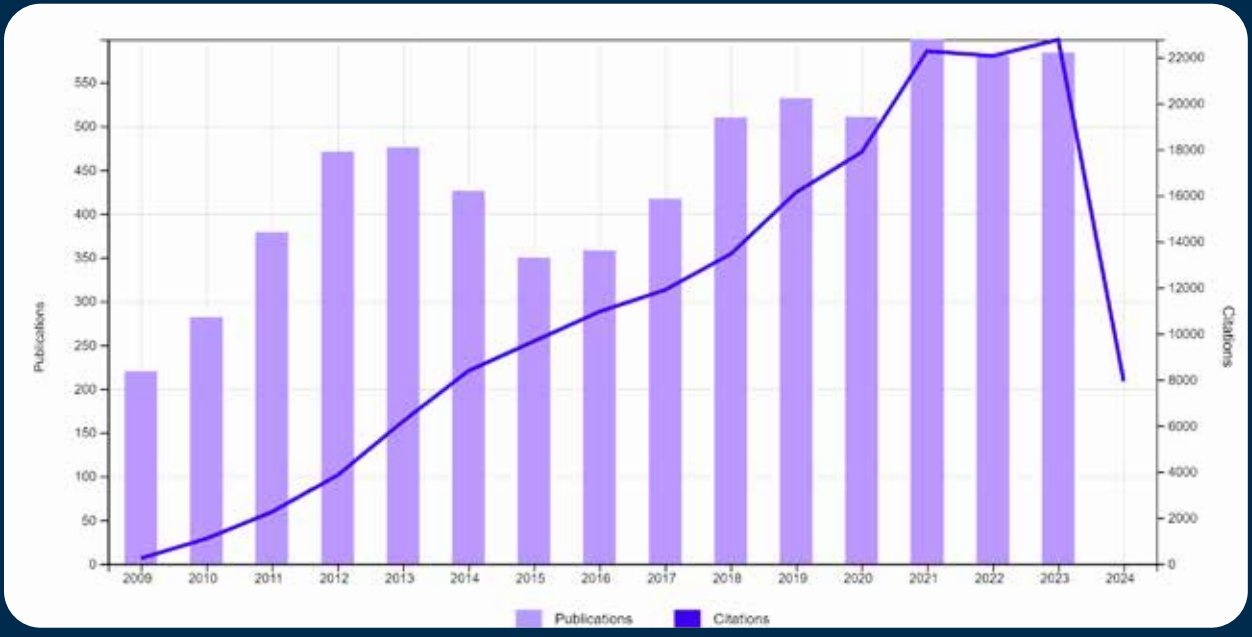


Figure 20: number of publications and citations from year 2009 to 2023

Figure 21: Top 15 affiliation of the authors

Figure 22: Top 15 countries of the organizations of authors

Figure 23: Top 15 research areas of the publications.

SCIENTIFIC OVERVIEW

Cineca is involved
in a wide scientific field
thanks to EU projects and
collaborations.

From artificial intelligence
to urgent computing,
bioinformatics and digital
humanities, HPC is used
to support them all.

LIFE SCIENCES

Towards personalized medicine: Ultra-fast Genomic Data Processing in a HPC Cloud-Secure Environment

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

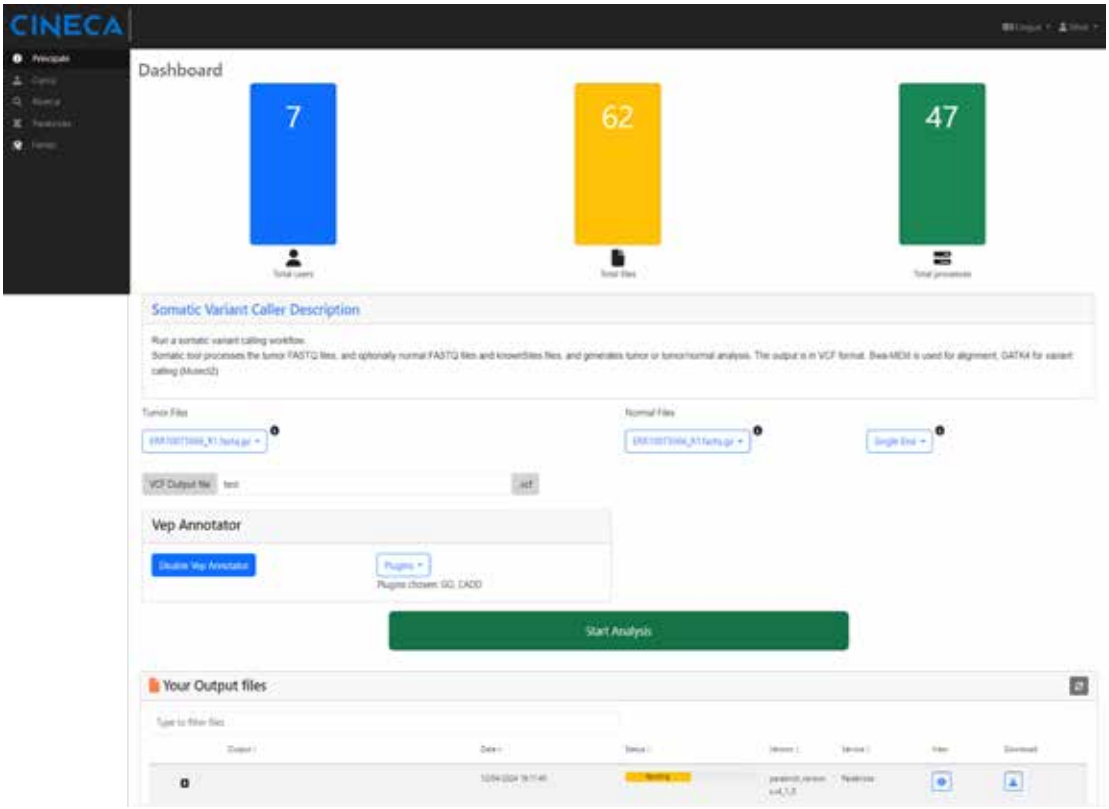


Figure 24: overview of Omics platform main Page. On the left side, the user menu is shown. It includes data upload, quality check/visualization and data analysis pipelines. On the right panel, a general usage statistics is shown, as well as the current status of a job.

The advancement of "-omics" technologies has led to an exponential surge in sensitive data volumes, necessitating secure storage and processing facilities. While these data hold potential for innovative progress in personalized and precision medicine, stringent security measures are mandatory to ethically safeguard individuals' data privacy. Traditionally, genomic data analysis implied significant time and computational resources due to data complexity and operations involved.

In this context, High-Performance Computing (HPC) and Graphics Processing Units (GPUs) emerge as indispensable tools for parallelizing and enabling ultra-fast variant calling analyses. For instance, NVIDIA Clara Parabricks is a software suite tailored for accelerated both individual and population genomics data analysis utilizing GPUs' parallel processing capability, which makes them ideal for analyzing extensive genomic datasets thus ensuring easy scalability of analyses. To streamline genomic data analysis while prioritizing user experience, data privacy, and accelerated processing through GPU technology CINECA is developing an intuitive and user-friendly interface, accessible at omics.cineca.it, which will facilitate Italian research centers to easily analyze patient's data. This service greatly allows the upload and processing of entire human genomes in approximately 45 minutes, a stark contrast to the roughly 30 hours required for 30x whole genome sequencing data.

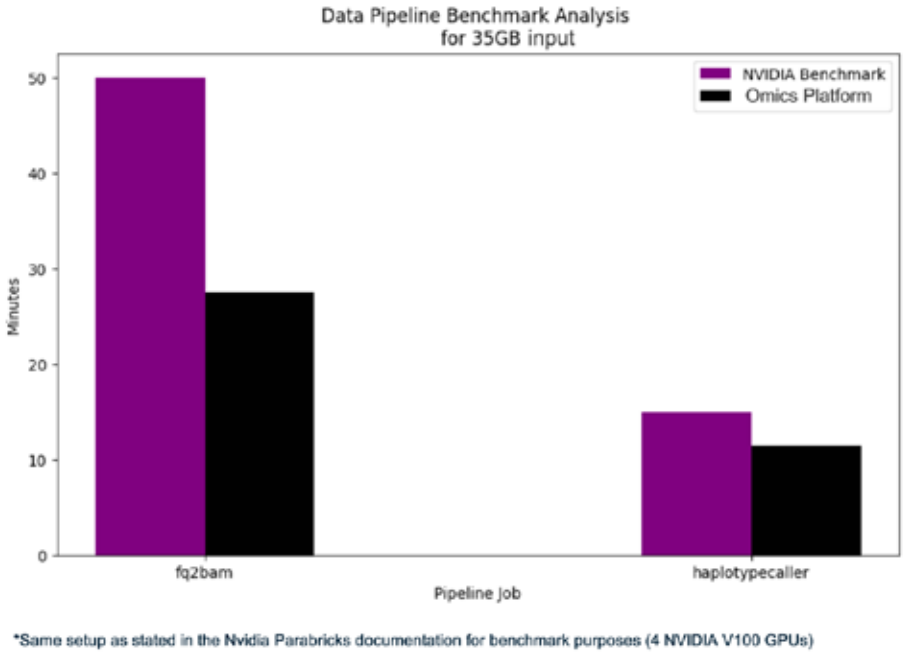


Figure 25: time benchmarks for two of the four data analysis pipilenes available on Omics platform, specifically FQ2BAM and Haplotypecaller. The tests have been performed with a 35GB paired-end input file and computing times are reported as minutes on y-axis. The purple bar-plots report the overall computational time reported on NVIDIA Clara Parabricks documentation and the black bars report Omics performances, thus showing that the second one can reach even better performances with CINECA's HPC Cloud infrastructure.

Further variant annotation for predicting Variant effect on human's health and phenotype is facilitated using downstream bioinformatics tools, as well as the experimental quality of data assessment. Upon completing analyses, users can both easily download final VCF files for data post-processing or explore genomic variants through CINECA's interface using an Interactive Genome Viewer tool for VCF visual inspection.

CINECA has adopted cloud computing solutions for the management of sensitive and personal data in the specific field of life sciences, extending the ISO 9001 (Quality Management Systems) and ISO/IEC 27001 (Information Security Management Systems) certifications to this area in November 2022 as well. Therefore, the new architecture under deployment is going to integrate into these security boundaries, thus ensuring data security and privacy compliance with EU GDPR regulations.

Specialised training in HPC for life sciences

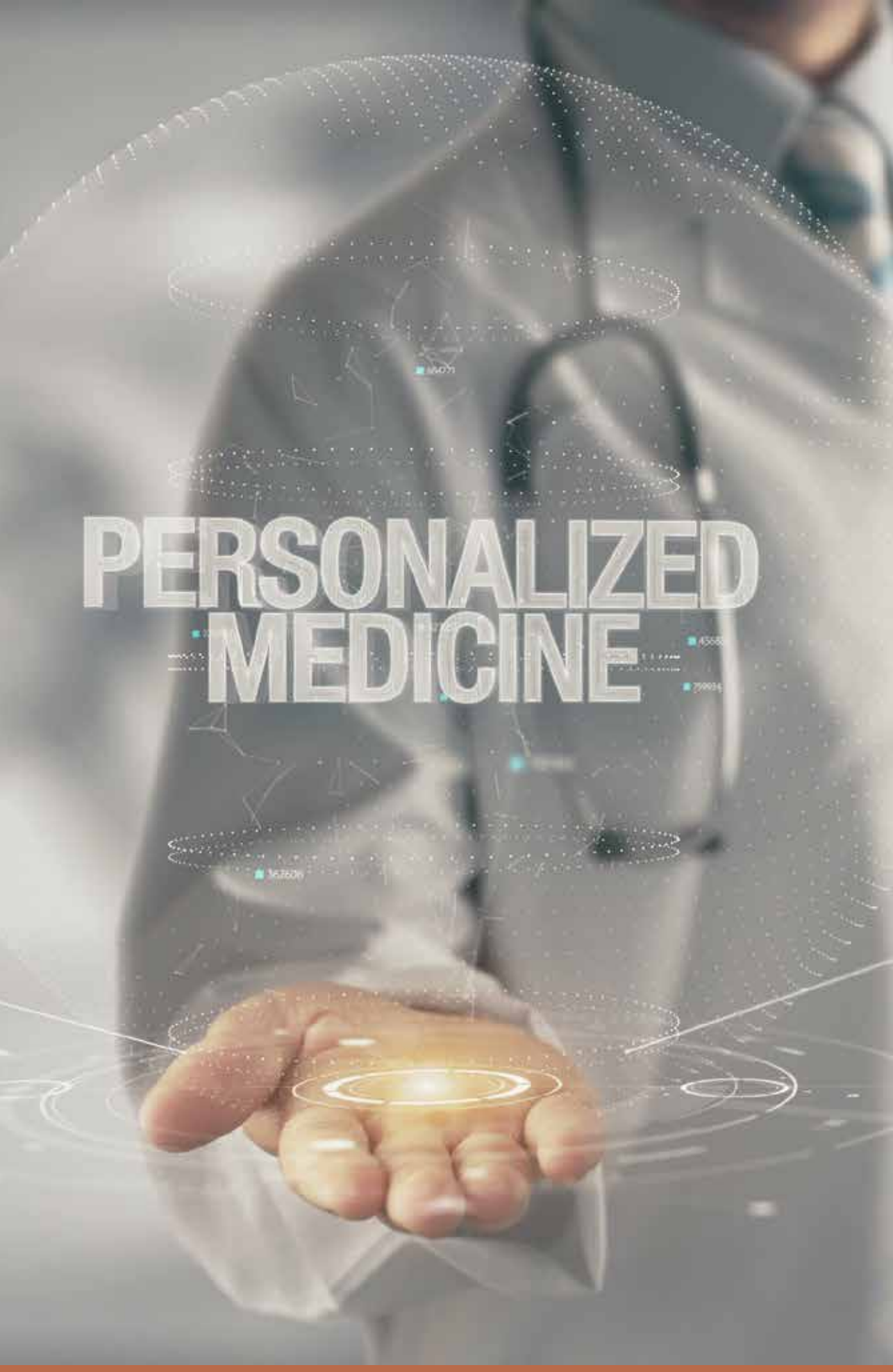
Balasubramanian Chandramouli, Silvia Gioiosa, Elisa Rossi
Cineca

The sequence-structure-function paradigm is paramount in biology. According to user's needs and understanding the importance of these topics for the research, the HPC Department offers two courses: High Performance Bioinformatics and Molecular Modeling. These courses address issues related to various optimization techniques for fine-grained simulations on HPC facilities.

The HPC Bioinformatics course introduces students to fundamental supercomputing concepts and technological solutions for processing and analyzing large-scale genomics and transcriptomics data on state-of-the-art HPC facilities. Molecular dynamics simulations, employed to study the dynamic behavior of biological macromolecules, pose computational challenges as well. These simulations involve solving complex equations of motion for thousands or millions of atoms over time, requiring advanced algorithms and significant computational power.

Finally, under patronage of ICSC National Research Centre for High Performance Computing, Big Data and Quantum Computing, an innovative training course has been launched in 2023, namely RE-TRAIN-ME, organized by CINECA supercomputing Centre, COMBINE Group, University of Catania, Concept Lab, Istituto Italiano di Tecnologia, Laboratorio di Tecnologia Medica and Istituto Ortopedico Rizzoli.

The project aims to provide post-degree training in biomedical computing to people with a previous master's degree in math, physics, chemistry, biology, engineering, needing to enforce their knowledge in biomedical computing. The main topics covered by the course include Introduction to Computer Science, Programming, Introduction to Biomedical Computing, Medical Informatics, Medical data science, Bioinformatics and In Silico Medicine.


PERSONALIZED
MEDICINE

EARTH AND ENVIRONMENTAL SCIENCES

Earth Science

Piero Lanucara
Cineca

The past year has strongly influenced the activities of the CINECA High Level Support Team (Earth sub-team) towards new sectoral and computational challenges. In particular, some of the most relevant outcomes include:

- **ChEESE-2P:** The first year of activity of the Center of Excellence (CoE) was focused particularly in the WP2 led by CINECA. The new task of performance portability (T2.2) has entailed significant learning skills towards some real challenging issues (HPC tools such as OpenMP offload, OpenACC, SYCL, Kokkos, to name a few, as well as libraries like PETSc). From a purely technical point of view, the first year provided an opportunity to start understanding the 11 flagship codes that are the focus of WP2 activities.
- **DT-GEO**: Activities have mainly focused on WP3 (support to the HPC+Cloud activities of the "vertical" Digital Twin Components) and some "porting" activities (such as in the WP5 "volcanos" and concerning the porting of code GALES to GPU using performance portability tools like the aforementioned Kokkos).
- **Geo-INQUIRE**: This project, particularly appealing for its characterization towards "curiosity-driven" science, has begun developing the HPC part of WP5, expecially in supporting the computational activities of the different infrastructures based on top of the ChEESE-2P codes and workflows. This activity is particularly relevant for the incoming Transnational Access (TA) calls targeted to Galileo100 and Leonardo booster, respectively equipped with CPU and GPU (NVIDIA A100).
- **DE360**: The project aimed at conveying activities of porting, tuning, and optimization regarding weather/climate models such as IFS, ecWAM (ECMWF "dwarf") through performance portable directive-based tools in the WP1. Where possible, support is provided for the use and improvement of the infrastructure for the POP based performance assessment of models (see Fig.1), with the methodology adapted to the codes and machines involved in the activity (primarily Leonardo booster, but not only).
- **Other projects**: Active support to other projects like EDITO-Model Lab (where the Earth team will be involved in supporting WP3 activities aimed at improving oceanographic models on EuroHPC architectures), ACROSS (specifically WP6 related to the use of particularly optimized weather/climate workflows for massive I/O), and EuPEX (improvement of a specific version of the flagship code SPECFEM3D on platforms "close" to the future European ARM-based processor) have just begun producing relevant outcomes.

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 nº 101058129.

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

DE_360: "Technology Adoption for DestinE" (Destination Earth)
ECMWF/DESTINE/2022/DE_360_A1

EO4EU

I-augmented ecosystem for Earth Observation data accessibility with Extended reality User Interfaces for Service and data exploitation

Francesco Maria Cultrera, Beatrice Chiavarini, Lucìa Rodriguez Munoz, Balasubramanian Chandramouli, Mattia Carello
Cineca

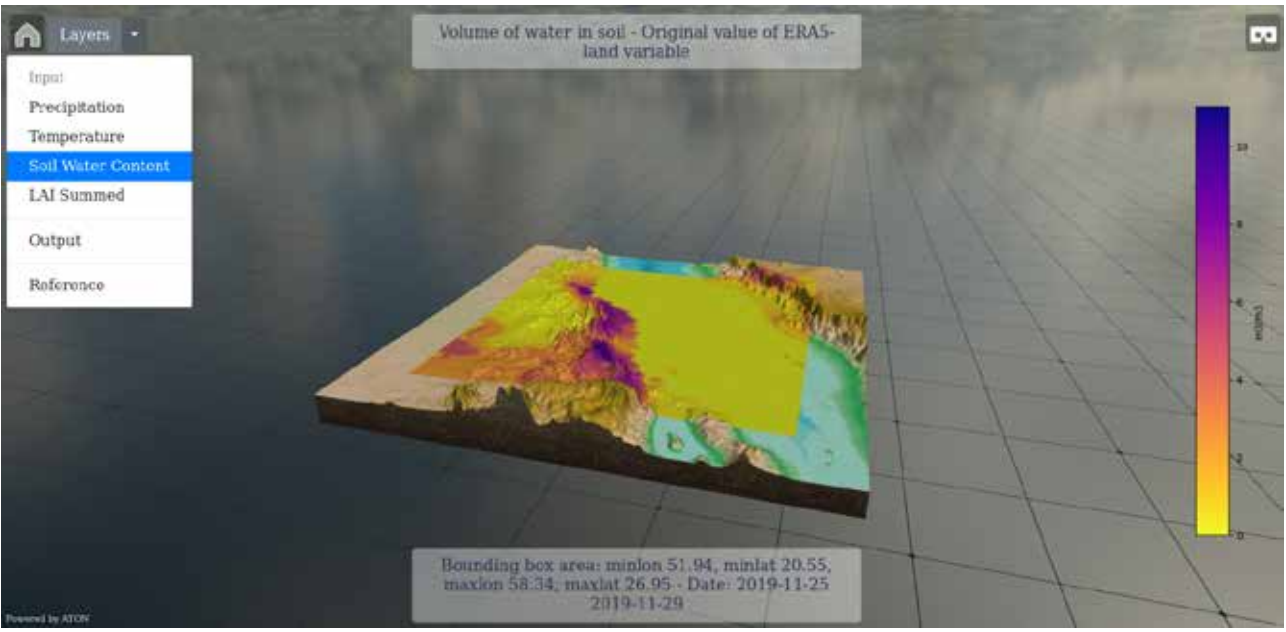
The EO4EU (Earth Observation for Europe) project is an initiative funded by the European Union's Horizon 2020 research and innovation program and it aims to enhance the accessibility and usability of Earth Observation (EO) data for various applications across Europe. The project focuses on creating a more efficient and user-friendly platform for accessing and analyzing EO data, which includes satellite imagery and other geospatial information.

Key objectives of the EO4EU project include:

- Data Integration: Combining different sources of EO data to provide a comprehensive view of environmental and climate conditions. This integration helps in better decision-making processes for various sectors such as agriculture, forestry, urban planning, and disaster management.
- Advanced Analytics: Developing tools and services that facilitate advanced data analytics, allowing users to extract valuable insights from EO data. This includes the use of artificial intelligence and machine learning to enhance data processing and interpretation.
- User-Friendly Platform: Creating an intuitive and accessible platform where users, including scientists, policymakers, and the general public, can easily access and utilize EO data. The platform aims to lower the barriers to entry for using complex EO data.
- Support for EU Policies: Assisting in the implementation and monitoring of EU policies related to the environment, climate change, and sustainable development by providing accurate and timely EO data.
- Collaboration and Knowledge Sharing: Promoting collaboration among various stakeholders, including academic institutions, research organizations, government agencies, and private companies. This collaboration aims to foster innovation and knowledge sharing in the field of EO.

The EO4EU platform relies on two different clouds, one of which is CINECA ADA cloud. CINECA is also tasked at creating the Observability platform, which allows to monitor every component and node in the entire EO4EU platform, and the XR component which allows to visualize the data in a virtual space.

The EO4EU Graphical User Interface includes, among the different visualization tools, an extended reality (XR) web-based application interface developed by CINECA, that enables the possibility to visualize and explore the selected EO data in a more immersive way. The XR interface includes a Virtual Reality (VR) web application and an Augmented reality (AR) web application. The augmented reality (AR) interface enables the visualization and the possibility to analyse the EO data in the physical world taking advantages of the geolocation system of the device where the user will run the application. The virtual reality (VR) interface enables the visualization of the EO data on 3D virtual scenes on hardware designed for supporting VR applications, such as cardboard and VR head-mounted devices. In the case of the VR head-mounted devices the user has the opportunity to move within the 3D scene and interact with the environment using the device hand controllers.



Screenshot of the current state of the Extended Reality (XR) application interface: The requested Earth Observation (EO) data are displayed as selectable layers jointly with a basic set of metadata on the 3D model of the portion of terrain related to their context.

Screenshot of the dashboard of the EO4EU portal from which it is possible to access the XR visualization

Funded by the European Union

EO4EU
Earth Observation Data & Service Accessibility

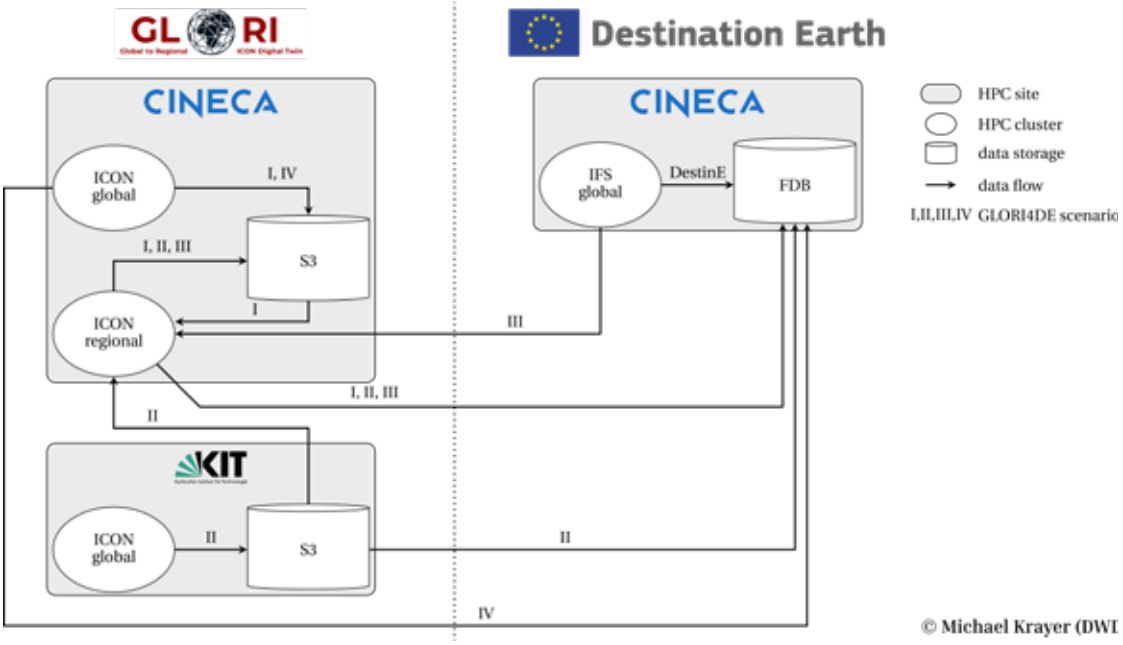
GLORI4DE

Massimo Gisonni, Fabio Di Sante, Gian Franco Marras, Gabriella Scipione
Cineca

GLORI4DE stands for: GLObal to Regional ICON (GLORI) for Destination Earth (DE). GLORI4DE is not just a crasis of two acronyms; its aim is to bring together and make interoperable two different weather and climate models: the ICON model from the COSMO consortium (ICOsahedral Nonhydrostatic weather and climate model) and the IFS model from ECMWF (Integrated Forecasting System).

The GLORI project aims to provide a configurable on-demand global-to-regional short-range high-resolution digital twin based on the prediction capability of ICON. The twin is designed to run on a heterogeneous GPU-CPU architecture and provides short-range global storm-resolving km-scale (~3 km horizontal) predictions using hybrid variational ensemble data assimilation.

Destination Earth is a flagship initiative of the European Commission to develop a highly accurate digital model of the Earth (a digital twin of the Earth) to model, monitor, and simulate natural phenomena, hazards, and related human activities.



GLORI4DE brings together expertise from CINECA, ECMWF, DWD, Arpae, CMCC, and AIM. CINECA's hosted cluster LEONARDO (Bologna, Italy) and KIT's hosted cluster HoreKa (Karlsruhe, Germany) are the two main HPC machines in use. The goal of the project is to run either the ICON or IFS global models and use the output to provide boundary conditions for a specific regional run, performed via the regional ICON model.

The regional model always runs on LEONARDO, while the global models run either on LEONARDO or HoreKa. In the latter case, boundary conditions are stored in an S3 (Simple Storage Service) bucket at KIT and fetched by LEONARDO. The final output of the enhanced regional runs is stored in an FDB (Field DataBase) server provided by ECMWF. In the context of Destination Earth, data stored in the FDB server will be made accessible to the entire community and users, with the possibility of querying data (temperature, wind, precipitation, etc.) along sections of four-dimensional data, including time.

The workflow is currently being tested on a specific regional area, the Alpine domain,

reproducing weather forecasts for the Emilia flood event of May 2023, and comparing the output of the GLORI4DE enhanced model with the vanilla models. Initial results, shown in the accompanying picture, demonstrate performance improvements for the GLORI4DE enhanced model.

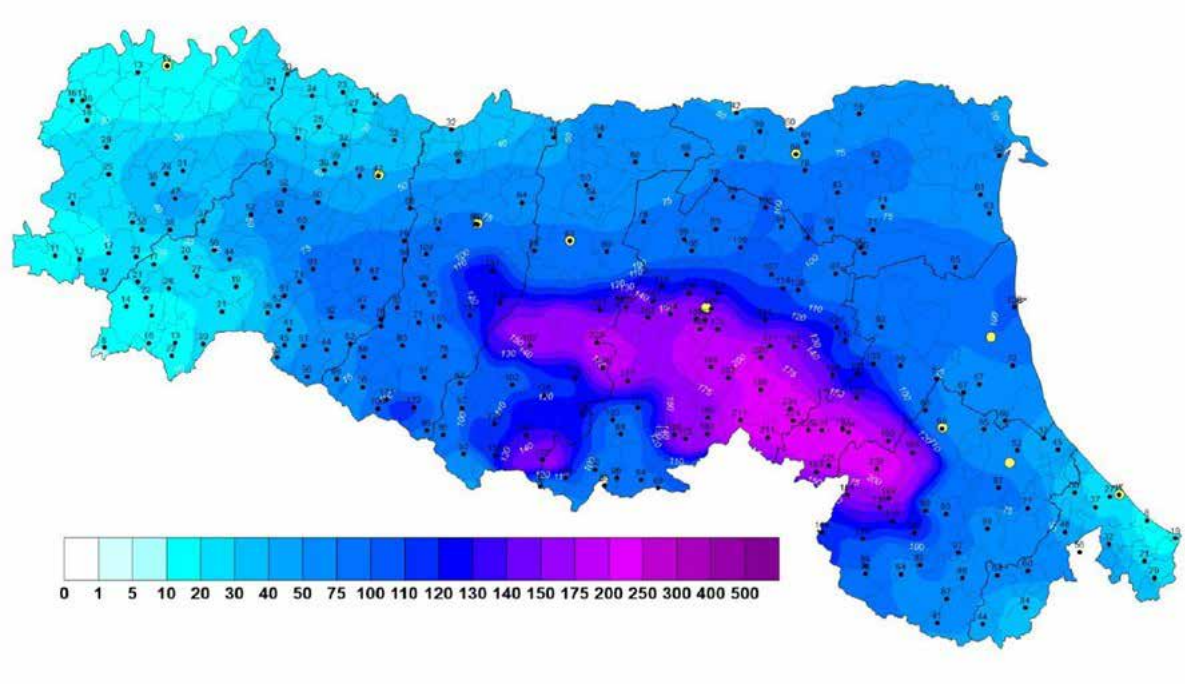


Figure 26: observed data from the Flood event of 1-4 May 2023.

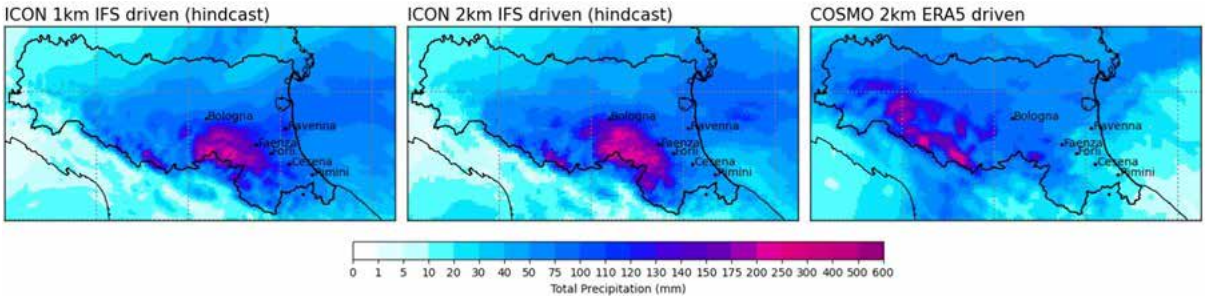


Figure 27: forecasts for the 1-4 May 2023: first and second picture show precipitations forecasted from the GLORI4DE workflow, third picture is from running default operational model.

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.

GLORI
GLObal to Regional | ICON Digital Twin

BIG DATA AND ARTIFICIAL INTELLIGENCE

Bridging AI and HPC: advancements in Access, Applications and Collaborations

Roberta Turra
Cineca

The advent of Artificial Intelligence (AI) has significantly impacted the field of High-Performance Computing (HPC), leading to changes in architecture, computational approaches and data management. AI has broadened the scope of HPC applications, introducing domains like deep learning, natural language processing, and computer vision. Hardware innovations, including accelerators like GPUs, are essential for AI workloads, prompting a shift towards heterogeneous computing environments and increased integration with cloud resources. New software frameworks like TensorFlow and PyTorch optimize AI tasks on HPC, while data management focuses on efficient storage, transfer and access for vast datasets.

While maintaining traditional HPC applications, Cineca's HPC infrastructure is designed to support AI and offers specific modules to set the environment for AI tools, data lakes to store high volumes of unstructured data and data catalogues to promote re-use and reduce data transfer.

These changes are enabling more sophisticated and efficient analysis, modelling, and problem-solving across a wide range of scientific, industrial, and technological domains. AI is being integrated into traditional HPC domains like physics, chemistry and climate modelling, accelerating simulations and improving model accuracy. Additionally, AI is emerging as a separate, data-driven approach within HPC, attracting new scientific and industrial users. Specific access channels have been opened to accommodate the new users' requirements through strategic partnerships (EuroHPC's AI calls, AI-BOOST and ALT-EDIC initiative), while ISCRA is still the main entry point and is being increasingly used to develop AI projects, especially on LLM (Large Language Models) and Computer Vision.

To provide computationally intensive projects with fast access to resources, several agreements have also been put in place with academic and industrial partners. Among the most relevant, the agreement with the Legislative Assembly of the Emilia-Romagna region to develop AI and LLM tools to support the drafting of regulatory texts and improve laws impact, with the University of Bologna to estimate the Healthy Life Expectancy using Retrieval Augmented Generation for systematic scientific literature review, with FBK to develop an Open Source State of Art Large Multimodal Model, with iGenius to jointly develop a new generation Foundational Large Language Model for the Italian language to be released under an open-source license, with Mistral and FAIR to provide the computational resources for foudational LLMs.

An increasing number of Machine Learning Weather Prediciton projects have been funded and are being developed. Among them the AI General Circulation Model with Illumia, Weather4Energy and OptimESM.

Cineca is also involved in AI funded projects on healthcare and precision medicine, on the cataloguing and meta-dating of artistic heritage by image and text analysis techniques (object detection, classification, captioning, labelling) and in designing massive data center graph models to improve energy efficiency.

The active participation to these projects enables Cineca to improve its competences and services, in particular on parallel training and scaling of AI models on HPC and on the fine-tuning of LLMs. By providing tailored access channels, support services, and training opportunities, Cineca aims to empower researchers, scientists, and practitioners to harness the full potential of HPC-enhanced AI for solving complex challenges and driving transformative discoveries.

ISCRA-C and AI: how Cineca is supporting the artificial intelligence revolution

Matteo Angelinelli
Cineca

We present a detailed analysis of the projects submitted to the ISCRA-C calls tagged as machine learning (ML) and artificial intelligence (AI) projects over the past year. We have explored emerging trends, identified areas of greatest interest and predicted possible future developments.
The first step focused on analysing the project titles, looking for the most recurring keywords to create a word cloud, which gave us a visual overview of the dominant topics, as shown in Figure 1.

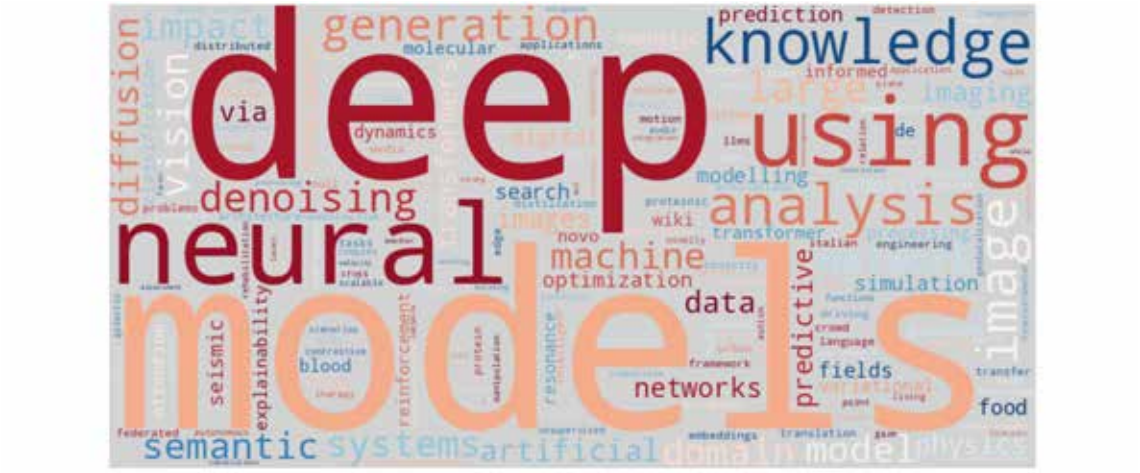


Figure 28: Cloud Word computed on ISCRA-C title.

We then analysed the project abstracts. We identified 10 distinct clusters, grouping projects with similar topics and objectives. The number of abstracts for each cluster and the two most recurring words in each cluster are shown in Figure 2.

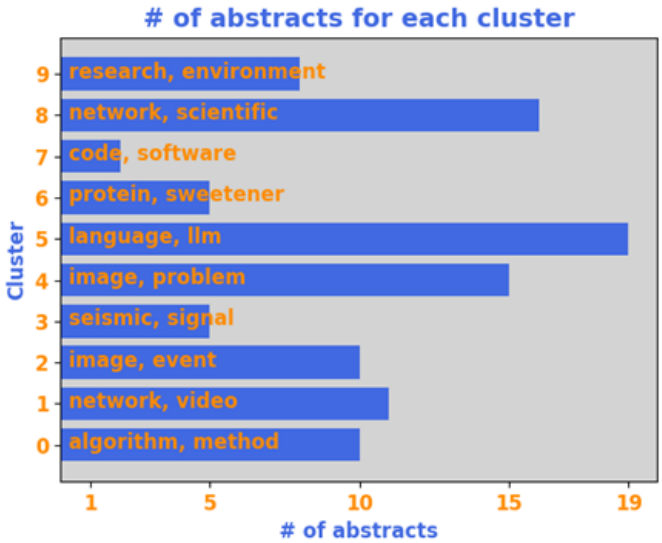


Figure 29: number of abstract and top 2 words for each cluster.

To better characterize and understand the topics of each cluster, we then extracted the 10 most frequent words. Figure 3 shows these 10 words for each cluster. Looking at the extracted words, we could assign a name to each cluster, describing all the abstracts belonging to that cluster.

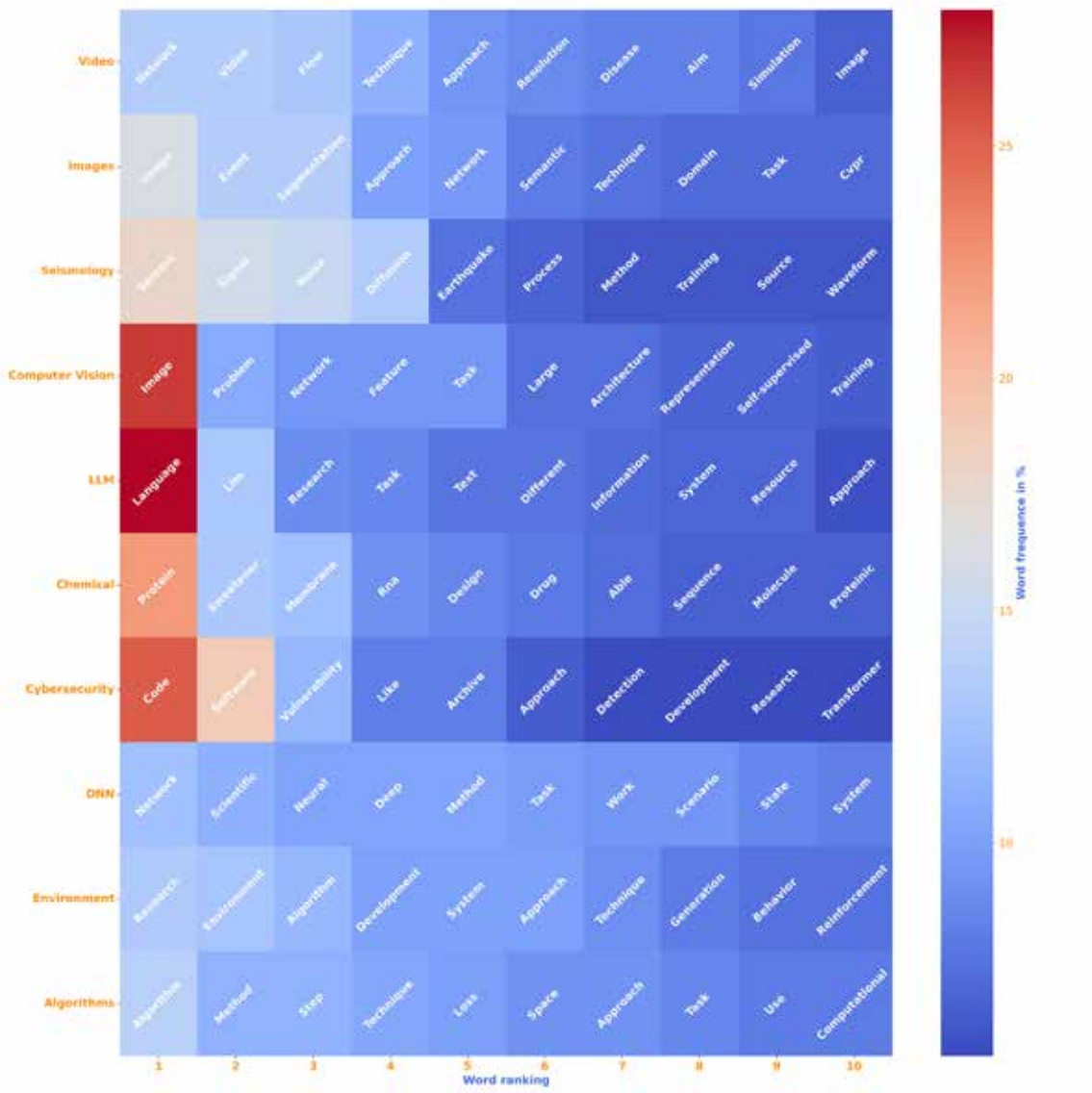


Figure 30: percentage of occurrence of the top 10 words for each cluster.

We then looked at how these clusters are distributed across the different calls of the year, to better understand the temporal distribution. In Figure 4, we present this analysis according to the months of the year, considering that the ISCRA-C calls are not published in August and December.

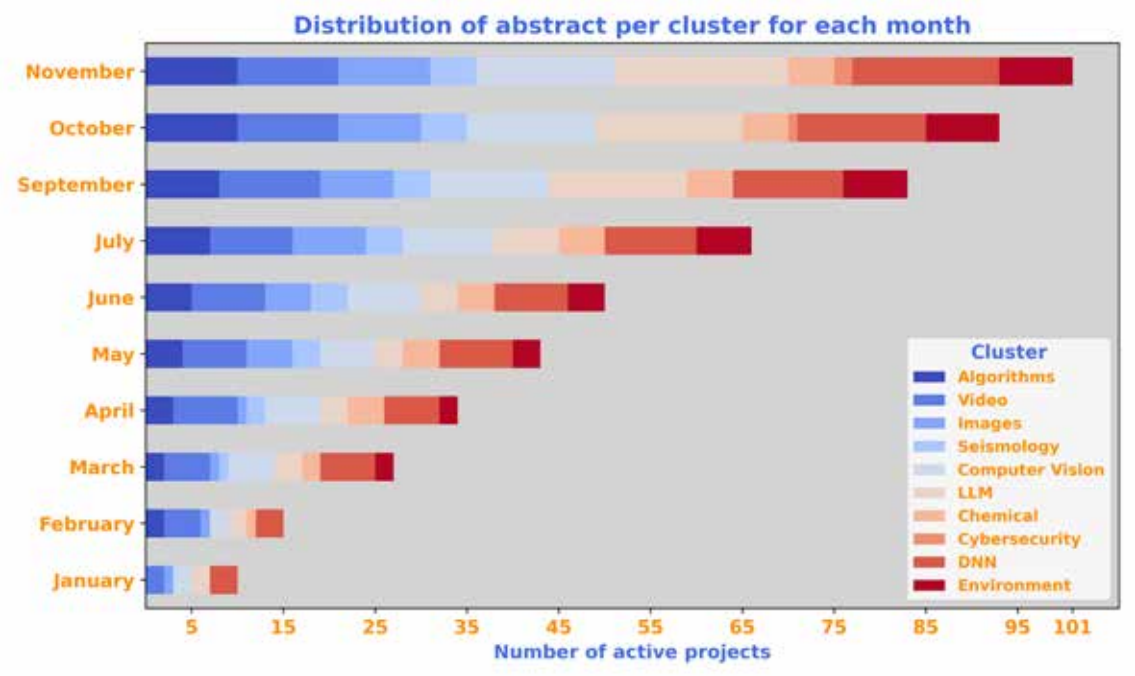


Figure 31: number of abstracts in each cluster, for each month of 2023.

The general increase in the number of abstracts for AI and ML is an obvious trend, with a clear predominance of topics such as computer vision and LLM. Our analysis revealed several keywords, related to computer vision, deep learning and natural language processing recurred with some frequency, resulting in a significant word cloud.

Dividing the projects into clusters allowed us to appreciate the breadth and diversity of the issues covered. We observed a rich variety of fields of study and research, demonstrating the complexity and multi-dimensionality of AI and machine learning. Looking at the distribution of clusters across the year, a dynamic and diverse picture emerged, suggesting that interest and engagement in AI and ML are constant and widespread over time.

For the coming year, we foresee a further increase in AI projects, with a focus on applications such as meteorology and natural language, with their implications and spin-offs, as well as cybersecurity projects. This analysis has provided a basis for understanding and trying to orient future strategies, identifying emerging opportunities in the increasingly dynamic context of AI and ML.

ArbörIA: knowing and classifying urban green

Arböria: An Open-Data Based Urban Forest Solution Improved By Ai Optimisation

Matteo Angelinelli
Cineca

ArbörIA is an experiment funded by the European project EUHubs4Data, which aims to address the growing need for accurate data on urban and forest trees in the context of climate change mitigation and urban ecosystem management. Led by Föra - forest technologies, in collaboration with technology partners such as CeADAR, Cineca and RISE, the project uses innovative technologies such as LiDAR, aerial RGB, infrared imagery and AI algorithms to automate the detection, localisation and species classification of trees in urban areas.

The primary objective is to create a dataset of 2000 images specific to the species Platanus x hybrida (Fig.1), using the urban tree inventory of Madrid to obtain information on tree location, species and height. However, the process is complex due to imperfect geolocation and the presence of multiple trees in some images. ArbörIA tackles these challenges by using object detection models (Fig.2) and binary classifiers, which ultimately generate a satisfactory dataset of images.

The second objective is to improve tree geolocation using LiDAR data and advanced point cloud processing techniques (Fig.3). The approach focuses on increasing the accuracy of urban tree identification, to contribute to the European Union's environmental objectives as outlined in the Green Deal. Through collaboration and technological innovation, ArbörIA is proving to be a vital resource for revolutionising the collection of urban tree data, which is essential for addressing environmental challenges and the sustainable management of urban environments.

The experiment is based on the idea of automating tree detection and classification using artificial intelligence algorithms, in line with wider efforts to promote environmental sustainability and climate change adaptation. The use of cutting-edge technologies such as LiDAR and image analysis enables accurate and detailed results, contributing to the understanding and management of urban ecosystems in a completely new way.

In conclusion, ArbörIA is a significant initiative that embodies the fusion of technology and sustainability, demonstrating how innovation can be a driver in addressing pressing environmental challenges. With its ability to collect detailed and accurate data on urban trees, ArbörIA will become an essential tool for urban planners, land managers and policymakers seeking to build greener and more climate-resilient cities.

The EUHubs4Data project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951771

EUH4D
EUROPEAN FEDERATION OF DATA DRIVEN INNOVATION HUBS

Figure 32: sample of Platanus x hybrida trees.

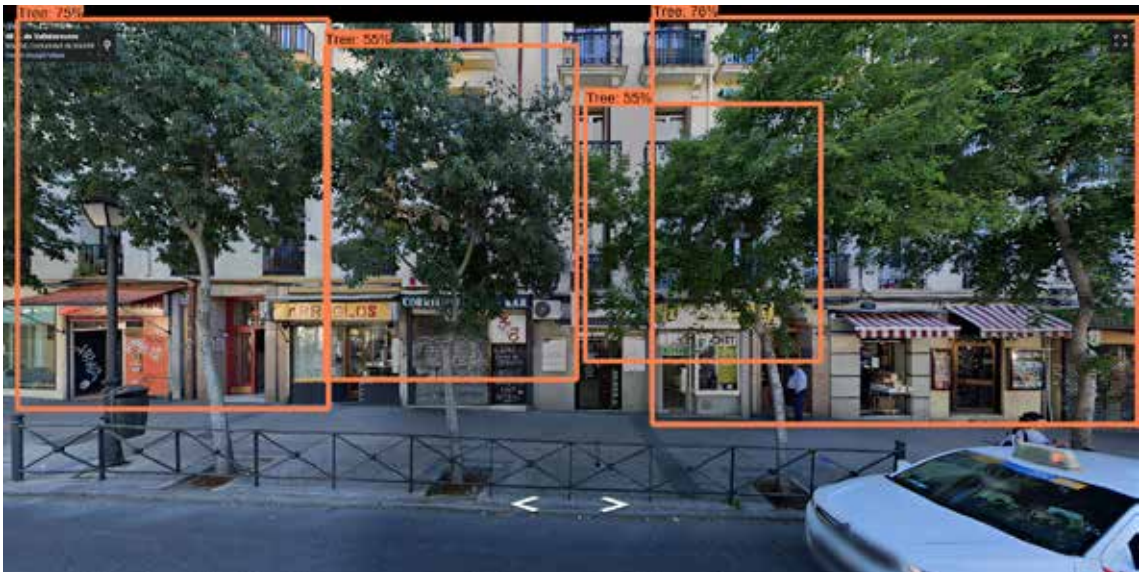


Figure 33:Object detection applied on Google Street View image.

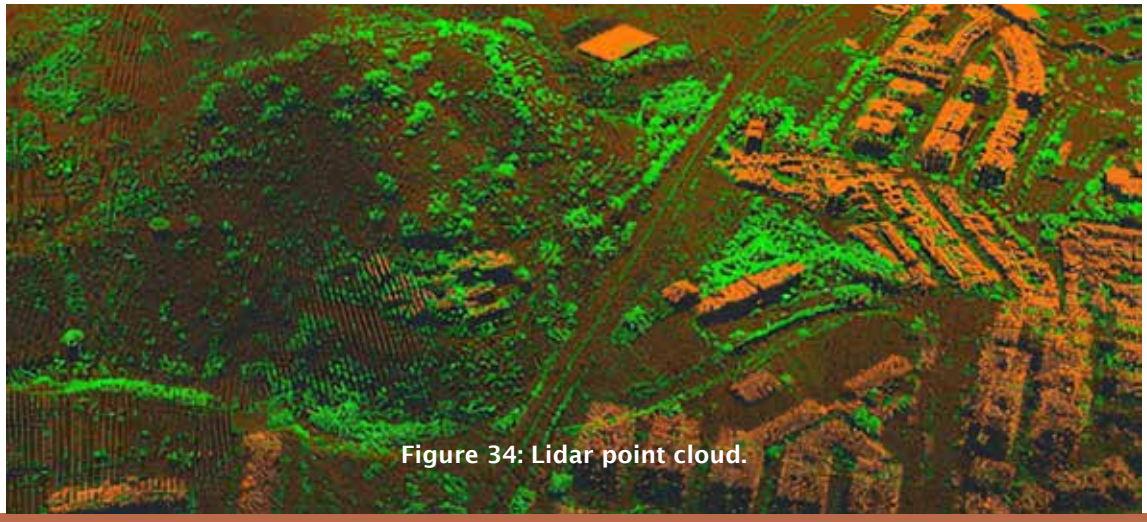


Figure 34: Lidar point cloud.

AI as a tool to innovate weather forecasting systems

Eleonora Bergamaschi, Matteo Angelinelli, Giacomo Masdato, Marcello Crosta
Illumia

Weather forecasting is crucial for many sectors, including agriculture management, green energy production, and natural disaster prevention. In recent years, particularly in 2023, AI-driven approaches, supported by advanced deep learning techniques, have revolutionized this field providing faster results and reducing prediction time from hours to seconds.

The project "The AI General Circulation Model (AIGCM)" aims to develop a Proof-of-Concept (POC) for an Italian deep-learning-based weather model, enabling hourly weather forecasts and laying the groundwork for the development of forecasting model for energy production and demand from renewable sources. An AI-driven local model over Italy can capture phenomena that influence the evolution of physical variables that global models with 30 km resolution cannot reflect (Fig.35).

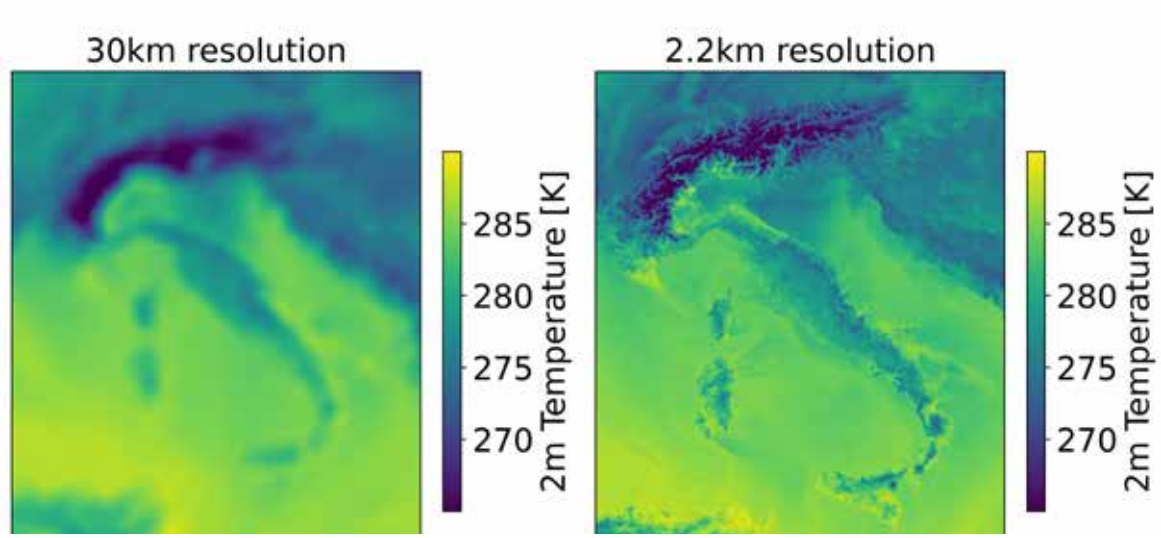


Figure 35: comparison between two maps showing the 2m Temperature of Italy at 30 km and at 2.2 km resolution. The image on the right represents a significant advancement in the accuracy of local phenomena prediction, and it was possible thanks to dynamic downscaling of ERA5.

The project is funded by IFAB-International Foundation Big Data and Artificial Intelligence for Human Development for 18 months and it was proposed by Illumia, an Italian company operating in the free electricity and natural gas market based in Bologna, in collaboration with CINECA, CMCC-Centro Euro-Mediterraneo sui Cambiamenti Climatici and Agenzia ItaliaMeteo.

The project involved setting up the infrastructure environment, selecting the deep learning techniques, training and refining the best model. The dataset used is the dynamical downscaling of ERA5 with a resolution of 2.2 km over Italy from 1981 to 2020 (developed within the CINECA project HIGHLANDER [1]).

Regarding the deep-learning techniques, two parallel approaches have been pursued. A U-Net architecture based on the idea of the model developed in collaboration with Microsoft used for sub-seasonal to seasonal forecasts [2] and a method based on spatio-temporal learning (STL) and Transformers architecture [3] (Fig.36)

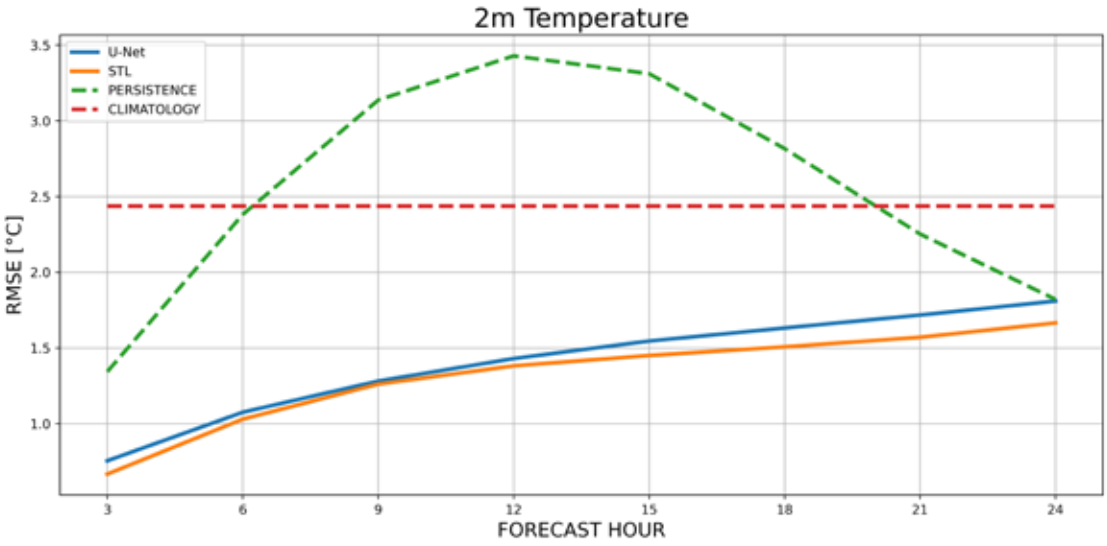


Figure 36: the plot shows the average RMSE, a standard metric used to evaluate the models' errors, computed for the models we trained. The figure displays the results obtained for the temperature at ground level on a forecast time of 24 hours. The dashed lines show the RMSE for climatology and persistence, which are considered standard for meteorological comparisons. Note that the 3-hour forecast for each of the models has an RMSE less than 1°C.

The project requires HPC resources due to the model complexity and the large amount of data involved, and the Booster partition of the supercomputer Leonardo is used to train the models. Parallelization and optimization of the training process are fundamental as raw data is around 3TB. Regarding the size of the models, the U-Net-based model has around 30 millions of parameters while the models developed with the STL approach contain from 4 to 12 millions of parameters.

The advantages obtained with AI-based models from a computational point of view are significant. The results being achieved with the project are very promising and lay the foundations for interesting future developments.

[1] Raffa, M., Reder, A., Marras, G. F., Mancini, M., Scipione, G., Santini, M., Mercogliano, P. (2021). VHR-REA_IT dataset: Very high resolution dynamical downscaling of ERA5 reanalysis over Italy by COSMO-CLM. Data. 6. 88. 10.3390/data6080088.
[2] Weyn, J. A., Durran, D. R., Caruana, R., & Cresswell-Clay, N. (2021). Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models. Journal of Advances in Modeling Earth Systems, 13, e2021MS002502. https://doi.org/10.1029/2021MS002502
[3] Tan, C., Li, S., Gao, Z., Guan, W., Wang, Z., Liu, Z., Wu, L., & Li, S. (2023). OpenSTL: A Comprehensive Benchmark of Spatio-Temporal Predictive Learning. ArXiv, abs/2306.11249.

Advanced Solutions For Automatic Text Analysis Of Cultural Heritage

Donatella Sforzini, Luca Mattei
Cineca

Text tagging, also known as text classification, is useful in various application: Information Organization, Search and Retrieval, Recommendation Systems, Text Summarization; overall, text tagging is crucial for structuring, analysing, and extracting insights from textual data.

In the field of Cultural Heritage, Named Entity Recognition (NER), a subtask of Natural Language Processing (NLP) which locates and classifies entities within text into predefined categories, plays a vital role in the digitalization and cataloguing process.

The task faces challenges due to the variability and ambiguity of natural language (entities can appear in various forms, due to spellings, abbreviations or linguistic variations), which led to the development of several methods over the years reflecting the technological evolution.

In our experiments at Cineca, we applied various NER tools on description cards about the monumental cemetery of the “Certosa di Bologna”.

We evaluated WikiNEuRal, it_core_news_lg (SpaCy), and spacy-dbpedia-spotlight, using models trained on Italian language data. The last one was dropped in favour of the other two models because of its high rate of wrong categories.

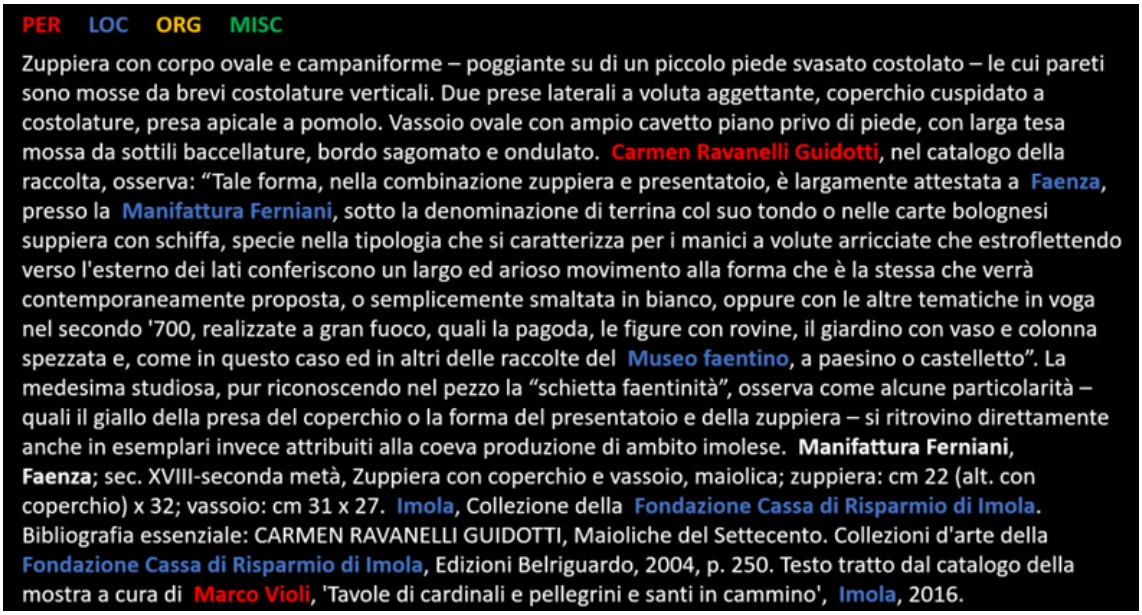
Due to the unavailability of labelled data, in order to improve our confidence on the results, we decided to take as results only the entities retrieved by both the models, with the same label, making our inference pipeline a very simple ensamble.

On the total of 1130 documents wikineural found a total of 27602 entities, while the SpaCy model 42805; combining the two models, the total amount of found entities is equal to 18150, which means an average of 16 entities per document.

	Accuracy	F1	Precision	Recall
it-core-news-lg	0.89	0.89	0.89	0.89
Wikineural	0.91	0.89	0.89	0.90

* the metrics are estimated as it follows: the common entities sharing the label are considered true positives; the common entities with different labels have been manually annortated and evaluated; the metrics above are the wheighted mean of those two.
NOTE: these metrics are highly uncertain estimates which don’t considerate the non-common entities found by the two models and the entities not found.

Here an example:



In conclusion, despite known limitations, AI tools are powerful aids, providing significant benefits in managing and preserving cultural heritage by enabling automatic extraction of information.

BIBLIOGRAPHY

[1] S. Tedeschi, V. Maiorca, N. Campolungo, F. Cecconi, and R. Navigli. 2021. WikiNEuRal: Combined Neural and Knowledge-based Silver Data Creation for Multilingual NER. In Findings of the Association for Computational Linguistics: EMNLP 2021, pages 2521–2533, Punta Cana, Dominican Republic. Association for Computational Linguistics.
[2] Honnibal, M., & Montani, I. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.
[3] Daiber, Joachim, et al. 2013. Improving Efficiency and Accuracy in Multilingual Entity Extraction. Zenodo (CERN European Organization for Nuclear Research), 4 Sept. 2013.
[4] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

Data science and environmental sustainability: could graphs be the answer?

Matteo Angelinelli
Cineca

The ubiquity of data analysis and processing in research, industry, commerce and government brings environmental challenges as well as benefits through energy consumption. The International Energy Agency (IEA) reports that data-related activities account for 1-1.5% of global energy consumption, with data centres and transmission networks responsible for 1% of associated greenhouse gas emissions. In addition, the rapid technological growth of data science is leading to the obsolescence of hardware and software, creating e-waste. The UN predicts 53.6 million tonnes of e-waste in 2019, rising to 74.7 million tonnes by 2030.

Initiatives such as the Graph Massivizer project are crucial in addressing these challenges, focusing on sustainable data analytics to improve efficiency and reduce energy consumption. Focusing on finance, news, automotive and data centres, Graph Massivizer aims to improve data analysis efficiency by 70% and reduce energy impact by 30%. It also aims to double the energy efficiency of data centres and reduce greenhouse gas emissions from database operations by more than 25%. A key aspect of the project is the 'Data Centre Digital Twin' use case, developed by Cineca and the University of Bologna, which aims to optimise the efficiency and sustainability of supercomputers.

The University of Bologna's landmark study, the result of a decade-long collaboration, provides insights into supercomputers and their data exploitation (https://www.nature.com/articles/s41597-023-02174-3). The paper presents EXAMON, a monitoring framework which provides the first comprehensive dataset from a Tier-0 Top10 supercomputer, shedding light on management, workload, facility and infrastructure data from the Marconi100 system. This dataset, the largest publicly available to date, facilitates further research through open source software modules.

In conclusion, while data science and computing pose environmental challenges, initiatives such as Graph Massivizer and research by institutions such as the University of Bologna and Cineca offer promising solutions. By prioritising sustainability and efficiency, these efforts demonstrate the potential of technology to drive environmental progress and mitigate the impact of data-driven innovation.

SAVIA: Artificial Intelligence in support of the lawmaking process

Laura Morselli
Cineca
Luisa Monti
Assemblea Legislativa - Regione Emilia Romagna

In May 2023, CINECA and the Legislative Assembly of the Emilia-Romagna region signed a framework agreement1 centered on the dissemination of the culture of information technology and ethics in the use of artificial intelligence systems. A pillar of the agreement was the development of tools exploiting Artificial Intelligence techniques in support of the lawmakers to ensure the quality of the laws.
In particular, the main goals of the project are: 1) improve the efficiency of the procedures connected to the drafting of regulatory texts, also using new technologies and artificial intelligence; 2) develop tools to support evaluation procedures of the effects produced by regional legislation, with IT solutions and artificial intelligence models.
SAVIA was born as a project to support regional councilors, public administration employees and citizens, by enhancing effectiveness, transparency and participation to the legislative processes.
To these aims, we explored the usage of open-source Large Language Models (LLMs) to support legal professionals, lawmakers, and citizens in accessing information on the current and past legislation of the Emilia-Romagna region. To obtain a model capable of understanding Italian language in the law domain and responding to questions related to laws enacted in the Emilia-Romagna region, we followed a multi-step approach. We started from an open-source LLM (e.g. LLaMAntino-2-7b-hf-ITA2, Mistral-7B-v0.13 and Mixtral-8x7B-Instruct-v0.14) adapted it to the legal language through unsupervised domain adaptation exploiting the corpus made of the regional laws of Emilia-Romagna, as well as the relative implementing acts at the regional level, the available reports on the expected and measured impact of a given law (e.g. “clausola valutativa”, “ex-ante and “ex-post” reports) and the mii-llm/gazzetta-ufficiale5 dataset on Hugging Face. The unsupervised domain adaptation was carried out exploiting the Quantized Low-Ranking Adaptation (QLoRA) technique. The resulting domain-adapted model was then fine-tuned for question-answering (on an instruction-based dataset prepared by domain experts for this purpose and made of a set of questions on the regional laws and their implementing acts and the relative answers. Finally, we implemented a domain-adapted retrieval model (based on the Retrieval Augmented Generation technique) to enrich the answers with relevant information from the law corpus. The evaluation of the answers provided by the final models was done by domain experts of the Legislative Assembly.
This tool is available to lawmakers for simple and fast consultation of regional law and administrative acts databases. The purposes are to support the Legislative Assembly and the Territorial Administrations in carrying out their institutional functions, but also to ensure transparency and information to the regional community regarding the legislative and administrative activities of the Region.

Regione Emilia-Romagna
Assemblea legislativa

The project is funded by the Assemblea legislativa, Regione Emilia-Romagna

SAVIA
INTELLIGENZA ARTIFICIALE PER LA QUALITÀ DELLE LEGGI

1https://www.cineca.it/index.php/area-stampa/comunicati-stampa/intelligenza-artificiale-accordo-emilia-romagna-tra-assemblea

2 https://huggingface.co/swap-uniba/LLaMAntino-2-7b-hf-ITA
3 https://huggingface.co/mistralai/Mistral-7B-v0.1
4 https://huggingface.co/cmarkea/Mixtral-8x7B-Instruct-v0.1-4bit
5 https://huggingface.co/datasets/mii-llm/gazzetta-ufficiale

MATERIAL SCIENCE

Material science

Mariella Ippolito
Cineca

CINECA has numerous national and international collaborations in the materials science domain and is a partner of two HPC Centers of Excellence (CoE) in this area namely MAX and TREX.

2023 was the last year of TREX, that is a CoE for Quantum Chemistry combining expertise in Monte Carlo algorithms with concepts from quantum mechanics, offering innovative solutions for modeling quantum systems and designing advanced materials. This project focuses on the porting, maintenance, and scaling of Quantum Monte Carlo (QMC) codes to multiple heterogeneous architectures; in this context CINECA played a crucial role, both providing support for the GPU porting of TREX codes and the architectures for benchmark and testing activities. In the last year CINECA has mainly focused on the GPU porting of the Quantum Monte Carlo Kernel Library (QMCKL), which is the core library of TREX: QMCKL aims to provide a high-performance implementation of the main kernels of Quantum Monte Carlo methods and is used by many TREX codes. Both OpenMP offload and OpenACC paradigms have been used to ensure broad portability between different architectures. CINECA, in collaboration with experts from Nvidia, dedicated to profile and optimize the most time-consuming QMCKL kernels in the atomic orbitals (AO) and molecular orbitals (MO) parts of the library and the offloaded kernels have been then benchmarked on the CINECA's clusters (Marconi100 first and Leonardo booster then). The CINECA application support team provided support for the installation of the TREX codes on the Leonardo booster partition, dedicating a particular effort to the TurboRVB code developed at the SISSA Institute (Italy), which is also part of the consortium that led the procurement of Leonardo.

As part of the partnership with TREX, CINECA also organized a GPU hackathon to support TREX code developers in porting their code on GPUs and to optimize their applications already running on GPUs. With the participation of experts from world-renowned organizations such as Nvidia and Intel, the Hackathon was a great opportunity to learn more about GPUs, develop performance improvements, and improve the portability of TREX flagship codes to new architectures.
The last year of TREX coincided with the first year of the third phase of MAX CoE. For many years MAX has played a crucial role in advancing the research and development of high-performance materials, supporting Density Functional Theory codes (i.e. QuantumESPRESSO, Yambo, Siesta, Fleur and BigDFT). Leveraging the powerful computational resources of exascale, the center aims to accelerate the discovery and optimization 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 third phase of the MAX CoE is to provide the European materials simulation community – both developers and end users – with the capabilities needed to exploit at the best the exascale resources to address scientific challenges hitherto considered prohibitive. To achieve this, in addition to foster the scaling of 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.
In this third phase of MAX, CINECA leads Work Package 3 for "Technical challenges towards exascale and post-exascale". It addresses the technical challenges posed by the exascale systems and provides solutions (in terms of prototypization, proof-of-concepts) to the code and workflow developers.

CINECA takes care of the continuous assessment and analysis of the parallel performance of the MAX flagship codes, pointing up the direction for the development aimed at the effective exploitation of the existing technology. To improve uniformity and homogeneity in the data produced for each MAX code and to have common metrics to evaluate the code's performance, we adopted a common

strategy for benchmarking and profiling activities of all MAX codes. In this regard, we decided to use the same benchmarking and profiling tools for all MAX codes, and, to coordinate these activities, we organize periodic meetings between CINECA experts and each MAX code community. Leonardo is the preferred system for the profiling and benchmarking activities, being the largest HPC architecture available within MAX and being accessible to all MAX code developers. However, within the CASTIEL project - promoting interaction and exchange between National Competence Centres – MAX developers should soon have access to other EuroHPC architectures.

Code characterization within MAX is pursued by using a modular benchmarking environment based on JUBE, developed at Julich Supercomputing Center, and providing a script-based framework to streamline the creation of benchmark sets, their execution on different computer systems and evaluation of the results. By taking advantage of the data format standardization obtained with JUBE, we are organizing a MAX benchmark repository so that the data can be easily extracted for visualization, to be automatized as well with appropriate tools. The effort to unify and automate data visualization will streamline the dissemination of results and their inspection by production users, thus contributing to minimizing the misuse of computational resources.

One other objective of the MAX CoE is to ensure that data produced and stored within the project are complying with the FAIR principles (making data Findable, Accessible, Interoperable and Reusable). To this end, it was agreed that EU HPC centers 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.
In 2023, thanks to a collaboration between CINECA and CSCS staff, a nightly backup of the Materials Cloud Archive has been set up from the CSCS servers to ADA Cloud, the CINECA HPC cloud infrastructure. The backup data are currently exposed on the mirror web site, where they are 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 in case the main CSCS repository were unreachable.

In addition to TREX and MAX CoE, CINECA is also involved in the ADMIRE project, finished in 2023 but extended by another 6 months. ADMIRE aims to implement an intelligent controller, able to tune the IO properties of applications from different domains to leverage its interaction with the filesystem on HPC clusters. In this context, CINECA supports the integration in the ADMIRE engine of QuantumESPRESSO, as a representative application for the material science domain, by bridging the ADMIRE developers with scientist and testing the integration of the intelligent controller on CINECA's clusters at large scale.

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.

The TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union's Horizon 2020 - Research and Innovation program - under grant agreement no. 952165.

This project has received funding from the European Union's Horizon 2020 JTI-EuroHPC research and innovation programme, with grant Agreement number: 956748

TribChem

Materials Modeling & Tribology

Tommaso Gorni
Cineca

A collaboration with the group of Professor Clelia Righi (University of Bologna) has started in late 2023. Professor Righi's group has been developing in recent years the TribChem code, a workflow manager for the ab initio calculation of tribological properties of materials, such as adhesion, shear strength, and charge redistribution.

The TribChem custom workflows are build via the atomate package, which on turns relies on VASP(+custodian) for the electronic structure calculations, pymatgen and materialsproject.org for managing the input structures, Fireworks as a low-level workflow manager and mongodb for storing the results.

The collaboration started from the group's request of assistance for the deployment of TribChem on Leonardo Booster, which required an ad-hoc solution to allow for a seamless interfacing with the mongodb instance hosted on the ADA Cloud service, and then evolved into a broader collaboration focusing on the optimization of TribChem workflows.

The deployment of TribChem on HPC clusters, necessary to leverage the GPU accelaration of VASP for the heavy electronic structure calculations, presents several challenges, mainly due to atomate's need of managing the workflow from a compute node, while retaining network access both to the materialsproject database via https and to the mongodb instance via ssh.

A previous deployment of TribChem on G100, with the mongodb residing on a VM on AdaCloud, could not be easily replicated by the users on Leonardo, where compute nodes are currently separated from the external network due to the lack of a NAT service. With the assistance of the user support team, we could finally deploy TribChem on Leonardo Booster by using an ssh tunnel to access the mongodb instance on ADA Cloud and a SOCKS proxy for the materialsprojects HTTPS requests.

As a second step, we helped TribChem developers in simplifying and updating the package dependencies, so resulting in a much easier installation process and allowing for the transition to the new materialsproject REST API.

Finally, a previously unnoticed bottleneck has emerged from TribChem runs on Leonardo, mainly due to the bad scaling with the system size of some interpolation steps, carried out with standard scipy routines. At present, in the framework of the EUROHPC EPICURE assistance program, we are helping profiling and optimizing these routines to try and mitigate such a bottleneck.

QUANTUM COMPUTING

Quantum computing

Daniele Ottaviani
Cineca

If 2022 was the year in which CINECA achieved the Hosting Entity status for one of Europe's earliest quantum computers, 2023 marked the official start of work on the integration of the new computing system with the Leonardo supercomputer. Upon responding to EuroHPC JU's expression of interest in March 2022, CINECA committed not only to physically installing the quantum machine at the new data center in Tecnopolo di Bologna but also to operating and maintaining it optimally, ensuring the seamless software and middleware integration of this new heterogeneous system into Leonardo's well-established HPC environment. This integration aims to allow end-users to utilize the new device in a manner similar to all other resources within Leonardo. CINECA's Quantum Computing Lab, collaborating with the other European quantum computing focused groups from other hosting entities, started to work in order to ensure a consistent and uniform integration process across all participating computing centers.

In pursuit of this goal, CINECA, alongside five other selected computing centers, jointly developed the 'EuroQHPC-I' proposal. This project proposal, tied to EuroHPC JU's request 'DIGITAL-EUROHPC-JU-2022-HPCQC-04-IBA,' features a comprehensive integration work plan agreed upon by all involved computing centers. As of this report's writing, the proposal awaits committee review, with updates expected shortly. Preceding and following the proposal submission, CINECA and other participating centers have organized and participated to workshops dedicated to this pioneering endeavor. This workshop series, known as 'EQS3,' held its initial two editions in January and July 2023, with the third scheduled for April 2024 in Copenhagen.

Simultaneously, CINECA's Quantum Computing Lab continued its educational and outreach initiatives. This included the annual "Introduction to Quantum Computing" school held in June 2023 and the sixth edition of the "High-Performance Computing and Quantum Computing - HPCQC" workshop, a staple every December since inception. In 2023, CINECA expanded its educational offerings by launching two doctoral programs in partnership with the University of Bologna in collaboration with Professor Elisa Ercolessi and Professor Tommaso Calarco, who recently returned from Germany (where he was serving at the Julich Research Center) to return to work at the University of Bologna .

Furthering its outreach efforts in 2023, CINECA continued its collaborations to support doctoral activities within the MCSA (Marie Curie Slodowska Action) project, specifically the MOQS grant and the AQTIVATE grant. For the former, a tailored quantum computing school was organized, catering to the needs of participants focused on quantum chemistry. For the latter, CINECA participated in the kickoff event as an invited speaker and is currently organizing a school with a similar thematic focus.

In the context of software development, the Quantum Computing Lab concentrated on two major projects throughout 2023: the National Center for Artificial Intelligence, Big Data, and Quantum Computing (ICSC), and the European HPCQS project. Research primarily centered around the integration of HPC and QC, with a focus on developing quantum computer emulators optimized for the capabilities of the Leonardo supercomputer. Special emphasis was placed on tensor network type emulators. Additionally, efforts were directed towards exploring computational methods to maximize the potential of forthcoming quantum machines in European computing centers. This included research on quantum computers utilizing neutral atoms qubit technology. Notable achievements in this area include the development and installation of QMatcha-Tea, a tensor network HPC emulator, in collaboration with the UniPD group led by Professor Montagero. Several algorithms were also developed for execution on neutral atoms computers, addressing challenges such as advanced embedding techniques, graph coloring problems, and QUBO formulations for molecular unfolding and docking problems, resolved through quantum annealing and QAA techniques. The latter efforts, particularly within the European project EuroCC in partnership with the Italian pharmaceutical company Dompé, have shown promising results.

A short video for communicating quantum physics and quantum computing to a broader audience

Maria Chiara Liguori, Antonella Guidazzoli, Giovanni Bellavia, Daniele de Luca, Paolo Zuzolo, Elena Saluzzi, Daniele Ottaviani, Riccardo Mengoni
CINECA
Nicoletta Tranquillo, Silvia Girardello
KILOWATT

In the coming months, Cineca will make a quantum computer available for research purposes. Motivated by this, the HPC department, through the VisIT laboratory, produced a short video dedicated to quantum physics and quantum computing.

Lasting approximately 14 minutes, the video was produced entirely in computer graphics, with the scientific supervision being entrusted to the department's quantum colleagues.

The complexity of the subject required lengthy preparatory work on the script. Although a short running time was initially planned, as the topics were defined, we became convinced of the importance of leaving adequate space for the explanation of such complexity.

Only after a thorough discussion with the scientific committee the text underlying the script was validated and it was possible to move on to the animatic and the actual making of the video.

In order to better define the two conceptual realms, i.e. the world of classical physics, which is known to us and surrounds us, and the quantum world, we chose to characterise them in contrasting ways. The former has been named the 'white world', an infinite blank sheet of paper on which formulas and concepts gradually appear. Through a series of portals, the viewer arrives instead at the 'black world', the space in which quantum physics unfolds along with the explanations that try to bring the viewer closer to its complexities.

The ambition of this work was to be able to reach as wide an audience as possible, without, however, displeasing the specialists. The use of correct scientific formulas to support the narrative, while also serving as decoration for the 'white world', was conceived precisely with this aim in mind.

The video, previewed during the High Performance Computing and Quantum Computing 2023, will become part of the communication content for the Bologna Technopole, where, in addition to the already present Leonardo, the Cineca quantum computer will be housed.

www.quantumcomputinglab.cineca.it/en/hpcqc-2023-en/

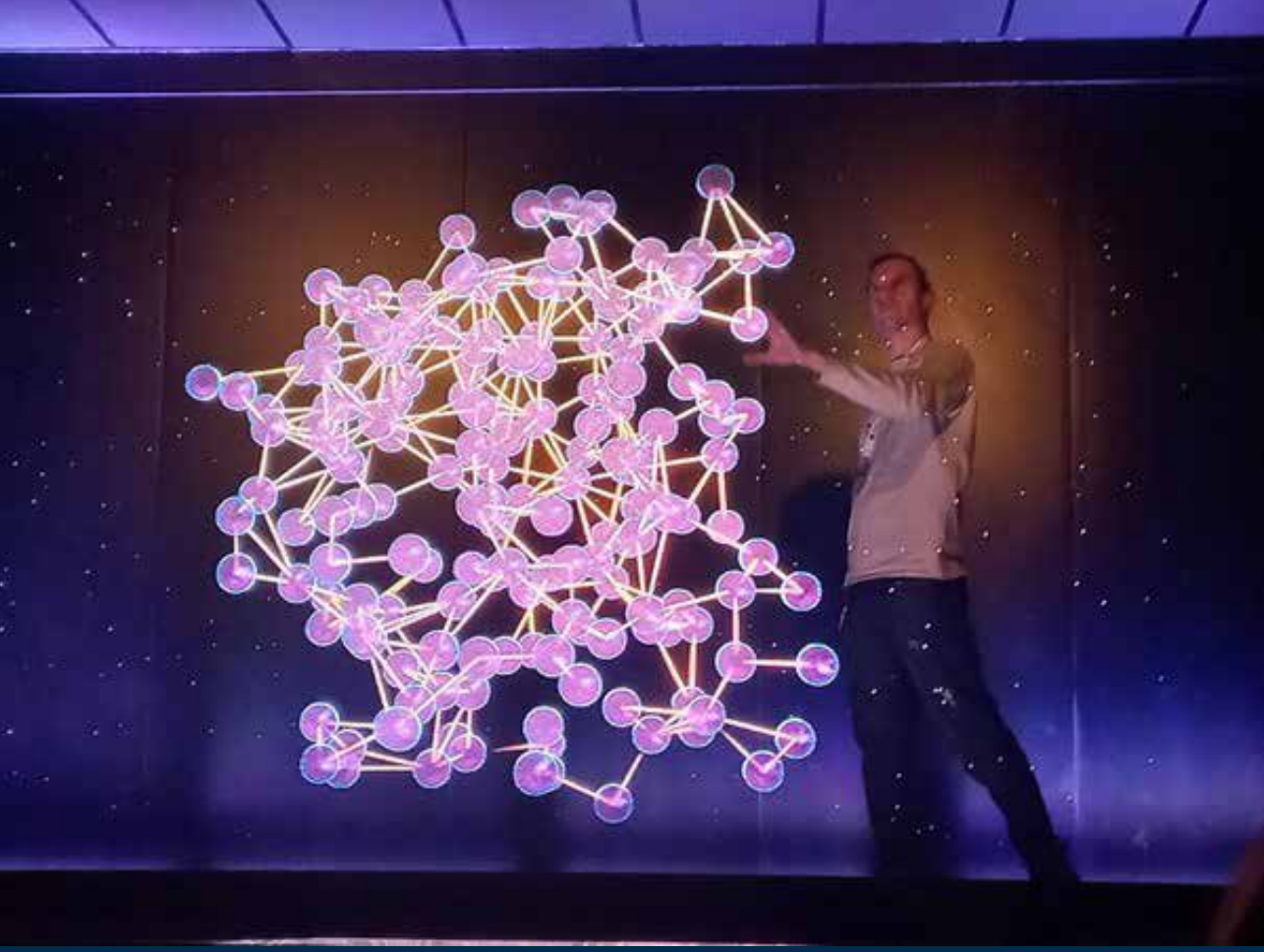


Figure 37: crossmediality. Elements created for the video can be used in other media. Holostage at INAF - Osservatorio Astronomico di Roma, Monte Porzio Catone.

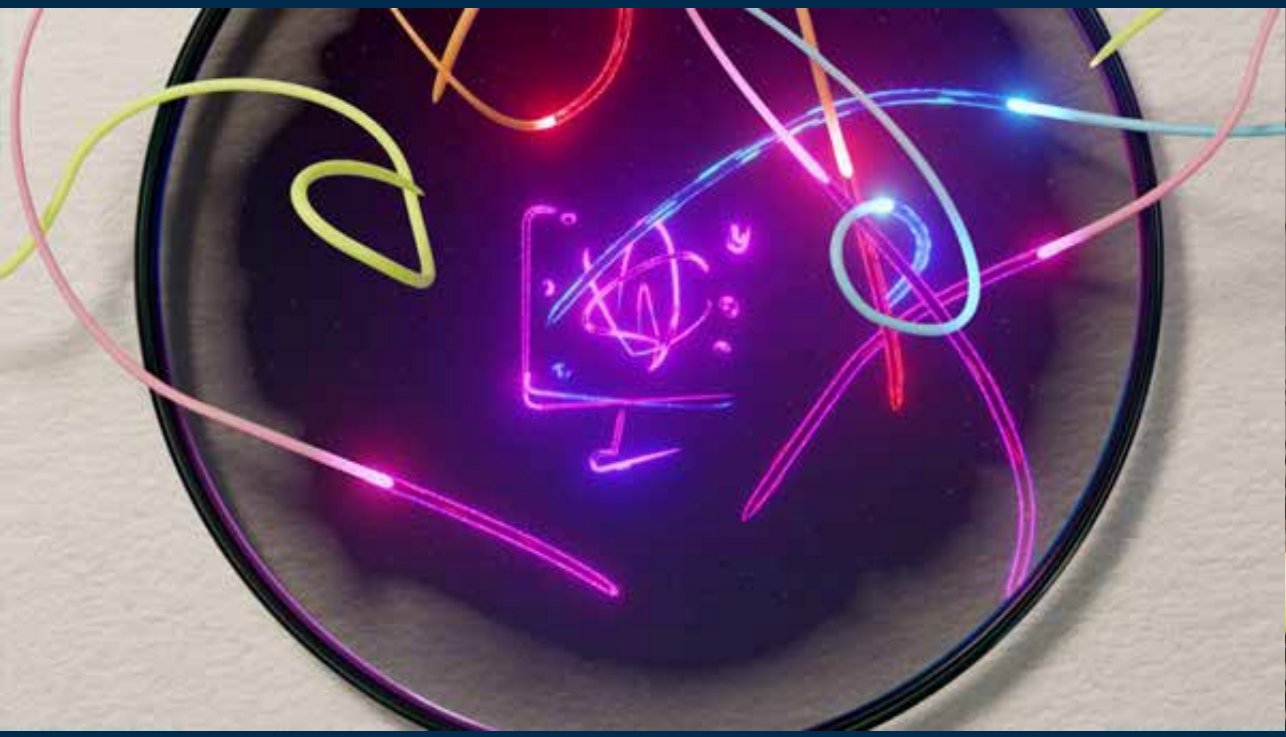


Figure 38: Transition between the two "worlds".

Figure 39: the "black world", where quantum physics is explained.

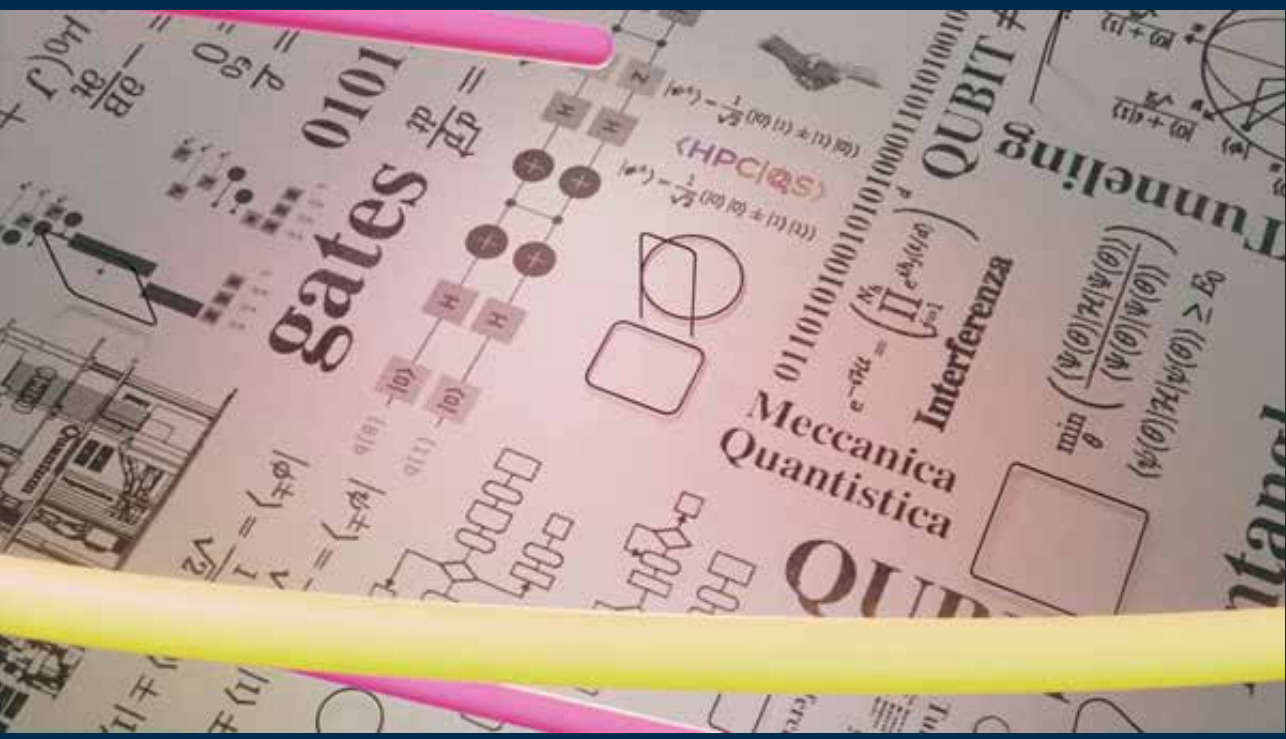


Figure 40: the "white world", populated by formulas and concepts.

HUMANITIES AND SOCIAL SCIENCES, DIGITAL ECOSYSTEMS

Bologna Digital Twin – Enhancing the public value of data

Stefania Paolazzi
Comune di Bologna
Marco Pistore
Fondazione Bruno Kessler
Mauro Bigi,
Fondazione Innovazione Urbana
Rebecca Montanari
Università di Bologna
Chiara Dellacasa,
CINECA

Bologna Digital Twin aims to create a new civic infrastructure to improve the quality of life for citizens and address the great challenges of our time, from environmental to economic and social challenges, in an inclusive and sustainable way. The project, promoted by the Municipality of Bologna and coordinated by Fondazione Bruno Kessler (FBK) in partnership with Cineca, University of Bologna, and Fondazione Innovazione Urbana (FIU), is centered on the concept of Urban Digital Twin: this is intended not only as an instrument for advanced analyses, predictions and decision support, but also as a strategic tool for generating awareness and capacity to use urban data, regulating its use for the benefit of citizens and promoting the generation of public value.

On the path toward building a virtual copy of the city, the Bologna Digital Twin project adopts an incremental approach to address the complexity of integrating heterogeneous data and analysis techniques related to urban services, processes and infrastructure. This integration should consider different spatial and temporal scales, reflecting considerable modelling and computational complexity.

A key objective for the first year of the project is to define the technological platform for the Digital Twin for governing a complex of data and services ecosystem, and for building tools for modelling, analysing, simulating and visualizing all the urban phenomena and entities (eg, processes, services, and infrastructure) necessary for the Digital Twin. The platform software stack is based upon Digital Hub platform, an integrated and interactive framework, developed by FBK, dedicated to the advanced management of data, services and AI models. Cineca provides its Cloud and High-Performance Computing (HPC) infrastructure, as well as support for advanced scenarios based on Big Data. and AI. By integrating data and AI management solutions and Cloud/HPC infrastructures, both specifically tailored for the Urban Dgital Twin, the platform offers a unique technical enabler towards the project goals.

In parallel to this technical development, University of Bologna provides its multidisciplinary expertise to support the development of the digital twin in all its aspects, starting with the values and methodological ones related to the processes of exploitation and enhancement of urban data. Finally, FIU oversees the engagement of civil servants and decision makers, citizens and other stakeholders according to the quintuple-helix approach.

The Bologna Digital Twin project, with its focus on democratic data governance and its advanced technological platform, combined with the inherent complexity that an urban context poses, is ambitious and challenging. However, the benefits that the use of advanced technologies, such as supercomputing and AI, can bring to address concrete environmental and social challenges such as green mobility and energy can be easily understood.

Figure 41: Bologna Digital Twin project teaser
Figure 42: Bologna Digital Twin 3D scenario

LiDAR and Orthophotos: Data for 3D Environmental Simulations in Urban Digital Twins

Antonella Guidazzoli, Alex Gianelli, Federico Andrucci
Cineca
Mirko degli Esposti
University of Bologna

The creation of an urban digital twin, as a digital replica of infrastructure, services, and processes—ranging from physical, environmental, and meteorological to social and economic—is both a fascinating research and technological challenge with significant impacts on land management policies aimed at sustainability and neutrality goals.

High-Performance Computing (HPC) to AI-based Data Analysis and Data Visualization are key technologies in the development and interaction with urban digital twins while the construction of a geographic, geometric, and potentially semantic model of the city marks an essential step in building urban twins

LiDAR data, high definition orthophotos, satellite data provide a novel perspective on the city at increasingly detailed scales, offering an incredible data resource for primarily 3D reconstruction of the city itself and opening new avenues for practical applications.

Data-driven model simulation using artificial intelligence—allow for the creation and simulation of complex phenomena at scales comparable to urban environments.

Following this approach two PoCs have been built for the city of Bologna

The digital terrain model of the city area was obtained through a cleanse and simplification of the detailed LiDAR data, later texturized using high-quality orthophotos and divided into various LODs. Using the basic information about the city's buildings, it was possible not only to create and apply textures to their walls but also to manage the extraction of the roof geometries and textures fusing LiDAR and photographic data. It is worth remember that extracting roof shapes is vital for both solar energy efficiency simulations and as a foundation for tools supporting the maintenance and enhancement of urban real estate.

Furthermore, the trees have been scattered around the city according to the censed data provided. This represents the first step towards the construction of a digital twin of the green urban infrastructure. We explore the use of AI-based tools to extract detailed information about vegetation, which can be instrumental for maintenance, identifying heat islands, improving air quality, and studying the impact on the quality of life.

Finally, all of this data - land, buildings, and trees - was used to construct the entire 3D scene of the city within Unreal Engine 5. This software not only allows for photorealistic rendering of the scene using the most modern rendering techniques, but also for visual simulations of the day/night cycle and rainfall, resulting in an extremely accurate visual model that closely mirrors the physical model. This aspect is crucial not only for the accuracy of the simulation but also for creating an engaging and intuitive user experience. 3D interfaces make it possible to translate complex data into clear and understandable visualizations, facilitating communication between experts and non-experts.

The process was particularly useful for an in-depth analysis of the data and its capabilities and flaws, not to mention to explore possible technologies for the realization of urban digital twins.

Figure 43: sunset in the city of Bologna simulated with Unreal Engine.
Figure 44: visualization of vegetation in the 3D environment of the city.

Figure 45: rain simulation on the Bologna Digital Twin using Unreal Engine,
Figure 46: first-person navigation of the digital twin of Bologna.

HOLOGRAPHIC PORTABLE DISPLAY

Daniele De Luca, Maria Chiara Ligtuori, Antonella Guidazzoli
Cineca

The new portable holographic display, acquired by Cineca, is based on LED technology and is composed of a four-ray rotor display that spins at 670 rpm, creating the optical illusion that shows 3D content floating in midair. The display can show, as a floating hologram, videos or a live stream directly from a computer.

The transparent protective hemisphere allows the display to be used in any context, even in public spaces, such as events and exhibits, where the public is intrigued and led to approach the device. During the “Notte europea dei ricercatori” (2023 edition), the display attracted attention and aroused much interest.

MAIN TECHNICAL SPECIFICATIONS:

- DIAMETER: 75 cm
- IMAGE RESOLUTION: 1000 x 1000 px
- INPUT: HDMI 1.4
- COMPATIBLE FILE FORMAT: video mp4@30fps with h264 - h265 codec
- WEIGHT (without accessories): 0,93 kg
- MAX BRIGHTNESS: 3000 Nit
- CONTRAST: ≥ 2000000:1

Figures 47, 48, 49, 50: a variety of visualizations on the Holographic Portable Display.

S+T+ARTS Studio at the CINECA AI Factory

Science, Technology & the ARTS

Violeta Vasileva, Luis Miguel Girao
Artshare
Veselina Aleksandrova
Rakovski National Defense College, Sofia, Bulgaria

ABSTRACT

The article highlights CINECA's engagement with the arts to democratize AI access across diverse disciplines and examines the S+T+ARTS initiative as a model for fostering interdisciplinary collaboration and entrepreneurship. Furthemore, it discusses the potential of the S+T+ARTS Studio concept to create successful ventures in AI Factories through artistic practices and high-performance computing.

LEONARDO: EMPOWERING GLOBAL AI EXPLORATION

It is no news that CINECA's Leonardo is at the forefront of advancing AI capabilities. It provides researchers and professionals around the world with powerful computational tools and resources, enabling groundbreaking AI research and applications.

Spaceship Earth, a concept popularized by Buckminster Fuller, emphasizes the interconnectedness of all living beings on our planet. This philosophy resonates with CINECA's mission to harness quantum computing and high-performance computing (HPC) to address global challenges. By leveraging cutting-edge technologies, CINECA facilitate exploration in AI, offering unprecedented opportunities for innovation across various sectors.

Figure 51: studio for GRIN Installation.

CINECA'S ENGAGEMENT WITH S+T+ARTS

CINECA's commitment to democratizing and humanizing AI extends beyond traditional scientific domains, reaching into the arts to foster creativity and interdisciplinary collaboration. The GRIN Residencies with Kilowatt in Bologna exemplify this approach, merging supercomputing, quantum computing, and digital twin technologies with artistic endeavors.

For example, Destination Earth by Salome Bazin, is a multi-sensory installation that highlights the intricate connections between the atmosphere, ocean, and human activities. Utilizing Leonardo's supercomputer technology, this project models ocean flows, sea mammal communication, and human impact in real time. The installation combines generative sonification and participatory interaction, encouraging audiences to understand and mitigate sound pollution and climate change through embodied experiences. The artwork guides participants through seasonal cycles, symbolizing the

rhythm of human actions and their effects on the ocean. By engaging with this installation, visitors gain a sensory understanding of the necessary changes in human behavior to ensure the planet's survival.

Fungi – Symbiotic Harmonies by Marco Barotti investigates the global underground ecosystem of mycorrhizal fungi, creating a digital twin of this biological network. The installation uses data sonification, 3D-printed sculptures, and fungal knowledge to explore the symbiotic relationships within fungal communities. Inspired by indigenous chants and biogeographical data, the soundscape evolves through machine learning, offering an immersive experience of fungal ecosystems and their ecological significance.

S+T+ARTS: A MODEL FOR CROSS-DISCIPLINARY COLLABORATION

S+T+ARTS (Science, Technology, and the Arts) is an initiative by the European Commission that fosters innovation through collaboration between artists, scientists, engineers, and researchers. By integrating artistic practices with technological advancements, S+T+ARTS aims to develop creative, inclusive, and sustainable projects.

Since 2015, S+T+ARTS has been successfully organizing 50+ residencies in top European innovation labs. These residencies have led to impactful projects like the S+T+ARTS Prize, Regional Centers, Academy, and Talks, creating a strong community of stakeholders from many different disciplines.

THE S+T+ARTS STUDIO: FOSTERING ENTREPRENEURIAL SUCCESS

The S+T+ARTS Studio concept was put out for the first time in 2020 and builds on the success of previous S+T+ARTS initiatives, proposing a startup studio model to commercialize innovative projects. By aligning investor and creator interests, the studio conceptualy aims to industrialize 10-15% of S+T+ARTS projects, transforming them into viable businesses. The studio model emphasizes a controlled creation process, internal resources, and a network of companies to optimize success rates and reduce failure risks. By leveraging the outputs of S+T+ARTS residencies, the studio concept aims to create new business opportunities across various sectors, demonstrating the true value of cross-disciplinary collaboration.

AI FACTORIES: ARTISTIC PRACTICES FOR HIGH-PERFORMANCE COMPUTING

AI Factories will be dynamic ecosystems to drive innovation in AI by combining computational power, data, and talent. These hubs will advance AI applications across key domains, including health, energy, manufacturing, and meteorology. The synergy

between AI Factories and Testing and Experimentation Facilities (TEFs) will enhance the ecosystem, providing pre-market validation for AI innovations.

The S+T+ARTS Studio model can significantly contribute to AI Factories by integrating artistic practices with high-performance computing. This approach fosters successful ventures by bringing together diverse talents and perspectives, ultimately driving advancements in AI and HPC.

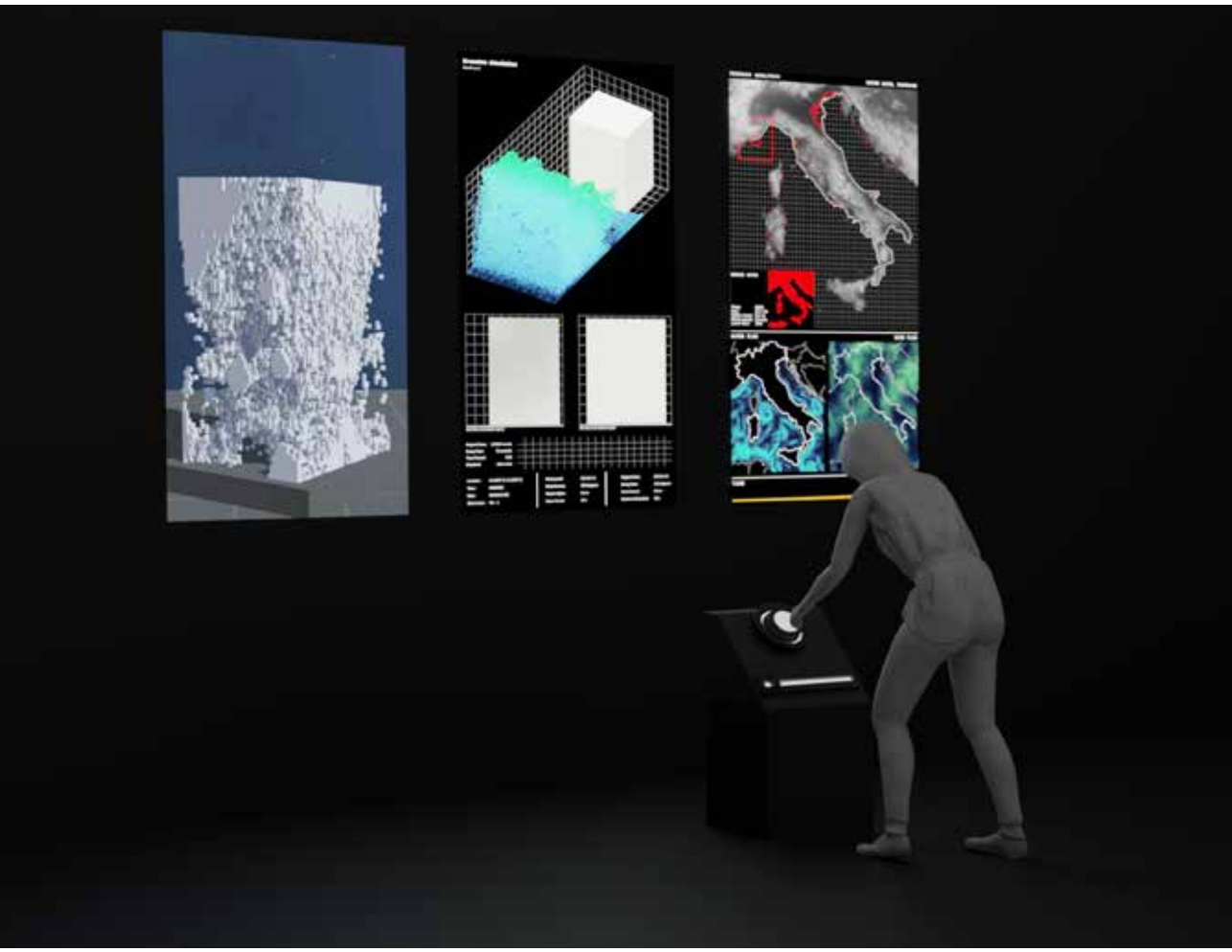


Figure 52; Calin Segal - Tales from the Receding Edge. Studio for the installation. Residency of the GRIN project at Cineca.

S+T+ARTS

European Commission

A Data Space for Italian culture

Luigi Cerullo, Antonella Negri
Central institute for digitization of cultural heritage – Digital Library, Ministry of Culture

The Central Institute for the Digitization of Cultural Heritage - Digital Library was established in 2020 within the Italian Ministry of Culture with the main objective of coordinating and promoting cultural heritage digitization programs. The Digital Library develops the National Plan for the digitization of cultural heritage and oversees its implementation. Within the framework of the National Recovery and Resilience Plan investment M1C3 1.1 "Strategies and digital platforms for cultural heritage", the Digital Library aims to create a digital ecosystem of culture based on an infrastructure that integrates and federates knowledge of cultural heritage and provides users with innovative access services.

I.PaC, the Infrastructure and Digital Services for Cultural Heritage, is the data space designed to preserve, manage, and enrich the digital cultural heritage of the country, in line with the main national and European strategies. It was born from the need to overcome the fragmentation of exploitation systems and manage stratified and heterogeneous data in terms of format, typology, domain of belonging, and protection policies, according to flexible and secure conceptual models.

The development of I.PaC is one of the strategic actions outlined in the National Plan for the Digitization of Cultural Heritage and is part of the broader digital transformation project promoted by the Ministry of Culture's Digital Library for the five years 2022-2026.

I.PaC is a complex system of advanced digital services based on innovative cloud-oriented technologies. It implements functions related to the management and enrichment of digital resources, based both on predefined models and schemes (rule engines and ontologies) and on artificial intelligence (AI) algorithms, and exposes a large catalog of APIs for application cooperation (read and write) related to domain and cross-domain data.

The platform strives to eliminate barriers to access to cultural information and to solve issues related to the management of heterogeneous data in terms of format, category and domain.

This digital space collects descriptive data and digital objects related to Italian cultural properties from archives, libraries, museums, and cultural sites across the country.

The comprehensive repository ensures that valuable cultural artifacts and their associated metadata are preserved for future generations and made accessible to researchers, educators, and the public.

The main services provided by I.PaC regard management and processing of digital assets (tools to preserve, process, and present digital objects linked to cultural heritage) and implementation of domain and cross-domain graphs to support the representation, querying, and retrieval of information about cultural entities and their semantic relationships. In these areas I.PaC is exploring the use of artificial intelligence (AI) models to improve, enrich, and extract data. These AI models are designed to enhance the accuracy and depth of cultural data, promoting continuous evolution in the way cultural information is managed and shared. By leveraging AI, I.PaC aims to facilitate more efficient data processing, uncover hidden connections between cultural entities, and provide users with richer, more contextualized information about Italy's cultural heritage.

Figure 53: Service Map I.Pac

Figure 54: human AI Collaboration framework for manuscripts' transcription in Cineca HPC environment.

CHEMISTRY

In Silico drug design with Ligate

LIgand Generator and portable drug discovery platform AT Exascale

Andrew Emerson
Cineca

The idea for the LIGATE project was conceived just before the Covid pandemic started circulating around the globe at the beginning of 2020. This was remarkable timing because the aims of the project are to provide a drug discovery platform to help accelerate the development of new therapeutic agents, especially against viruses. In fact, the covid-19 pandemic has demonstrated conclusively that a top priority for society, now and in the future, is to be able to respond quickly to diseases with effective treatments. The central goal of LIGATE is to create and validate a leading application solution for drug discovery in High Performance Computing (HPC) systems up to the exascale level. It aims to deliver an integrated solution for drug discovery using HPC, including resources such as LiGen (the High Throughput virtual Screening of Dompe' pharmaceuticals co-designed by CINECA and other partners), Gromacs (for molecular dynamics simulation), HyperQueue (a flexible scheduling system from IT4I) and a new AI engine, combined with improved management techniques of large data sets.

The LIGATE is now in its final year and the various components of the project are being integrated into the final CADD (Computer Aided Drug Design) workflow. A crucial contribution by CINECA in the last year has been a molecular dynamics campaign involving the simulation of over 16,000 complexes using GROMACS and the GPU resources of Leonardo. This large undertaking has been completed in less than 4 months, accelerated by the fast pre-exascale resources of Leonardo. These results are being used to construct an accurate machine learning module which can be attached to LiGen to provide a more accurate assessment of potential drug molecules.

CINECA is also involved in one of the final tasks of the project, namely the implementation of the CADD workflow on Leonardo. This is almost ready and is expected to be employed soon by scientists in the project to look for new molecules against a range of viral diseases.

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.

ASTROPHYSICS

SPACE: Scalable Parallel Astrophysical Codes for Exascale

Nitin Shukla

Cineca

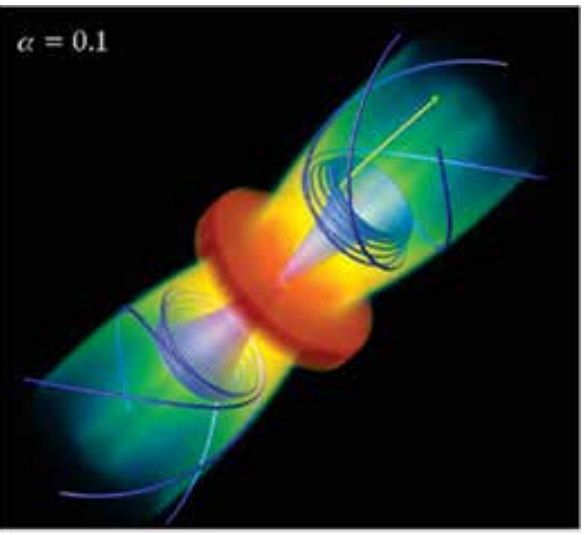
Astrophysics tackles the grand mysteries of our cosmos, from the birth and death of stars to the vast expanse of galaxies. To comprehend these complexities, researchers rely on immensely powerful supercomputers for large-scale simulations. However, the problem complexity is such that current supercomputers equipped with the most powerful processors struggle to keep pace. But a new era of supercomputing is dawning: exascale computing. These novel exascale machines can perform 1018 (this is 1 followed by 18 zeros) calculations every second leveraging the full power of what are called accelerators: Graphical Processing Units (GPUs), Field Programmable Gate Arrays (FPGAs) etc. Thus, these novel machines can help researchers overpass the current computational limits and perform larger and more detailed simulations to unveil the mysteries of our universe. CINECA stands at the forefront of this endeavour, providing cutting-edge HPC (High Performance Computing) systems and exceptional support to scientists.

CINECA is a key partner of the Centre of Excellence SPACE financed by Euro HPC JU in 2023 through a very competitive call. One of the aims of SPACE, short for "Scalable Parallel Astrophysical Codes for Exascale," is to make seven selected astrophysical codes ready to take full advantage of the exascale era. The role of CINECA in SPACE entails assisting scientists with this challenge. We collaborate with the scientists who developed these codes and provide our expertise in profiling, optimising performance, and accelerating their codes on European pre-exascale machines equipped with GPUs. The task is not simple, as many of these codes were not written in a GPU-friendly fashion. Many times, more parallelism needs to be exposed and codes must be re-factored. After each of these modifications, codes must be validated again, and performance checked.

Besides helping SPACE scientists in improving their codes and empowering them to do better science, we also share our knowledge with them by organising online and in presence dedicated HPC trainings and schools. And finally, we provide to them computational resources otherwise not available.

Since the beginning of the project, we have been providing support to profile and assess the existing status of SPACE codes using a tool called Extra-E. We have also been helping in improving the performance of two codes: Pluto and OpenGadget and enabling them to run on GPUs. In the case of Pluto, the current GPU version runs 30-40 times faster than the CPU version on a single node (see Fig1:left). When using the maximum number of GPUs available (1024 across 256 nodes), the speedup is 10 to 20 times faster than using CPUs. On the other hand, OpenGadget GPU implementation produced a speed-up of 4X with respect to the CPU-only code when 32 nodes have been used (so, using 128 GPUs). This result has been obtained by simulating the gravity driven evolution of a cosmological box of nearly 9×109 particles in a volume of about 0.13 Gpc (see Fig2:right). We are still in the process of improving such performance and to exploit all the computational possibilities that GPU and exascale machines can provide to cosmology.

Thanks to the dynamic efforts of the newly established AstroCINECA team, we've seamlessly integrated into various activities and delivered outstanding results. CINECA's participation in prestigious conference has been well-received at PASC and HiPEAC.



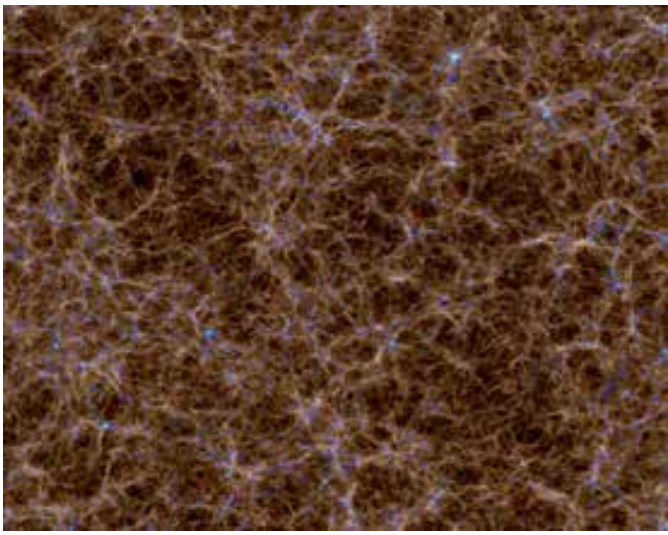


Figure 55: on the left side: a visualization of a simulation produced by Pluto. The orange part is an accretion disk from which, by effect of magnetocentrifugal mechanism, a magnetized plasma jet is ejected along the two axes (N and S). The blue lines depict the magnetic field lines of the jet. The yellow line, on the other hand, represents the trajectory of a hypothetical particle anchored to the plasma. The parameter α regulates the level of viscous turbulence in the disk. With α=0 there could be no dissipation of the angular momentum of the disk, let alone the creation of jets (a.k.a. outflows) along the rotational axis of the disk. On right side: a visualization from the Magneticum simulation set done with Gadget. The shown region spans a total size of 3800 Mpc (1 Mpc, which stands for one MegaParsec, is about 3x10^19 km). At redshift z = 0.0 (basically our present time) it contains a total number of 1.86×10^11dark matter, gas, star and black hole particles. Visualized is the gas which fills the space between the galaxies (colour coded according to its temperature from cold/brown to hot/light blue) together with the galaxies and stars forming in the simulation (coloured in white).

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.

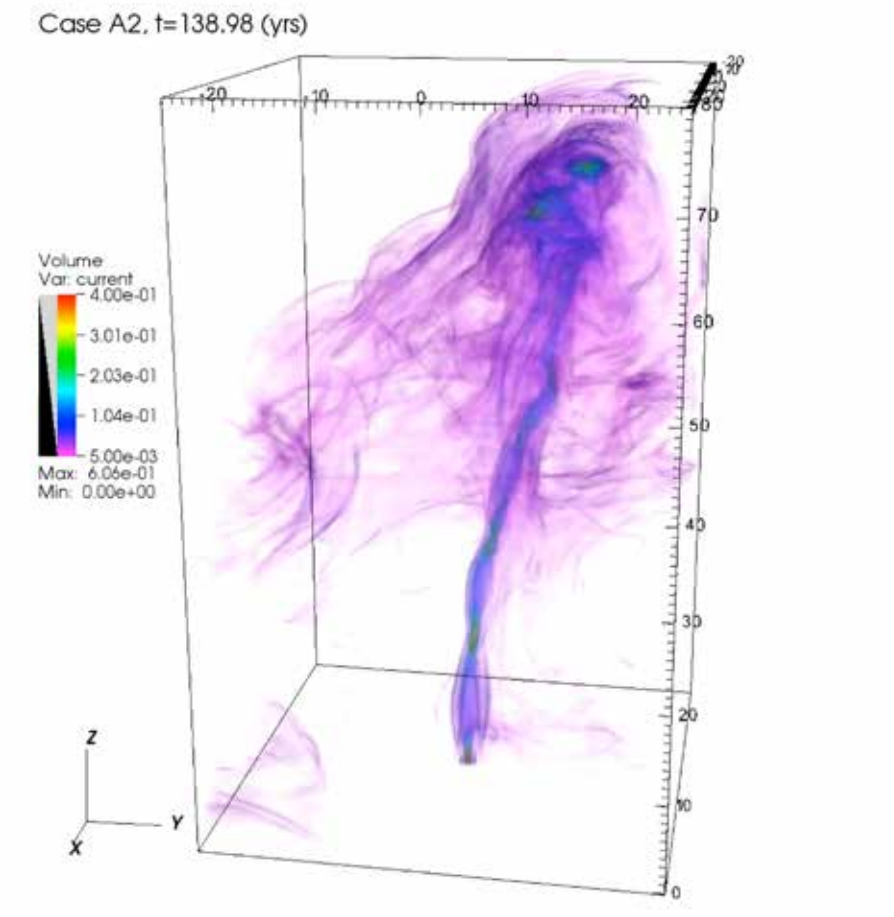
The astrophysical codes PLUTO and OpenGadget in SPACE CoE

Andrea Mignone
University of Turin
Luca Tornatore
INAF

Thanks to SPACE CoE, significant efforts are being undertaken in order to deliver a completely renewed, exascale ready version of the PLUTO code. The upcoming version - gPLUTO - will keep the fundamental structure and user-friendliness of its predecessor, but with GPU-optimized computational kernels extensively based on the OpenACC paradigm. gPLUTO will allow researchers to investigate astrophysical environments - such as extragalactic jets, magnetic reconnection and accretion disks around powerful black holes - to unprecedented levels of detail with more sophisticated simulations. For a field like astrophysics, which is largely inaccessible to laboratory experiments, the significance of this development cannot be overstated.

In the case of high-energy astrophysical sources (e.g., Active Galactic Nuclei, Pulsar Wind Nebulae, Gamma Ray Bursts), for instance, non-thermal radiation is emitted primarily through synchrotron and inverse Compton processes, but it requires efficient particle acceleration mechanisms in order to operate. Charged particles may be present, among others, in the inner regions of fast-spinning accretion disks where powerful magnetized supersonic outflows are produced. The figure below represents the output of a numerical simulation showing the electric current being dragged by such an outflow. These – and several other - challenging questions have been puzzling scientists for decades.

The figure 56 shows the electtic current dragged by an outflow simulated with the PLUTO code.



The N-body hydrodynamical code OpenGadget will be deeply revised within the SPACE CoE. Our aim is to deliver a modern, modular, easy-to-maintain and easy-to-use code, natively able to adapt to the very complex and diverse architectures that will populate the HPC panorama.

The design we have in mind relies upon more cache-friendly data structures that could lead to high vectorization efficiency and upon re-designed algorithms implementations that execute efficiently on many-cores CPUs as well as on cutting-edge GPUs.

OpenGadget will come with state-of-the-art physical modules, as refined stellar evolution and chemical enrichement, black holes accretion and feedback models, advanced radiative non-equilibrium chemistry and cooling, sophisticated star formation and magnetohydrodynamics that will provide researchers with an invaluable tool for exploring the evolution of the Universe linking large cosmic structures with star-forming sites within galaxies.

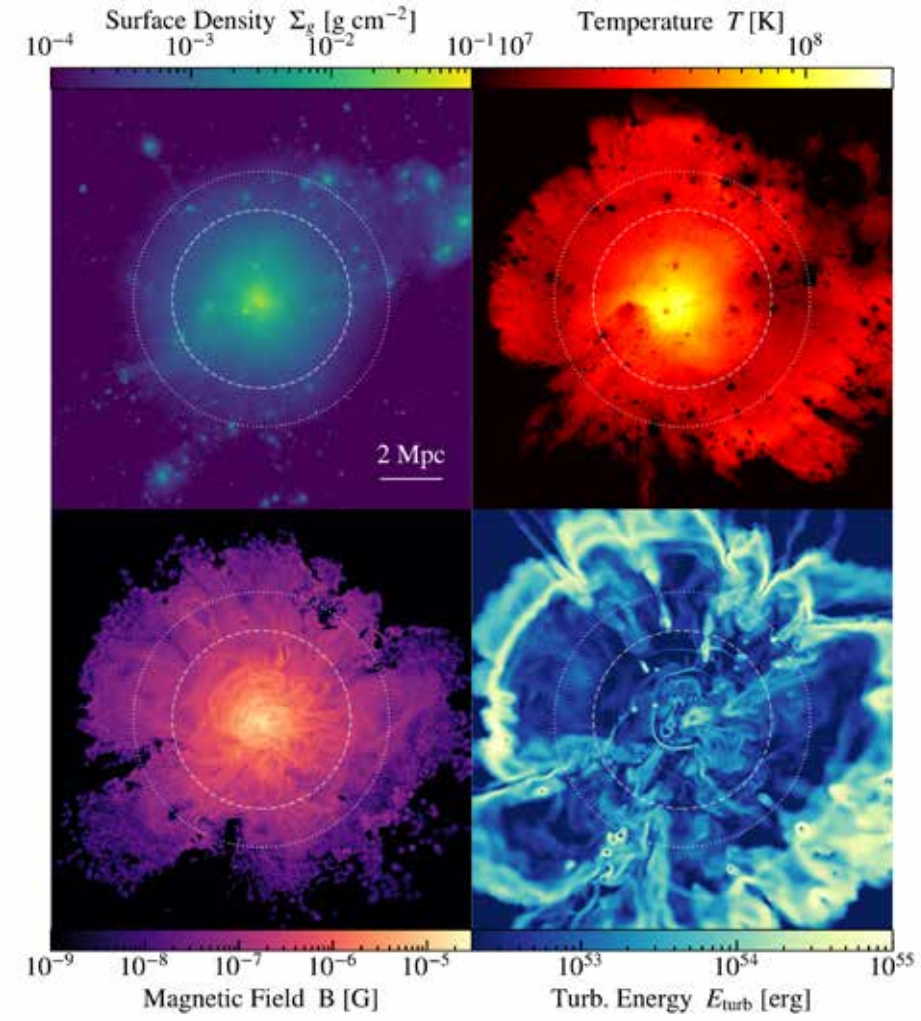


Figure 57: a galaxy cluster region from a extreme resolution simulation with magnethydrodynamics of the OpenGadget code. The cluster has a mass of 2x10^15 solar masses and is resolved with 5 billion particles with a spatial resolution of 250 pc (then, with a dynamic spatial range of ~4 orders of magnitude just for the object itself). The maps shows with impressive richness of detail fundamental pysical quantities as the surface gas density, linked with the stellar density, and temperature, linked with the X-ray emission, and, in the bottom row, the magnetic field and turbulent energy (appreciate the correlation between the two quantities). Including multiple physical processes in the simulations is of paramount importance in the era of multi-messenger astronomy; that allows to cross-correlate observables in different bands and to exaplain the sophisticated interplay among several complex physical processes. Figure from Steinwandel et al., 2023.

Energy transfer and space–time properties of plasma turbulence from magnetohydrodynamics to electron kinetic scales

Francesco Califano
Dipartimento di Fisica E. Fermi, University of Pisa, Italy
Davide Manzini
Dipartimento di Fisica E. Fermi, University of Pisa, Italy,) Laboratoire de Physique des Plasmas, CNRS, École Polytechnique, Sorbonne Université, France
Oreste Pezzi
Istituto per la Scienza e Tecnologia dei Plasmi, Consiglio Nazionale delle Ricerche, Bari, Italy

Modelling plasma turbulence in magnetised, (almost) collisionless plasmas is one of the most challenging fundamental physics problems still under discussion today.

With respect to fluid dynamics, the main difficulty lies in the presence of three different physical regimes, (magneto)fluid, ion kinetic and electron kinetic separating the injection scale from the dissipative ones. These regimes are characterized by different properties of the turbulent cascade, which reflect a change in the nature of the underlying physical processes. Indeed, diverse linear modes and local (micro)-instabilities can emerge at different scales, strongly affecting the global dynamics and in particular the cross-scale transfer of kinetic and magnetic energy. On the top of that, magnetic reconnection, probably the most famous plasma instability, plays a key role in particular by competing with the standard hydrodynamic-like turbulent cascade in non-locally transferring the large-scale energy input directly at ion kinetic scales. Space plasmas, in particular the solar wind and the Earth's magnetosphere, are probably the best natural laboratories for investigating fundamental plasma physics, in particular turbulence, by exploiting in-situ spacecraft measurements. From the theoretical side, the computational approach is at the moment the most useful and powerful tool to advance our knowledge on plasma turbulence.

Because of the interplay of these processes at play in the different regimes, both ion and electron Vlasov equations in the entire six-dimensional phase space, self-consistently coupled with Maxwell equations, must be numerically integrated for determining the correct energy transfer rate along the full spectrum. However, due to the huge amount of characteristic scales and frequencies involved, this cannot be done even on the most powerful supercomputers at disposal and one must focus on a limited part of the spectrum.

As of today, the fluid and the fluid-ion kinetic regimes have been studied numerically. Full-kinetic simulations have been recently performed by using the Lagrangian PIC approach. In this case however, due to the intrinsic numerical noise at small scales, the electron dynamics is poorly described on the last part of the spectrum. For these reasons we have recently developed and ported to GPU-based infrastructures ViDA, a fully-kinetic Vlasov-Maxwell code which employs the Darwin approximation to get rid of light waves.

We are currently exploiting the ViDA code by performing productions runs on LEONARDO supercomputer at CINECA through ISCRA projects.

Typical 2D-3V and 3D-3V production runs are made using a mesh size of 20482 x 493 and 2563 x 493, respectively. The memory requirement is of the order of ~10 TB on 64 nodes for a total of 256 GPUs. A weak scaling of the code on Leonardo is shown in Figure 1. The code is run for a few 100k steps.

We have already obtained very interesting results that will be collected in a peer-reviewed publication. In Figure 2 we show as an example the shaded iso-contours of the electron velocity for a simulation aiming at investigating the turbulence decay at electron scales. In particular, we found that the electrostatic interactions dominate with respect to electromagnetic ones. Patterns of wave modes are clearly visible in the figure. These waves are at the basis of the electrostatic spectrum that dominates on the electromagnetic one and may have important implications in the processes of energy conversions and dissipation active at electron scales.

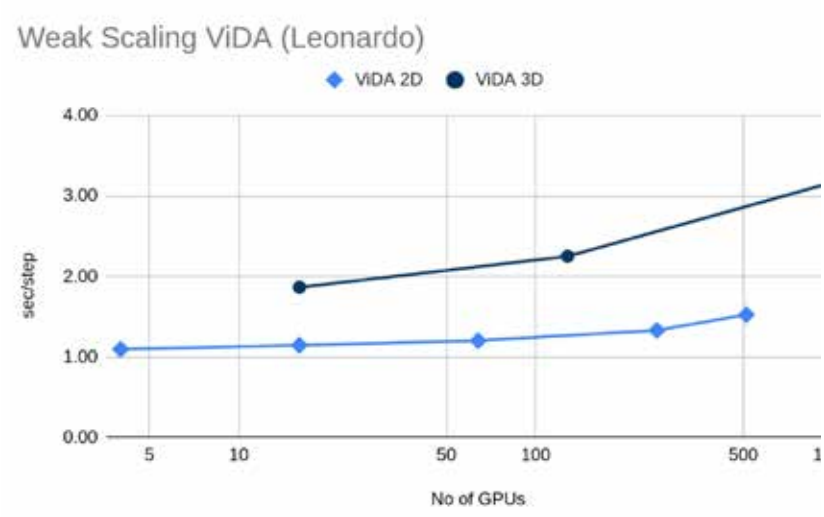
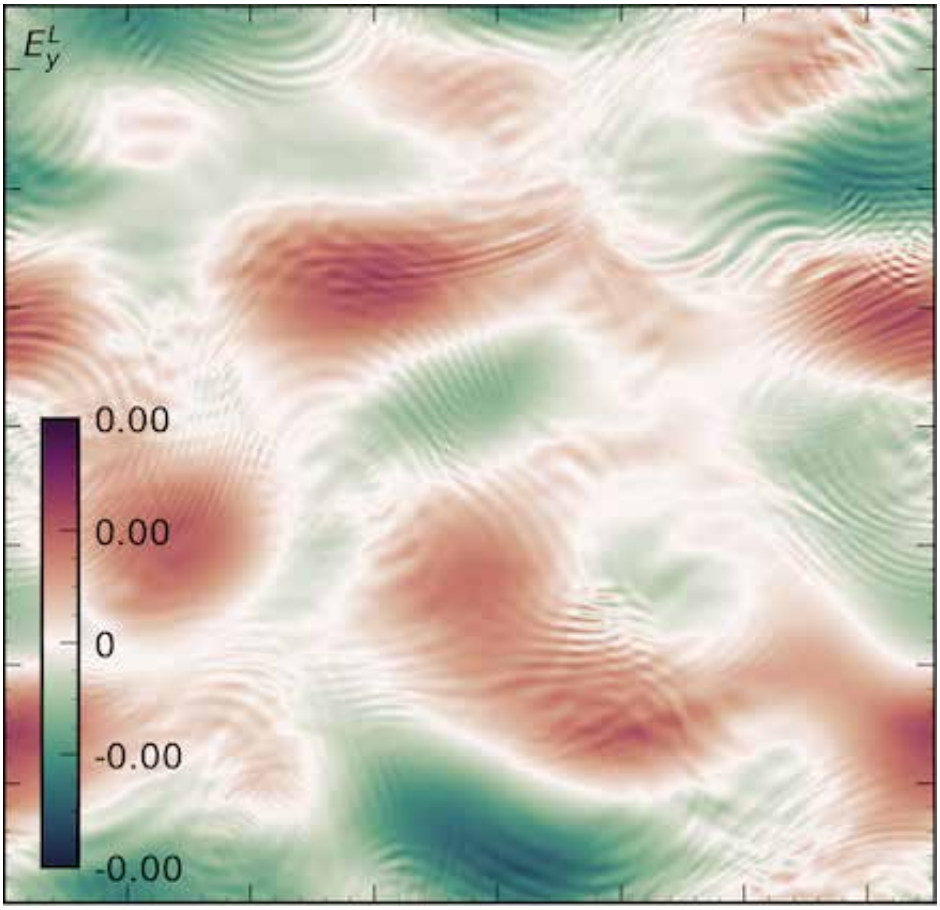


Figure 58: weak scaling of the ViDA code on Leonardo: wall time per single time step as function of number of GPUs . The size of the problem is increased to keep the memory per GPU constant.

Figure 59: electron scale plasma turbulence: the electron velocity.



References

Birdsall, C. K. & Langdon, A. B. 2004, Plasma physics via computer simulation, CRC press
Pezzi O., Califano F., et al., Journal of Plasma Physics vol. 124, Issue 12, 1-22 (2019)
Kaufman, A. N. & Rostler, P. S., 1971, The Darwin Model as a Tool for Electromagnetic Plasma Simulation, Physics of Fluids 14 (2), 446–448.

NATIONAL RECOVERY AND RESILIENCE PLAN

Italian Research Center on HPG, Big Data and Quantum Computing (ICSC)

Mirko Cestari, Massimiliano Guarrasi, Maurizio Ortali
Cineca

The High-Performance Computing, Big Data e Quantum Computing Research Centre (ICSC), created and managed by the ICSC Foundation, is one of the five National Centres established by the National Recovery and Resilience Plan (NRRP). The Center operates on a "Hub and Spoke" model, where the central Hub oversees the validation and management of the research program. The activities are executed by the Spokes, which consist of Italian universities, research bodies, and both private and public operators. The National Center comprises a transversal Spoke known as "Supercomputing Cloud Infrastructure" (Spoke 0), along with 10 thematic spokes focusing on specific areas of scientific interest.

The National Center has two primary goals:

1. to establish a comprehensive national computing infrastructure, consolidating existing High-Performance Computing (HPC), High Throughput Computing (HTC), Big Data, and network infrastructures. This will be supplemented with new targeted resources funded internally. The aim is to furnish the scientific and industrial communities with a versatile and standardized Cloud interface.

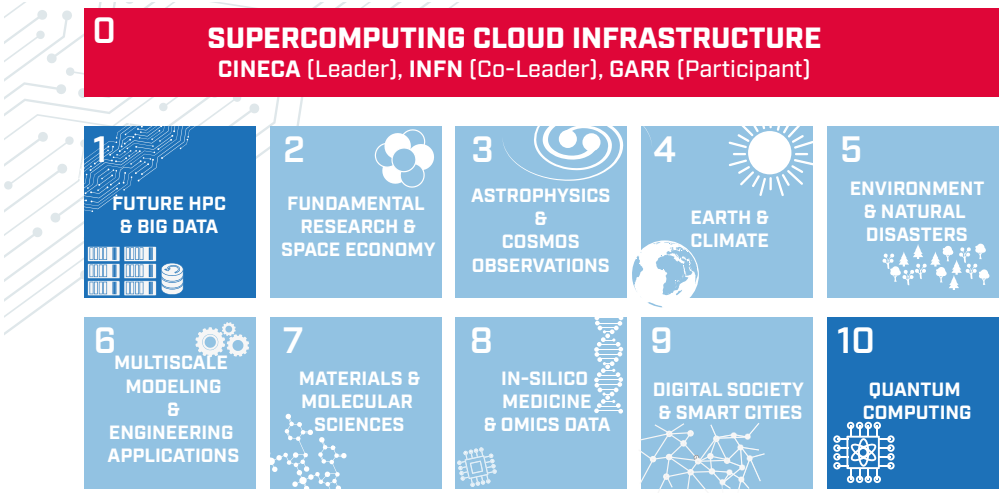
2. To cultivate a vibrant ecosystem around this infrastructure, fostering collaboration between academia and industry. This ecosystem will promote the utilization of IT resources and the advancement of new cutting-edge computing technologies.

From the technological and infrastructural standpoint, the main steps include upgrading the CINECA Leonardo supercomputer and the distributed computing infrastructure of the INFN, purchasing a quantum computer to be placed inside the Bologna Technopole, expanding the GARR-T network (which will raise national network capacity for education and research to multiples of Terabits/second) and several interventions for the creation of thematic satellite Centres in other Italian locations.

Cineca in the Center is clearly a key actor, being among the founders of the Hub, Spoke Leader of the Infrastructure component and Affiliated party (partner) in two other Spokes (Future HPC and Quantum Computing).

In the first period of the project, CINECA has coordinated the activities of the spoke 0 related to National Center Infrastructure, in close collaboration with co-leader INFN and with affiliate partner GARR. Major activities have covered preliminary allocation resource models and related functioning within the National Center, preparatory activities to the acquisition of the HPC supercomputers needed by the Center (in Bologna and in the new Data Center in Naples) and background activities with JU EUROHPC to recover the co-funding components needed for the infrastructure assets foreseen within Spoke 0 (Leonardo upgrade - LISA and first Italian Quantum computer).

In addition to the planned Supercomputing procurements initially foreseen, it is under acquisition a new integrated data lake storage as part of the procurement of a data processing system ("Upgrade of the Galileo 100") and the acquisition of a new general purpose quantum computer operating digitally using a "gate model" system and qubits technology based on superconductors with a relatively small QPU to be installed in one of the CINECA data centers, mainly for educational purposes.



Fondazione ICSC

Fondazione Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

D3 4-Health

Digital Driven Diagnostics, prognostics and therapeutics for Sustainable Health care

Elisa Rossi
Cineca

D3 4 Health 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 aims, coordinated by La Sapienza University of Rome and composed of 28 partners, at enabling new technologies for data collection and analysis in order to provide personalized 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 advancementwill be represented by the development of a Digital Twin and a Biological Twin.

D3 4 Health consists of 4 spokes which the aim to develop new solutions for the diagnosis, monitoring and therapy of 5 different diseases: Colon cancer, liver cancer, central nervous system cancer, type 1 diabetes, and multiple sclerosis.

CINECA HPC and cloud infrastructure will be a key asset for the proposal, as the appropriate usage of HPC architectures has already expressed high potentiality in the field of personalized medicine where urgent computing is critical. For the project, the Research and Simulation Lab is managed by the CINECA where HPC resources will be available for the management and analysis of data and for training and optimization of Al’s models. This allows to greatly accelerate computing processes.
https://sites.google.com/uniroma1.it/d3forhealth/home

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

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

GRINS

Giorgio Pedrazzi
Cineca

The GRINS (Growing Resilient, Inclusive, and Sustainable) Consortium, coordinated by the University of Bologna's Department of Economics, has made considerable progress in its first 18 months. It deployed AMELIA (Data Platform for Knowledge Transfer and Statistical Analysis), an online platform that offers high-quality data and analytical tools for various applications. Hosted on Cineca's AdaCloud, AMELIA operates on a robust infrastructure featuring 240 vCPUs, 1800 GB of vRAM, and 4.3 TB of storage, with planned expansions. Moreover, a pilot project on Leonardo, in collaboration with Exprivia, has started with the aim to leverage Cineca's high-performance computing resources to train AI models, expanding AMELIA's capabilities.

AMELIA's architecture is built on three core principles: SAFE data analysis for sustainable, accurate, fair, and explainable machine learning; adherence to FAIR open data principles for optimal transparency and collaboration; and rigorous data protection and ethics to meet stringent security and ethical standards. These foundational guidelines set the stage for Cineca's contributions in Spoke 0, which encompass the study of anonymization techniques and secure multi-party computation (SMPC). Cineca has focused on improving data privacy and has produced two detailed reports analysing the features, as well as the pros and cons, of available open source and commercial solutions like ARX, Amnesia, SdcMicro and Sharemind. These reports discuss the deployment of these advanced technologies on the platform to enhance data privacy and security within collaborative settings. Cineca, moreover, continues to support other project's Spokes, aiding in data acquisition and the development of specialized analytical tools.

Looking forward, the consortium aims to complete the platform's development, expand data integration, and start extensive testing phases. These efforts are set to accomplish the collaborative and analytical needs of its stakeholders, fostering significant contributions to research innovation and policy development.

The computational platform of the National Biodiversity Future Center (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 Center (NBFC) aims to monitor, conserve, restore, and valorize (MCRV) Italian and Mediterranean biodiversity. It has been created to address interdisciplinary and frontier research and innovation activities devoted to the knowledge monitoring, conservation, restoring and valorization of Italian biodiversity as a National Champion of R&D based on Key Enabling Technologies 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 form the NBFC, it has been decided to aggregate these four Platforms in a single National Access Point to Biodiversity Data and Tools.

In other words, CINECA has in NBFC a double role: 1) to build the NBFC Platform that includes Biodiversity Collections, Molecular Biodiversity, Bioresources, Ecosystem Function and Monitoring; 2) to provide Cloud and HPC resources with its infrastructures, including Galileo 100 and the pre-exascale system Leonardo.

To develop the Platform, CINECA has decided to follow the containerized microservices structure 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 Retrieve 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.

COMPUTING PROVISIONING SERVICES

Bringing HPC-fueled innovation to the EU industry

Claudio Arlandini
Cineca

Startups and SMEs can access to Cineca HPC services mainly through calls for innovation projects: Cineca is the beneficiary of several projects aimed to develop Proof of Concepts (PoC) and innovation studies, funded by the EU and national agencies. Startups and SMEs can contact the Cineca team for tech-transfer to identify which funded project best meets their needs and to gain the insight to submit a solid proposal. Once selected, the winning proposals are supported by Cineca, whose role may range from supplying the HPC resources, to entirely provide the expertise the winners may be lacking to reach their target. This test before invest approach fosters a virtuous model of innovation and enables SMEs and startups to discover how HPC might support their innovation path, minimizing the risks.

Cineca has a particular focus on funded projects for industry related to Computational Fluid Dynamics (CFD) simulations, Artificial Intelligence (AI) and Big Data.

Cineca has a long and rich history of developing PoC involving CFD simulations, thanks to which Cineca gained a robust reputation among the engineering community. We mention the projects EXCELLERAT P2, developing tools and frameworks for the engineering at the exascale level, and exaFOAM, preparing the most utilised open source CFD toolkit OpenFOAM for novel architectures.

Concurrently, the recent decade has witnessed the AI breakthrough, clearly producing a loud echo across the industrial environment as well. This is evident from the increase of SMEs approaching Cineca with the aim of incorporating the AI benefits in their own workflow. Indeed, several AI-innovation studies were co-developed with Italian SMEs within the framework of the FF4EuroHPC project, renewed in early 2024 as FortissimoPlus.

The technology is at disposal and Cineca offers the necessary guidance to leverage it. However, SMEs often needs to develop the expertise to independently manage it. For this reason, a consistent amount of Cineca's effort is devoted to training events focused on HPC for industry, i.e. the ones funded by EuroCC2.

In the following, we will show an interview to a startup who significantly grew by accessing to Cineca AI-services through a 100% funded project. We will devote a paragraph to ENI, the biggest industrial partner of Cineca, then we will illustrate two success stories related respectively to CFD and Data Lake implementation, the latter giving an example of how HPC can be useful for Big Data.

Axyon Interview: boosting AI model for finance with HPC

Eric Pascolo
Cineca
Giacono Barigazzi
Axyon

We met Giacomo Barigazzi, founder of Axyon. Axyon is a company that has gone through an HPC technology transfer path with the support of Cineca. With a background in Business Administration and an advanced master's degree in Econometrics, Giacomo Barigazzi founded Axyon AI, guiding it in creating strategic partnerships and establishing it in the field of AI innovation. He contributes to EU AI policies focusing on SMEs, and has previous entrepreneurial experience in ICT.

WHAT DOES YOUR COMPANY DO? WHY DID YOU START A TECHNOLOGY TRANSFER PATH IN THE FIELD OF HPC?

Axyon AI stands at the forefront of revolutionizing the asset management industry by harnessing advanced AI and Deep Learning technologies. We are dedicated to empowering asset managers with cutting-edge tools to forge superior investment strategies and our core mission centres on exploiting alpha opportunities and broadening the spectrum of alpha diversification for our clients, utilizing AI-driven insights to forecast asset performance with remarkable precision. The foundation of our expertise lies in a deep fusion of technological prowess and comprehensive financial sector experience, positioning us as trailblazers in delivering innovative solutions.

Embarking on a strategic journey towards technology transfer in the field of High-Performance Computing (HPC) was a calculated move, driven by the intricate computational demands of our AI models. HPC systems are indispensable in managing the data volumes and complex computations our AI algorithms require. This seamless integration significantly enhances the speed and efficiency of our model training and operational processes, marking a leap forward in our capability to deliver superior outcomes.

The rapid evolution of AI and deep learning within the financial sectors underscores the necessity for a robust computational infrastructure capable of navigating the complexities of financial time series and predictive analytics. By anchoring our operations on HPC, we provide a solid foundation for addressing these intensive tasks, enabling us to supply our clients with avant-garde solutions that are not only precise but also timely. Our venture into HPC technology transfer underlines our unwavering commitment to continually explore the limits of AI's potential in investment management, ensuring our position as leaders in a competitive and dynamic industry.

WHAT WERE THE MAIN CHALLENGES AND OPPORTUNITIES YOU ENCOUNTERED DURING THIS PROCESS?

While the technology transfer process in the field of HPC presented its challenges, the collaboration with CINECA ultimately offered significant opportunities for growth and development. It allowed us to expand our expertise, enhance our product offerings, and solidify our standing as a leader in AI-powered investment solutions.

The main challenges and opportunities faced during the implementation process have been both technical and organizational.

Technical Challenges:

1. Integration: One of the main technical hurdles was integrating our advanced AI models with CINECA's HPC infrastructure. Ensuring compatibility and optimizing performance across different computing environments required significant effort.

2. Scalability: Another challenge was scaling our AI solutions to leverage the full potential of HPC systems effectively. This involved adapting our algorithms to work efficiently across multiple nodes, taking into account the parallel computing architecture of HPC.

3. Data Handling: Managing and processing vast volumes of data in a high-performance computing environment presented challenges in terms of data storage, transfer, and privacy concerns. Implementing efficient data pipelines and ensuring secure data practices were crucial.

Organizational Challenges:

1. Collaboration: Bridging the knowledge gap between our team and the HPC experts at CINECA required fostering a collaborative environment. It was essential to ensure smooth communication and effective knowledge sharing.

2. Training and Skills Development: The specialized nature of HPC meant that our team had to undergo significant training to fully utilize the capabilities of CINECA's resources. Developing the requisite skills for HPC took time and resources.

Opportunities and Benefits:

1. Enhanced Computing Power: Access to CINECA's HPC facilities provided us with unprecedented computing power, allowing us to tackle more complex problems and accelerate our AI models' training and execution times.

2. Advanced Research Capabilities: The collaboration opened new avenues for research and development, enabling us to explore innovative AI applications that were previously beyond our reach due to computational limitations.

3. Knowledge Gain and Expertise: Working closely with CINECA's experts allowed our team to gain valuable insights into HPC, enhancing our technical capabilities and enabling us to incorporate best practices in HPC into our operations.

4. Market Competitiveness: Leveraging HPC has significantly improved our solutions' performance, making them more attractive to clients and strengthening our competitive position in the market.

IN WHAT WAYS HAS THE APPLICATION OF SUPERCOMPUTING TO YOUR PLATFORM PROVIDED ADDED VALUE?

Supercomputing is not just an enhancement to our platform; it's a cornerstone of our value proposition. The financial sector is increasingly data-driven, with competitive success hinging on the ability to quickly analyze and act on vast amounts of information. HPC's computational power is fundamental to Axyon AI's ability to provide these capabilities at scale. It underpins our innovative approach to financial analytics, enabling the processing and interpretation of complex data sets that traditional computing resources could not handle. This technological advantage is integral to our business model, allowing us to deliver advanced AI-powered investment solutions that set us apart in the market.

The application of supercomputing, facilitated through our partnership with CINECA in High-Performance Computing (HPC), has significantly enhanced Axyon AI's platform by providing added value in several key areas:

1. Enhanced Analytical Capabilities: Supercomputing has enabled us to process vast datasets and perform complex calculations at unprecedented scale. This has sharpened our analytical capabilities, allowing for the generation of deeper insights and more predictive accuracy in financial market trends. By leveraging these enhanced capabilities, our solutions offer asset managers a competitive edge in identifying alpha opportunities and executing investment strategies.

2. Innovation Acceleration: The high computational power afforded by HPC has accelerated our innovation cycles, enabling us to develop, test, and deploy new AI models and features much more rapidly. This agility allows us to quickly adapt to market changes and client demands, reinforcing our position as market leaders in AI-driven investment solutions.

3. Scalability and Performance: Supercomputing has allowed our platform to scale operations without compromising performance. We can now accommodate more clients and manage larger datasets, all while maintaining the high performance of our AI applications. This scalability is integral to our growth strategy, helping us to attract a broader client base and enhance our market presence.

4. Cost Efficiency: By optimizing our computational processes, we have achieved significant cost efficiencies in data processing and model training. These savings are passed on to our clients, making our solutions not only more powerful but also more cost-effective compared to traditional financial analysis tools.

5. Risk Management: The application of HPC enables more sophisticated risk analysis models,

offering our clients advanced tools for managing financial risks. This supports better informed, data-driven decision-making, which is crucial in the volatile world of investment management.

DID OTHER USEFUL ASPECTS EMERGE DURING THE TECH TRANSFER PATH, BESIDES THE APPLICATION OF THE TECHNOLOGY?

During our technology transfer journey with CINECA, we uncovered a wealth of benefits that transcended the straightforward application of the technology itself. Our collaboration with CINECA evolved into a dynamic partnership, sparking technological progress, forging strategic alliances, and promoting a rich exchange of knowledge.

Engaging with CINECA extended beyond merely tapping into their High-Performance Computing (HPC) prowess; it fostered a symbiotic relationship that championed mutual innovation and learning. CINECA's mastery in navigating intricate HPC infrastructures, coupled with their profound insight into computational intricacies across diverse scientific arenas, significantly enriched our efforts in financial analytics with HPC. Through a series of regular interactions, workshops, and collaborative projects, we exchanged insights and best practices, elevating both entities' operational and strategic capacities.

A pivotal element of our alliance with CINECA was the opportunity to connect with the academic world. Leveraging CINECA's robust connections with research institutions and universities, particularly with the University of Modena and Reggio Emilia (UNIMORE), opened new avenues for Axyon AI in academic collaboration. This collaboration encompassed joint research initiatives, internships, and PhD sponsorships, positioning us at the innovation forefront in AI and machine learning for financial services. These academic partnerships not only kept us aligned with the latest technological breakthroughs but also enriched our team with emerging talent, offering fresh insights and specialized expertise.

An unforeseen yet immensely valuable outcome of this collaboration has been our active role in cultivating a community centred around HPC and AI applications. Through participating in and contributing to conferences, workshops, and forums alongside CINECA and our academic allies, we have both contributed to and reaped benefits from an expanding ecosystem dedicated to technological and knowledge exchange. This initiative in community building has not only bolstered our reputation as thought leaders but also provided us with broader perspectives from various fields, enhancing our methodologies and solutions for addressing complex challenges.

WHAT ARE YOUR GOALS FOR THE FUTURE?

Looking towards the future, Axyon AI is firmly positioned to continue its pioneering work at the intersection of artificial intelligence and finance, with the goal of leading the path in developing cutting-edge AI models for predictive analytics through refining the accuracy, efficiency, and applicability of our models to various asset classes and investment strategies.

Supercomputing stands as the cornerstone of our ambitions and the realization of our vision. It underpins our research and development efforts in AI, empowering us to analyze extensive datasets, refine complex models, and implement high-frequency trading strategies with unparalleled speed. As we explore the frontiers of AI's potential within finance, the advanced capabilities of supercomputing become increasingly indispensable. It allows us to expedite the innovation process, rapidly transforming novel concepts into actionable solutions, and to manage the growing complexity of our models and the expanding volume of data, ensuring that our solutions consistently lead the market.

Moreover, supercomputing facilitates robust, complex results, providing our clients with a substantial competitive edge. It not only enhances our existing strengths but also paves the way for new avenues of innovation and development, unlocking extraordinary levels of predictive accuracy and operational efficiency. In summary, supercomputing doesn't just bolster our current operations; it propels us into new realms of possibilities, enabling us to harness unprecedented predictive insights and achieve greater levels of performance and client satisfaction.

INDUSTRIAL APPLICATIONS

Industrial applications: Oceanwings®: giving automated wings to motor vessels

Francesca Gebbia
Cineca

AYRO created a patented rigid wing technology, named Oceanwings® that solves the environmental and economic dilemma faced by many ship owners: reducing their environmental footprint.

The Oceanwings® consists of a two-element articulated wingsail providing aerodynamic lift. It is controlled by an automated AI-based software that optimizes the wing's angle of attack and camber in real time to maximize thrust and thereby can save up to 45% of a ship's fuel consumption.

The optimal configuration of Oceanwings® is different for each type of vessel in terms of size and position because the wind flow is different for every ship. Hence, the challenge of AYRO was to optimize the Oceanwings® technology and to broaden its customer base to cover any kind of ship.

However, AYRO lacked both expertise in the specific numerical simulation tools and the computing resources to run the simulations for different aerodynamic characteristics at the required scale.

PoliMi, AYRO and CINECA developed a robust tool consisting of a command line interface that can be used to automatically launch CFD simulations on HPC systems. This replaced the previously used commercial software with a highly scalable solution, which is easy to use and allows thousands of aerodynamics calculations to be performed in a few hours. Simulation results are included in ToolsPole's BREVA software environment, which finds the best trim of the wings with respect to optimal thrust and generates large databases to feed the AI-based control software.

Direct benefits

- Additional 300,000€ annual turnover expected for AYRO by selling improved control software to existing clients.
- Wing sails related turnover is projected to increase by 300% until 2030.
- ToolsPole will be able to leverage improved technology toolchains to increase its sales by up to 40,000€ annually.
- CINECA could gain additional business as HPC supplier approximately 20,000€ per year.

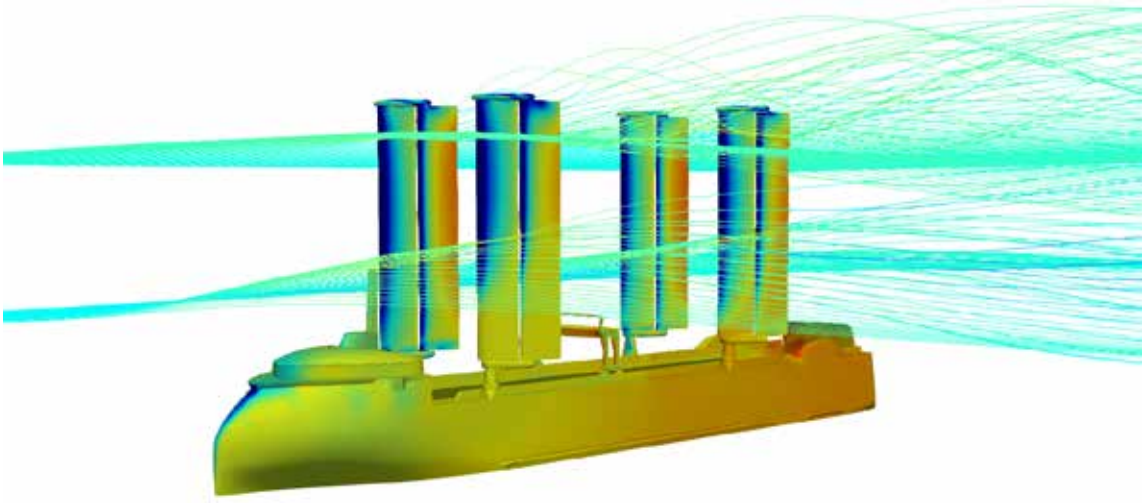


Figure 60: source: https://www.ff4eurohpc.eu/en/success-stories/2023102011561452/hpcbased_design_of_wind_assisted_propulsion_technology

Yuppies: Data Lake and HPC in constructions sector

Eric Pascolo
Cineca

YDMS (Yuppies Data Management System) is a project focused on creating a Data Lake infrastructure specifically for managing, storing and organizing datasets related to surveys of buildings and other civil infrastructure.

To accomplish this, the project strategically employs the power of data lakes and high-performance computing (HPC). The YDMS is hosted on Cloud as a service and on HPC for computational part, the integration is possible thanks to a dual storage system that maps the Parallel Filesystem to an object storage accessible via S3 API.

A centralized data lake enables YDMS to store diverse datasets from surveys, including images, sensor readings, and technical reports, with the flexibility traditional databases lack. The integration of HPC, provides the computational muscle needed to process these complex datasets quickly. This allows Yuppies to perform sophisticated analytics, simulations, and modelling that would be difficult or time-consuming on standard computers.

By combining the scalability of data lakes with the power of HPC, YDMS establishes a robust infrastructure for storing and organizing survey data in a way that optimizes analysis for energy management, facilities management, and the preservation of the public heritage as a whole. The project enables Yuppies to use the stored data for various purposes, potentially including identifying patterns, predicting potential issues, and optimizing maintenance.

The project has demonstrated that introducing new digitalization techniques, such as the introduction of a Data Lake and HPC, in a field like construction can open up new possibilities in terms of both management and optimization of complex structures such as buildings, leading to significant savings.

Source: https://eosc-dih.eu/ydms/

Eni's HPC6: A Leap in High-Performance Computing with Cineca's Contribution

ENI S.p.A.

In 2023, the collaboration between Cineca and Eni made significant strides with the groundwork for the operational management of Eni's new high-performance computing system, HPC6. This system aims to enhance Eni's computational power from the existing 70 PFlop/s of HPC4 and HPC5 to a remarkable peak of over 600 PFlop/s.

TECHNOLOGICAL LEAP

HPC6 is being built on state-of-the-art technology, employing HPE Cray EX4000 systems and HPE Cray ClusterStor E1000 technologies. It integrates AMD EPYC™ CPUs and AMD Instinct™ GPUs, supported by the HPE Slingshot Interconnect for handling exascale-class workloads. This setup promises unparalleled computational efficiency and versatility, positioning HPC6 among the world's most advanced supercomputers dedicated to industrial applications.

COMMITMENT TO SUSTAINABILITY

Eni's new system will operate within the Green Data Center, featuring an innovative liquid cooling system to boost energy efficiency and reduce carbon emissions. This aligns with Eni's commitment to sustainability, reflecting a balance between high performance and environmental responsibility.

CINECA'S ROLE IN HPC MANAGEMENT

Cineca will help Eni manage HPC6, and will support software engineering and porting, under a longstanding framework agreement with the company. Leveraging many decades of experience in managing large-scale computing systems and complex data centers, Cineca's HPC system managers will contribute to ensure the supercomputer's full functionality both onsite and remotely. This includes working with Eni in developing and optimizing the software running on the supercomputer, ensuring it meets the highest performance standards.

TECHNICAL HIGHLIGHTS

- Computing Power: HPC6 achieves over 600 PFlop/s peak and 400 PFlop/s sustained performance.
- Node Composition: Each node includes a 64-core AMD EPYC™ CPU and four AMD Instinct™ MI250X GPUs.
- System Size: Comprising 3472 computing nodes and 13,888 GPUs, organized into 28 racks.
- Network: The HPE Slingshot Interconnect network ensures fast, reliable communication.
- Cooling System: Direct liquid cooling technology dissipates 96% of heat, enhancing efficiency.
- Power Consumption: Maximum power consumption of 10.17 MVA, reflecting high efficiency.

In 2023, Eni took critical steps towards the future operationalization of HPC6, and Cineca continued to play a crucial role in managing Eni's systems. This partnership showcases the synergy between industry and research institutions, enhancing Eni's computational power and aligning with broader goals of innovation, sustainability, and energy transition. The advancements made set a benchmark for future projects, demonstrating the potential of high-performance computing in addressing global challenges.

Future Plans

Trends for the future

Sanzio Bassini
Cineca

Our institution's mission is to support scientific research for knowledge discovery and technological innovation for economic development and quality of life. We achieve this by designing and implementing advanced solutions for numerical computing and data processing. In the scientific process, mathematical models validated by data are used to formalize new theories. This highlights the crucial role of increasingly powerful supercomputing and data processing systems in scientific discovery and innovation. Doing science and innovation requires supercomputing, big data processing, including data lakes, and machine learning to extract value from data. This has led to a highly competitive landscape; a global race where individual countries struggle to keep up. This intense competition, however, drives continuous innovation in methods and tools.

No other technological field has seen the scale-up of power and innovation witnessed in supercomputing and big data. Supercomputing performance grows by a factor of 10x between generations. The difference in computing power between CINECA's first supercomputer and the state-of-the-art Leonardo system is in the billions. Predicting future trends is challenging, but we can identify disruptive trajectories to address.

Big data is a key area. When the Bologna Tecnopolo project was conceptualized in 2016-2017, it was envisioned as a "big data valley." Supercomputing systems remain central, with the Tecnopolo being one of the world's largest supercomputing concentrations. However, the value of data and its processing is the primary driver of development.

We achieved this by:

- Conceptualizing the Leonardo system for processing data for artificial intelligence applications.
- Equipping CINECA's computing system with cutting-edge technologies to handle big data's variability, volume, and velocity.

Therefore, a future development trajectory will focus on broad data management. This includes creating high-capacity data lakes with flexible access protocols, high bandwidth, and security/confidentiality features.

Another priority is the access model for computing resources. Numerical models have evolved into complex workflows combining computing with machine learning and AI. This has led to a convergence between cloud service provider and high-performance computing system architectures. The future will require a focus on federated access models between tier0 and tier1 systems with shared high-bandwidth access to large data lakes. This will address needs such as training large language models and production-based inferential methods.

Finally, the disruptive impact of artificial intelligence based on large language models (LLMs) needs to be addressed. The “big data valley” will need to transform into an “AI factory” centered on a supercomputing system capable of supporting LLM training workloads.

This may look like a self-fulfilling cycle, but significant challenges and fierce competition lie ahead. To succeed, we need to interpret signals from the technological innovation and R&D worlds, alongside mobilizing significant economic resources. However, the most crucial resource is a competent and consistent staff structure. In this paradigm shift, software development skills will become equally important as hardware, placing people and skills at the center of the scene.

CINECA HPC Annual Report 2023/24

EuroHPC
Joint Undertaking