

CINECA

HPC ANNUAL REPORT

CINECA HPC Annual Report 2022/23

2022/23



EuroHPC
Joint Undertaking

ISBN 9788886037464

Cineca HPC Report 2022-2023
website: www.hpc.cineca.it
mail: info-hpc@cinca.it

Contents are © Cineca 2021- 2022 are released under
the Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)
Creative Commons license, version 3.0.

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, Claudio Arlandini, Marco Ballarin, Sanzio Bassini, Laura Belletani, Domenico Bonanni, Neva Besker, Susana Bueno Mininguez, Domenico Bonanni, Simona Caraceni, Cinzia Caroli, Federico Cattorini, Daniele Cesarini, Mirko Cestari, Beatrice Chiavarini, Patrizia Coluccia, Maurizio Cremonesi, Chiara Della Casa, Daniele De Luca, Francesca Delli Ponti, Andrew Emerson, ENI S.p.A, Francesco Falciano, Gabriele Fatigati, Federico Ficarelli, Daniela Galetti, Bruno Giacomazzo, Silvia Gioiosa, Silvia Girardello, Luca Govoni, Massimiliano Guarrasi, Antonella Guidazzoli, Daniel Jaschke, Massimo Infunti, Mariella Ippolito, Maria Chiara Liguori, Nicola Loglisci, Marco Masseroli, Riccardo Mengoni, Simone Montangero, Giuseppa Muscianisi, Salvatore Orlando, Maurizio Ortali, Daniele Ottaviani, Eric Pascolo, Enrico Prati, Giorgio Pedrazzi, Elisa Rossi, Monica Santini, Gabriella Scipione, Donatella Sforzini, Fausto Tomei, Nicoletta Tranquillo, Claudia Truini, Roberta Turra, Marco Venturini, Giulia Villani, Cinzia Zannoni.

CINECA

2022 - 2023

Dear colleagues,

To introduce the annual report for 2022/2023, first of all, a warm welcome to the newcomers hired in the Supercomputing Application & Innovation Department staff during the last twelve months. A brief presentation section on the new colleagues already hired (around 25) is present in this annual report. This action to expand the staff of the HPC department is progressing further with a new selection underway for an additional 30 positions, and a new call for candidatures is expected before the end of 2023.

After an important development action relating to the supercomputing infrastructure, including the production launch of the EuroHPC Leonardo system, which was confirmed as #4 in the TOP500 ranking of May 2023, the development strategy focused on expanding the staff. The goal is to double the size of the staff, going from 100 to around 200 figures by 2024, both in terms of the number of personnel and in terms of profiles and domain skills. This expansion is necessary to cope with the growing central role of CINECA in the HPC, Big data, and Quantum computing scenario of the European international system and the Italian national system. Just to recall some positioning elements: CINECA HPC is currently the core partner of 40 research and innovation projects funded by Horizon Europe and EuroHPC. It has also signed 20 joint development agreements with major academic and institutional structures of the national research system, has 20 ongoing collaborations for industrial innovation, and participates in 8 projects of strategic relevance within the missions of the Italian PNRR. The report contains in-depth sections on many of the actions referred to, providing summarized synthetic data that highlight the relevance of our role and the responsibility of constituting the keystone of an ecosystem at different levels—European, National, and territorial—where advanced research, technological innovation, and socio-economic development converge and progress.

While the current period has focused on strengthening specialist support and staff development, a new infrastructural evolution plan has already been launched, with CINECA HPC as the primary implementer of an extraordinary commitment resulting from the aggregate mobilization of various economic flows of funds. The funds available to CINECA, the funds from various European subjects such as EuroHPC (for the upgrade of the Leonardo system and the installation of a Quantum Computing system) and EUROfusion (for the update of the system dedicated to this community), as well as national funds (for the deployment of tier1 systems, including partitions dedicated to mission-critical services, such as the system for the newly established Italia Meteo agency), amount to a total of approximately 125 million Euros. In addition to these investment funds, there are also economic resources allocated to cover operational costs, which are already consolidated and of a similar magnitude, considering a production cycle of 5 years. Operationally, we will be engaged in 6 procurement actions: 2 in support of EuroHPC (LISA and HPCQS-Italy projects) and 4 directly by CINECA (for the acquisition of the PITAGORA system intended for the EUROfusion community, Tier1 Tecnapolo of Bologna, Tier1 CINECA HPC Naples, and GALILEO 100+ Cloud and data lake system, as an upgrade of the already in-production Galileo100 cloud system).

You will find a section at the end of the report that discusses the upcoming trends in high-performance extreme computing, high scalable throughput computing, and high-performance data processing for the period 2023/2024.

However, this will be the subject of the next report, in which we will share what we will be able to achieve in the next intense twelve months of commitment.

Best regards

A handwritten signature in blue ink, appearing to read "Sergio Follini".

| | | | |
|----|--|---|----|
| | 2022 IN NUMBERS | | |
| 2 | HPC Infrastructures | Integrating Artists into the Scientific Loop: The GRIN Project | 62 |
| 6 | Users' statistics | CHEMISTRY | |
| 12 | Resource allocation: projects, agreements, collaborations | In silico drug design with LIGATE | 64 |
| 13 | PRACE - EuroHPC | CoE - CENTERS OF EXCELLENCE | |
| 15 | Iscra | EXCELLERAT | 66 |
| 16 | EUROFusion | MAX | 68 |
| 18 | Agreements | chEESE-2P | 70 |
| 19 | Training HPC | NATIONAL RECOVERY AND RESILIENCE PLAN | |
| 22 | Staff | ICSC | 72 |
| 29 | Events | NBFC | 73 |
| 30 | HPC H-index | ECOSISTER | 75 |
| 33 | SCIENTIFIC OVERVIEW | GRINS | 76 |
| | LIFE SCIENCES | D-3-4 Health | 76 |
| 36 | Projects in “-omics,, science field | Digital Strategy and Platform for Cultural | |
| | Genomic Computing and CINECA for genomic research and precision medicine | Heritage | 77 |
| | EARTH AND ENVIRONMENTAL SCIENCES | NEW TECHNOLOGIES AND HARDWARE | |
| | HIGHLANDER user-tailored Downstream Application and (pre-)Operational Services (DApOS) | Cineca HPC Department Unveils New Tech Wall Project | 72 |
| 39 | | ASTROPHYSICS | |
| 42 | | General relativistic simulations of massive black hole binary mergers in magnetized environment | 84 |
| | BIG DATA AND ARTIFICIAL INTELLIGENCE | Modeling Cassiopeia A | 86 |
| 46 | Enhancing Climate Risk Assessment in Agriculture Insurance | COMPUTING PROVISIONING SERVICES | |
| 48 | MobilitySquare: Innovating E-Bike Rental | Bringing HPC-fueled innovation to the EU industry | 90 |
| | Exploring the Integration of Artificial Intelligence in Artwork Analysis | ENI | 91 |
| 51 | | Dompé and Cineca | 93 |
| | MATERIAL SCIENCES | EuroCC | 95 |
| 48 | Material science getting ready for exascale | NEXT TRENDS | |
| | QUANTUM COMPUTING | Leonardo Early access | 97 |
| 51 | Cineca quantum computing perspective | What's next | 98 |
| | HUMANITIES AND SOCIAL SCIENCES, DIGITAL ECOSYSTEMS | | |

2022 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.

HPC infrastructures

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

In year 2022 the Cineca HPC infrastructure was composed by four clusters, the Tier-0 systems Marconi and Marconi 100, the Tier-1 system Galileo 100 and the DGX system, an accelerated system for AI based on NVIDIA A100 GPUs. Finally, in addition to aforementioned systems, other infrastructures are present and dedicated to special uses or to specific communities. These includes:

GATEWAY: 88 nodes (Skylake Intel Xeon 8160) with 48 cores each, connected with a Intel OmniPath network. It is a dedicated interactive cluster for EUROfusion.

CNAF: 265 Broadwell nodes (Intel Xeon E5-2697) increased to 517 on April 2022. They are 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.

During this year, the main event affecting the HPC infrastructure has been the enhancement of Galileo 100 system. This HPC system, co-funded by the European ICEI (Interactive Computing e-Infrastructure) project and available since August 2021,

has been expanded on November 2022 with the introduction of 82 additional nodes (2 racks).

Also, during 2022 the Tier-0 cluster Marconi A3 was reduced of 64 nodes that were dedicated to expanding the Gateway system, a dedicated interactive cluster for EUROfusion.

Moreover, WLCG has been expanded by 252 nodes taken from Meucci (3 racks) and Galileo (1 rack).

The Tier-0 cluster Marconi 100 and the special system DGX remained unchanged during 2022.

In year 2022, also started the deployment of Leonardo cluster, which is composed by two main computing modules: a Booster module, that was deployed in October 2022, which purpose is to maximize the computational capacity through NVIDIA Ampere based GPUs, and a Data Centric module, that will be ready in 2023, aiming to satisfy a broader range of applications via Intel Sapphire Rapids CPUs.

| | CPU | Total nodes/ cores | Memory/ node |
|---------------------------|--|--|---|
| GALILEO 100 (DELL) | Dell Dual Socket PowerEdge + NVIDIA Volta 100 | 554 nodes --> 636 nodes since November 2022 26.592 cores 2x36 GPUs | 348 GB + 3.0TB Optane on 180 fat nodes |
| MARCONI 100 (IBM) | IBM Power AC922 + NVIDIA Volta 100 | 980 nodes 31.360 cores, 4x980 GPUs | 256 GB |
| MARCONI A3 – SKL (LENOVO) | Intel SkyLake 2x Intel Xeon 8160 @2.1GHz 24 cores each | 3.124 nodes 149.952 cores | 192 GB |

| Supercomputer | Resources (Core Hours) | Performance capacity (Exaflops) |
|---------------|------------------------|---------------------------------|
| Marconi A3 | 1.313.579.520 | 315.259 |
| Marconi 100 | 274.713.600 | 988.969 |
| Galileo 100 | 238.075.776 | 63.031 |
| Total | 1.826.368.896 | 1.367.259 |

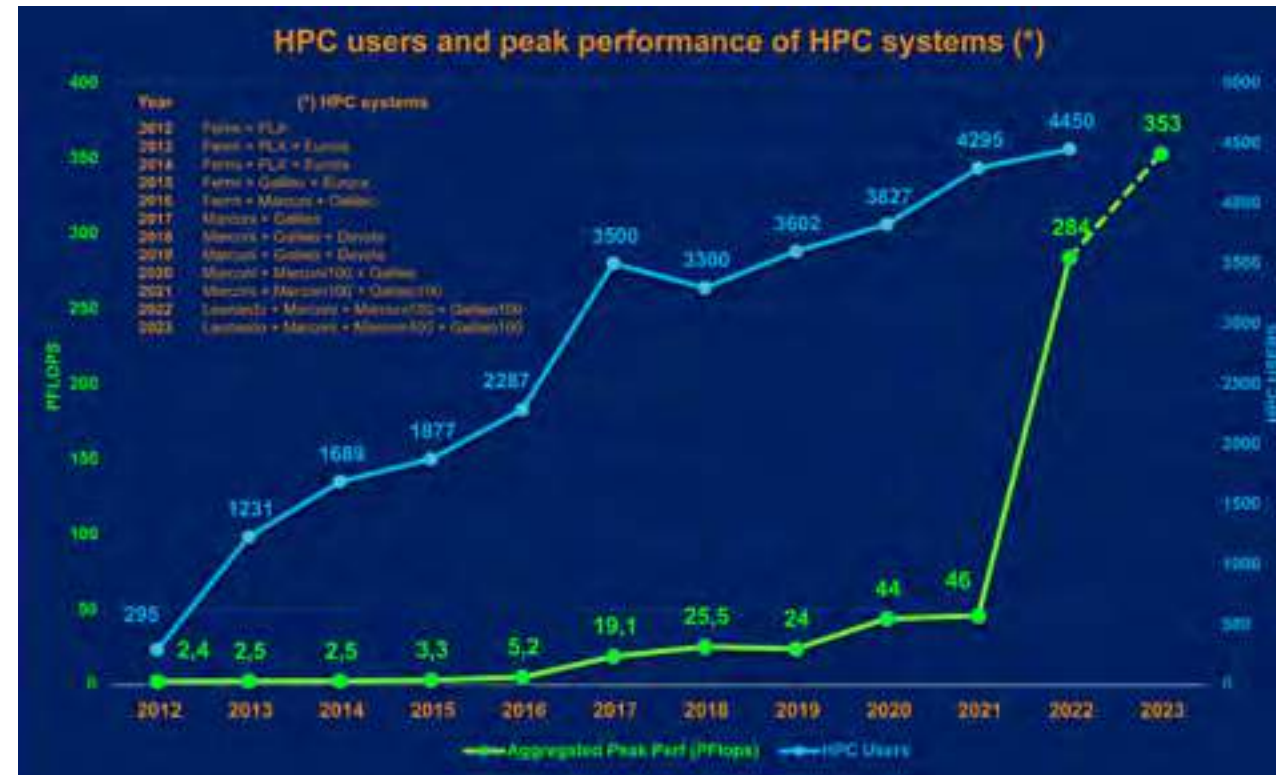


Figure 1: in this figure it is represented the Peak Performance (PFlops) of all HPC Systems available for production since Fermi to Leonardo. With the arrival of Leonardo, the peak performance of our HPC systems increases significantly by reaching a peak performance of 284 PFlops. In the same figure we also present the number of HPC users that shows a trend that grows together with the aggregated performance of our supercomputers.

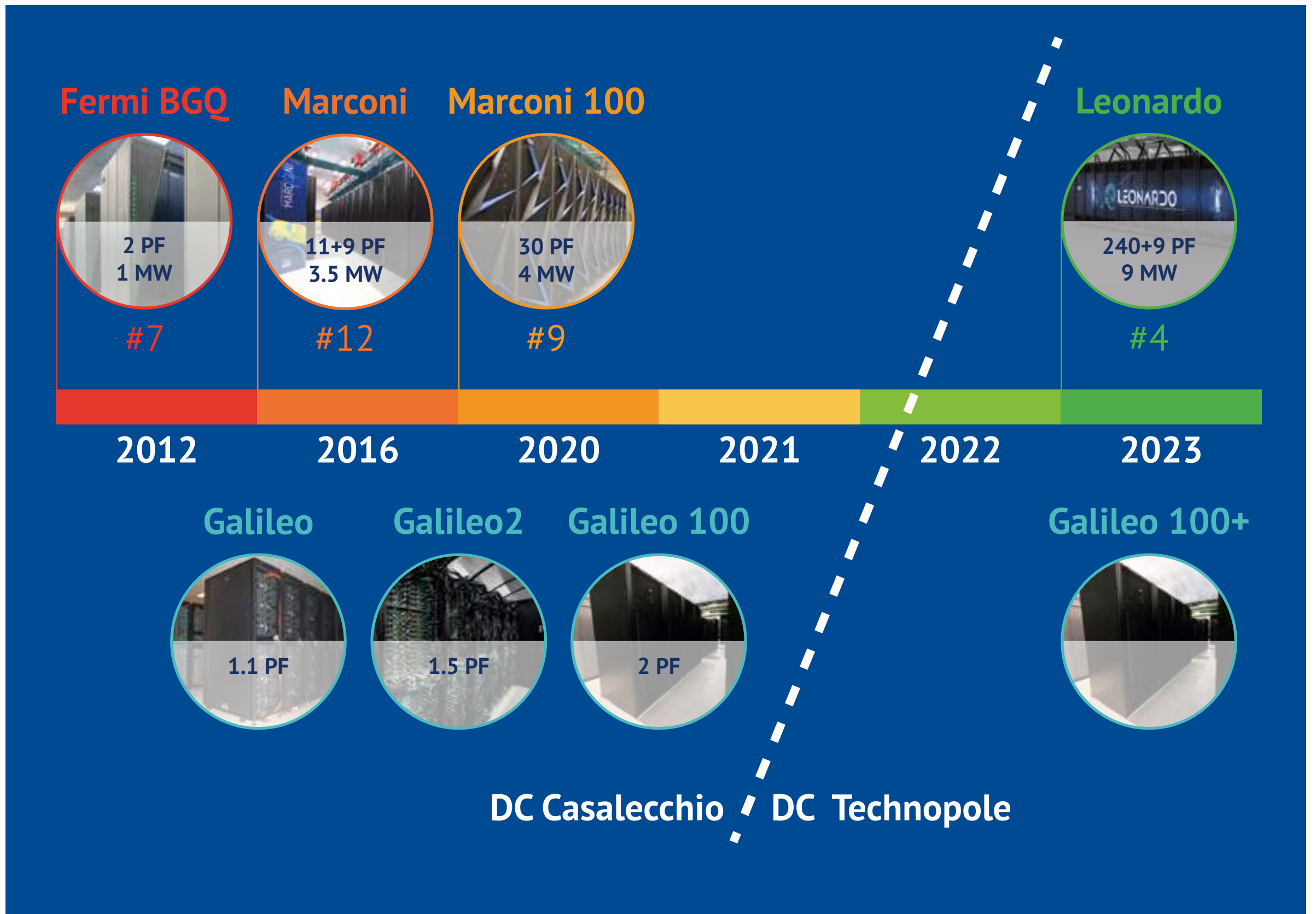


Figure 2: HPC timeline: from 2012 to nowadays the Cineca Supercomputers Tier-0 with the relative position in the TOP500 list. Below the Supercomputers Tier-1, the different configurations of Galileo.

Users Statistics

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

On our HPC infrastructure, the active users at the end of 2022 are 4450, 155 more than the year before. A large part of the users are males (78%), 41% of them are aged between 31 and 45 years old while 36% are less than 30 years old.

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

The more represented foreign countries are Germany with a 5% and United Kingdom, France and Spain with 3% each, mainly working for universities and public or non-profit organizations.

Our computational resources are used by scientists within all scientific disciplines, the most represented three are Computational Chemistry, Condensed Matter Physics and Computational Fluid Dynamics with 11% each one, followed by Nuclear Fusion (10%), Computational Engineering and Astrophysics and Plasma Physics with 9% each one.

Figure 3: new HPC users per year.

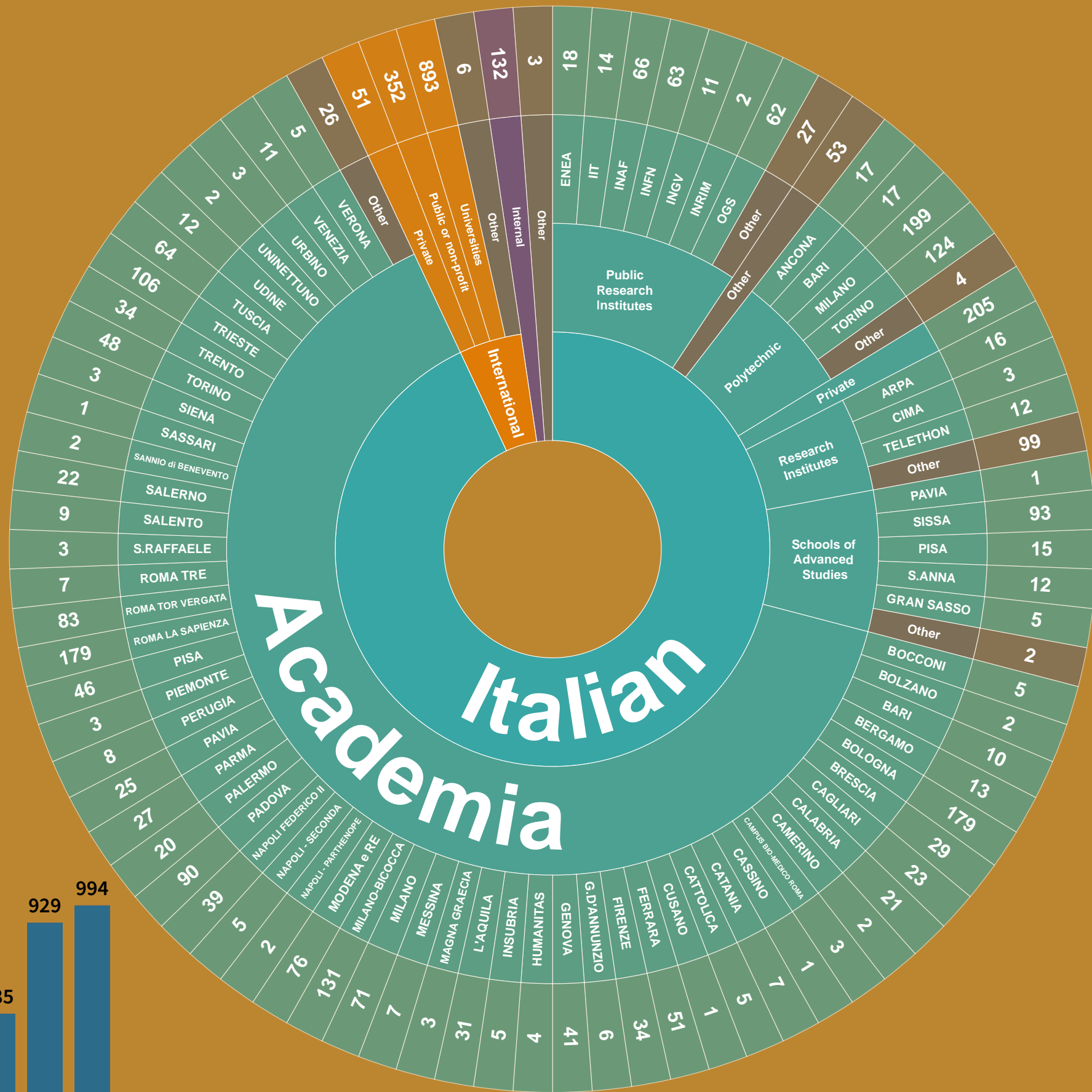
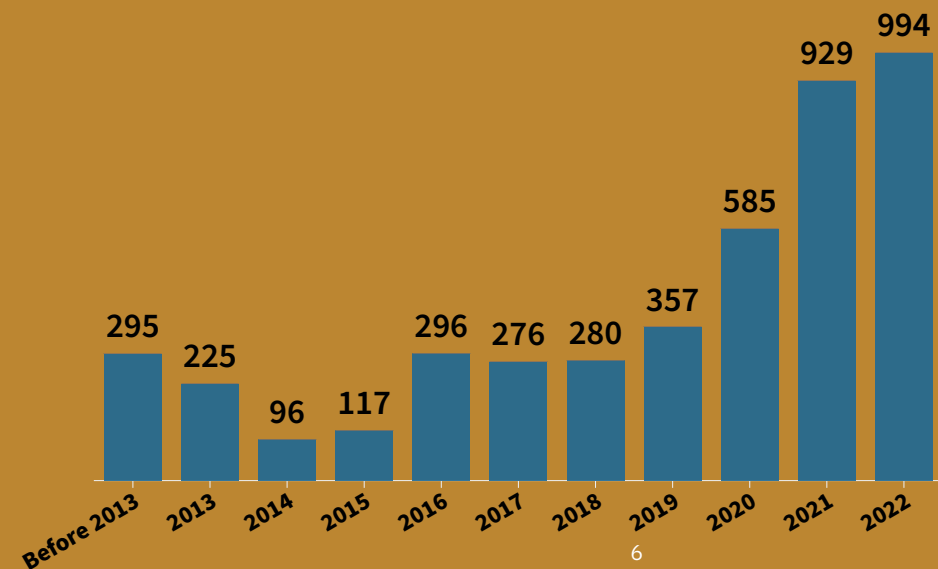


Figure 4: classification of users' institutions.

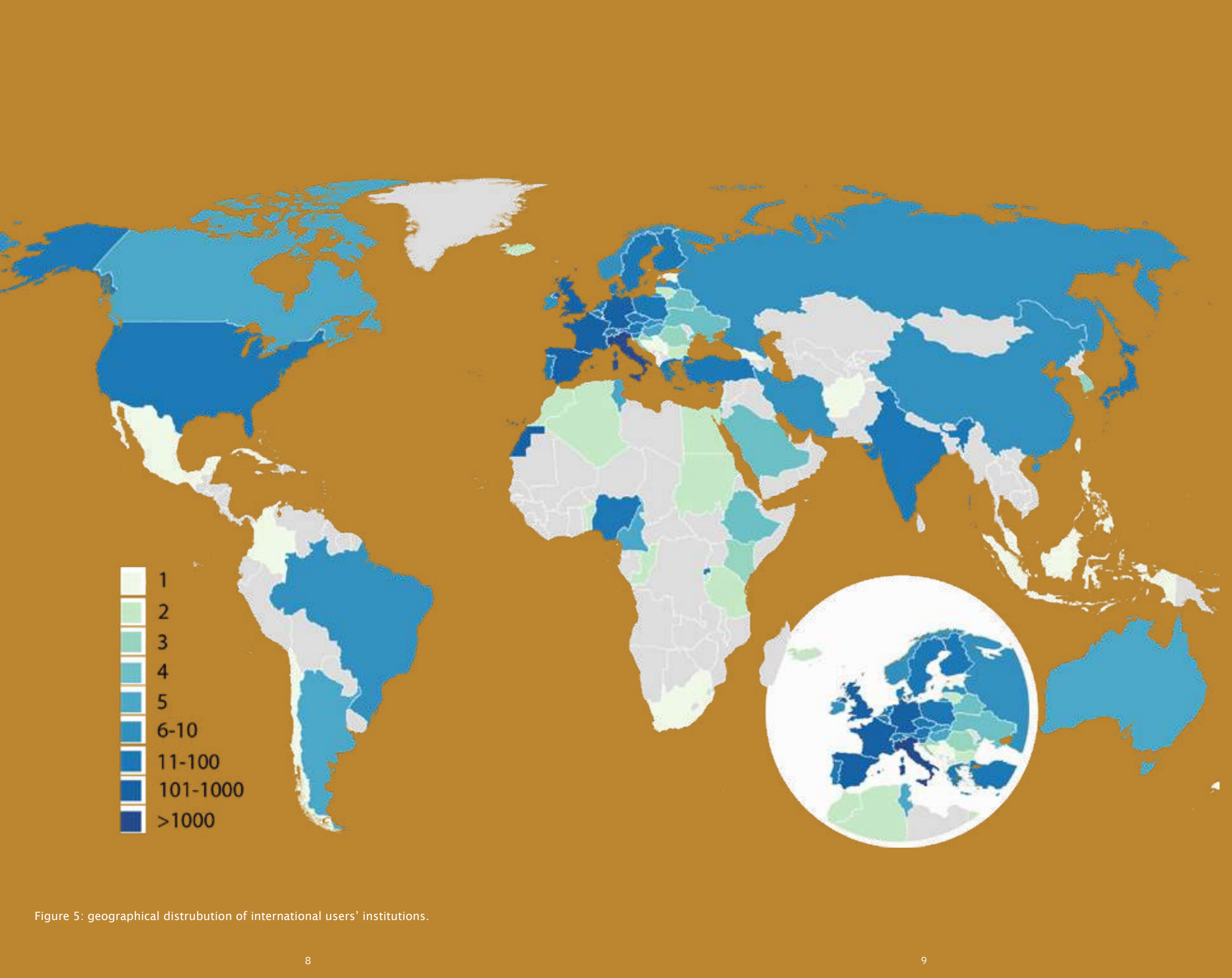


Figure 5: geographical distribution of international users' institutions.

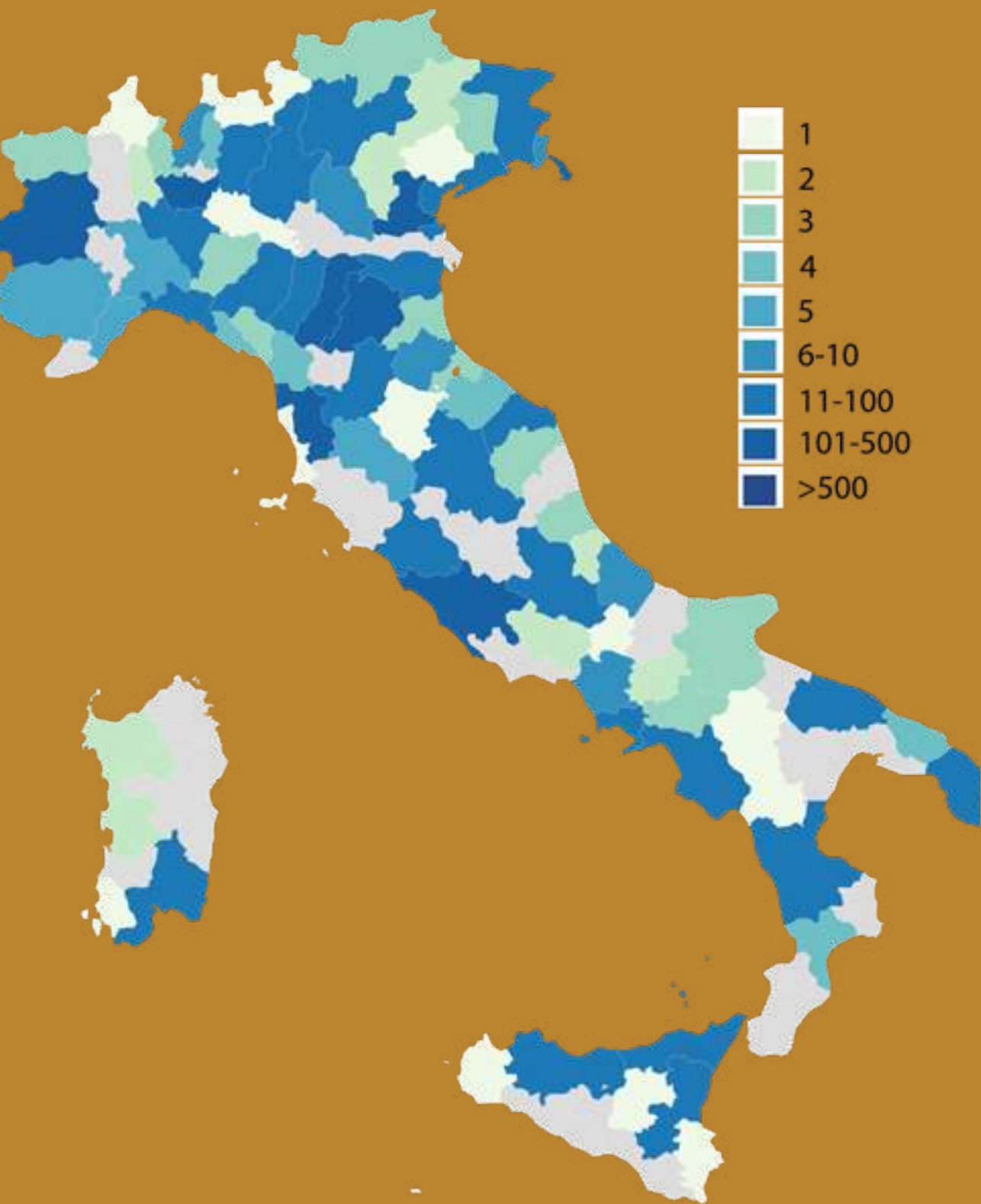


Figure 6: geographical distribution of Italian users' institutions.

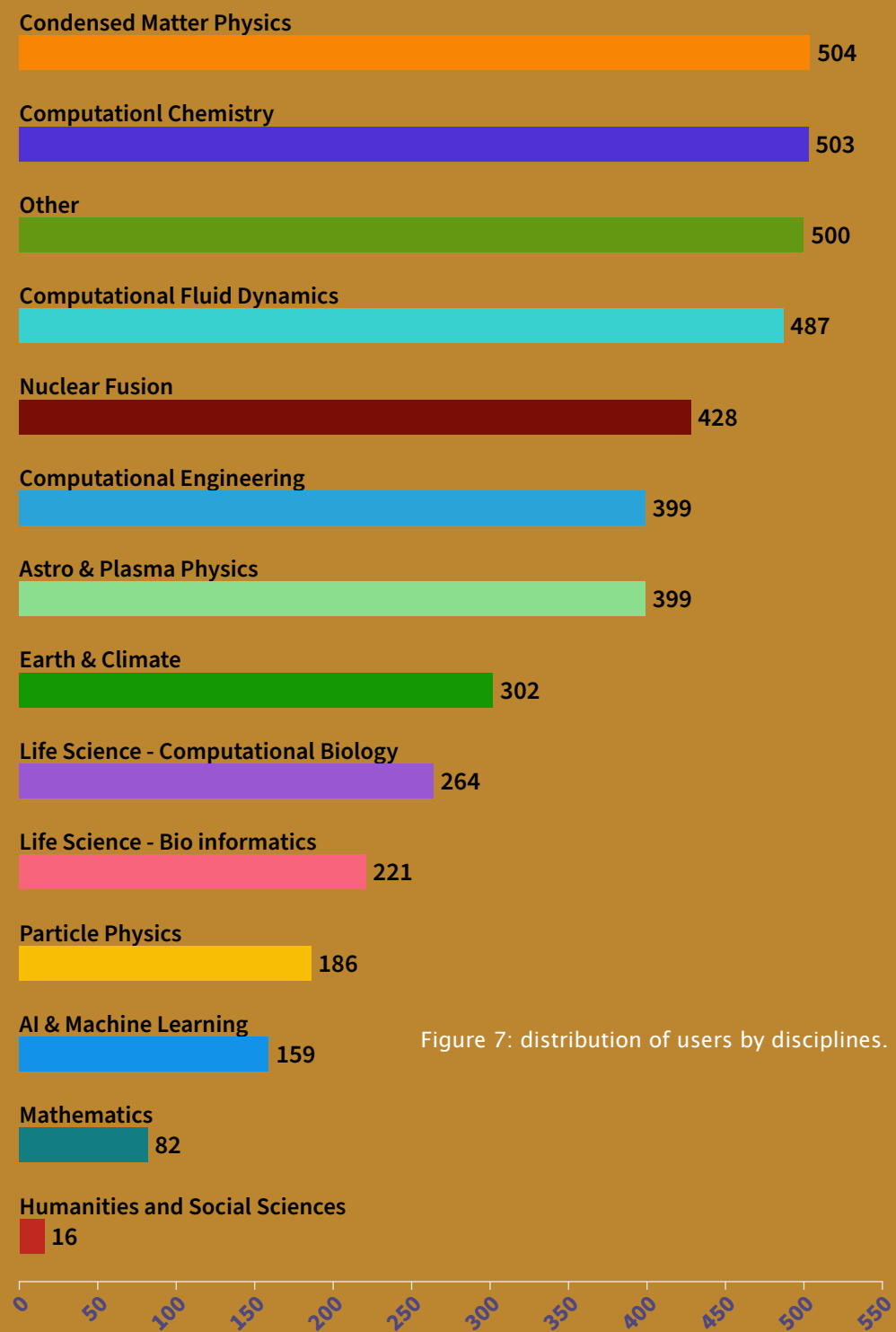


Figure 7: distribution of users by disciplines.

Resource allocation: projects, agreements, collaborations

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

A significant part of the allocations on Cineca's system are based on a peer-review process to ensure the highest scientific value of the selected projects.

In fact, Italian and European researchers take advantage of the HPC clusters in Cineca to perform their computational research by applying for "computational projects" which allow of access to the HPC facilities.

PRACE and ISCRA programs allocate most of computational resources on Marconi100 and Galileo100 and they are both based on the peer-review mechanism at the European and national level, respectively.

On the other hand, Institutions which want to use HPC infrastructures can sign an "Agreements" with Cineca as well as the "Industrial" users. 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) with several collaborators. Along with these wide classes of projects, there are dedicated partitions on Marconi A3 and Galileo100 for weather and biogeochemical forecast activities and industrial projects while some special partitions of Marconi100 and Marconi A3 are dedicated to the EUROfusion consortium. For these partitions, the allocation is made autonomously by the dedicated committee.

During 2022 were available on overall Cineca HPC infrastructure a total of 532 million core-h whilst the total allocation was of 627 million core-h, meaning that the systems were overallocated at 117,8% due to the huge request for computing resources by the scientific community.

The most significant part of the resources in terms of core hours has been dedicated to peer reviewed projects. These resources were distributed as in the following: 35% ISCRA, 20% to PRACE (with a lower number of allocated resources with respect to previous years being 2022 the last year for the project), 47% agreements and less than 2% to industrial projects.

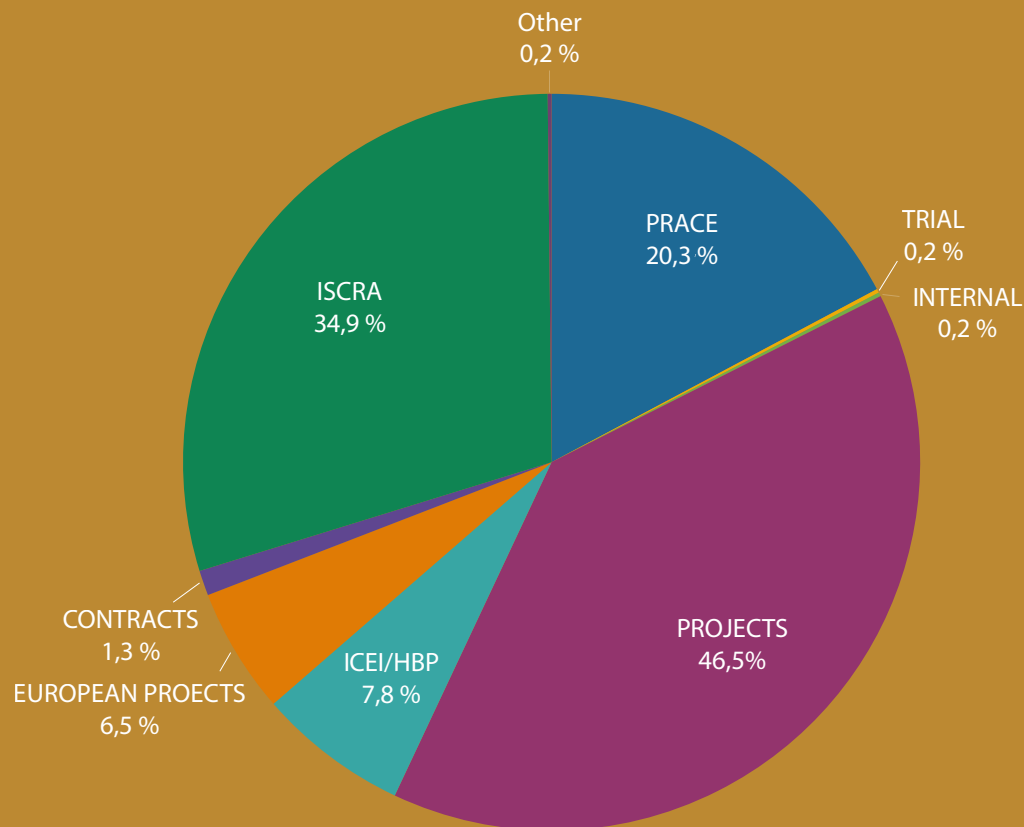


Figure 8: HPC allocation (Core hours) by project type.

PRACE – EuroHPC

Partnership for Advanced Computing in Europe

Andrew Emerson

Cineca

2022 was year of significant change for PRACE since many of its activities, such as the PRACE implementation phases and the Calls for Project Access, came to an end. Some of the duties of PRACE, such as resource allocation, have been handed over to EuroHPC which, thanks to the many years of refinement achieved in PRACE, has been able to inherit efficient tools and processes for the HPC users in Europe. CINECA has been a part of the PRACE story right from the beginning, both as a hosting entity, providing Tier-0 resources for large European projects, and as project partner in the seven PRACE implementation phase projects.

Tier-0 resource allocation on Marconi 100

Being the last year for PRACE Project Access, in 2020 only one call for resources, call 24, was issued for Marconi100. For this regular call a total of 17 projects were allocated on Marconi100, 4 of which with Italian PIs. According to the PRACE-2 agreement, as in the previous calls, 5% of the PRACE dedicated computational resources were awarded to Centre of Excellence (9 projects). The call distributed a total of 7,500,000 hours (corresponding to 60,000,000 Local Core Hours on Marconi100).

The distribution of projects is given in the figure.

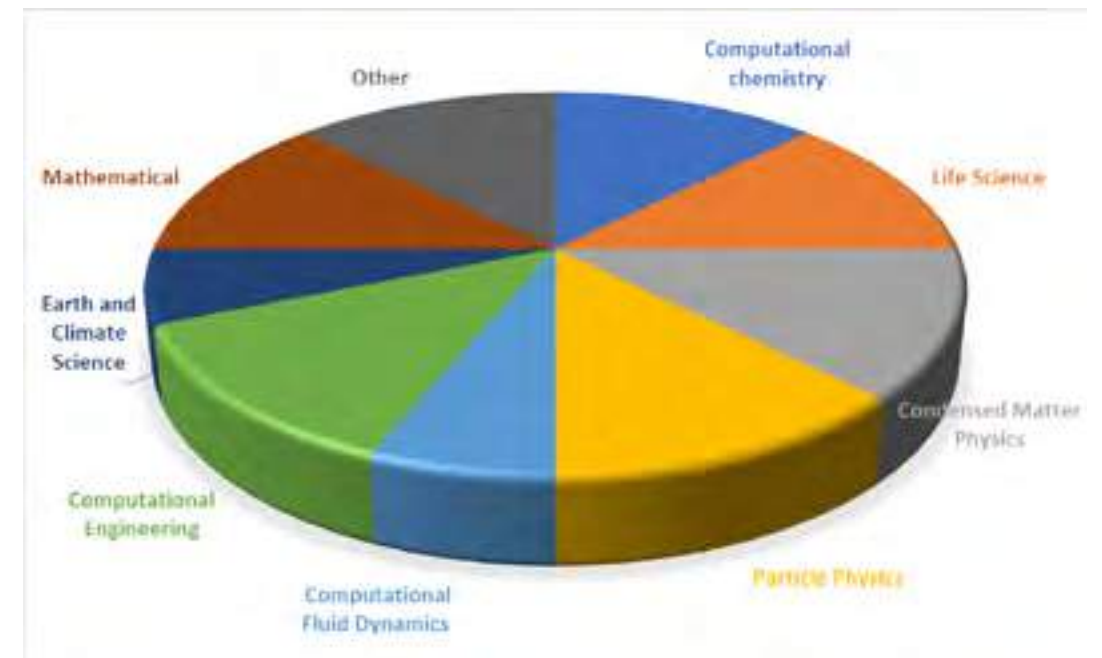


Figure 9: the awarded projects cover the whole range of scientific and engineering disciplines which traditionally use HPC resources.

Tier-1 Resources (DECI)

For small projects, not requiring the extensive resources required in the Tier-0 calls, PRACE provided Tier-1 resources via the DECI (Distributed European Computing Initiative) calls. These worked on a different model to that of the Tier0 system: instead of applying to for a particular supercomputer in Europe, applicants could request a particular computer architecture. If successful, the DECI committee would allocate the project on the most appropriate machine. In 2022, Italian PIs submitted a total of 6 successful projects requesting about 14M hours, which were allocated at 6 HPC sites over Europe. A similar allocation was given to PIs from outside Italy on CINECA's supercomputers.

Training Courses

As a PTC (PRACE training centre) the CINECA HPC department has organised training events and courses, supplementing the usual teaching activities. Before the COVID-19 pandemic, these events were in-person but in 2022 some restrictions were still in place and with the difficulties in travel, most of the activities were moved on-line. Despite the challenges, 2 schools and 4 classes were successfully held as part of PRACE.



Iscra

Italian SuperComputing Resource Allocation
Paola Alberigo
Cineca



ISCRA (Italian SuperComputing Resource Allocation) grants computational resources and technical support to researchers affiliated to Italian institutions, in order to support the Italian research. It provides HPC resources through Class C projects (for code testing, pre-production and benchmarking) and Class B projects (full production). IscraC projects are evaluated and selected on a continuous basis, IscraB projects twice a year.

In 2022 ISCRA CINECA issued the hundredth class C call!

The ISCRA Calls B projects are evaluated on a technical basis and scientific merit. In 2022 the Board of Directors of CINECA nominated the member of the ISCRA Scientific Evaluation Panel, that will remain in office until May 2024. The members are professors at Italian Universities, experts in many different scientific domains:

Marco Aldinucci Università degli Studi di Torino, AI&Machine Learning

Roberto Benzi, Università Roma Tor Vergata, Computational Engineering

Mauro Ferrario, Università di Modena, Condensed Matter Physics

Marco Ferretti, Università di Pavia, AI&Machine Learning

Rossella Ferretti, Università degli dell'Aquila, Earth and Climate Science

Giovanni Peres, Università di Palermo, Astrophysics and Plasma Physics

Silvano Simula, Università degli Studi Roma Tre, Particle Physics

Alfredo Soldati, Università di Udine, Computational Fluid Dynamics

Paolo Ruggerone, Università di Cagliari, Life Science – Computational Biology

Claudio Zannoni, Università di Bologna, Computational Chemistry

In 2022 Cineca provided to the Italian research community about 130 Mcore-h on Galileo100, Marconi100 and DGX systems. Two IscraB calls were launched as reported in the table below (calls 25 and 26).

| | call 25B 14 April, 2022 to 14 April, 2023 | call 26B 20 October, 2022 to 20 April, 2024 |
|--|--|--|
| Accepted | 41 | 49 |
| Rejected | 30 | 32 |
| Total number submitted projects | 152 | |
| Total accepted projects | 90 | |

| | Call 25B | Call 26B | Local Core Hours |
|--------------|--------------------|------------|------------------|
| DGX | 105.000 | 10.000 | 115.000 |
| G100 | 32.275.194 | 33.779.913 | 66.055.107 |
| M100 | 29.403.248 | 32.574.640 | 61.977.888 |
| Total | 128.147.995 | | |

EUROfusion

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

Paola Alberigo, Patrizia Coluccia
Cineca

Cineca strengthens the role of Italy and the Tecnopolo of Bologna as an international hub for high-performance computing, signing an agreement with Eurofusion and ENEA to build a supercomputer dedicated to the international nuclear fusion research community. The system will have an aggregate power of 47 million billion operations per second, and the planned investment is 24M Euro for 5 Years.

Casalecchio di Reno, 10 May 2023 - Cineca, Eurofusion and ENEA have signed an agreement to install a supercomputer dedicated to nuclear fusion in Italy. The new system, which will be installed at the Tecnopolo of Bologna in early 2024, will have an aggregate power of about 47 Petaflops and will be dedicated to support European and international research in numerical simulation of thermonuclear plasma physics and structural analysis of advanced materials for energy applications. Installed alongside the ECMWF center for weather, and the Leonardo supercomputer, the EuroFusion system confirms the effectiveness of the strategies put in place by Cineca to guarantee the scientific community the persistence in Italy of a world-class ecosystem for high-performance computing, in continuity with previous experiences. This agreement, in fact, is the continuation of a collaboration begun in 2016 to make a partition of the Marconi supercomputer available to the scientific community of the nuclear fusion. In 2020,

the system was ranked among the top 10 most powerful supercomputers in the world, paving the way for the implementation of the Leonardo supercomputer project, now ranked fourth in the world.

The agreement for the five-year period 2024 - 2028, with an investment of 63 M euros provides for the installation and operation of a dedicated supercomputer consisting of a conventional and an accelerated partition, equipped with a dedicated fast data memory and capable of performing a total of more than 46 million billion operations per second: 33.7 Petaflops for the system configured with accelerated architecture, 13.6 Pflops for the system configured with conventional architecture. Installation is planned in early 2024, and system startup in production by the first half of 2024. The management of an additional gateway system, and all support activities to enable the fusion science community to make the best use of the systems in their research areas, is also planned as part of the agreement.

ENEA is the National Agency for New Technologies, Energy and Sustainable Economic Development, with 9 research centers throughout Italy and more than 2,600 employees. In its role as coordinator of the national fusion research program and partner of the European agency Fusion for Energy (F4E) and the EUROfusion Consortium, ENEA contributes to major international programs in the field: in addition to ITER (International Thermonuclear Experimental Reactor) - the fusion machine being built in France, in Cadarache - also DEMO, the reactor that will have to feed electricity from nuclear fusion into the grid after 2050, and Broader Approach, the Euro-Japanese agreement aimed at faster execution of the global nuclear fusion research program. www.enea.it

EUROfusion is the European consortium for nuclear fusion research to which 29 research institutions from 27 countries belong. It is led by Ambrogio Fasoli of Italy, head of the Swiss fusion program and professor de l'École Polytechnique Fédérale de Lausanne. Established in 2014, it manages and funds research activities in nuclear fusion on behalf of Euratom. www.euro-fusion.org.



Figure 10: 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

Cineca computing platforms may be accessed also by agreements for collaborations.

During year 2022 about 30 agreements were active with Universities and Research Institutions.

These agreements grant access to computing resources from one to a few years, without the need for researchers of submitting a project proposal, so they are particularly suited for special projects and small experimental activities. The research and academic institutions with active agreements in 2022 are listed below.

In total more than 229 Mcore-h were

allocated to agreement projects. The largest budget corresponds to INFN, at top position for some years. Università degli Studi di Milano Bicocca, SISSA and ICTP follow with roughly half resources each. INAF also holds important agreements.

Apart Milano Bicocca, the other Universities and Polytechnics sum up a total of more than 10 Mcore_h, while the institutions in the field of Life Science research all together sum up more than 4 Mcore-h grant.

The main computing resources used in these agreements are Marconi A3, Marconi100, Galileo100.

| Institution | Award resources (Mcore-h) |
|---|---------------------------|
| INFN (Istituto Nazionale di Fisica Nucleare) | 80,68 |
| Università degli Studi di Milano Bicocca | 42,79 |
| SISSA (Scuola Internazionale Superiore di Studi Avanzati) | 40,72 |
| ICTP (International Centre for Theoretical Physics) | 36,12 |
| INAF (Istituto Nazionale Astrofisica) | 13,36 |
| Università degli Studi di Trieste | 2,92 |
| Università di Bologna | 1,99 |
| AIRC | 1,80 |
| Politecnico di Milano | 1,79 |
| Università degli Studi dell'Aquila | 1,72 |
| Elixir (Distributed Infrastructure for Biological Data) | 1,50 |
| Università degli Studi di Milano | 0,95 |
| Università degli Studi di Genova | 0,60 |
| INMI (Istituto Nazionale Malattie Infettive Spallanzani) | 0,49 |
| Università degli Studi di Brescia | 0,47 |
| Ospedale Pediatrico Bambin Gesù | 0,46 |
| IIT (Istituto Italiano Tecnologia) | 0,38 |
| Telethon | 0,22 |
| Università Luigi Bocconi | 0,14 |
| Scuola Superiore Sant'Anna | 0,12 |
| Politecnico di Torino | 0,07 |

Training HPC

Francesco Falciano
Cineca

Training has always been a hallmark of the support activities carried on by Cineca for the Italian research community and users. Starting from 2012 it was extended to European researchers, Cineca has being recognised as a PRACE Training Center (PTC) in HPC.

In 2022 the emergency for the pandemic COVID-19 was partially ended and with some security precautions, the courses could have been developed in presence. Pandemic forced us to deepened the knowledge necessary in developing on line courses and made the them very popular since that was the only way for students for attending educational events.

During the 2022 we pushed ahead the numbers of students accepted, opening the courses to the largest number of people admitting some “official students” with full access to theory and practicals and all the rest in the waiting list as Auditors allowing them to access to the whole course excluding just from the support during practicals.

Furthermore that way of teaching allowed easily the recording of the sessions. They were published on the courses web site, building a rich library of teaching videos, many of them held in English language, making them available for the people worldwide.

The forcing in this new way of teaching increased our competence in terms of organization and in using learning platforms.

Students expressed in many fashions the fact that we fully reached our goals in converting process and organization.

Many comments in the surveys expressed their likelihood in following the course being at home, not having to plan their travels, with no expenses and with the possibilities to join the class at the very last minute.

For this reasons we discovered that a new way of teaching was born and that the in-presence fashion became a secondary choosen respect the on-line way.

Of course on line events have some disadvantages: the interactions between

students and the support during practicals have limitations, but we tried to improve them suggesting teachers to experiments some open rooms where students could interact each others.

For this reasons in 2022 the main teaching system was online mood leaving in presence only the events where interactions between people was a main goal.

The teaching platform we used was again Microsoft Teams and this was used as exclusive for on line educational events in 2022.

In 2022 we reached these numbers: 14 courses on 10 different topics (one hybrid online and in presence) and 5 schools (one in presence, 4 online). 45 people in the department were involved in teaching, for a total of 76 days of lessons and highly specialised training. In total, more than 1331 Italian and European researchers took advantage of our training program, with high satisfaction levels (8,6/10), as shown by the surveys we regularly administer.

In addition to the Cineca HPC Department classes, our expert were involved in partnership with external societies or universities, giving their support in organization or teaching in events like Masters and Doctorate schools, workshops, tutorials about HPC and Big data topics.

Other important initiatives: Summer of HPC (a PRACE initiative offering young students from all over Europe to work with HPC teams); European Researchers Night (Cineca is the coordinator of Society, a European project organising the Night in Bologna and surroundings in 2018-2019); HPC-Europa3 (we are the coordinator of the project offering transnational access to researchers visiting HPC centers).

The Training activity is regulated by two specific procedures that refers to the ISO standard: ISO 27001:2013 “Information Technology - Security techniques - Information security management system”. These are the “numbers” of our training activities:

TRAINING NUMBERS



Training 2023 PLAN

For 2023 our objective is to get back for some courses to in classroom fashion, especially where interactions between people were the main goals, but the many comments in the surveys expressing likelihood for online event and the possibility to teach to a wider number of students made us to take the decision to keep the online training as our main way of teaching.

Of course we will improve the way of interactions between students and trying to minimize the disadvantages 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.

Figure 11: numbers in training.

STAFF

In the HPC department of Cineca work 123 people, distributed among the three premises: 98 in Bologna, 8 in Milan and 17 in Rome.

Several new colleagues were enrolled in 2022 and 2023: we wish them a very fruitful and interesting job.

Angela Acocella

I obtained my PhD in Chemical Sciences at the University of Bologna. I worked in the field of computational chemistry as post-doc and then as fixed-term researcher at the Department of Chemistry "G. Ciamician". My research activity focused on the investigation of complex processes in molecular and solid systems, at different length- and time-scales, by means of quantum mechanical and classical numerical simulations. I currently work in the HPC Department of Cineca, in the Material Sciences group.

Matteo Angelinelli

I obtained a bachelor's degree in astronomy, a master's degree and a PhD in astrophysics from the University of Bologna. The main topic of my PhD was to improve our knowledge of galaxy clusters, analysing modern numerical simulations and obtaining information about their formation and evolution over cosmic time. I am currently working as an HPC Data Engineer in the Cineca HPC Data Analytics group.

Filippo Barbari

I obtained my bachelor's degree in Computer Science and Engineering at University of Bologna by implementing in CUDA a well-known graph-exploration algorithm. In CINECA I work as an HPC Technology Specialist with a focus on benchmarking and optimizing high-performance applications.

Eleonora Bergamaschi

I received a Bachelor's degree in Mathematics from the University of Bologna and a few months ago a Master's degree in Stochastics and Data Science from the University of Turin. After a post-graduate internship at CINECA during which I started learning about HPC, I currently work in the HPC department in the Data Analytics team on Machine Learning and AI projects.

Francesco Bottau

I got my master's degree in aerospace engineering at the university of Bologna. I worked for 4 years for a race car manufacturer as CFD engineer and aerodynamicist. There I started using and developing CFD tools for industrial fluid-dynamics applications. I am currently an HPC application engineer in the Industrial POC team of Cineca's HPC department, with particular focus on CFD applications.

Gabriella Bettonte

I got my PhD in Mathematical Computer Science at the University of Paris-Saclay, working in the French research center CEA. My thesis investigated the potential of Quantum Computing to analyze Worst-Case Execution-Times in real-time applications. In Cineca I work in the HPC Departement, in the Quantum Computing Lab group.

Caterina Caravita

I got a degree in Astrophysics and Cosmology, and a

PhD in Astrophysics, at the University of Bologna. My research was in the theoretical extragalactic field, focusing on numerical models of stellar dynamics in galaxies. After an internship with Cineca, I work now in the User Support and Production Team.

Mattia Carello

I got my PhD in "Infrastructure, Resource and land use Engineering" at the University of Bologna, providing a camera calibration framework for underwater surveys through Structure from Motion technique. My postdoc research involved the study of vertical transport in the Mediterranean Sea at the ISMAR La Spezia. At Cineca I work in the Data Management group as support to visualize weather forecasting and innovative systems for visualizing data from physical models and not only.

Alessandro Casalino

After obtaining the MSc in Physics at the University of Trento and Imperial College London, I got my Ph.D. in Physics in Trento in Theoretical and Computational Physics. In particular, the thesis was about cosmology and the possible computational methods to constrain new models with experimental data. Then I worked as a Post-doc at the University of Bologna, investigating alternative cosmological theories of gravity using N-body simulations. In Cineca, I am a programmer working on chemistry software and CI/CD implementation.

Antonio Costantini

I got a Ph.D. in Theoretical Physics from University of Salento, with a thesis on theoretical and phenomenological implications of conformal theories. After a year as post-doc in Lecce, I moved to Bologna for a two-year post-doctoral fellowship with INFN. During this period, I worked mostly on phenomenological models for High Energy Physics, testable at the LHC at CERN. These models require huge numerical simulations, runned on clusters, and a deep postprocessing analysis. I also worked in Louvain-la-Neuve as a post-doc, for eight months, together with my supervisor's group and researchers from other Institutions. Before joining the Data Management Group in CINECA I was a teacher in high school, in Bologna, for about one year.

Tommaso Gorni

I graduated in Physics at the University of Modena and Reggio Emilia, followed by a PhD in Condensed Matter Physics at SISSA (Trieste) and 5+ years as a postdoctoral researcher in Paris. The core of my research has been the modelling of the electronic and magnetic properties of materials by means of large-scale numerical simulations. In Cineca I work in the SCAI department providing support to the Material Science community within the High-Performance Computing domain.

Juan Mata Naranjo

I received a bachelor's degree in Physics from the University of Santiago de Compostela. After several years working for different financial institutions all around Europe, I decided to return to the university to complete my master degree in Data Science at the university of La Sapienza, during which I did an internship at CINECA in the HPC Bioinformatics team, where I am currently working in.

Attilio Marcelli

I earned a Bachelor's Degree in Astronomy and a Master's Degree in Astrophysics and Cosmology at University of Bologna with a thesis on galaxy formation

and evolution. I worked as IT Business Analyst for CRIF, a multinational company located in Bologna, while now in Cineca I'm part of the HPC User Support group.

Mattia Mencagli

I am a computational astrophysicist. I acquired HPC competencies during both my master's in astrophysics and my PhD in SISSA. I mastered those competencies during the master in HPC in the second year of my PhD. My scientific work was to develop a GPU N-body code that simulates merging black holes, sources of gravitational waves, within star clusters. Now, I am working in the HPC specialist group in CINECA and I work with various European researchers to accelerate their codes using GPUs.

Laura Morselli

I got my PhD in Astrophysics and Cosmology from the Technical University of Munich in Germany, working on the evolution of the morphology of galaxies across cosmic time. I came back to Italy as a Post-Doc researcher at the University of Padova and then worked one year for Associazione Big Data in Bologna, focused on industrial projects involving Artificial Intelligence and HPC; these topics remain my main focus in CINECA.

Anna Nikishova

I received my PhD degree in Computational Science from the University of Amsterdam in the Netherlands. My research has been devoted to Uncertainty Quantification of Multiscale Models. Then, in my postdoctoral studies I deepened my experience in Data Science and Artificial Intelligence. In CINECA I am a member of the HPC department, where I deal with data analysis and machine learning.

Marco Puccini

I got my 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. I grant a Scholarship at CNR INM (Institute of Marine Engineering) working on data analysis from experimental fluidodynamics. I taught at High School Math and Physics meanwhile collaborating at Molecular Medicine Department of La Sapienza with data analysis for novel PET scintillator crystals testing. I grant a Research Fellowship in the HPC laboratory at ENEA working on Big Data managment and DevOps in a wide range of projects. Now I'm an HPC Data Engineer at the HPC Data Management group at CINECA.

Alessandro Romeo

I got a bachelor's degree in Physics in Messina, then I moved in Padova to get a master's degree in Astronomy and finally I obtained my PhD in Astronomy at the Università degli studi di Bologna. I worked mainly in performing cosmological simulations for 21 cm radiation weak gravitational lensing, implementing a pipeline which had the aim of producing dark matter images from a statistical reconstruction of the underlying lens mass taken at several redshifts at the epoch of reionization. Then I have been teaching Optics, Physics, Statistics and Aberrometry at Istituto B. Zaccagnini in Bologna and Milano for six years, and I currently work as a HPC specialist in Cineca within the Space Center of Excellence, in order to profile astrophysical and physical simulations involved in the project and optimize their GPU porting.

Francesco Talpo

I got a Bachelor's and Master's Degree in Chemistry from the University of Ferrara. For my master's thesis

work I used computational methods based on Density Functional Theory to study the thermodynamics of deactivation and recovery of Ni based catalysts for catalytic methane decomposition. At CINECA I'm part of the HPC User Support group.

Orlenys Troconis

I got a PhD in nuclear astrophysics at Universidade Federal Fluminense in Brazil and my bachelor's degree in theoretical physics at Universidad Central de Venezuela. In my research activity I investigated the magnetic field effects in the structure of neutron stars from the point of view of general relativity. During my stay at the International Center of Theoretical Physics (ICTP-Trieste, Italy) I attended the courses of the master in high performance computing. In CINECA I am working in the HPC User Support group to assist users to efficiently run their codes in our HPC system.

Matteo Turisini

I studied Physics at La Sapienza University of Rome and earned my PhD at the University of Ferrara with a thesis on custom electronics for single photon detection. I have been working for more than a decade in R&D projects for particle physics experiments and industrial applications in the USA (JLAB) and Europe (CERN). Recently I expanded my field of study to high performance computing and machine learning. I really love music, nature and build big things.

Tommaso Zanelli

I obtained my master's degree in aeronautical engineering from Politecnico di Milano, with a specialization in aerodynamics and a thesis on Large Eddy Simulation turbulence modelling. I subsequently joined Nextflow Software, a startup based in Nantes, France, now part of Siemens DISW, where I worked on the development of numerical methods for simulating multiphase flows. After three years I moved back to Italy and joined CINECA's HPC POC solutions team, where I work on porting fluid mechanics simulation software to Exascale platforms.

Flavia Zanon

I am part of the project management office of the High Performance Computing Department of CINECA. I am a research manager with a long experience in supporting complex organisations in the financial and legal management of collaborative projects. My fields of expertise includes project management, open science, research integrity, intellectual property and technology transfer. I have a background as public policy analyst and I hold a PhD in international relations.



Events

The European Researchers' Night of Society project

Simona Caraceni
Cineca

The European Researchers' Night (ERN) in the Emilia Romagna Region was organized once again by the SOCIETY consortium in Bologna, Cesena, Faenza, Ferrara, Forlì, Predappio, Ravenna, Rimini (Figure 1). The SOCIETY consortium is formed by researchers from CNR (Consiglio Nazionale delle Ricerche – Area Territoriale di Ricerca di Bologna), the University of Bologna, CINECA (Consorzio Interuniversitario – Dipartimento di Supercalcolo), 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), together with two Ltd companies ComunicaMente and Naxta. The SOCIETY consortium has been organizing the ERN in the Region since 2015 in Bologna and all the nearby towns that host campuses of the University of Bologna. Since 2021, a collaboration with the University of Ferrara and with the local branches of INFN and INGV granted the involvement of the town of Ferrara. Considering all venues, a total of 186 activities were proposed to the public by more than 500 researchers. The effort was rewarded by a great turnout, with more than 6000 people attending, despite the bad weather. Table 1 shows the estimated number of people attending the ERN at each venue.

The theme proposed in 2022 was riPENSACi, which means literally think again (or think twice), an invitation to review our own knowledge and understanding of the world, changing perspective and opening our minds to alternative interpretations of reality. The ERN can be a fruitful occasion to revisit our place in our world and in our society and to assess the impact of our actions on economy, society and environment.

In order to think again, the research activities presented during the Night were grouped around three different perspectives: explore the present, interpret the past, and imagine the future. The 2022 edition of the ERN featured two significant innovations. The first is a substantial change of the main venue. After

many successful editions held downtown, at the heart of the academic district, the ERN could take advantage of a large covered plaza that once housed one of the city's general markets. Recently renewed and named after the popular singer Lucio Dalla, the plaza was inaugurated at the beginning of the summer, and is now destined for entertainment and cultural events. Located near the new train station and the new town hall, during the summer of 2022 Piazza Lucio Dalla was equipped with reception and refreshment facilities. It is not far from the city centre, yet it belongs to a suburb and gathers a different kind of public.

The exploitation of such a vast space required some planning, to grant an enjoyable access to all the activities proposed. To ensure the best result, we engaged Accademia delle Belle Arti, the tertiary higher education institution for the study of Fine Arts in Bologna. The Academy's students acquired all the necessary information about the event and proposed 6 possible layouts, among which the SOCIETY consortium could choose the most appropriate. We granted a stand to the Academy students, who could present to the public all the proposed projects and the ideas behind them.



Figures 12, 13: European Researchers' Night in Bologna.



Events

The European Researchers' Night of Net project

Neva Besker
Cineca

Cineca with all the major local (Lazio) Research Institutes and Universities, such as CNR, Ispira, INAF, INFN, INGV, ENEA, University of Rome Sapienza, University Tor Vergata, University of Tuscia and Telematic University UniNettuno forms a partnership NET- sci-eNCE 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 2022, the third edition of the European Researchers' Night in Viterbo and Rome at Città dell'Altra Economia (CAE). The European 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 2022 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 super-

computer", "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, 16: European Researchers' Night in Rome.



Events

Supercomputing 2022, a sparkling return

Daniela Galetti
Cineca

Going through all the phases and rituals that lead to the departure for the USA and which we had neglected for two years to follow the event in virtual mode, was a great feeling.

Starting from the choice of our most relevant results to share with other HPC experts on the booth walls, moving to the comparison between different potential gadgets, to the design of the exhibition coordinating with INFN team, accompanied us with enthusiasm to the day of departure.

Even with all the necessary precautions, being able to attend again in person at the most important event in the HPC world would have been a great thing in itself, but this year we also brought with us a great hope: Leonardo, the cluster of pre-exascale class installed by Atos and Cineca in record time at the Bologna Tecnopolo as part of the EuroHPC project could have been ranked among the top 5 most powerful supercomputers in the world...

And all our efforts to have Leonardo ready for SC22 was awarded by the best possible rank. Leonardo came up fourth in the top500 of november 2022!

Figure 17: Leonardo delegation at Supercomputing.



HPC H-index

Web of Science <https://www.webofscience.com/>
 Database: Web of Science Core Collection
 Search: (FT=Cineca or FO=Cineca) and (PY=2022-2023)
 Date: 10.05.2023



Results found:
767



Number of Citations:
1664



Average Citations per Item:
2.17



H-index:
15



Figure 18: number of publications and citations from year 2008 to 2022.

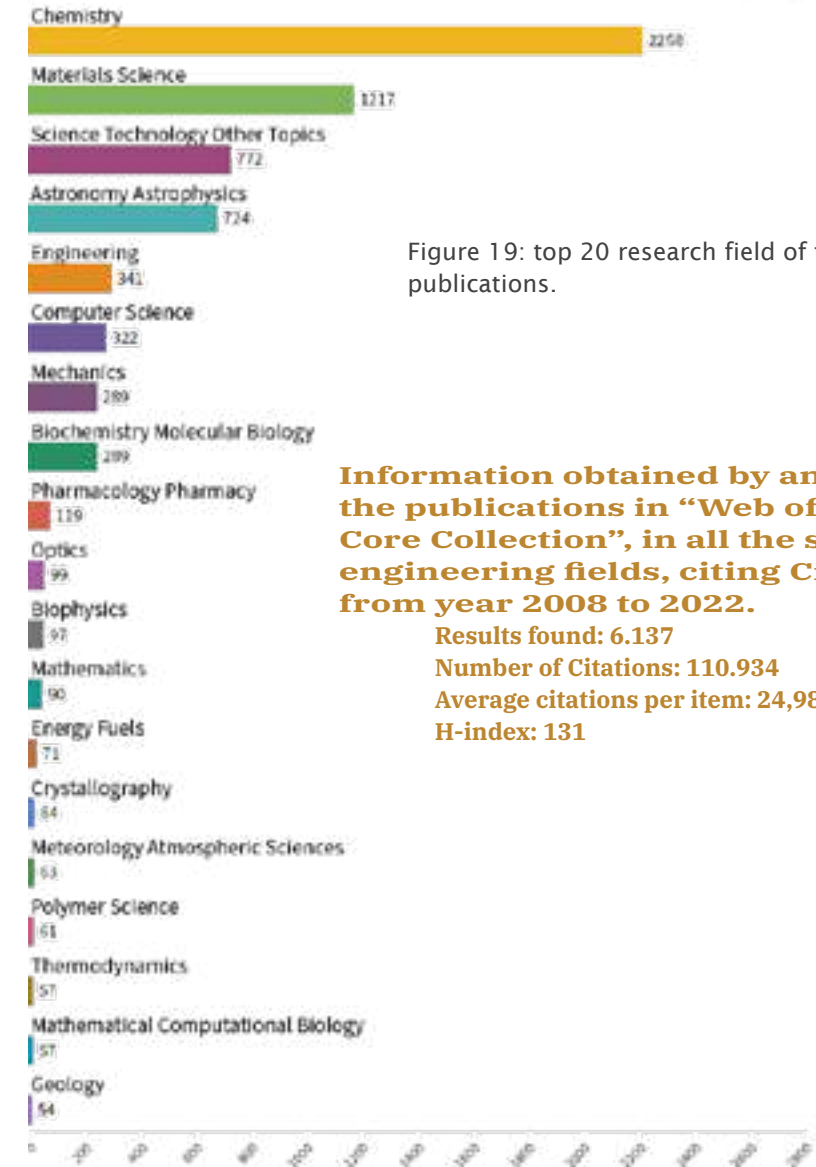
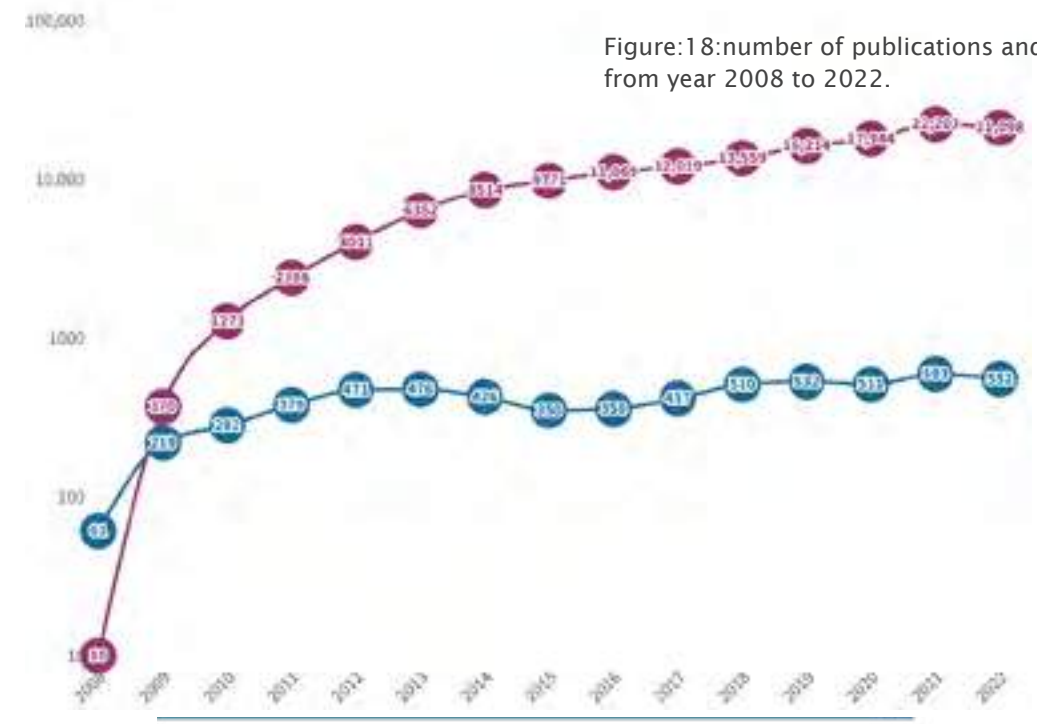


Figure 19: top 20 research field of the publications.

Information obtained by analyzing the publications in “Web of Science Core Collection”, in all the science and engineering fields, citing Cineca, from year 2008 to 2022.

Results found: 6.137
 Number of Citations: 110.934
 Average citations per item: 24,98
 H-index: 131

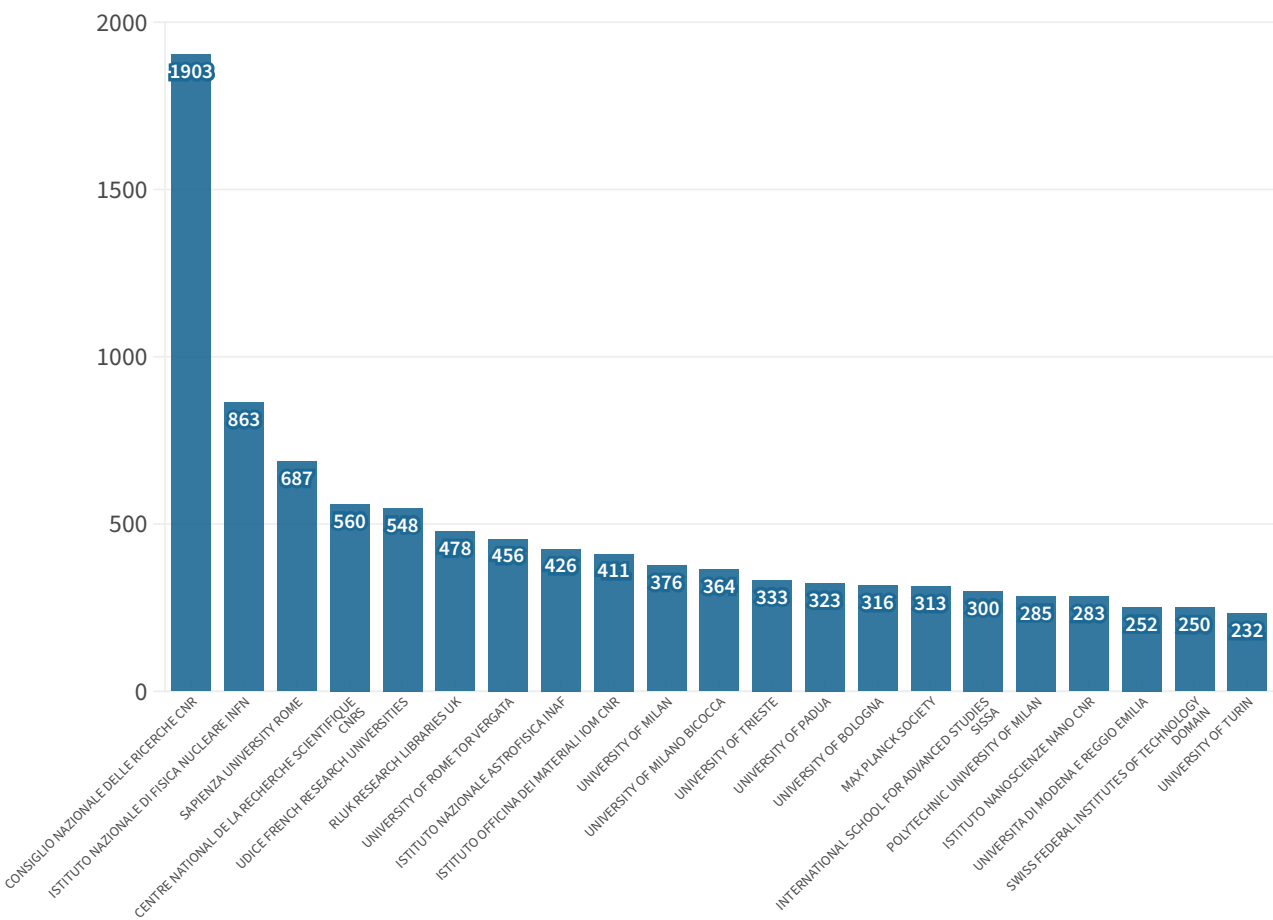


Figure 20: main affiliations of the authors.

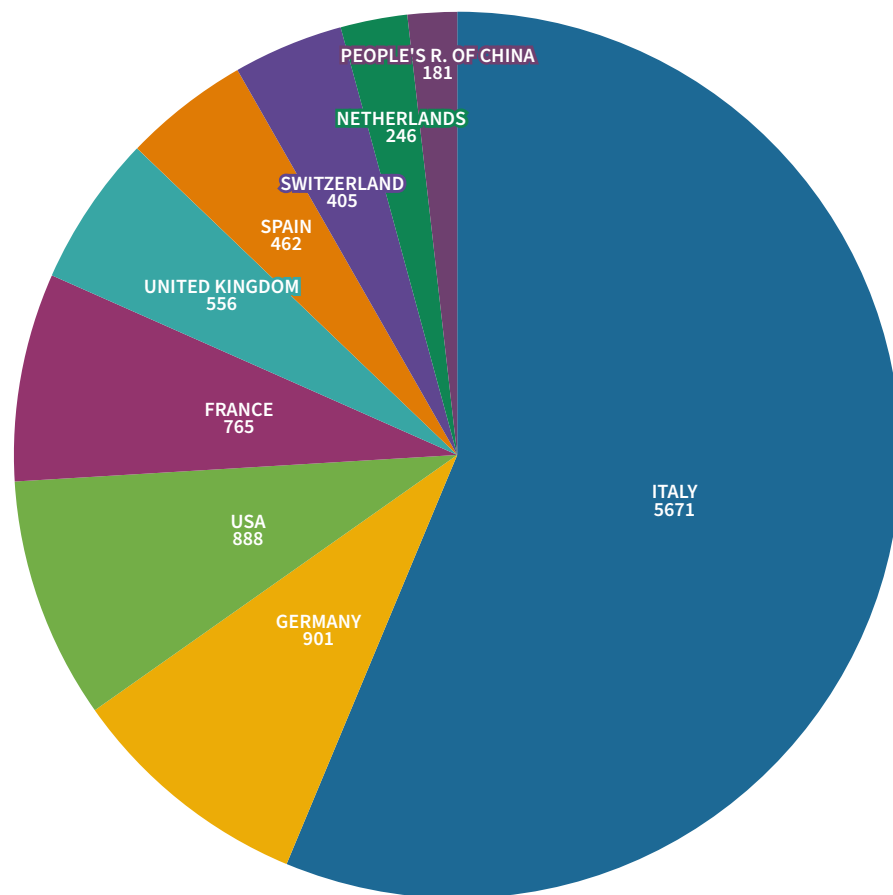


Figure 21: top 10 nationalities authors' organization.

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

Projects in “-omics” science field

Elisa Rossi, Silvia Gioiosa, Beatrice Chiavarini, Giuseppa Muscianisi
Cineca

Computational research requires increasing computing power for the analysis of ever-larger amounts of data. The advancement of “-omics” technologies, in particular, has led to exponential increases in the amount of sensitive data produced that require secure storage and processing storage. These data hold the potential to lead to unimaginable breakthroughs in the field of personalized and precision medicine, but at the same time, very high standards of security are needed to best ensure the privacy of the individuals with whom the data are associated.

To meet this need, Cineca has adopted cloud computing solutions for the management of sensitive and personal data in the specific field of life sciences, extending its ISO 9001

(Quality Management Systems) and ISO/IEC 27001 (Information Security Management Systems) certifications to this field as well, starting from November 2022. The implementation and the great work aimed at obtaining this certification has enabled Cineca to create a working environment that complies with the standards required for information security management and, at the same time, to offer HPC services in the cloud to partners who decide to entrust Cineca with the management of their data in full compliance with the GDPR rules on personal data protection.

Some projects involving the management and analysis of omics data by CINECA are shown below.



Figure 22: <https://www.raiscuola.rai.it/laboratori/ildna>

NIG (Network for Italian Genomes)

The NIG (Network for Italian Genome) project aims to create a centralized repository for human genomic variant data produced by massively parallel sequencing.

CINECA has implemented and made available only to users of NIG centers and in possession of an appropriate digital certificate for the client workstation, an online platform that allows the upload of input files in fastq.gz format and their associated metadata. To ensure the security and privacy of the sensitive -omics data of the subjects whose exomes have been sequenced, the files uploaded are stored on Virtual Machines with encrypted volumes accessible only by internal personnel through 2FA and access log tracking. CINECA has also been appointed as data processor by specific agreements signed with each single center.

Once the upload is finished, there is a Workflow Management System (WMS) based on Snakemake and orchestrated by Celery that dispatches the data in parallel jobs, allocating for each step of the analysis pipeline the necessary resources in terms of vCPUs and RAM.

The massive and continued processing of data produced by massively parallel sequencing with High Throughput Sequencing mode will allow the creation of the first centralized repository for human genomic variant data referring to the Italian population.



GENERA

GENERA Project, **Genoma mEdiciNa pERsonalizzatA** promoted by the Network for Italian Genomes (NIG), in response to the calls of Piano Operativo Salute - Trajectory 3 of the Ministry of Health, envisages the creation of a national network dedicated to personalized medicine, aimed at standardizing processes for the collection of clinical and genomic data and their storage, in a National and European context.

The platform is aimed at generating knowledge for health planning and scientific research and, prospectively, studying the genetic predisposition of clinical conditions with high morbidity and morbidity impact for public health systems such as complex multifactorial diseases (oncological, neurological and cardiological, etc.). GENERA aims to define a point of aggregation and alignment of genomic and informatics analyses in order to lead health care toward precision medicine. Integration of clinical data and other “-omics” technologies will help in data interpretation and limit false predictions. To this end, GENERA will implement new approaches, for the analysis of large masses of data, based on Machine Learning and Artificial Intelligence.

GenoMed4ALL

GenoMed4All - Genomics and Personalized Medicine for all through Artificial Intelligence in Haematological Diseases is an initiative funded by the European Commission under the Horizon 2020 Research & Innovation program.

GenoMed4All develops a platform for high-throughput, confidential and secure federated learning-based management and sharing of genomic, clinical and other “-omics” health data. The goal is to use artificial intelligence (AI) model resources for research in the field of personalized medicine for rare hematologic diseases, combining already established clinicopathologic parameters with advanced genomic profiling, and thus create synergies for diagnosis, prognosis and innovative therapies.

For the EU Genomed4All project, CINECA provides its computational resources and expertise for a number of use cases, ranging from MRI image processing through distributed machine learning to the execution of complex analysis pipelines on genomic sequences.



GENOMED4ALL has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017549.



ORCHESTRA

ORCHESTRA (CONNECTING EUROPEAN COHORTS TO INCREASE COMMON AND EFFECTIVE RESPONSE TO SARS-COV-2 PANDEMIC) is a three-year international research project aimed at tackling the coronavirus pandemic, led by the University of Verona and involving 26 partners (extending to a wider network of 37 partners) from 15 countries: Argentina, Belgium, Brazil, Congo, France, Gabon, Germany, India, Italy, Luxemburg, Netherlands, Romania, Slovakia, Spain, Venezuela. The project is funded by **the European Union's Horizon 2020 research and innovation programme**

The project uses innovative techniques in microbiology, genomics, statistics (artificial intelligence and mathematical models) and platforms for data collection among European countries.

Cineca is the coordinator of WP7 "Data Management." Together with the other 2 European HPC centers (Cines for France and HLRS for Germany), it is responsible for the design and implementation of the platform that collects data from the various cohorts involved, taking into consideration data size, security, privacy.

To this end, CINECA has implemented a centralized platform to collect sensitive personal data in the Italian National Hub (NH) in compliance with GDPR and Italian regulations. Clinical data and outcomes of serological and microbiological analyses are collected using a dedicated instance of REDCap provided in the infrastructure itself. Data collection is performed using pseudonymized codes.

The hosted data come from the centers that have appointed CINECA as the data controller under individual agreements.

In addition to the necessary security measures, a Vulnerability Assessment and Penetration Testing (VAPT) is performed periodically to ensure that the required level of security is met.



ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167.



Virtual Machines in autoprovisioning

This solution is designed for experienced users, who already have on board figures capable of taking charge of managing a virtual machine both in terms of installation and software maintenance. The user logs into their dashboard on ADAcloud and sets up all the stack they need, while also taking care of creating the users enabled for access on the machine itself. Access to VMs is allowed only to the administrator user of his VM, who then remains the only one able to process and analyze his own -omics data.

Genomic Computing and CINECA for genomic research and precision medicine

Marco Masseroli
Politecnico di Milano

Big biomolecular data from DNA sequencing are increasingly readily available and promise to guide precision medicine in the fight against diseases of genetic origin, including cancer. Thanks to Next-Generation Sequencing (NGS) technologies, it is now possible to generate rapidly and at low cost large quantities of multiple types of biomolecular data that describe not only the sequence of an individual's entire genome and its variations (mutations) with respect to a reference genome sequence, but also the activity (expression) under different conditions of the genes present in the genome, as well as the modifications of the proteins that interact with the DNA by regulating gene expression in different ways.

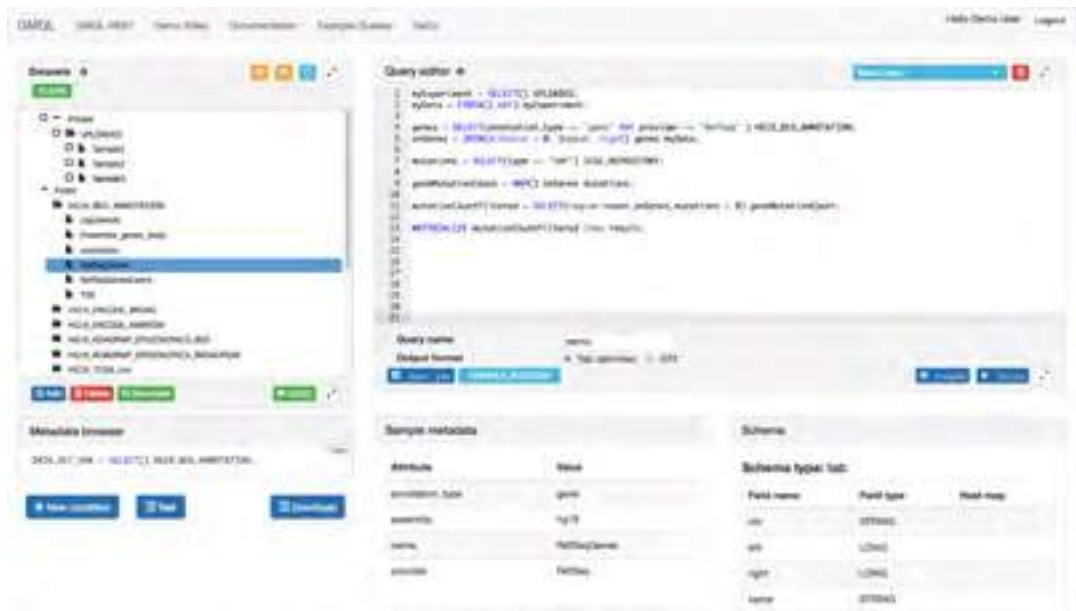
The availability of large public collections of such big genomic data produced by international consortia has stimulated the development of computational technologies that have made it possible to take advantage of the large number of such data to extract statistically significant information, fundamental both for the understanding of the biomolecular phenomena underlying the development of genetic pathologies, and for the identification of possible specific therapies for each type of patient.

The group of Politecnico di Milano led by Prof. Stefano Ceri, thanks to the "Data-driven Genomic Computing (GeCo)" project (<http://www.bioinformatics.deib.polimi.it/geco/>) funded by the European Research Council (ERC), has developed both an abstract model of genomic data (the Genomic Data Model - GDM), which guarantees interoperability between the various formats usable for such data, and a specific query language for such data (the GenoMetric Query Language - GMQL); the latter one, by taking advantage of the GDM data model, allows for easy and integrated queries of the various types of data of genomic regions and their clinical-biological metadata. Within the GeCo project, the same group has also implemented an innovative open-source system (<http://www.bioinformatics.deib.polimi.it/geco/?try>)

based on GDM and GMQL, which allows querying and processing multimodal big data of genomic regions and their metadata to extract information that helps giving them "biological meaning"; this allows better understanding, for example, of the origin and development of serious genetic pathologies, and their dependence on environmental factors. This result is at the research forefront in the sector worldwide, as the countless scientific publications produced by the GeCo project demonstrate (<http://www.bioinformatics.deib.polimi.it/geco/?publications>).

Thanks to the significant collaboration between Politecnico di Milano and the Italian Interuniversity Consortium CINECA, the developed GMQL system is available to the scientific community on the CINECA Cloud resources. The system includes a scalable Genomic Data Management System (GDMS), for genomic region-based query of numerous genomic data sets and their metadata; this GDMS is based on Apache Spark technologies (<https://spark.apache.org/>), an engine for large-scale data analysis, and Apache Hadoop Distributed File System (<https://hadoop.apache.org/>), for distributed and scalable cloud computing of large datasets. The system also includes a large integrated collection of various types of public genomic data downloaded from the main large international databases in the field, including the Encyclopedia of DNA Elements (ENCODE - <https://www.encodeproject.org/>), The Cancer Genome Atlas (TCGA) at the Genomic Data Commons (GDC) Data Portal (<https://portal.gdc.cancer.gov/>), and the Roadmap Epigenomics Project (REP - <http://www.roadmapepigenomics.org/>).

The GMQL system is easily accessible through both an easy-to-use web user interface and a set of REST web services that allow programmatic access. There are also available an R/Bioconductor package (RGMQL - <https://bioconductor.org/packages/release/bioc/html/RGMQL.html>) and a Python library



(pyGMQL - <https://pypi.org/project/gmql/>) for the programmatic use of the GMQL system integrated with the numerous bioinformatics and data science processing and analysis libraries available for R and Python languages.

The availability of the GMQL system on the CINECA high-performance Cloud lets the efficient execution of complex processing of integrated genomic big data, allowing performing numerous scientific researches of high relevance. Among these some in the onco-genomics field, related to the identification of candidate tumor biomarker genes for a more effective diagnosis, and to the characterization of molecular subtypes of breast cancer for a more precise prognosis and the identification of patients who more likely can respond to drug treatment. All results of these researches have been published in high-impact international scientific journals, highlighting the relevance of the collaboration between Politecnico di Milano and CINECA, which made it possible to make the GMQL system an important and unique successful service for genomic research and personalized precision medicine.

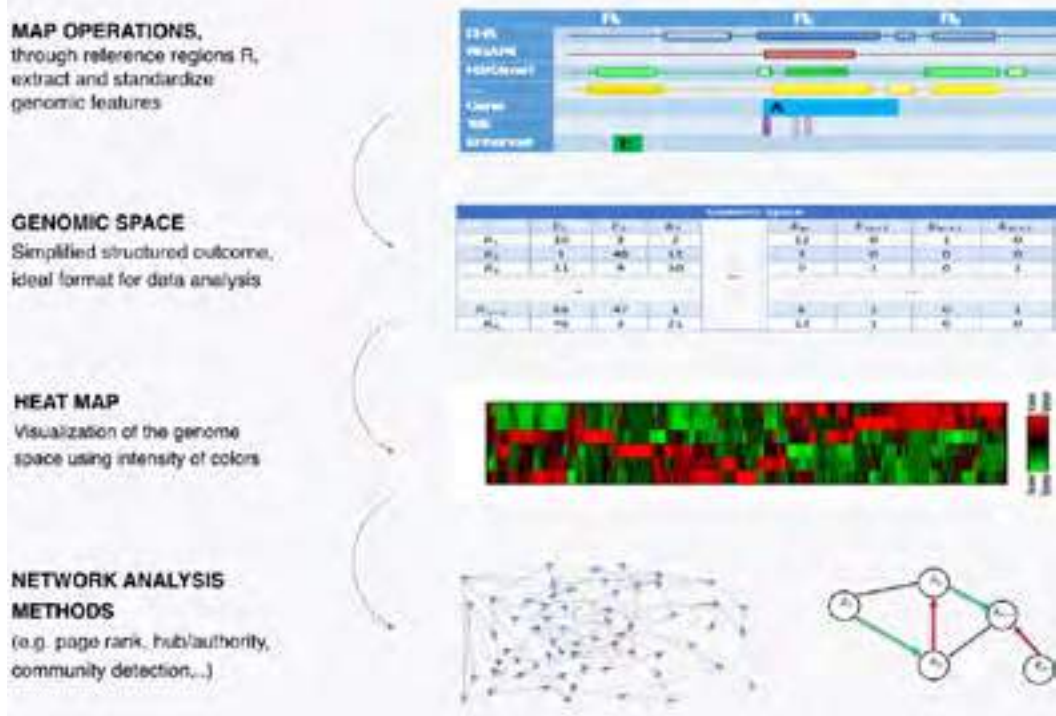


Figure 23: GMQL web interface.

Figure 24: : GMQL MAP operation applied on heterogeneous genomic region data allows extracting and standardizing the related information, structuring it in the matrix form of “genome space”; this is ideal for subsequent data analysis, visualization (e.g., as heat map), and network analysis (with the genome space representing the network adjacency matrix).

EARTH AND ENVIRONMENTAL SCIENCES

HIGHLANDER user-tailored Downstream Application and (pre-)Operational Services (DapOS)

Fausto Tomei, Giulia Villani

ARPAE

Nicola Loglisci

ARPAP

Monia Santini

CMCC

Chiara Della Casa, Cinzia Caroli, Gabriella Scipione

CINECA

HIGHLANDER project has designed and implemented a multidisciplinary framework of highly detailed and harmonized data to support land services. Combining remote and in-situ monitoring, analytical tools, numerical models to machine learning algorithms, the project creates a valuable information that can be directly exploited by a wide range of users. Leveraging the capabilities of Cineca's HPC infrastructure it was possible to feed climate simulations and other integrated territorial observational data into several applications and services, the so-called Downstream Application and (pre-)Operational Service (DapOS). The DApOS can enable a more effective and rapid integration of information into decisions, strategies, and planning of land management activities at different interacting spatial-temporal scales and sectoral levels.

Short-term forecasts (weekly onwards) to medium-term projections (decades onwards) of climate variability and climate-related hazards and impacts' indicators/indices at high spatio-temporal resolution are used in the services related to:

- Crop water requirements
- Land suitability for vegetation
- Forest fire potential
- Soil erosion
- Water cycle and sustainability of competing uses (hydropower, domestic, agriculture, ecological)
- Human wellbeing in rural and urban areas, respectively

All the services designed and developed in the project are available to the final users through the HIGHLANDER data portal and web platform at <https://dds.highlanderproject.eu/>.

Crop water requirements

This DApOS is a climate service addressed to the agricultural sector that provides sub-seasonal forecasts of irrigation needs for crops and impact studies of crop irrigation under climate change projections. Therefore, this DApOS includes two different rationales according to the forecast/projections used as input.

The sub-seasonal irrigation forecasts are weekly produced by ARPAE on three Irrigation Consortia of Emilia-Romagna Region, whereas the irrigations projections are developed in Emilia-Romagna (Faenza plain area) by ARPAE and then applied to three pilot areas: Capitanata Irrigation district in Apulia, in part served by the Ofanto river whose hydrological droughts have been also considered (in collaboration with CMCC), Piana Rotaliana in Trento province (in collaboration with FEM) and Alessandria, Carmagnola and Saluzzo areas in Piedmont (setup and developed by ARPAP).

In general terms, this DApOS combines information on agricultural land use from satellite data, observed weather data, climate weather series, HIGHLANDER sub-seasonal forecast/future projections, and a one-dimensional agro-hydrological model developed by ARPAE (CRITERIA-1D).

Figure 25: water: sub-seasonal (4 weeks) irrigation forecasts for the Renana Consortium in Emilia-Romagna.



Land suitability for vegetation

This DApOS, developed by ARPAP in collaboration with CIA and UNITO, evaluate how climate change might affect regional vegetation.

Two sectors have been taken into consideration: agriculture and forests.

Land suitability for agriculture is very important in Piedmont region and viticulture and mais cultivation make up a good portion of the Piedmontese economy.

On the other hand, the distributions of many terrestrial organisms are currently shifting in most parts of the world and mountain areas are particularly sensitive to these shifts.

In the light of climate change, policymakers, and environmental managers need analytical tools to develop reliable and solid predictions to address management activities such as forest production, biodiversity conservation and reliability of disciplines.

HPC resources developed in HIGHLANDER project have been used to this scope.



Figure 26: land suitability for grapevine over Piedmont Region from 2021 to 2050

Forest fire potential

The DApOS “Forest fire potential” is being developed by ARPAP in collaboration with UNITO. The Mediterranean basin is considered a biodiversity hotspot and, more than other regions of the globe, is affected by the changes of climate and land use. Exploiting historical data, past and future climatic data, the high-resolution models will predict the future dynamic trends of Mediterranean forests (mainly the alpine ones). The results will be analysed at the level of specific composition and alteration of the regimes of natural disturbances.

Human wellbeing

The DApOS “Human wellbeing” was developed by CMCC. Climate change threatens the individuals’ wellbeing, especially in the context of large urban settlements where the microclimate is highly influenced by the physical processes related to the occupation of soils by artificial areas. The high spatio-temporal resolution and last generation climate simulations developed in HIGHLANDER allowed to reproduce the land-surface interactions taking into account the land cover with high spatial detail, especially thanks to information derived from satellite-based land use distribution maps, which allow to distinguish the land use distribution for civil, industrial and agricultural purposes from naturally vegetated areas.

The DApOs exploited data from intense HPC use to derive an ensemble of indicators of human wellbeing at hourly and daily statistics level, for two times horizons (1991-2020 and 2021-2050) and under different climate forcing (downscaling of reanalysis and of projections).

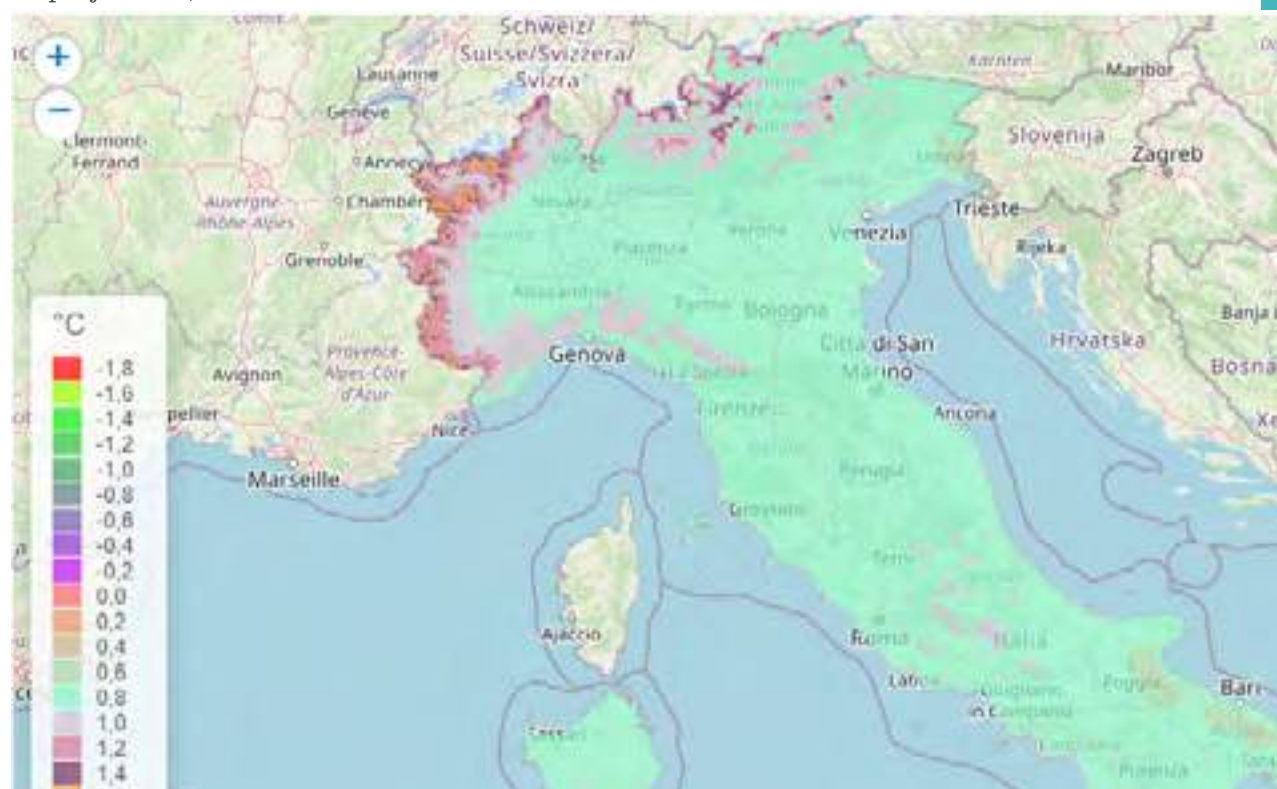


Figure 27: wellbeing: future variation in the Discomfort Index (°C) over Italy.

Enhancing Climate Risk Assessment in Agriculture Insurance

Marco Venturini

Amigo

Roberta Turra, Giorgio Pedrazzi,

Cineca

In the agritech and insurance sectors, low-resolution climate data presents a significant challenge in accurately assessing risks associated with agrifood production. With climate change increasingly affecting the agricultural sector, it has become crucial to have precise and reliable climate information to assess and manage risks. Amigo's goal is to address the challenge of insufficient climate data by leveraging AI predictive models and multi-source data enrichment products. One such effort is the ClimateByte experiment – Climate risk assessment for agriculture insurance using Big data and artificial intelligence funded by the EUHubs4Data project (<https://euhubs4data.eu/experiments/climate-risk-assessment-for-agriculture-insurance-using-big-data-and-artificial-intelligence/>) – which focuses on climate risk assessment for agriculture insurance using big data and artificial intelligence. By utilizing Earth Observations and Model Climate, ClimateByte produces higher quality and spatial resolution data tailored to the user's needs. Main achievements of the experiment are:

- Up to 300m of spatial resolution from an initial 25km of raw data.
- A novel AI algorithm that deals with the lack of high resolution targets
- The integration of different data structures maintaining daily and sub-daily original resolution
- 3 climate variables (minimum temperature, maximum temperature and precipitation) integrated
- 8 regions covered
- A web application that plugs into

proprietary Amigo SaaS.

During the 9-month duration of the experiment, the Marconi100 HPC was used to produce 2.1TB of new data starting from 500GB of raw data, with a total of 6571 cluster hours consumed. The use of HPC has enabled the parallelization and optimization of the model by selecting the best cost-effective parameters and set of predictors. The chosen model was a Stacked Ensemble that includes several executions of different decision tree models, resulting in a correlation to the target variable of about 90%. To validate the model for the regions of interest it was used one main dataset from EUHubs4Data catalogue: the MISTRAL meteo-hub. It offers a single access point repository of ground weather station data alongside the Italian country. Data from 3 different regions, the most dense in terms of data, has been downloaded.

The experiment conducted two pilots in Trentino Alto-Adige region in Italy for assessing the risk of wine production during mid-April due to cold spells, and in Marche region (Italy) for the forage production during the spring, to understand the lack or excess of precipitation. These pilots have shown that the ClimateByte experiment has the potential to provide accurate and relevant data to assess climate risks in agriculture insurance, benefiting both farmers and insurance companies.

The results of the pilots were very promising, especially for the temperature downscaling model which is highly correlated with the cold spells events provided by the customer.

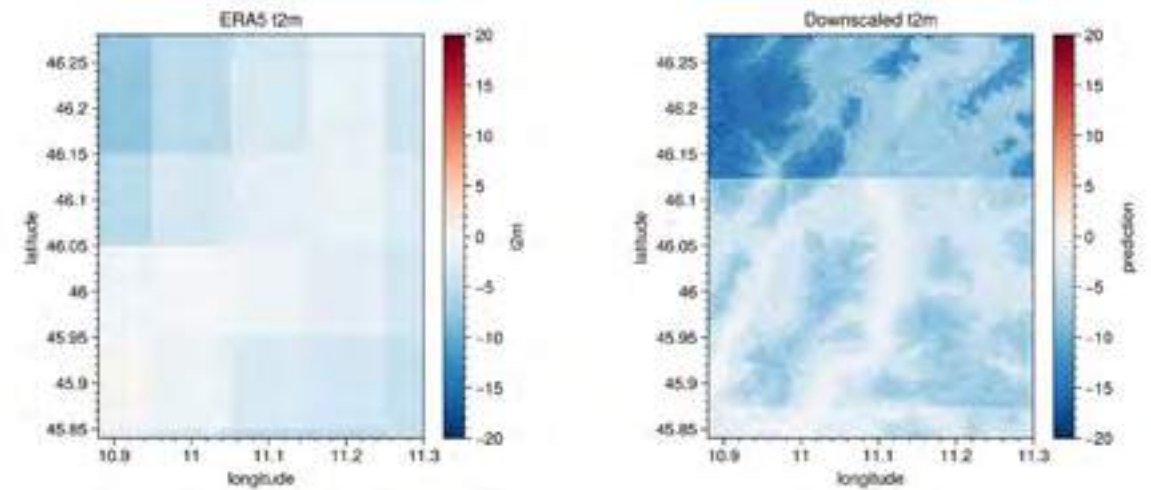


Figure 28: minimum Temperature (Trentino Alto-Adige). In the left-hand-side panel, the target temperature at low-resolution (9km). In the right-hand-side panel, the downscaled temperature at high resolution (300m).

Figure 29: A screenshot of one page of the Web Application.



MobilitySquare: Innovating E-Bike Rental

Giorgio Pedrazzi, Donatella Sforzini

Cineca

Massimo Infunti

iMpronta

MobilitySquare, funded by the European project EUHubs4Data, is an experiment in which the Cineca Data Analytics Team serves as a coach and service provider. The experiment was proposed by iMpronta, an Italian IT consultancy company, in partnership with BikeSquare, a company specializing in designing personalized cycling routes for tourists and cyclists. The main objective of the MobilitySquare experiment is to develop an e-bike rental recommendation engine that can accomplish the following:

- Predict cancellations of bookings based on weather forecasts and relevant user characteristics.
- Recommend alternative destinations and exciting activities to users who are highly likely to cancel their bookings.

The goal of MobilitySquare is to enable BikeSquare to provide even more personalized and tailored services to its customers.



The experiment primarily focused on forecasting cancellations due to weather conditions and other factors to propose alternative destinations to users who were likely to cancel or postpone their reservations. In the second phase, a significant amount of bike location data was collected, which will be utilized to recommend unique and interesting routes to users. The Cineca Data Analytics team utilized historical rental data provided by BikeSquare to create a Machine Learning model for predicting cancellations. The team conducted preliminary exploratory analysis of the data and carried out various steps, including defining new variables and selecting a suitable subset of records for the modelling phase. The data presented a challenge due to its highly unbalanced nature, with a significantly

lower number of cancellations compared to confirmed bookings. To select the best machine learning algorithm for the project, the team explored various strategies such as class balancing, Ensemble methods, and threshold tuning. While these efforts resulted in a high accuracy for the model, other performance parameters, specifically precision and recall, were given high priority during model evaluation, recognizing their critical role over accuracy. The primary goal is to provide users with a reliable experience by delivering timely and accurate notifications while minimizing the likelihood of false positives. Therefore, achieving the highest precision possible in the model is prioritized.

To find the best model, the team utilized the Python version of the H2O library. Initially, the AutoML function of H2O was experimented to obtain indications on the models to be used and their parameters. The cross-validation strategy employed used 75% of the observations, with 80% allocated for training and 20% for cross-validation testing. The remaining 25% of observations were reserved for testing purposes. The cross-validation metrics for the main model were computed by pooling the holdout predictions, which were predictions for every row of the training data made by a model that had not seen that row during training. The resulting “holdout prediction” was then scored against the true values to compute the overall cross-validation metrics.

Table 1 presents the performance of various algorithms, including the winning algorithm from AutoML (based on the GBM algorithm), Random Forest, XGBoost, and GBM. Multiple models were trained using different parameters, and the performance of the winning AutoML GBM model is compared to that of other algorithms. The optimal parameters estimated by grid search are highlighted in bold font.

| Name | Algorithm | Parameters | Test Cross-Validation | | | Test Error Avg | | |
|----------|---------------------------|---|-----------------------|--------|----------|----------------|--------|----------|
| | | | Precision | Recall | Accuracy | Precision | Recall | Accuracy |
| GBM 2023 | Gradient Boosting Machine | max_depth=10, min_child_weight=1, subsample=0.8, colsample_bytree=0.8 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 |
| RF 2023 | Random Forest | n_estimators=100, min_samples_split=10, min_samples_leaf=5, max_depth=10, colsample_bytree=0.8 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 |
| XGB 2023 | XGBoost | max_depth=10, min_child_weight=1, subsample=0.8, colsample_bytree=0.8 | 0.93 | 0.93 | 0.93 | 0.93 | 0.93 | 0.93 |
| GBM 2022 | Gradient Boosting Machine | max_depth=10, min_child_weight=1, subsample=0.8, colsample_bytree=0.8 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 |

Table 1: Comparison of different algorithms results.

The models are compared using standard classification measurements such as accuracy, precision, and recall. Given the priorities of the MobilitySquare experiment, the model with the highest precision on the test data, GBM2023 (an evolution of the GBM AutoML model created by AutoML), is considered the most favourable.

Furthermore, H2O provides a range of model explainability methods that are valuable in understanding the inner workings of the model, identifying critical variables, and ensuring accurate predictions. This capability of explaining a model’s decisions is beneficial for MobilitySquare in selecting the most suitable model. It goes beyond examining performance measures by considering the relevant variables. This enables iMpronta to enhance its data knowledge and make more informed decisions based on the insights provided by the model. For example, the importance of variables across multiple models can be illustrated using a variable importance heatmap, providing a visual representation of their significance, as in the following figure X



The EUHUBS4DATA project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 951771



Exploring the Integration of Artificial Intelligence in Artwork Analysis

Giorgio Pedrazzi, Donatella Sforzini, Gabriele Fatigati
Cineca

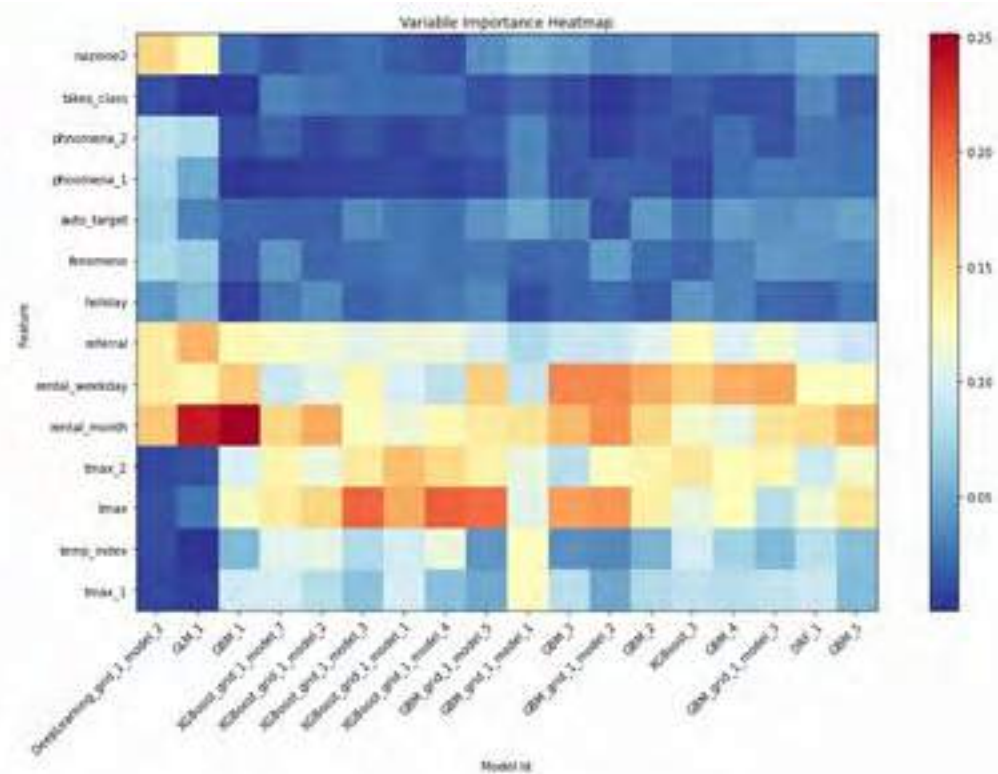


Figure 30: variable importance heatmap.

After applying the selected model, MobilitySquare utilizes the geographical data collected by the GPS Kit to identify the route chosen by the tourist and recommend points of interest. The team has been working with the GPS Position dataset, which is cross-referenced with the dataset of Points of Interest served by one of their “official” routes. This allows them to deduce the most similar route that a tourist has taken during their tour. To achieve this result, a Jaccard Similarity Algorithm has been employed, which compares all the “official routes” with all the Points of Interest visited by an e-bike. By leveraging the geographical data and the previously developed algorithm, Mobility-Square aims to provide tourists with personalized recommendations and enhance their overall experience.

In recent years, Cineca has recognized the potential of AI for cognitive applications and has focused on integrating AI technologies into their data analytics efforts. As Data Analytics team inside the HPC department we would like to leverage the power of AI algorithms and methodologies to our analysis. By harnessing AI for cognitive applications, we can develop models or use pre-trained models. We explore areas such as deep learning, natural language processing (NLP), and computer vision to develop AI-driven solutions tailored to the specific needs of academic and research communities.

In the domain of visual arts, we undertook a research endeavor to explore the integration of artificial intelligence (AI) in the field. Our objective was to investigate the potential of AI in classifying artwork, a subject that had garnered significant attention from art enthusiasts and experts alike. We delved into cutting-edge technologies such as object detection and zero-shot detection models. These advanced AI systems possessed the remarkable ability to identify and categorize objects within artwork without extensive prior training. With the aid of high-performance computing, these models could process large volumes of images rapidly, opening new possibilities in the world of art.

Our efforts were focused on automating the tasks of cataloging and annotation, which traditionally required significant manual labor. We identified three primary areas of interest: classification, object detection, and zero-shot detection.

In the classification domain, our goal was to assign artworks to specific categories based on predetermined criteria such as authorship or artistic style. To achieve this, we trained models on carefully curated datasets, enabling them to learn how to accurately classify new artworks based on their characteristics.

Object detection aimed to identify and locate specific objects within artwork,

such as people, animals, or symbolic elements. We trained our models on annotated images, teaching them to detect and precisely localize objects within the artwork. We utilized tools like Detectron2, which facilitated object detection in new images based on pre-annotated examples. We employed a diverse dataset comprising thousands of annotated images from various sources, provided by projects dedicated to visual arts. Rigorous preprocessing and training phases were executed to ensure optimal performance of the models. However, we encountered challenges with object detection due to the high variability and specificity of the predefined classes in our project. Some classes were more general, while others required more precise annotations. Additionally, the limited number of annotated images hindered the model’s accuracy. Creating a new dataset with different classes would have been time-consuming and required significant manual effort.

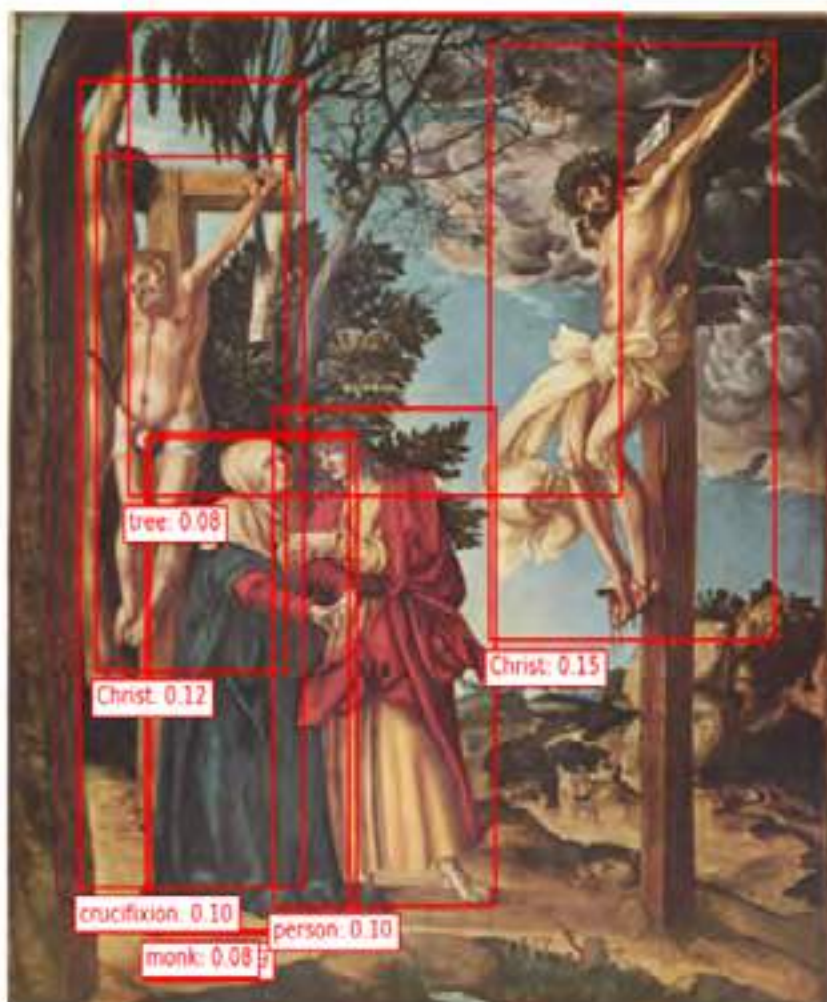
Undeterred by these challenges, we turned our attention to zero-shot detection, an intriguing machine learning task that involved detecting objects or concepts in an image without prior training on those specific items. We leveraged the power of advanced language models and transformers to analyze images and comprehend their contents with remarkable accuracy. This technology garnered significant attention for its ability to automatically detect and classify objects.

Willingly embracing zero-shot detection, we explored models such as CLIP, a versatile model capable of assigning probabilities to classes (freely defined) present in an image. We also investigated extensions of CLIP, such as OWL-ViT, which allowed for determining bounding boxes and corresponding classes for detected objects. In both CLIP and OWL-ViT models, a vector of strings is used to establish a connection between textual descriptions and visual content. This vector, often referred to as the

“text prompt,” is a sequence of words or phrases that describe the desired visual attributes or concepts. In CLIP, the text prompt serves as a conditioning input to the model. It allows the model to associate the textual information with the corresponding visual features in an image. Similarly, OWL-ViT also employs a vector of strings as a text prompt to guide its visual recognition capabilities. By combining the information from the text prompt with the image features, OWL-ViT can localize and classify objects within an image. The model learns to associate the textual descriptions with specific regions in the image. In this phase, to the set of classes derived from the annotation of the SGoaB dataset, we added further terms identified during periodic comparisons with a team of domain experts, the digital humanists of the MiC (Italian Ministry of Culture) for a total of 75 terms



text_vector
 ["angel", "apple", "arrow", "banana", "banner", "bird", "boat", "book", "camauro", "cat", "chalice", "crown", "crown of horns", "crozier", "deer", "devil", "dog", "dove", "dragon", "eagle", "elephant", "horn", "horse", "key of heaven", "knight", "lance", "lily", "lion", "mitre", "monk", "nude", "orange", "palm", "person", "prayer", "rooster", "saturno", "scroll", "serpent", "sheep", "shepherd", "skull", "stole", "sword", "tiara", "tree", "trumpet", "zebra", "zucchetto", "Christ", "Virgin Mary", "saints", "evangelist", "bow", "armour", "halo", "the baby Jesus", "hat", "clamshell", "cross", "flower", "fruit", "jewel", "lute", "mantle", "cow", "cloud", "landscape", "shield", "snake", "mirror", "violin", "crucifixion", "hand", "head"]



We tested the model on a diverse range of images, including grayscale and color variations, to assess their performance across different scenarios.

Impressed by the capabilities demonstrated by these AI tools, we envisioned a future where they could support manual annotators in the cataloging and annotation processes, streamlining the organization and understanding of artwork in various collections and repositories.



Figure 31: “The Crucifixion, by Cranach, Lucas the Elder (1472-1553). Oil on wood”. image with OWL-ViT mod google/owlvit-base-patch32 annotation.

Figure 32: MiC Digital Library website.

MATERIAL SCIENCES

Material science getting ready for Exascale

Materials at Exascale

Fabio Affinito, Laura Belletani

Cineca

Computational material sciences have been always playing a crucial role in Italy. Physics and Chemistry of the solid state have a long history and their users' communities have been the major customers of CINECA clusters since the first HPC infrastructures were installed. The collaboration with SISSA and the availability of one of the first HPC systems gave momentum to the development of the famous Car-Parrinello method. Still today, CINECA is a pivotal point for all the computational scientist in all Italian academic institution, permitting to work on new computational methods and on the exploitation of our systems for challenging simulations.

During the years, CINECA consolidated a collaboration with SISSA for the support and the development of Quantum ESPRESSO, one of the most popular codes for density-functional based methods. As part of the Quantum ESPRESSO Foundation, we have been promoting the evolution of this suite and its further development. In particular, during last year we increased our effort and engagement in the GPU porting of more codes from the Quantum ESPRESSO suite. This permitted to have a more flexible and efficient code, opening to the users the usage of GPU-enabled systems.

Together with SISSA, CNR and University of Modena and Reggio Emilia, we are partners of the European Center of Excellence MaX. This is the third phase of a project originally started in 2015, in which CINECA supports the evolution of

material science codes and their co-design under a HPC perspective. This phase of the project focuses not only on performance and efficiency on heterogeneous architectures, but also on the implementation of a high-throughput ecosystem where codes interoperate in exascale workflows.

As partners of the TREX Center of Excellence, we supported the development of Quantum Monte Carlo codes, by providing HPC resources, and by organizing an Hackathon, hosted at CINECA premises, to port TREX applications on GPUs.

In the context of the activities funded by the national PNRR, we are supporting the developers and users' communities in the domain of computational materials, in the ICSC and ECOSISTER projects. Thanks to these resources, we increased our staff in the high-level support team, opening up new opportunities for support and development activities.

Looking to the future, CINECA aims to strengthen its position in computational material sciences and HPC through continued investments in infrastructure, research funding, and interdisciplinary collaboration. By fostering international collaborations, and leveraging advanced computational tools, we are poised to make significant contributions to the development of novel materials, sustainable technologies, and advancements in various industrial sectors.



MAX has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 824143.



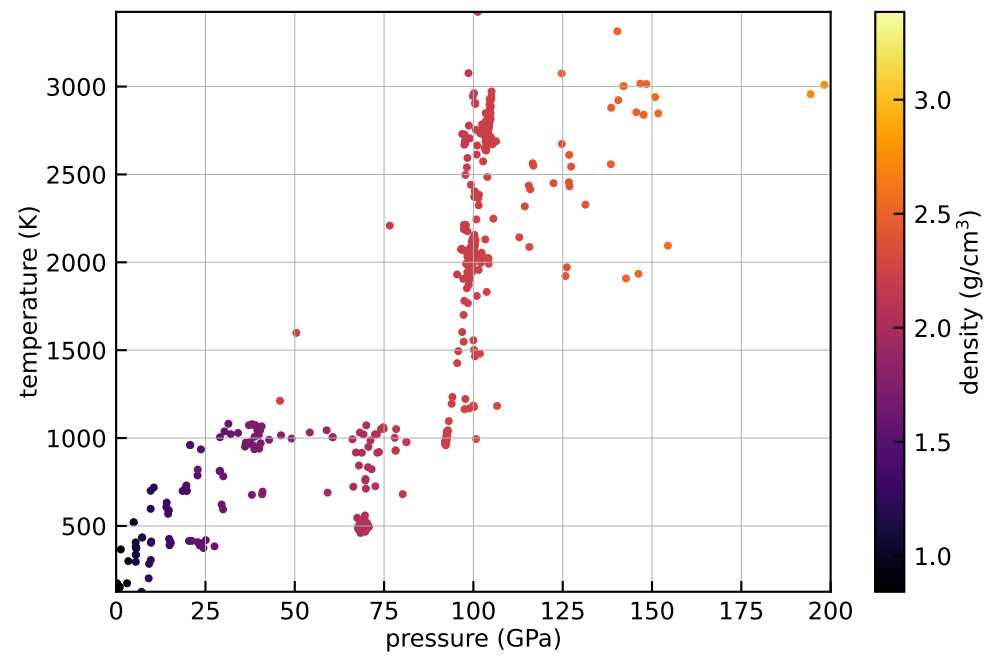


Figure 33: equation of state diagram of Ammonia, computed with Car-Parrinello method in Quantum ESPRESSO (courtesy of R. Bertossa, SISSA).

QUANTUM COMPUTING

Cineca Quantum computing perspectives

Daniele Ottaviani, Riccardo Mengoni, Christian Fiori, Emanuele Triuzzi, Gabriella Bettonte

Cineca

Enrico Prati

Milan University

Simone Montangero, Daniel Jaschke, Marco Ballarin

Padua University

Domenico Bonanni

Dompé



2022 was a very important year for quantum computing at CINECA: it will forever be remembered as the year in which CINECA, leading an international consortium comprising Italy, Germany and Slovenia, won the role of hosting entity for one of the firsts European quantum computers.

In March 2022, in fact, the EuroHPC JU launched a call for the selection of a certain number of European supercomputing centres able to host a quantum computer co-financed by the European entity. CINECA participated by submitting an expression of interest and, in October 2022, was selected as one of the hosting entities.

This means that the Leonardo supercomputer will soon be equipped with a latest-generation quantum accelerator: once integrated into the HPC system, the quantum computer will in fact make the Leonardo one of the first supercomputers in the world to be able to exploit a quantum accelerator with hybrid HPC-QC codes, opening up numerous possibilities for Italian and European research, which will be able to have access to the new quantum

machine exactly as already happens today with all the supercomputing resources that CINECA makes available to its users.

In addition to our infrastructure advancements, we have also dedicated substantial efforts to the development of various quantum computing applications. Through collaboration with our esteemed partners and researchers, we have explored the potential of quantum computing in diverse fields. These collaborations have yielded promising results, and we are eager to present some of the notable applications that have emerged from our collaborative efforts.

As we move forward, we remain committed to advancing the frontiers of quantum computing and its practical applications. Our goal is to leverage this emerging technology to address complex problems and revolutionize computation in science, industry, and society at large. With our ongoing efforts and the support of our collaborators, we are confident in the bright future that lies ahead for quantum computing at CINECA.

Application: Molecular Docking

Within the EuroCC Initiative from EuroHPC Joint Undertaking, CINECA Quantum Lab, Politecnico di Milano and Dompé tackled the Molecular Docking problem by taking advantage of the original Weighted Subgraph Isomorphism on Quantum Annealers approach.

Molecular Docking (MD) is an essential step in the drug discovery process involving the detection of three-dimensional poses of the ligand inside the active site of the protein. In such work, the team addressed the entire Molecular Docking search phase by directly formulating the problem in QUBO terms, suitable for a quantum annealing approach. More in detail, ligands were represented via weighted graphs which embed geometrical properties of the molecule and were mapped into a weighted spatial grid resembling a discretization of the 3D space region inside the protein pocket. Ligand poses were then evaluated in terms of an optimal weighted subgraph isomorphism between the ligand graph and the space grid, expressed as a Quadratic Unconstrained Binary Optimization (QUBO) problem.

This approach granted significant speedups in translating the original Hamiltonian in a QUBO problem and related Time to Solution (TTS) with respect to classical simulated annealing method.

Application: Quantum Machine Learning on different quantum computing platforms

The scholarship began in July 2022 with the study of quantum machine learning algorithms on different quantum computing platforms. In September 2022 the main topics of the initial study have been highlighted with a presentation, with major focus on Quantum Support Vector Machines. The following months were dedicated to researching and implementing a quantum SVM through QUBO problem solving; sampling tests were done using the D-Wave Quantum Annealer, comparing it to traditional machine learning approaches. The problem involved classifying targets in hyperspectral images with 200 channels per pixel. Quantum SVM showed comparable accuracy with respect to the classical SVM, but superior metrics (AUPRC, AUROC). This is due to the inherent statistical nature of solutions obtained through annealing: furthermore, it has been assessed that the combination of multiple models results in a composite predictor with higher scores of each metric. A scientific article is being written to document the work. The problem was later approached using the emulation software Pulser, developed by Pasqal, exploring different tasks and datasets. Research was then conducted on solving the Maximum Independence Set problem on unit-disk graphs and related applications. Currently, the focus is on efficient information transmission in IoT networks without centralized infrastructure. It is being formalized using unit-disk graph topologies, in order to leverage the Pasqal quantum computer. Solving the Maximum Independence Set problem via quantum algorithms can significantly improve information transmission efficiency and reduce interferences among devices.

Application: Developing an HPC Tensor-Network Quantum Computing Emulator in collaboration with the University of Padua

The quantum research group at the university of Padova led by Professor Simone Montangero and CINECA collaborate on several subjects and have established a fruitful knowledge transfer in both directions, where students profit from internships, tutorials, and hackathons organized by Cineca while researchers offer themselves tutorials on specialized topics of quantum computing. Moreover, there is an ongoing joint effort in writing the Quantum TEA library, a software suite to simulate quantum systems with tensor networks: the goal is to fully exploit the capabilities of the CINECA clusters. In particular, the CINECA team is helping Padova in the parallelization of the code through OPENMP, MPI, and GPU

Figure 34: first quantum computers in Europe.

usage. Most focus the effort are the low-level libraries and quantum matcha TEA, a tensor network emulator for quantum circuits. The HPC clusters at Cineca are an essential resource for the simulation of many-body quantum systems and enabled cutting-edge research in condensed matter, lattice gauge theories, and quantum technologies. One example is the publication “Entanglement entropy production in quantum neural networks”, accepted in Quantum, which is a collaboration of the university of Padova and Pavia as well as Cineca and was run on Galileo100.

Workshop: High Performance Computing & Quantum Computing (HPCQC) 2022 – Fifth Edition

On 15 December 2022, the High Performance Computing & Quantum Computing (HPCQC) was held. The annual conference focusses on the relationship between HPC and quantum computing and it was held by Cineca Quantum Computing Lab, this year in collaboration with the Politecnico di Milano and the University of Milan.

After the success of the previous years, HPCQC reached its fifth edition, finally back in attendance, and like every year it has brought together the community of researchers and companies working in the fields of quantum computing and HPC. The edition was trilled by the recent approval by EuroHPC JU of the first six European quantum computers, one of which to be hosted at Cineca

The edition involved both the European and Italian ecosystems, with a look at the large quantum infrastructure that is being created with the continuation of the HPCQS project. A special look was devoted to the new startups aiming at creating quantum hardware in Europe, in addition to more established companies like IBM and D-Wave.

Last part of the workshop was dedicated to research scientist who reported about recent updates from both the algorithmic, the application and the hardware integration sides.

Italy confirms as on of the most active Countries in the field of quantum computing and the initiative provides a key networking moment of the European Quantum Computing and HPC ecosystem

HUMANITIES AND
SOCIAL SCIENCES,
DIGITAL
ECOSYSTEMS

Integrating Artists into the Scientific Loop: The GRIN Project

Antonella Guidazzoli, Maria Chiara Liguori

CINECA

Nicoletta Tranquillo, Silvia Girardello

KILOWATT

In February 2023, the GRIN project (Art-driven innovation for digital and green transition in European Regions) was launched, a preparatory action of the STARTS regional centres aimed at developing effective solutions for the digital and ecological transition in Europe. Five Regions are involved: Upper Austria, with the Johannes Kepler Universität Linz, which is coordinating the project; Aveiro, in Portugal, with Artshare; Porto District, also in Portugal, with Coliseu Porto Ageas; the Finnish Oulu University of Applied Sciences, representing the North Ostrobothnia region; and KILOWATT and CINECA, for the Emilia-Romagna Region in Italy.

12 artistic residencies are foreseen for the project in order to promote the creation of multi-disciplinary teams of scientists, technology providers, engineers, designers, and artists, and to foster the development of 12 new artworks/solutions in line with a human-centred approach to innovation.

The call for the residencies proposes a series of possible challenges as the main source of inspiration for the artists. For the Emilia-Romagna region and its 3 residencies, Kilowatt and Cineca have developed a series of themes, starting with explorations around the concept of digital twin as a synthetic representation of the complexity of reality. Drawing on CINECA's HPC facilities and

expertise in meteorological and climate data management, AI and scientific visualisation, artists will be invited to explore:

the design of technology to enhance the human experience and promote sensory and cognitive capacities;

how to promote non-anthropocentric perspectives to decolonise our relationship with nature, society and ecosystems;

the ethical considerations that underpin the creation and implementation of digital twins to avoid reinforcing existing power structures or creating new ones;

and, in general, how art can contribute to the development and use of these technologies to promote a more just and sustainable system.

GRIN also aims to scale up new business ventures and create five new STARTS Regional centres, while raising public awareness and knowledge on content, technology, and processes that empower all citizens to act in the green and digital transitions. The project framework will create a network of partners to work together on common European and global challenges for ecological transition and sustainable well-being, with a consortium of the six core partners.

The project will end in September 2024.

S + T + ARTS

SCIENCE + TECHNOLOGY + ARTS

GRIN S+T+ARTS regional centres preparatory action under grant agreement LC-01984769 aims to develop effective solutions for Europe's digital and ecological transition.



CHEMISTRY

In Silico drug design with Ligate

Andrew Emerson
Cineca

It is well known that the development of new drugs is extremely difficult and expensive – a development time of 12 years and an investment of 2 billion dollars is not uncommon. The failure rate is high, with some estimates suggesting that only 12% of molecules that enter clinical development are approved for market use. Although, the rate of discovery has remained constant over the last few years the coverage is not even, with oncological therapies and biologics (medicines based on genetic information) far surpassing new antibiotics or antivirals. The difficulty with finding molecules for combatting viruses was highlighted by the very recent Covid-19 pandemic – very effective vaccines against the virus were rapidly developed, but pharmaceutical companies had much more difficulty finding successful drug treatments.

The use of computer aided drug design (CADD), i.e. using in silico methods, to accelerate drug development is increasingly widespread and can be useful in limiting the use of animal models in pharmacological research, aiding the rational design of novel and safe drug candidates, and for repositioning marketed drugs, as well as supporting researchers during drug discovery. An important CADD technique is high throughput virtual screening (HTVS) which involves binding or “docking” candidate drug molecules (or “ligands” from a database against a “target”, normally a protein which is involved with the disease or condition. In the case of viral infection, for example, the protein could be an enzyme needed for the virus to infect or replicate within the host. The act of binding a small molecule to the protein may affect its function, e.g., by inhibiting the enzyme and hence reducing the viral infection. The top-scoring molecules, i.e. those which bind in the most effective way with the target, can then be passed on to the laboratory for further analysis.

This procedure thus acts like a filter, selecting only those molecules which are like to have a successful outcome. HTVS can be done very efficiently in a modern supercomputer since the process is inherently parallel, with each binding event being independent of the other ligands. For example, in the European-funded Exscalate4Cov project, in the so-called “Big Run”, the binding

capabilities of more than 70 billion ligands were evaluated on 15 possible binding sites of 12 SARS-Cov-2 viral proteins in 60 hours simultaneously on two entire clusters, the CINECA M100 and the Eni HPC5 systems. In the experiment, using the Ligen software jointly developed by CINECA, DOMPE’ pharmaceuticals and other partners, all the available CPU cores and Nvidia GPUs were used to perform more than 1 trillion binding evaluations. This is, to our knowledge, the largest such experiment ever performed.

The EXscalate4Cov project finished in mid-2021 but by that time the Ligate project had started. Funded by EuroHPC, it has some of partners in common with Exscalate4Cov, including CINECA, Dompe Pharmaceuticals, Politecnico Milano and KTH (Sweden). The aim of the project though is different to that of Exscalate4Cov - whereas the latter project was a multidisciplinary consortium aimed at fighting the SARS-Cov2 virus, LIGATE aims to deliver an integrated solution for drug discovery using HPC. Resources employed include Ligen, the GROMACS molecular dynamics simulation software developed at KTH (Sweden), Hyperloom (from IT4I) for enabling pipelines and workflows, and an AI-based engine for evaluating binding interactions. This last aspect of the project, i.e., a machine learning model for ligand-protein interactions will be a key outcome of the project. In fact, the use of machine learning in general is expected to be an important driver in drug discovery. Other features of the project include the conversion of the software components to SYCL, a C++-like language which is architecture-independent and will ensure that the software will run on any hardware. Given that an increasing number of TOP500 are now using AMD GPUs (e.g. LUMI, number 3 in the TOP500 list in June 2023), this enhanced portability will be important for rapid deployment of the workflow. The Ligate project is now in its final year and is making extensive use of the computational resources available from CINECA and other sites in the EURO HPC ecosystem to complete and validate the drug discovery pipeline. Hopefully, with tools like those that will be available from Ligate, we will better prepared for future pandemics.

[1] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7014862/>



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.



COE
Centers of
Excellence

EXCELLERAT: Advancing High-Performance Computing for Industry Transformation

Claudio Arlandini
CINECA

High-performance computing (HPC) has become a cornerstone of innovation and progress across numerous industries, enabling complex simulations, data analysis, and modelling at an unprecedented scale. Recognizing the potential of HPC for industrial applications, the EXCELLERAT1 Centre of Excellence for Engineering Applications project was initiated to push the boundaries of HPC technologies and revolutionize various engineering sectors. Its phase 1 ended in May 2022, and a new 4-year project (Phase 2) started January 2023.

The EXCELLERAT project, co-funded by the European Union through the EuroHPC JU Programme, was launched with the primary aim of creating a single point of access for expertise on how data management, data analytics, visualisation, simulation-driven design and Co-design with high-performance computing (HPC) can benefit engineering, especially in the aeronautics, automotive, energy and manufacturing sectors. The goal of EXCELLERAT is to enable the European engineering industry to advance towards Exascale technologies and to create a single entry point to services and knowledge for all stakeholders (industrial end users, ISVs, technology providers, HPC providers, academics, code

developers, engineering experts) of HPC for engineering. In order to achieve this goal, EXCELLERAT brings together key players from industry, research and HPC to provide all necessary services.

EXCELLERAT Phase I developed 11 use cases, based on 6 popular engineering software (Nek5000, Alya, AVBP, TPLS FEniCS and CODA), whose performances were enhanced and new features added. Most use cases were related to the Aerospace and Automotive sectors. CINECA worked on two of them, in collaboration with KTH, on High fidelity simulation of rotating parts on Nek5000, and the Adjoint optimization in external aerodynamics shape optimization with FEniCS.

CINECA was also responsible for Workpackage 5 “Centre Implementation”, that supported the implementation of the defined EXCELLERAT services, including training portfolio organization and procurement of computational resources for the technical work packages. Another focus of WP5 was on managing services, users’ access and related issues, both for internal general services and for services towards the users. Finally, WP5 was responsible for managing the service provision, including the integration of infrastructure services for

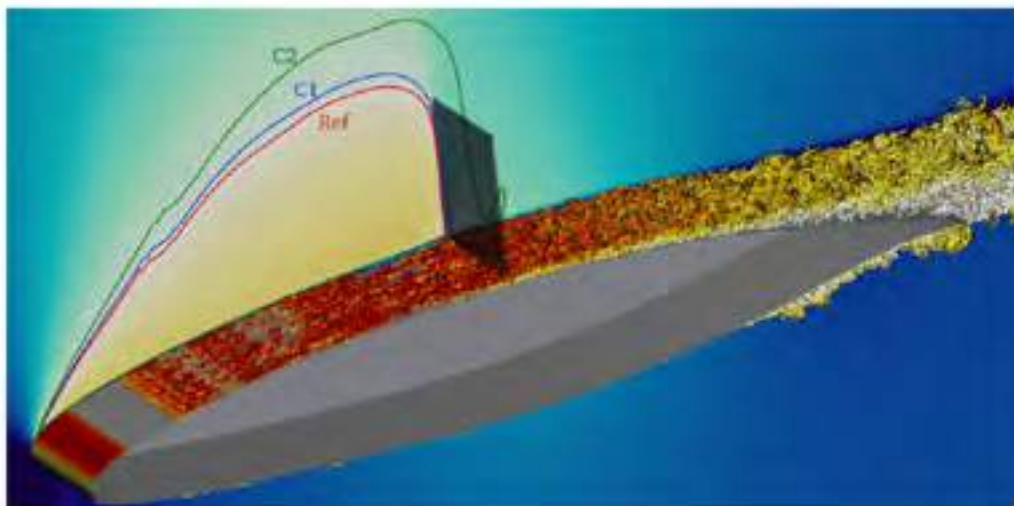
EXCELLERAT activities, to be coordinated with the provision services of the HPC centres. For that purpose, one of the main products of CINECA was the development and deployment of the EXCELLERAT service portal (<https://services.excellerat.eu>).

During Phase 2 the service portal will be renewed and enlarged, and CINECA is working on another relevant use case, in collaboration with La Sapienza University. In this task, CINECA is supporting a team lead by Prof. Sergio Pirozzoli of the Mechanics and Aerospace Engineering Department of La Sapienza, to improve the code FLEW (Finite volume Wing) for simulating active control for drag reduction of transonic airfoils.

In aeronautical applications, several drag-reduction strategies have been developed, but few of them have been applied due to an unfavourable cost-benefit ratio. However, some of them are very promising: increasing the aerodynamic efficiency of a wing with a positive net power saving will easily lead to a reduction in fuel consumption, with cost and environmental benefits, especially in terms of CO2 reduction. Among them, a promising flow control consists of streamwise-traveling waves of spanwise velocity applied on a portion of the suction side of the wing. Unfortunately, only the use of computationally expensive high-fidelity simulations (DNS) provides a reliable tool to fully investigate the effects of the control strategy, since all flow variables and physical correlations are easily accessible. Preliminary first tests on Leonardo show already an excellent scalability.

In the project is involved also the Italian company E4, in charge of a task responsible for supporting co-design and testing of the project tools and software on novel computing architectures.

Figure 35: flow over supercritical airfoil. Lines denote the position of the mean sonic line, with no actuation (Ref), and with moderate (C1), and strong (C2) wall actuation.



This project has received funding from the European High Performance Computing Joint Undertaking (JU) and Germany, Italy, Slovenia, Spain, Sweden, and France under grant agreement No 101092621.



MaX – European Centre of Excellence

Fabio Affinito, Laura Bellentani, Andrew Emerson and Mariella Ippolito
CINECA

MaX (Materials design at the eXascale”) is a European Centre of Excellence that focuses on the development and application of high-performance computing (HPC) for materials science and innovation, bringing together a consortium of research institutions, universities, and industry partners. It is funded by the European Union since 2015 and in January 2023 started its third phase. The primary objective of MaX is to advance the field of materials science by harnessing the power of HPC and data analytics. By leveraging advanced computational methods, simulations, and data-driven approaches, MaX aims to accelerate the discovery, design, and development of new materials with tailored properties for various applications.

CINECA is a partner of MaX together with SISSA, CNR and the University of Modena and Reggio Emilia, as the other Italian partners, confirming its commitment to the progress of material science in Europe. Within MaX CINECA supports the evolution of codes in the consortium and their co-design on emerging exascale architectures. Among the codes involved in MaX, we are actively contributing to the further development and GPU offload of QuantumESPRESSO, one of the most important programs for electronic structure calculation and material modelling based on density functional theory methods. In

particular, the HLST of CINECA worked in strict collaboration with SISSA developers to enable more codes from the suite on GPUs, to improve their portability across different architectures, and to ensure their optimization for NVIDIA cards, in order to fully exploit the computing power provided by the upcoming Leonardo.

In this phase of the project, our work will focus not only on code porting, but also on designing the interoperation for complex workflows. The availability of a pre-exascale (and in the short-term of exascale) system, enables the convergence of HPC and High-Throughput calculations. Thus, together with the performance of HPC codes, it is very important to ensure the robustness of workflow managers, schedulers and the data management. The goal is the implementation of a high throughput ecosystem of interoperable codes to solve scientific problems that cannot be tackled without the computing power provided by exascale machines. By leading the workpackage for technical challenges towards exascale, CINECA will address the technical needs of the codes, ranging from the assessment and analysis of their parallel performance to the adoption of advanced programming models and runtime environment, with a special focus on workflows and resilience, to the deployment of the codes on EuroHPC machines.

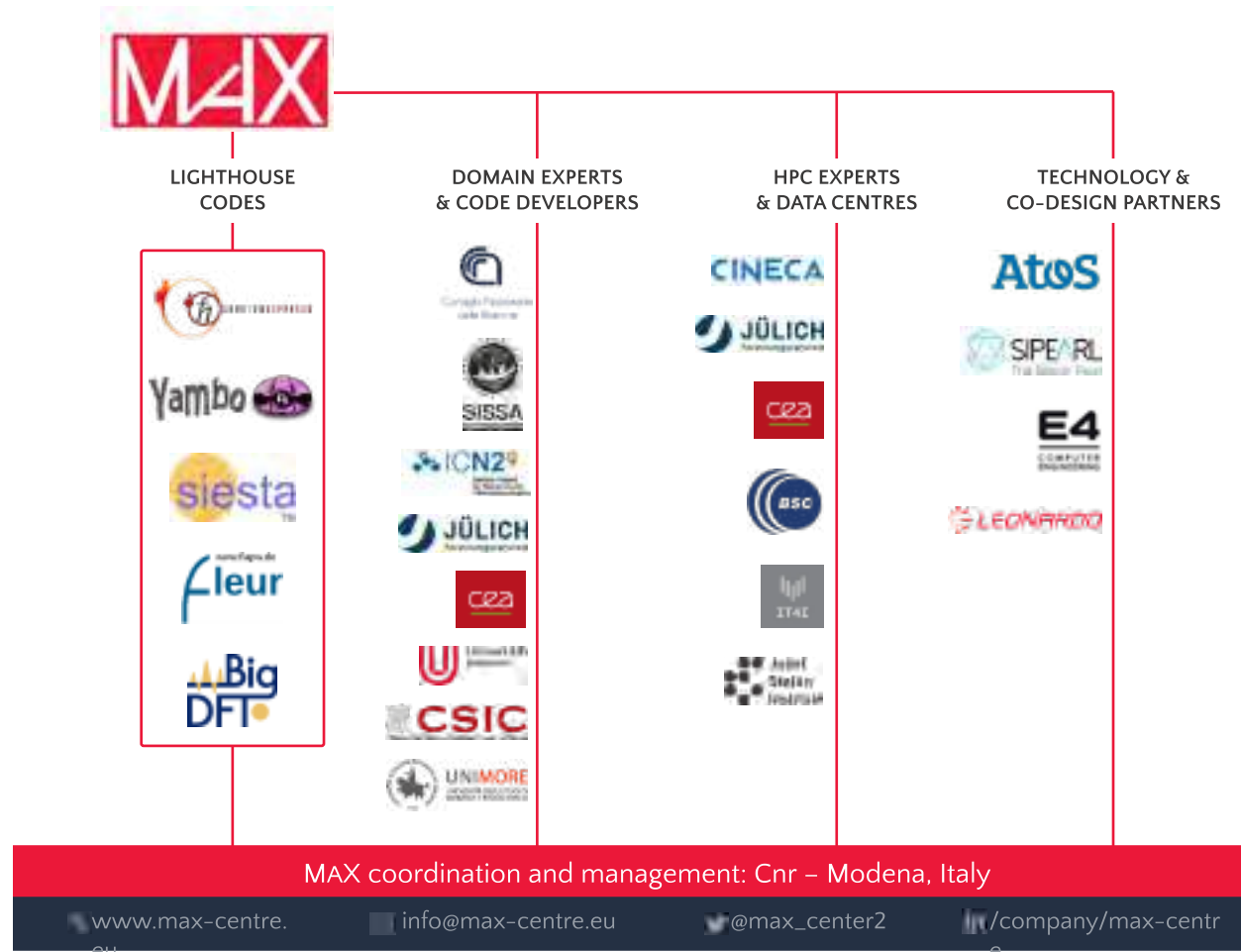


Figure 36: MaX partnership, third edition.

The second phase of the ChEESE project is underway. The potential of High Performance Computing and the effective use of pre-exascale machines like Leonardo at the service of emergencies and the estimation of the danger of earthquakes, volcanoes, tsunamis.

ChEESE-2P, the follow-up of the successful ChEESE project, is on track to address challenges in the field of geosciences in support of civil protection actions at European level and based on the help of supercomputers to exascale, very fast computers that make it possible provide “urgent computing” services during emergencies, early warning, and assessment of the hazards of relevant geological events. For exascale computing we mean the use of supercomputers capable of processing up to a billion billion mathematical operations per second: a real scientific and technological challenge.

ChEESE-2P represents the second phase of the ChEESE project, a Center of Excellence for the use of Exascale Computation in Solid Earth Sciences, during which were successfully developed and upgraded codes and procedures to provide operational risk mitigation products.

The current project is coordinated by the Spanish National Research Council (CSIC) and will be financed by the European Commission and the funds of the 10 participating countries with approximately 8 million euros.

CINECA's role in the project is, approximately the same as in the first ChEESE round, crucial to address the computational challenges and bottlenecks that can arise in the 11 (10 in the former ChEESE project) flagship codes and that can limit the efficiency and the performance of the codes when using pre-exascale hardware like CINECA Leonardo pre-exascale system. This hard job is done in the WorkPackage 2 of ChEESE-2P, lead by CINECA and with the active participation of all partners. Particular emphasis is put on performance portability across different architectures, with a view to the future of High Performance Computing, and

European processor initiatives like EuPEX, EPI, EuPILOT.

ChEESE-2P covers the Earth Science disciplines: seismology, tsunami study, magnetohydrodynamics, physical volcanology, geodynamics and glaciology. Among the different methodologies used in the project to achieve its purposes they are used seismic tomography, artificial intelligence, numerical simulation of the phenomena studied, and uncertainty quantification.

“The continuation of the ChEESE initiative will further enhance the capabilities of the European Solid Earth community, enabling scientists to enable services related to urgent calculation, early warning, risk assessment, which can be used by European civil protections” explains Arnau Folch, Professor of the CSIC and coordinator of the project. The impact of the initiative has allowed a winning synergy between research in the geosciences and that of High Performance Computing giving life, among other things, to further European projects such as DT-GEO (Digital Twin for GEOphysical Extremes), which is creating a “digital twin” for natural hazards in the context of the European initiative Destination Earth.

The ChEESE-2P project involves European academic and industrial partners at an international level such as BullAtos, Barcelona Supercomputing Center, Center National de la Recherche Scientifique, CINECA, CSC - Tieteen Tietotekniikan Keskus Oy, National Institute of Geophysics and Volcanology (INGV), Institut de Physique du Globe de Paris, Johannes Gutenberg-Universität Mainz, Ludwig-Maximilians-Universität Munich, Ruder Boskovic Institute, Sorbonne Université, Stiftelsen Norges Geotekniske Institutt, Technical University of Munich, Universidad de Málaga, Universität Stuttgart and Vedurstofa Islands.

NATIONAL RECOVERY AND RESILIENCE PLAN

National Research Center for HPC, Big data and Quantum Computing (ICSC)

Maurizio Ortali, Massimiliano Guarasi, Mirko Cestari
Cineca

The High-Performance Computing, Big Data e Quantum Computing Research Centre, created and managed by the ICSC Foundation, is one of the five National Centres established by the National Recovery and Resilience Plan (NRRP), covering designated strategic sectors for the development of the country: i.e., simulations, computing, and high-performance data analysis, Agritech, development of gene therapy and drugs via RNA technology, sustainable mobility, biodiversity.

The National Center provides a fundamental opportunity for the national system at the scientific, industrial and economic level to face present and future scientific and social challenges, strengthening and expanding existing skills and infrastructural resources. The Center is structured according to the hub and spoke model: the Hub is responsible for the validation and management of the research program, whose activities are elaborated and implemented by the “Spokes” and their affiliated institutions, as well as through open tenders. The Hub is also starting to implement all activities in education and training, entrepreneurship, knowledge transfer, policy and awareness raising.

The Hub and the Spokes are made up of universities, research bodies, private and public operators. The National Center includes a transversal Spoke (Spoke 0 “Supercomputing Cloud Infrastructure”) and 10 thematic spokes.

The National Center has two main objectives:

- 1) create a national computing / computing infrastructure, similar to a Datalake, grouping the existing High Performance Computing (HPC), High Throughput Computing (HTC), Big Data and network infrastructures, with new targeted resources, activated through internal funding, in order to provide the scientific and industrial communities with a flexible and uniform Cloud interface;
- 2) create an attractive ecosystem around the infrastructure that supports the academic world and the industrial system, favoring the exploitation of IT resources and the development of new advanced 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 months 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 GARR. Major activities for the starting period have covered preliminary allocation resource models and related functioning within the National Center, preparatory activities to the acquisition of the HPC assets needed by the Center 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)

The Italian National Biodiversity Future Center (NBFC)

Antonio Costantini, Gabriella Scipione
Cineca

The National Biodiversity Future Center (NBFC) aims at addressing 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.

The NBFC has a Hub-Spoke structure, involving 48 affiliates among Universities, Research Centers and Private Companies. CINECA is directly involved as a partner of Spoke 7 and will ensure the link with the National Center for High Performance Computing, as platform for storage and data management of NBFC. CINECA's leading role in the NBFC is to collaborate with partners of all other Spokes to the creation of four digitalization platforms. These platforms will make available and regularly updated the huge amount of data that the NBFC will generate. These data will come from the coordination among all the eight Nodes, as well as the existing and newly formed collaborations at national and at international level.

These four platforms are dubbed Biodiversity Collection, Molecular Biodiversity, Bioresources and Ecosystem Function (BEF) and Monitoring.

Biodiversity Collection platform target the development and implementation of digitization procedures together with the enhancement of Italian scattered naturalistic collections; one of the main goals of Molecular Biodiversity platform is the database rationalization of reference specimens and already produced DNA-base data, collection of reference specimens (plants and animals) and genotyping of new and already known species endemic to Italy; Bioresources platform aims to develop a database of biological source form national biodiversity, a national database for process and bioprocess to valorize biomasses and biological sources, a database of bioactive molecular and similar activities; BEF platform is more devoted to analysis: testing plant species and varieties to be used to restore damaged areas, data monitoring near real-time for operational services in selected areas, operationalization of models for national level based on climatic and remote sensing data, future projections, as well as other national data products.

The interoperability of these platforms and their connection with well-known International Services used by the Biodiversity Scientific Community are challenging opportunities that CINECA is willing to accept.

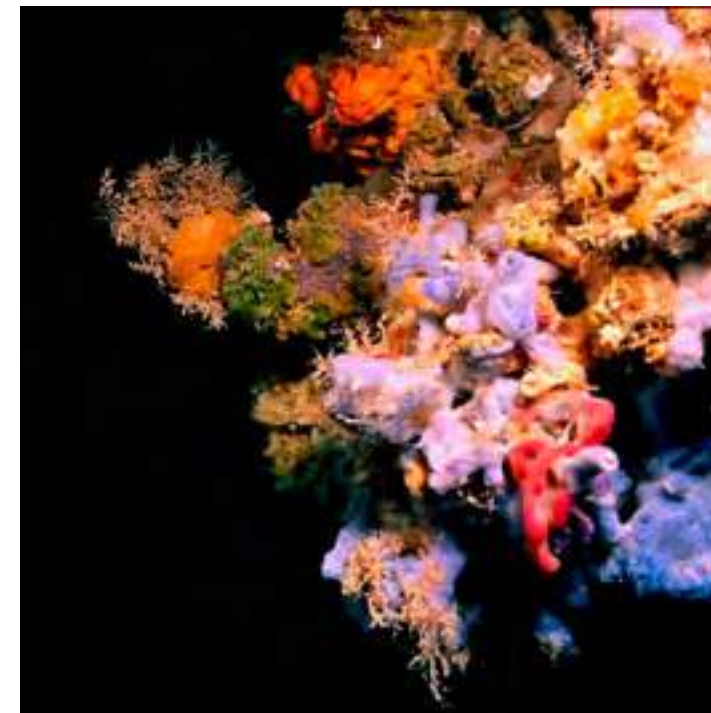


Figure 37: marine ecosystem.
Figure 38: musks and lichens.
Figure 39: vertical wood.



ECOSISTER – Ecosystem for sustainable transition of Emilia-Romagna

Maurizio Ortali, Fabio Affinito
Cineca

The regional “Innovation Ecosystems” are networks of state and non-state universities, Research publics, local public bodies and other highly qualified public and private entities aimed to intervene in areas of technological specialization consistent with the industrial and research vocations of the reference territory, promoting and strengthening the collaboration between the research system, the production system and local institutions.

To support his strong and highly competitive production system, coordinated by University of Bologna and Art-ER, in 2022 some Universities and Public Research Bodies of Emilia-Romagna decided to build an integrated innovative ecosystem, in association with Universities, Research Centers and other territorial players.

In order to maintain a leadership role in the international context and remain anchored to the vocations that characterize the Emilia-Romagna region, the project intends to support the ecological transition of the regional economic and social system through a process that transversally involves all sectors, technologies and skills by combining digital transition and sustainability with the work and well-being of people and the protection of the environment, in line with the objectives of the Pact for Work and Climate, and integrating with regional, national and European programs.

CINECA is affiliated to the University of Parma and is collaborating with the Spoke leader for the setup phase of the Spoke6 activities. In particular CINECA performed a requirement analysis among the partners, in order to define which High Performance Computing resources will best suit the technical needs of the project.



GRINS – Extended Partnership

Roberta Turra, Giorgio Pedrazzi
Cineca

The GRINS (Growing Resilient, INclusive and Sustainable) Consortium is an extended partnership coordinated by the University of Bologna - Department of Economics. Its goal is to develop AMELIA (dAta platforM for the transfER of knowLedge and statistical Analysis), an Online Data Platform that provides access to high-quality data and analytical tools for a wide range of applications.

GRINS AMELIA aims to support fundamental and applied research for businesses, households, and policy analysis for public administrations. It will generate knowledge and disseminate it to private and public stakeholders, as well as the national research system, in line with the principles of the Italian National Research Plan (PNR) and the broader EU-NRRP objectives of fostering resilient, inclusive, and sustainable growth.

In the 36 months of the lifespan of the project, Cineca will participate in the activities of Spoke 0, with the aim to support the development of the ICT and HPC component needed to the project, by aggregating its innovative skills in High Performance Computing, Data Science and Technology.

During the initial stages of the Spoke 0 activities, Cineca has provided access to its cloud infrastructure to facilitate the deployment of the AMELIA platform that will be integrated with Cineca's HPC resources during the project. Furthermore, Cineca is actively collaborating in the selection of statistical and Artificial Intelligence software to be included on the platform. Additionally, Cineca is contributing to the Spoke 0 activities, which focus on addressing data processing challenges while strictly adhering to privacy regulations.

D3 4-Health

Elisa Rossi
Cineca

Officially started on Dec. 1, 2022 with the kick off meeting on March 2023, D3 4 Health project is 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 at enabling new technologies for data collection and analysis in order to provide personalized medicine.

The hub of the project is the D3 4 Health Foundation, coordinated by La Sapienza University of Rome and composed of 28 partners.

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.

The approach involves the prevention, the diagnosis and therapy of oncological, neuropsychiatric and metabolic diseases through the creation of a "digital" and a "biological twin" of patients.

Digital Strategy and Platform for Cultural Heritage

Cinzia Zannoni, Antonella Guidazzoli
Cineca

The Central Institute for the Digitisation of Cultural Heritage - Digital Library ICDP, established within the Ministry of Culture (MIC), is in charge of coordinating and promoting the digitization programs of cultural heritage under the jurisdiction of the Ministry of Culture.

It aims to increase, organize and integrate the immense digital heritage produced over the years by archives, libraries, museums, and cultural sites in general, to offer citizens, cultural operators, universities, research institutes, schools, and economic operators new experiences of exploitation, and at the same time to involve cultural operators and innovative start-ups in the development of new digital services.

This significant digitization effort requires combining the cultural dimension with the managerial extent to guide a transformation of the systems for exploiting cultural heritage from a perspective based on "products" to one based on the development of "services", based on interoperability between systems that will act as a means of transversal sharing between different sectors (museums, archives, libraries, cultural activities).

Within this strategic framework, the Institute is called upon to identify a series of structural actions that will be developed over the years based on a long-term vision, guided by the National Digitisation Plan, where cultural heritage is the main asset for generating a knowledge-based economy and building innovation in service design.

Investment 1.1 is planned as part of M1C3, entitled "Digital Strategy and Platforms for Cultural Heritage", proposed by the Ministry for the implementation of the National Strategy for the Digitisation of Cultural Heritage, to create a digital cultural ecosystem based on a coordinated and interdependent set of digital infrastructures and platforms, natively cloud-based, for the creation and management of services for the production, collection, preservation, distribution, and use of digital cultural resources.

Cineca is involved as an executing body to contribute with infrastructure and skills to the realization of the Cultural Heritage Software Infrastructure, investment 1.1.4.

Cineca's role and activities:

- Support the gathering of platform requirements
- Supporting the creation of the national ecosystem and the implementation of research and development activities
- Creation of infrastructure environments and provision of related services to support the development and testing process
- Design of pre-production infrastructure and HPC infrastructure to support research in the field of digital cultural heritage.

NEW TECHNOLOGIES AND HARDWARE

Cineca HPC Department Unveils New Tech Wall Project

Antonella Guidazzoli
Cineca

The Tech Wall Project is a new initiative by the HPC department to effectively communicate the complexity and importance of HPC projects to society, research, and industry. As part of the project, a large projective wall by Sony has been installed at the entrance of the HPC department. The wall features very high resolution, enabling visualization and perceptibility of various types of developed projects through rendering and advanced graphics tools.

The Sony wall is equipped with Crystal LED displays, which offer super high contrast and brightness images with either a 1.2mm- or 1.5mm-pixel pitch. Its outstanding contrast ratio of over 1,000,000:1 is achieved through each Micro LED occupying a small fraction of the mounting area.

In addition to the large projective wall, the Tech Wall Project also incorporates 3D visualization capabilities and audio equipment to enhance the overall experience. Additionally, it will support real-time applications such as the Digital Twin Leonardo supercomputer at Bologna Technopole.

The Tech Wall Project is a significant step forward in the HPC department's efforts to communicate the importance of HPC to a wider audience. The project's innovative use of technology will enable the department to effectively visualize and explain the complex concepts of HPC to a variety of stakeholders and audiences.





Figure 40: the Tech Wall setup. The Isabella d'Este Virtual Studiolo shows its bright colours even in daylight.

Figure 41: magnetic fields in astrophysics. Scientific visualisation enhanced with the large screen.

Figures 42,43: two views from the documentary on Quantum Computing as seen on the Tech Wall.

Figures 44,45: thanks to techwall, Isabella d'Este Virtual Studiolo can be perceived almost at a 1:1 ratio.





ASTROPHISYCS



General relativistic simulations of massive black hole binary mergers in magnetized environment

Federico Cattorini, Bruno Giacomazzo

Department of Physics, University of Milano-Bicocca

According to current theoretical models of galaxy formation and evolution, massive black hole binaries (MBHBs) should form frequently and may be common in the center of galaxies. Coalescing MBHBs release tremendous amount of energy in the form of gravitational radiation and will pose key sources for future space-based gravitational-wave detectors such as the Laser Interferometer Space Antenna (LISA). Furthermore, these systems are expected to emit bright electromagnetic signals due to the interaction of the orbiting massive black holes with the gas-rich environment in their vicinity. The concurrent (multimessenger) detection of both gravitational and electromagnetic radiation will provide unparalleled advances for exploring fundamental physics and testing cosmological models.

One major challenge in observational endeavors searching for these systems is the scarcity of firm predictions for the electromagnetic emission arising before, during, and after merger. Without accurate predictions, it will be challenging to know what to look for and where to search in the electromagnetic spectrum. Limited knowledge of the unique spectral and timing features of such signals would frustrate future multimessenger efforts. Thus, next-generation multimessenger astronomy relies on detailed theoretical understanding of the mechanisms that may give rise to electromagnetic signatures of MBHBs, aiming to determine distinctive features that will help distinguishing them from other events.

To address this issue and produce sharp predictions, numerical relativity plays a key role. Numerical relativity provides computational methods suitable to solve the general-relativistic magnetohydrodynamic (GRMHD) equations that govern the evolution of plasma and magnetic fields in strong-gravity with the aid of parallel codes and HPC clusters, allowing to model the behavior of these sources in great detail. Our group employs numerical

relativity techniques to explore these intriguing systems. We produced several three-dimensional GRMHD simulations of MBHBs embedded in magnetized clouds of matter, modeling the final stages of binary inspiral and merger using the publicly available Einstein Toolkit framework (www.einsteintoolkit.org). All our calculations were run on the Cineca cluster Marconi A3 through a CINECA-INFN agreement. Our typical simulation required computational resources distributed on hundreds of cores and weeks of runtime. These calculations provide theoretical predictions of both gravitational and electromagnetic radiation emitted by a massive binary system and help shedding light on the behavior of matter and magnetic fields in such an extreme regime.

A major challenge in numerical simulations of MBHBs lies in the need to cover a broad parameter space, encompassing a wide range of physical properties such as black hole spins, mass ratios, orbital separations, and magnetic field strengths. Our research stands out for its exploration of this parameter space, enabling us to capture a wide range of binary configurations. By varying these parameters, we have obtained precious data that elucidates the behavior of accretion flows in different scenarios. In particular, our simulations were the first to consider merging binaries of spinning black holes immersed in a gas cloud. We focused on the effects that spin exerts on the magnetized gas, finding that there may be a tight connection between the spins of binary massive black holes and the timing features of the associated electromagnetic signal, which can exhibit a “chirp” analogous to the gravitational one.

The finding might have far-reaching implications. Linking a particular electromagnetic signal to a binary spin configuration can provide a new diagnostic to probe black hole spins observationally. Also, it can provide a more targeted approach for electromagnetic observatories, improving their chances of detecting and characterizing the electromagnetic

According to current theoretical models of galaxy formation and evolution, massive black hole binaries (MBHBs) should form frequently and may be common in the center of galaxies. Coalescing MBHBs release tremendous amount of energy in the form of gravitational radiation and will pose key sources for future space-based gravitational-wave detectors such as the Laser Interferometer Space Antenna (LISA). Furthermore, these

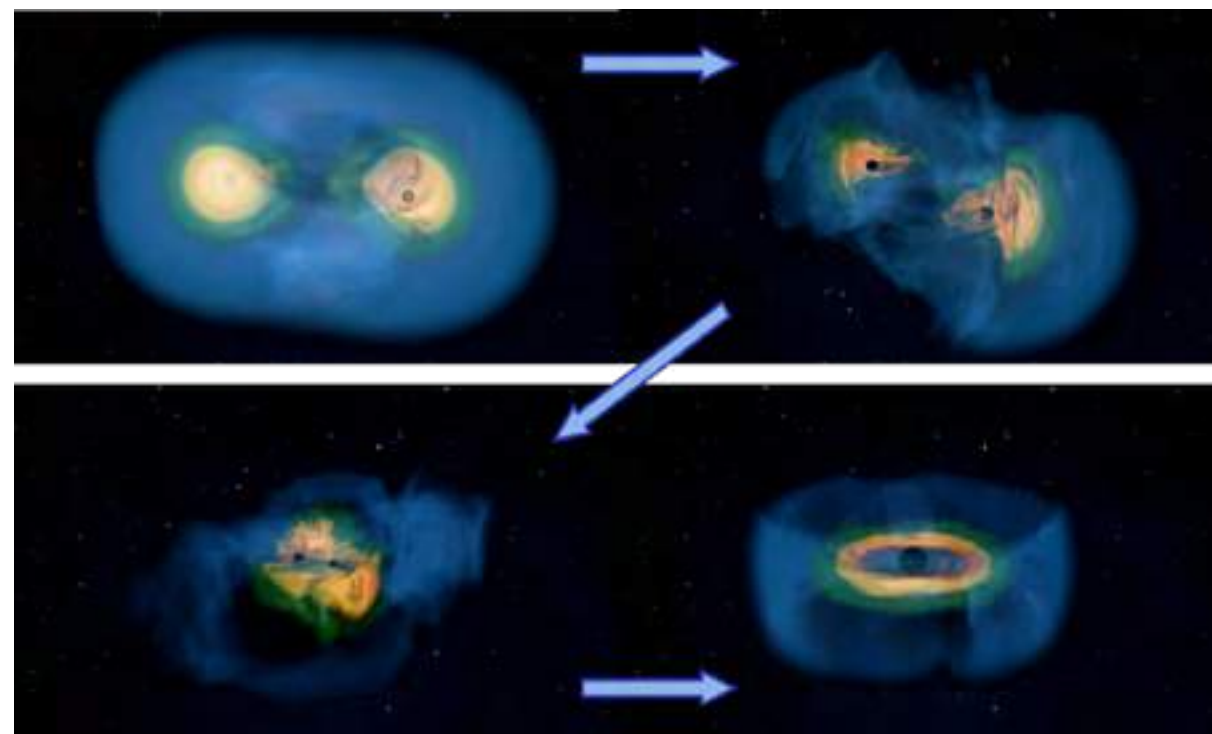


Figure 46: the figure displays three-dimensional snapshots of the rest-mass density distribution of matter around a massive black hole binary. Across the binary orbital evolution, the magnetized gas is twisted and stirred as it is accreted onto the black holes. After the coalescence, it settles down to a disk-like structure around the massive black hole remnant.

Modeling Cassiopeia A: The full fledged remnant of a neutrino-driven core-collapse supernova

Salvatore Orlando
INAF

Massive stars explode in core-collapse supernovae (SNe), forming supernova remnants (SNRs) with complex structures and intricate distribution of stellar debris. Young SNRs (<5000 years old) provide insights into the inner processes of SN explosions, including nucleosynthetic yields and asymmetries. Additional features stem from the progenitor star's internal structure at collapse and the interactions between the remnant and the circumstellar medium (CSM), which is shaped by the mass-loss history of the progenitor. Hence, understanding the connection between young SNRs, parent SNe, and progenitor stars is crucial for studying SN physics and massive star evolution. In recent years, advanced 3D magnetohydrodynamic models and the availability of adequate computing resources have improved our ability to investigate this connection. Cassiopeia A (Cas A), a well-studied SNR in the Milky Way, is an ideal target for studying the structure and composition of stellar material ejected during a SN event. Observations suggest that Cas A's asymmetries are largely from the early stages of the explosion, making it an important laboratory for studying the processes at work during the complex phases of the SN and the nature of progenitor stars.

Recently, we started a project with the goal of investigating the extent to which Cas A keeps memory of the asymmetries that develop in the neutrino-heating layer soon after the core-collapse as a result of hydrodynamic instabilities, such as convective overturn and the standing accretion shock instability (SASI). Additionally we investigated whether a prior interaction between Cas A and a massive, asymmetric shell of the CSM could explain the features observed in the remnant that cannot be explained as the result of phenomena occurring soon after the SN explosion.

We used a 3D hydrodynamic model coupled with 3D MHD simulations to study the

formation of Cas A. Our simulations spanned approximately 2000 years, incorporating all relevant physical processes involved in the evolution of the SN and the interaction between ejected debris, stellar wind, and a progenitor's circumstellar shell. The simulations were conducted using the PLUTO code, which is designed for astrophysical applications and high Mach number flows. We received computational resources from CINECA on Marconi through IS CRA Award N.HP10BARP6Y and the "Accordo Quadro INAF-CINECA" framework.

We discovered that the asymmetries and ejecta features in Cas A can be explained by the interaction between the reverse shock and initial large-scale asymmetries resulting from early stochastic processes during the SN explosion. Small-scale structures within large-scale iron-rich plumes, formed during the initial stages of the SN, combined with subsequent hydrodynamic instabilities, result in ring- and crown-like structures within the shocked ejecta. The mature remnant displays voids and cavities within the innermost unshocked ejecta, connected to ring-like features of shocked ejecta in the primary shell. These features primarily arise from the expansion of iron-rich plumes and the decay of radioactive elements. Additionally, the interaction between the remnant and the circumstellar shell reveals that observed asymmetries in the reverse shock can be produced if the shell has an asymmetric structure, with the densest region located on the (blueshifted) nearside toward the northwest. We propose that this shell likely resulted from a massive eruption of the progenitor star occurring between 10,000 and 100,000 years prior to core-collapse, with an estimated total mass of around 2 solar masses.

References. Orlando, S., Wongwathanarat, A., Janka, H.-T., Miceli, M., Nagataki, S., Ono, M., Bocchino, F., Vink, J., Milisavljevic, D., Patnaude, D. J., Peres, G., (2022), Evidence for past interaction with an asymmetric circumstellar shell in the young SNR Cassiopeia A. *Astronomy & Astrophysics*, vol. 666, A2.

(<https://www.aanda.org/articles/aa/pdf/2022/10/aa43258-22.pdf>)

Orlando, S., Miceli, M., Bocchino, F., Peres, G., (2022), Connecting supernova remnants to their parent supernovae. *Mem. Soc. Astron. Italiana*, vol. 93, 182

(<https://memsait.oa-roma.inaf.it/wp-content/uploads/2022/11/2022MmSAI..2-3..182O.pdf>)

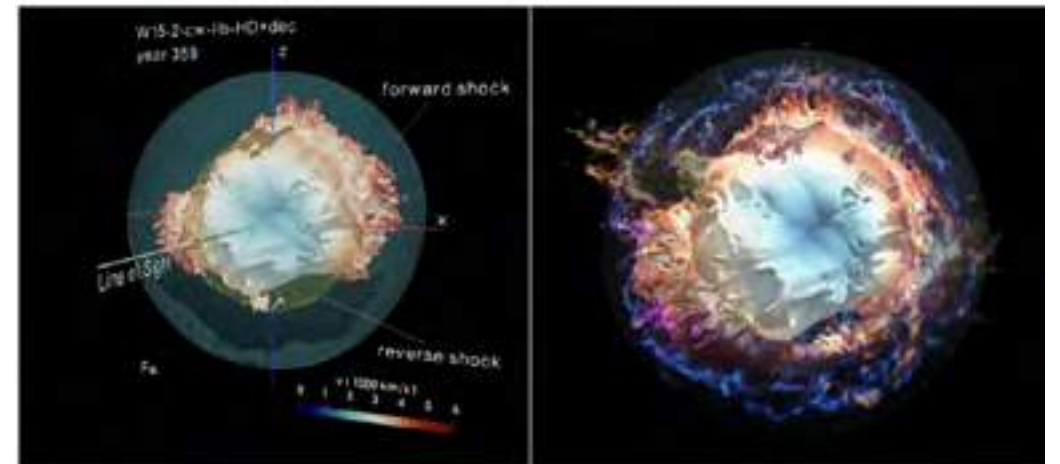


Figure 47. the iron distribution within the remnant of a neutrino-driven SN is depicted based on the W15-2-cw-11b-HD+dec model (adapted from Orlando et al. 2022, Mem. S.A.It. Vol. 93, 182). The left panel showcases a colored isosurface representing the distribution of iron following 359 years of evolution, which is nearly equivalent to the age of the Cas A. The opaque irregular isosurface indicates a value of iron density at 5% of the peak density, while the colors correspond to the radial ejecta velocity in units of 1000 km/s on the isosurface (color coding defined at the panel's bottom). Additionally, semi-transparent quasi-spherical surfaces denote the forward shock (green) and reverse shock (yellow). The Earth's vantage point is positioned along the negative y-axis. The right panel presents a comparison between the modeled distribution of iron-rich ejecta, represented similarly to the left panel, and actual observations of Cas A. The transparent image passing through the center of the remnant comprises a composite of optical observations (Hubble) and X-ray observations (Chandra) of the remnant. For an interactive 3D experience of this model, please visit the following links: <https://skfb.ly/6TKRK> and <https://skfb.ly/o8r7V>.

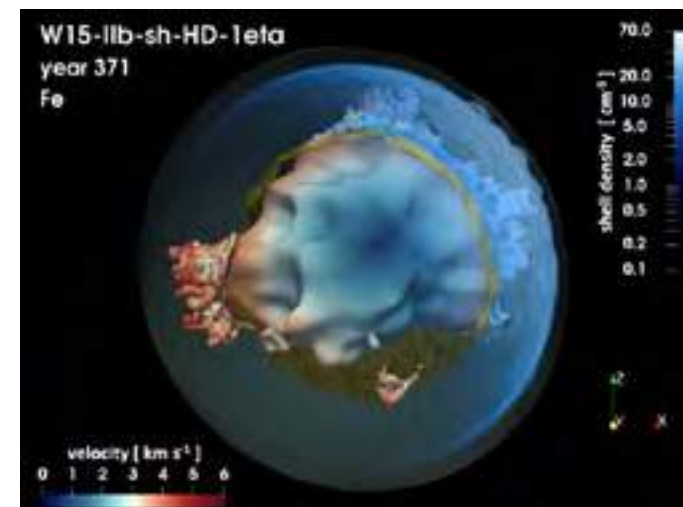


Figure 48. isosurface of the distribution of iron (corresponding to a value of iron density which is at 5% of the peak density) at the age of Cas A for model W15-11b-sh-HD-1eta (adapted from Orlando et al. 2022, A&A 666, A2); the colors give the radial velocity in units of 1000 km/s on the isosurface (color coding defined at the bottom of the panel). The semi-transparent clipped quasi-spherical surfaces indicate the forward (green) and reverse (yellow) shocks. The shocked shell is visualized through a volume rendering that uses the blue color palette (color coding on the right of the panel); the opacity is proportional to the plasma density. A navigable 3D graphic of this model is available at <https://skfb.ly/o8FnO>.

Bringing HPC-fueled innovation to the EU industry

Claudio Arlandini
Cineca

In 2022, the interest of the industries for Cineca as a trusted innovation partner saw a steadfast increase, with new companies seeking our support in introducing HPC-enabled innovation. Our focus is on Italian SMEs that face a significant technology gap compared to their EU counterparts. By partnering with Cineca, these SMEs can reap a host of benefits.

Firstly, Cineca actively participates in an expanding network of innovation actors. This network includes European Digital Innovation Hubs, Competence Centres for Industry 4.0, and Industrial Associations. By being part of this network, SMEs working with Cineca gain access to a diverse range of resources, collaborations, and opportunities. This interconnected ecosystem facilitates knowledge sharing, fosters synergies, and provides SMEs with a broader platform for exposure and growth. Training is a key factor of this activity. Thanks to the EuroCC project, in 2021/2022 we offered 12 courses in 7 Italian cities, involving 380 participants from 130 companies.

Moreover, Cineca offers the opportunity of “test-before-invest.” This service allows SMEs to experiment with new digital technologies, such as software, hardware, and business models. SMEs can test these innovations in a controlled environment without the need for upfront investments. This significantly reduces the financial

risks associated with introducing new technologies. Furthermore, SMEs can leverage the expertise of technicians, who are co-funded by the EU or other funding agencies. This collaboration further enhances the success rate of innovation projects, ensuring that SMEs receive the necessary support to make informed decisions. In 2022, we estimate that Cineca support was key to obtain over 2M€ of funding to Italian SMEs in cascading grants. Working with Cineca also provides SMEs with access to state-of-the-art infrastructures. This includes advanced high-performance computing (HPC), artificial intelligence (AI), and big data technologies. SMEs can harness these powerful tools to drive their innovation initiatives, enhance their competitiveness, and accelerate their growth. The availability of such infrastructures is a significant advantage that empowers SMEs to explore new possibilities and push the boundaries of their industry.

Industries might access our resources and competences through value services or initiatives (co)funded by the EU or national funding agencies. Among our notable new customers of 2022, we provided computational resources for 2 million core hours to Risk Engineering and Design srl for updating catastrophe risk models for heavy rain events.

Eni's supercomputing power for the decarbonization path

ENI S.p.A:

For over 20 years, Cineca has been collaborating with Eni (the most important Italian energy company) in the set up of the company's HPC ecosystem, implementing the most advanced algorithms developed by Eni researchers, developing applications and managing the company HPC infrastructure. From 2007 until 2012, Cineca hosted Eni supercomputers in its own datacentre. In 2013, Eni set up its own HPC infrastructure in its Green Data Centre (GDC) and Cineca continues to help manage Eni's HPC infrastructure in GDC.

The collaboration with Eni allowed Cineca staff to face the problems and the typical needs of industrial production, to grow in the understanding of the specific domain and in the ability to work in highly specialized and effective multidisciplinary teams.

HPC4 and HPC5 computational power installed at the Eni Green Data Center (GDC) combined with a new generation reservoir

simulation platform, allow to concurrently run thousands of reservoir simulations in extremely short times, reducing the time-to-market of key-asset projects.

Thanks to continuous improvements, Eni's mainframe supports the Company in its path towards energy transition through the definition, modeling and optimization of Carbon Capture and Storage. Moreover, the high-speed computational power of Eni supercomputers is used to improve technologies related to renewable energies. The Green Data Center also contributes to R&D activities, leading to the acceleration of technological innovation.

It has in fact provided a cluster of more than 1,000 cores to supply the computing power necessary to solve the increasing request for Computational Fluid Dynamics simulations in a timeframe compatible with industrial development needs. Computational Fluid Dynamics is used to solve fluid flow

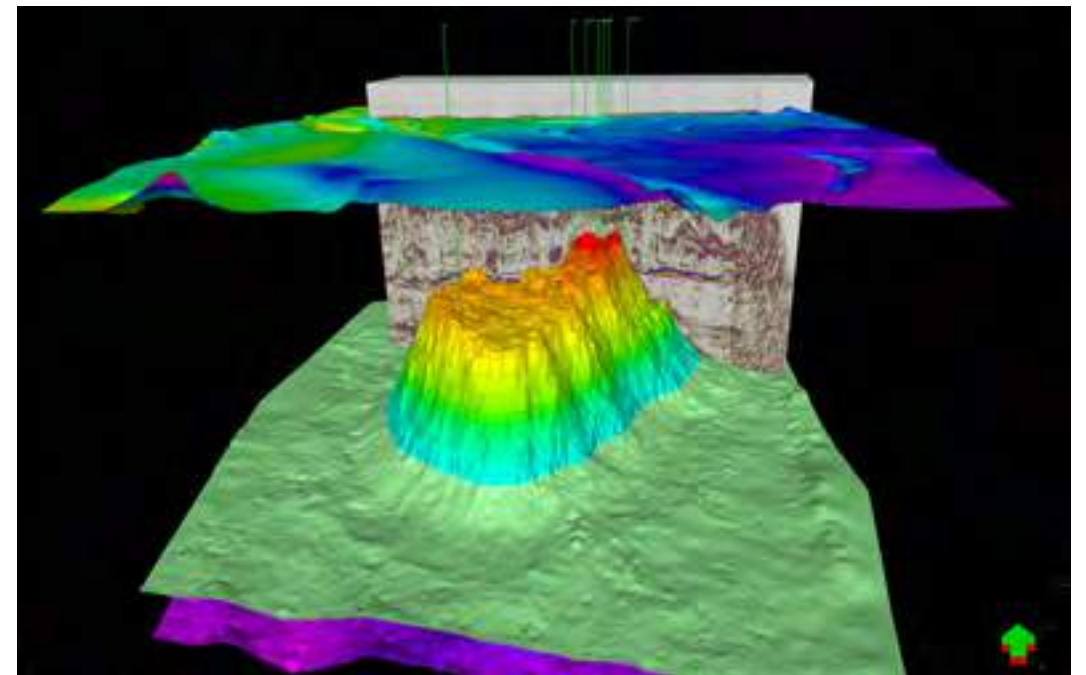


Figure 49: Zohr reservoir – 3D image of the Top Carbonate

problems in complex geometry, allowing the development of a virtual prototype to improve equipment performances.

Eni-GDC runs bespoke programs dedicated to magnetic confinement fusion: the development of superconducting magnets and the study of plasmas in the Tokamak reactor. The use of Eni HPC is not confined to applications in energy production and decarbonization goals: for example, during the Coronavirus pandemic it was also made available into a consortium dedicated to new drugs design and will be fundamental to develop new research programs on artificial intelligence.

Dompé and Cineca: Extreme Scale Drug Discovery

Federico Ficarelli
Cineca

In-silico, high-performance computing (HPC) simulations play a crucial role in the discovery and development of anti-viral drugs and have revolutionized the field of drug discovery by significantly reducing costs, time, and resources compared to traditional experimental approaches. The long-standing collaboration between Dompé Farmaceutici and Cineca have seen our institutions work for the past 20 years in pursuing technological and scientific innovations in this field, with several breakthroughs also in times of global pandemics [1] [2]. The collaboration between Cineca and Dompé is ready to take a step further: the recently signed deal not only strengthens their partnership but also paves the way for the establishment of a cutting-edge joint research center in Napoli and marks a milestone in the field of scientific research and drug development. Additionally, the inclusion of a dedicated HPC data center in the same location further enhances their capabilities and potential. The prospect of this collaboration brings forth a world of possibilities and promises significant advancements in the discovery and development of new drug discovery solutions. By combining the expertise of Cineca in HPC and Dompé's deep understanding of the pharmaceutical industry, the joint research center will become a hub for innovation and scientific excellence. The new site in Napoli will serve as a collaborative platform where scientists, researchers, and experts can work together, leveraging state-of-the-art computational resources and techniques. This convergence of knowledge and technology will enable the exploration of complex scientific challenges and expedite the development of novel drugs. Moreover, the presence of a dedicated data center further amplifies the potential of this collaboration: with access to vast computational power and storage capabilities, researchers will be able to conduct large-scale simulations, process massive amounts of data and, more importantly, extract knowledge from it. This infrastructure empowers them model drug-target interactions and explore new areas in the compound's chemical space. Moreover, the interdisciplinary nature of the collaboration will require scientists from diverse backgrounds from both institutions to work together, bridging the gap between computational sciences and pharmaceutical research.

The workhorse of our joint innovation efforts is the Ligen framework, an Exascale-ready, comprehensive simulation software that sees its birth almost two decades ago by the joint work by Dompé scientists and Cineca HPC specialists. During the years it has been proven

as a valuable tool for both pharmaceutical research and as a technology driver given its proven scaling capabilities. Cineca recently set up a dedicated, multi-disciplinary team that includes computational chemists, physicists and engineers to drive the future developments of the platform towards the open-ended future HPC has ahead in terms of technology and architectures. Valuable contributions are also coming from several EU-funded projects that chose Ligen as their main application driver, from Ligate [3], aiming at integrating AI methods and large scale cloud infrastructures, to EUPEX [4], one of the European pre-exascale pilots aiming at kickstarting an ecosystem of true European technological sovereignty in HPC and beyond. Another goal for the next future is providing scientists at Dompé with a comprehensive big data analysis platform, empowering researchers to process and analyze vast amounts of data across different hardware platforms. This platform is going to enable efficient exploration of the expanding chemical space, accelerating discovery, and optimizing drug design processes. By leveraging high-performance computing capabilities and diverse data analysis techniques, researchers can uncover insights, predict compound properties, and identify promising drug candidates more effectively.

The collaboration between Cineca and Dompé, along with the establishment of a joint research center and a dedicated data center in Napoli, holds immense promise for the advancement of scientific research and the discovery of innovative drugs.

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

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

3 <https://www.ligateproject.eu>

4 <https://eupex.eu>



Figures 50, 51: the event “EuroCC: dove l’industria incontra il supercalcolo”.



EUROCC-Italy:

Empowering Industries and Public Administration through High-Performance Computing, Artificial Intelligence, and Big Data Analysis

Eric Pascolo
Cineca

EuroCC-Italy, a competence centre within the EuroCC 2 network consisting of 33 centers across Europe, is an ambitious project funded by the European joint undertaking EuroHPC. With renewed funding until 2025, EuroCC-Italy has achieved remarkable success, solidifying its position as a frontrunner in the fields of high-performance computing, artificial intelligence, and big data analysis.

At the core of EuroCC-Italy’s success are its partners, three institutional entities: CINECA, the Italian supercomputing center; BI-REX, a center for industry 4.0 initiatives; and IFAB, a foundation dedicated to financing projects that bridge the gap between technology and human development. Additionally, EuroCC-Italy has forged partnerships with two companies that have long utilized HPC: Leonardo SpA and Dompé Farmaceutici.

The overarching goal of EuroCC-Italy is to facilitate the transfer of competencies and technologies in the domains of supercomputing, AI, and big data analysis to industries and the Public Administration. By leveraging the expertise of its partners, EuroCC-Italy is fostering a collaborative ecosystem that empowers organizations to unlock the full potential of these transformative technologies.

The three pillars driving EuroCC-Italy’s activities are focused on:

- driving innovation,
- upskilling the workforce,
- fostering collaboration with academia and other EUROCC centre

Through innovation projects, EUROCC-Italy encourages organizations to explore and implement groundbreaking solutions that harness the power of high-performance computing, AI, and big data analytics.

This promotes the development of novel approaches, unleashing economic growth and technological advancement throughout Italy.

Recognizing the importance of upskilling, EUROCC-Italy is committed to equipping professionals with the necessary knowledge and competencies to navigate the complexities of these advanced technologies. Through training programs, workshops, and educational initiatives, EUROCC-Italy is narrowing the skills gap and empowering individuals to harness the potential of high-performance computing, AI, and big data analysis in their respective fields.

Collaboration lies at the heart of EUROCC-Italy’s mission. By forging strong partnerships with academia and the other 32 EUROCC centres, EUROCC-Italy is fostering an environment where knowledge exchange and collective efforts can thrive. This collaborative approach stimulates innovation, promotes best practices, and ultimately leads to groundbreaking advancements in high-performance computing, AI, and big data analysis.

With renewed funding and its expertise, EUROCC-Italy is ready to help industries and public administration with digital transformation. The project’s partners, CINECA, BI-REX, IFAB, Leonardo SpA, and Dompé Farmaceutici, bring together a wealth of expertise and experience, ensuring that EUROCC-Italy remains at the forefront of technological innovation. Together, over the next 3 years, we will continue to drive innovation and support companies and PA in adopting enabling technologies such as HPC, AI, and big data analytics, engaging more players through innovation projects and upskilling pathways.



This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under Grant Agreement 951732



Future Plans

Leonardo Earl Access LEAP

Paola Alberigo
Cineca

Leonardo Earl Access (LEAP) In November 2022 the EuroHPC JU, the Italian Ministry of Research and CINECA, as hosting entity, launched the Leonardo early access program, to grant computing time during the pre-production phase.

The call aimed to offer selected users' early access to the Booster module of EuroHPC Leonardo exascale precursor system. This partition module is expected to provide a computational performance over 240 Pflops.

The Leonardo Early Access Program (LEAP) call was open to all fields of science, industry, and the public sector. This call targets projects expected to have a high scientific impact by leveraging extremely large computing resource availability.

The Leonardo Early Access Program call has closed on December 4th, with more than 80 proposals submitted for a total of 85 million GPU hours of calculation requested, corresponding to more than 70% annual accelerated computing capacity.

The projects have come from 11 Countries and involve more than 400 researchers in various scientific fields, including Weather/Climate, Artificial Intelligence, Solid Earth, Computational fluid dynamics and Bioinformatics.

The proposals went through a technical and scientific selection process in early 2023 and in April 2023 the first ranked proposals started to run on Leonardo machine.

The thirteen selected proposals were in many different scientific fields such as: Computational Fluid Dynamics, Fundamental Physics, Chemistry, Astrophysics, Geophysics and Machine Learning.

We are very interested in the forthcoming results of their research.

Evolution of the computing infrastructure and next trends

Daniele Cesarini, Mirko Cestari
Cineca

In the forthcoming years, Cineca will continue to adopt its three-pillar strategy to extend the capabilities and performance of its computing and storage infrastructure. In doing so, Cineca can now leverage the availability of two new data centres: the new Big Data Technopole data centre currently hosting Leonardo, and a new data centre in Naples in the renovated area of San Giovanni a Teduccio. The geo-distribution of the infrastructure allows to improve the resiliency of the compute and storage services, providing the capability to host critical workloads. The evolution of the infrastructure will benefit from multiple projects involving action like the recently founded national centre ICSC, established Italian research institutions like CNR, INAF, OGS, INGV, and national agencies such as ACN (National Cybersecurity Authority) and AIM (Agenzia Italia Meteo). All of them are HPC stakeholders, relevant at national and international level, bringing significant expertise in the field and providing new requirements for the national compute and data infrastructure. The infrastructure evolution investments amount to more than 125 million euros.

Tier-0 HPC pillar.

. Cineca will upgrade Leonardo system with the Lisa (from Monna Lisa) project. Lisa aims to complement the HPC services of Leonardo by adding new computing partition featuring state-of-the-art processing unit solutions. With Lisa It is envisioned to support workloads extremely benefitting from high memory bandwidth (> 1TB/s) and from new generation of GPUs that tightly couple host and accelerator device (such the AMD Mi300-a or the NVIDIA Grace-Hopper CPU-GPU solution). In close succession, the new Quantum Computing infrastructure provided by the EuroQCS-Italy project will be installed in the Big Data Tecnopole and interfaced to Leonardo, to provide a single system able to execute simulations based on conventional and quantum algorithms or their combination. This will make Leonardo one of the first hybrid HPC/QC (Quantum Computing) systems in the world. The quantum infrastructure, a digital/analog system with qubits technology based on neutral atoms, is part of a larger EuroHPC JU initiative that entails the installation of six different European-manufactured NISQ quantum computers in selected supercomputing centres. Each computer is chosen to be unique in terms of qubit technology and operational mode.

Tier-1 HPC pillar.

Cineca will procure and provide multiple Tier-1 class systems for specific workloads and scientific communities. Belong to this class, the new system dedicated to the EUROfusion community named PITAGORA, that will replace MARCONI-FUSION approaching to its end-of-life. To be noted, that while being a Tier-1 system if compared to LEONARDO's performance, PITAGORA with its 70 PFlops peak, would currently rank 12th in the last top500.org list, and it is expected to make it well into the top 30 once installed. Moreover, in a new data hall beside LEONARDO, it will be deployed a Tier-1 system focused on material design, astrophysics and weather forecast. Design requirements for the specific use cases are progressing, and the procurement is expected to be launched by the end of 2023. A similar trajectory is following the Tier-1 system that will be hosted in Naples data centre. A technical working group is analysing the provided use cases and oversees drafting the system requirements to address very diverse workloads: from pharmaceutical exploration to machine learning, from fundamental research to cybersecurity analysis.

Tier-1 Cloud and data services pillar.

Cineca plans to procure a new infrastructure aiming at upgrading the cloud and data services already available. This project will bring a significant advancement in the diversification of computational resources made available to researchers. The goal for Cineca is to provide users with a capable and multi-protocol data repository, over 200,000 virtual processors and modular compute and data services to help the user communities in creating dedicated compute platforms. Data management services will be completely re-designed, including backup, archiving, data replication and so on. Therefore, the virtualization and data services will be significantly enhanced, through a research cloud infrastructure competing at European level, and able to complement the computational services provided by the main supercomputing systems (Tier-0 and Tier-1 HPC systems). The infrastructure procurement is underway, targeting the end of the year for its conclusion. The new services will be made available in the second half of 2024.

National Cybersecurity Authority:

The realm of cyberspace has progressively become the sphere in which individuals' essential rights to access information, express themselves, and connect with others are realized. It is also where the pursuit of transparent public policies and the drive for efficient public administrations take place, enabling growth, innovation, and the reshaping of our world as we know it. Every day, billions of connections are established across geographical boundaries, facilitating the exchange of knowledge and ushering in unprecedented transformations at a rapid pace. The interconnected nature of networks, the inherent imbalance of cyber threats, and the pervasive presence of cyberspace in our daily lives all necessitate a comprehensive approach and collaborative endeavors from all relevant stakeholders. Only through such collective actions can we guarantee an appropriate level of security in the digital realm.

For this reason, the National Cybersecurity Agency (ACN) was established by Italian Decree Law No. 82 of June 14, 2021, which redefined the national cybersecurity architecture, aiming to streamline and simplify the existing system of national competences while further enhancing cybersecurity and cyber resilience aspects, including for the purpose of safeguarding national security in cyberspace. ACN is responsible for safeguarding national interests in the field of cybersecurity and its mission is to protect security and resilience in cyberspace. It focuses on preventing and mitigating cyber-attacks while promoting the achievement of technological autonomy. In 2023, the ACN will join forces with the CINECA consortium to initiate a collaboration aimed at developing a dedicated supercomputer to handle the cybersecurity workload of the agency. This supercomputer will be installed in Cineca's datacentre located in Naples.



EuroHPC
Joint Undertaking