



**Finanziato
dall'Unione europea**
NextGenerationEU



D1.2.4.3

CAPACITY PLANNING

TeRABIT

THIRD RELEASE
DECEMBER 2025

Mauro Campanella (INFN-GARR)
Silvia Calegari, Giacinto Donvito, Claudio Grandi (INFN)
Massimo Carboni, WP leader; Claudia Battista, Direttrice (GARR)
Mirko Cestari, Giuseppa Muscianisi, Debora Testi (CINECA)
Giorgio Bolzon, Stefano Salon (OGS)

Eccetto dove diversamente specificato, il contenuto di questo sito è
concesso in licenza con Attribuzione Creative Commons - 4.0 internazionale.



Executive Summary

The TeRABIT project has the objective to further develop the capacity of each Research Infrastructure involved and create a strong synergy between the three RIs. The document assesses their status at the time of writing, and the achieved and expected capacity increase according to the results of the tenders, and the requirements to serve the respective users' communities. The capacities in ICT resources increased more than planned due to the positive results of the tenders. The modularity of the architecture computing system procured, the availability of fibres and of an open optical line system will permit to continue to scale without special architectural modifications according to users' requirements. The final section updates on the ongoing effort to engineer and develop common services. The capacity plan will not be further updated before the project ends.

Table of Content

1	<i>CURRENT DEVELOPMENTS, OFFERED CAPACITIES AND THEIR USE</i>	4
1.1.	GARR-T Network	4
1.1.1	Optical fibres	5
1.1.2	Packet equipment	6
1.2	PRACE-Italy	7
1.2.	HPC-BD-AI	9
2	<i>CAPACITY PLANS</i>	10
2.1	GARR-T	11
2.2	PRACE-Italy	12
2.3	HPC-BD-AI	14
3	<i>JOINT SERVICES PLANS</i>	15
4	<i>REFERENCES</i>	16

List of Figures

Figure 1: GARR-T Capacity evolution	4
Figure 2: GARR-T fibres in Sardinia: High level and detailed view	6
Figure 3 - Commitment for the HPC computing resource allocation on Galileo100 for 2023	7
Figure 4. Allocation breakdown for year 2024 by class of projects.	8
Figure 5 – ADA cloud virtual CPUs allocation in 2024	8
Figure 6. Cloud allocation breakdown by class of projects.	9
Figure 7 - The INFN HPC-BD-AI infrastructure	10
Figure 8 The upgraded GARR-T network to the end of 2025.	11
Figure 9: Planned GARR-T Capacity growth	12
Figure 10: the new GAIA system	13
Figure 11- The INFN Cloud-based HPC-BD-AI Continuum from Edge to Cloud to HPC	15

List of Tables

Table 1– WP2 optical fibre implementation status at end of August 2025	5
Table 2 – WP2 optical equipment implementation status at end of August 2025	5
Table 3 – WP2 packet equipment implementation status at end of August 2025	7
Table 4 - The projected computing capacity	13
Table 5: Nov 2025, Ranking of the CNAF HPC Bubble in the TOP500 list of the world's most powerful supercomputer	15

1 CURRENT DEVELOPMENTS, OFFERED CAPACITIES AND THEIR USE

The three Research infrastructures are constantly evolving. The specific developments by the project represent a substantial increase in capacity in selected geographical regions and central ICT computing. These developments by TeRABIT benefit by concurrent contribution to GARR-T and HPC-BD-AI by the ICSC project and to PRACE-Italy by ICSC and the MEET project, coordinate by INGV (National Institute for Geology and Vulcanology). Section 1 reports the main achievements and changes since the last version of the capacity plan.

All the upgrades of the infrastructures have been installed and are entering in production.

1.1. GARR-T Network

Since the last release of the capacity plan, the GARR-T network backbone capacity in the Italian continent has increased from 30 Tbps in 2024 to 38.70 Tbps in late 2025.

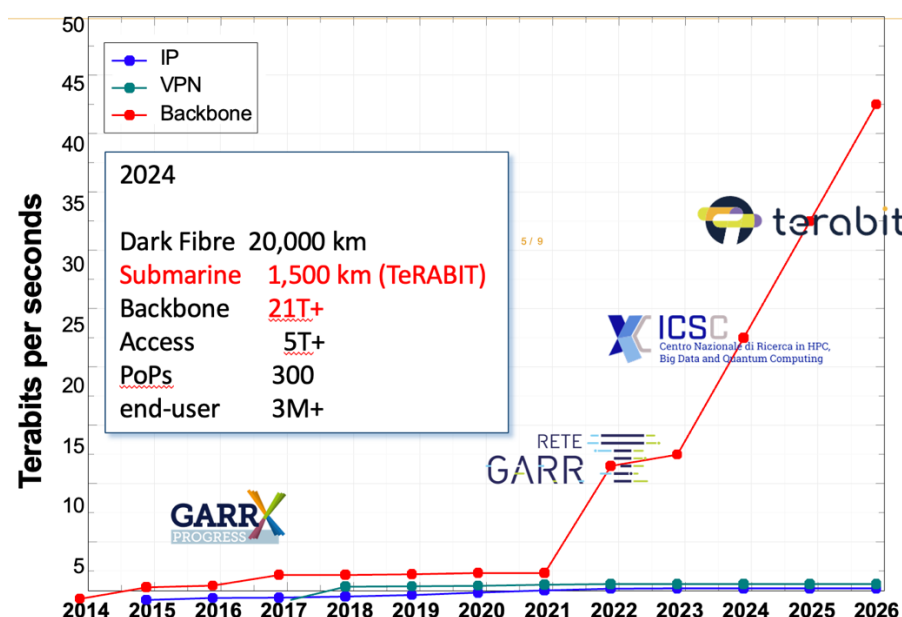


Figure 1: GARR-T Capacity evolution

The activity in Sardinia and Sicily completed installation all three components addressed by the tenders: optical components (fibres and active equipment) and packet network equipment.

- In Sardinia, the acquisition of optical fibers and submarine spectrum has enabled the creation of a robust regional network of nodes and amplification sites, fully interconnected with the mainland backbone via spectrum services over open submarine cables towards Rome and Milan. A key achievement of this expansion is that advanced connectivity has been successfully established at the Sos Enattos site, which has been shortlisted as a candidate to host the Einstein Telescope, the next-generation European gravitational wave observatory.
- In Southern Sicily, the deployment of state-of-the-art open line systems and coherent

DWDM technologies has successfully upgraded the optical transport infrastructure. High-capacity data center interconnection (DCI) services, ranging from 100 Gbps up to 400 Gbps, are now fully operational, providing robust connectivity to regional research and scientific facilities.

In parallel, the TeRABIT project has successfully enhanced the national Packet Network layer and implemented a unified overlay management system, based on automation and telemetry, enabling seamless monitoring, programmability, and orchestration of services across all architectural layers.

1.1.1 Optical fibres

The extension development of the infrastructure is reported in Table 1. All objectives have been reached and installation and entrance in production of all links and equipment will be achieved by end of 2025. The activity operated through the long-term acquisition (20 years IRU contracts) of terrestrial optical fibers in Sardinia and the interconnection of this terrestrial Infrastructure to GARR-T network in the mainland through a long-term acquisition (also 20 years IRU contracts) of optical spectrum over submarine open-cables up to Milan (900GHz) and up to Rome (900GHz). The extension of the GARR-T network in the Sardinia region consists of 6 nodes and 11 amplification sites, while for sites in the southern regions of Italy there are few fiber adaptation interventions (tender n. 2301 Lot 1) (see *TeRABIT Report Deliverable D2.1.1 GARR-T fiber network expansion and open line system (OLS)*).

KPI Project: TeRABIT Tender: 2301- Lot1	Total planned in Project	Aug 2025	Achieved %
Dark Fiber Span delivered	18	18	100%
Spectrum Over Subsea Cable (900GHz)	2	2	100%
Hardware Shipped (%)	100%	100%	100%
Delivered Nodes (stand-alone acceptance)	20	20	100%
PoP Nodes activated and onboarded	20	20	100%
PoP Nodes operational	20	20	100%

Table 1– WP2 optical fibre implementation status at end of August 2025

KPI Project: Terabit Tender: 2302 - Lot1	Total planned in Project	Aug 2025	Achieved %
Hardware Shipped (%)	100%	100%	100%
Delivered Nodes (stand-alone acceptance)	24	24	100%
PoP Nodes activated and on-boarded	24	24	100%
PoP Nodes operational	24	24	100%
Capacity activated (Gbps)	5500	5500	100%

Table 2 – WP2 optical equipment implementation status at end of August 2025

Upgrade of the active components of the optical network is reported in Table 2. It extends a state-of-the-art open line system in the south of Sicily and provides access bandwidth (Data Center Interconnection) from 100Gbps up to 400Gbps via coherent DWDM transmission systems (tender n. 2302 Lot 1) (see *TeRABIT Report Deliverable D2.2.1 GARR-T transport capacity (DCI) expansion and GARR-X Progress OLS upgrade*). All the hardware equipment (100%) has been acquired via the tenders n. 2301 Lot 1 and n. 2302 Lot 1 has been delivered and installed within the national territory.

This permits to increase both the IP backbone capacity bringing it in many cases to values equal to 100Gbps and 400Gbps and the access capacity of some users' sites which require high throughput such as INFN Tiers and other institutions. A notable example is the Einstein Telescope proposed site in Sos Enattos (see *TeRABIT Report Deliverable D2.3.1 expansion and enhancement of the GARR-T Packet Network - ICSC and TERABIT Projects*)

Figure 2 shows a high-level and detailed view of the deployment of fibres in Sardinia (green terrestrial and red submarine in the picture at the right).

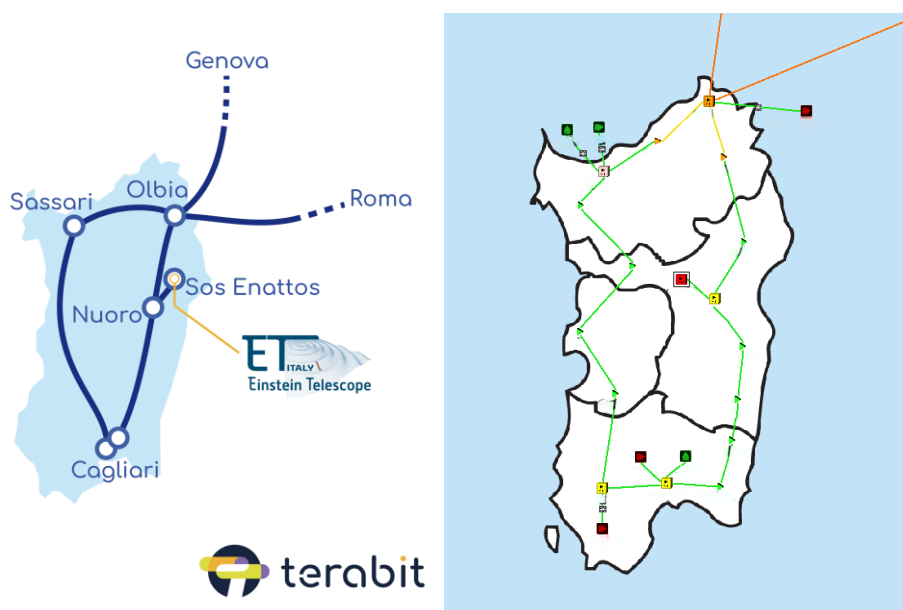


Figure 2: GARR-T fibres in Sardinia: High level and detailed view

1.1.2 Packet equipment

The status of the Packet Network implementation plan is shown in Table 3. All the hardware equipment acquired and ordered through the tender n. 2303 Lot 1 has been delivered and installed in the PoPs.

KPI Project: Terabit Tender: 2303 – Lotto1	Total planned in Project	IO2.3.3	Incremental from project start	%
Hardware Shipped (%)	100%	100%	100%	100%
PoP Nodes tested and operational	15	15	15	100%

Table 3 – WP2 packet equipment implementation status at end of August 2025

The creation of a uniform and consistent national packet network layer has been completed, through the expansion of the router platforms in operation (including new line cards and optical interfaces). The IP backbone capacity has been increased—reaching 100 Gbps and 400 Gbps per link in many cases—and higher-capacity access has been extended to user sites with demanding throughput requirements. (Reference: TeRABIT Report Deliverable D2.3.1 “Expansion and enhancement of the GARR-T Packet Network – ICSC and TeRABIT Projects”).

1.2 PRACE-Italy

The capacity of PRACE-Italy has been based in the Tier-1 HPC system Galileo100 (G100), hosted at CINECA and supporting scalable and interactive processing workflows with nodes equipped with cutting edge technology. G100 features a high-capacity storage system providing a data-centric approach to the entire infrastructure. As a reference, for 2024 it approaches 200 million core-hours. Figure 3 reports the plot of allocated resources for the year 2024. The plot shows that the system's availability significantly decreased in November 2024 due to a data centre incident¹ caused by severe flooding in the Bologna area. The allocation nearly reaches the system's maximum capacity but stays just below it, which helps the system respond more quickly to user requests.

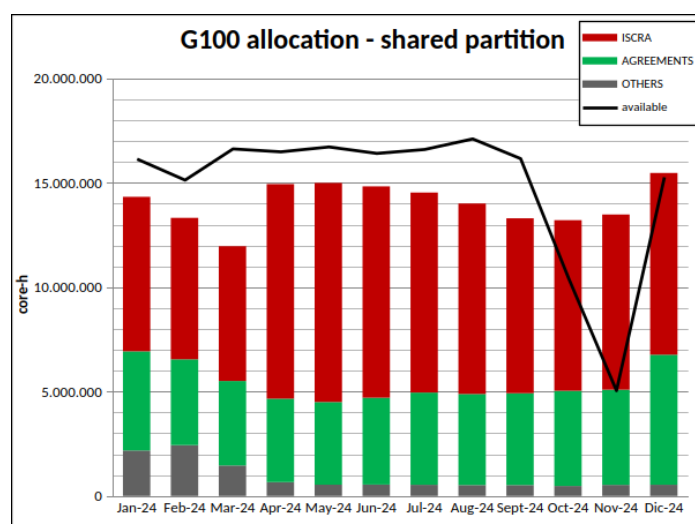


Figure 3 - Commitment for the HPC computing resource allocation on Galileo100 for 2023

¹ <https://www.hpc.cineca.it/newsletter/weather-emergency-cineca-update-access-to-leonardo-restored/>

The allocation breakdown by project class is shown in Figure 4. Nearly 60 million in resources were allocated to the national scientific allocation program ISCRA, which provides excellence-based access to the infrastructure.

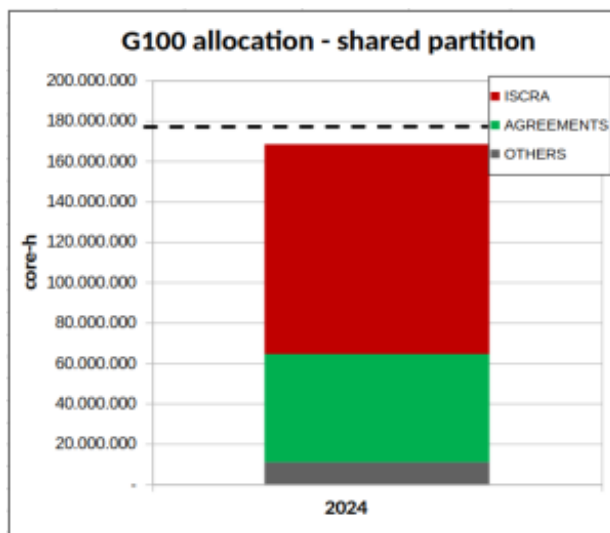


Figure 4. Allocation breakdown for year 2024 by class of projects.

The Cloud computing monthly capability (part of the PRACE-Italy Tier1 system) is reported in Figure 5. The infrastructure is capable to provide a maximum capacity of 7200 vCPUs.

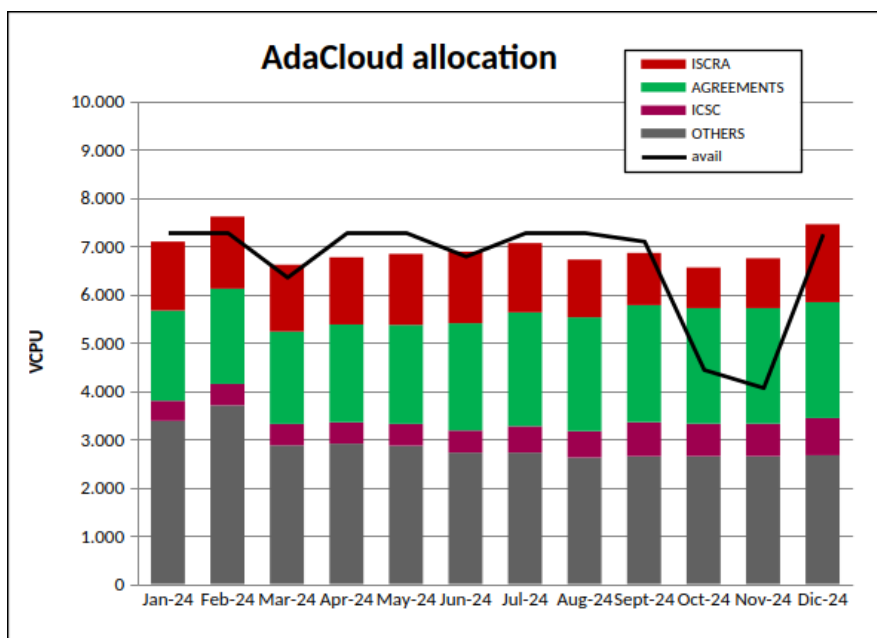


Figure 5 – ADA cloud virtual CPUs allocation in 2024

Allocation approaches full capacity of the cloud infrastructure. A similar drop in availability is shown due to the flooding incident reported above. Figure 6 reports the allocation breakdown by class of

projects. To be noted the allocation of 20% for ISCRA projects and 8% of ICSC projects.

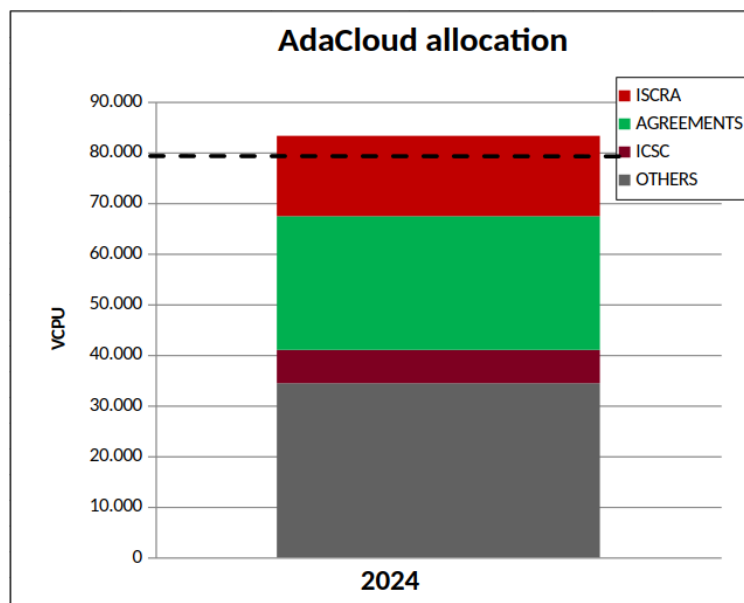


Figure 6. Cloud allocation breakdown by class of projects.

The GALILEO100 Tier-1 system integrates a cloud HPC infrastructure, which completes the CINECA HPC ecosystem providing a tightly integrated solution that covers both high-performance and high flexible computing.

1.2. HPC-BD-AI

The INFN distributed computing infrastructure includes a first level (called "Tier-1") computing centre in Bologna and 9 second level (called "Tier-2") computing centres in Turin, Milan, Padua/Legnaro, Pisa, Rome, Frascati, Naples, Bari, Catania. In addition, a new data centre at the Gran Sasso National Laboratories and a second centre at the Frascati National Laboratories have been implemented in the context of the ICSC National Centre and are now part of the distributed infrastructure. The current overall capacity is about 240,000 CPU cores, 230 GPUs, 240 PB of net disk storage and 250 PB of tape storage. The centres are connected to the GARR network, with capacities ranging from multiples of 10 Gbps to multiples of 100 Gbps the Tier-1 centre in Bologna tested a DCI connection with the CERN laboratories at 1.2 Tbps.

The HPC-BD-AI infrastructure is mainly used for data and compute processing by the INFN high energy, nuclear and astro-particle physics experiments, particularly the four collaborations working at the CERN LHC collider. However, the infrastructure is also serving several other activities related to the many collaborations coming from national and international projects to which INFN participates. These collaborations range from life science to medicine, from earth observations to smart cities and many more.

The software infrastructure is composed by both Grid and Cloud middleware.

The current topology is reported in Figure 7.

In the same figure, the HPC “HPC Bubbles systems” have been shown, as procured and installed by the project.



Figure 7 - The INFN HPC-BD-AI infrastructure

2 CAPACITY PLANS

The upgrades of the three RIs match and exceed the planned objectives of the project. The upgrades are in production, or in the final stage of testing and are providing services either in the form of connectivity or of computing.

The upgraded capacities are a substantial addition to previous capacity. The plans foresee, after the project formally ends, to investigate the best way to provide adequate resources to researchers. The requests for use of computing, in its various forms, from traditional HPC to resources adequate for AI workloads continue to grow. The recent evolution of Research Infrastructures devoted to computing represents a complementary and significant increase in the European offer of capacity to AI -based workloads. Recent European grants and initiatives aim at creating a landscape of large “AI factories” and a possible funding for private to public partnerships for even larger AI Gigafactories.

2.1 GARR-T

The GARR-T network performed a large evolutionary steps using grants provided by three projects: GARR-T itself, ICSC (National HPC Centre) and TeRABIT. Figure 8 shows the upgraded GARR-T network at the end of 2025.

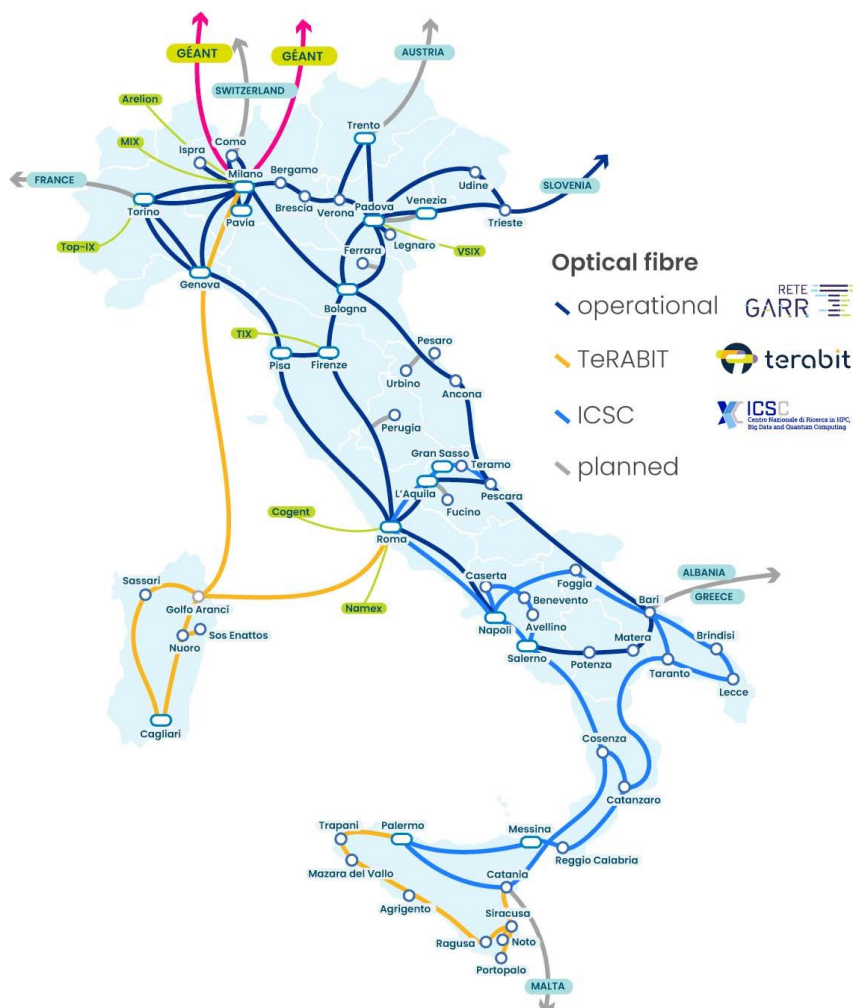


Figure 8 The upgraded GARR-T network to the end of 2025.

The upgrades provide greater reachability, lowering geographical digital divide, and faster capacities for the end user. The local loop upgrades are agreed with the users to ensure that the users' network is capable of benefitting of the larger capacity offered (usually an initial factor of two, and up to ten, is considered).



Figure 9: Planned GARR-T Capacity growth

The GARR-T network will seamlessly continue to upgrade its infrastructure according to requirements, like the international capacity towards GEANT and towards users or sites e.g. the Technopole in Bologna.

2.2 PRACE-Italy

WP3 conducted a survey for PRACE-Italy users to quantify their needs, specifically in terms of HPC / Cloud resources and usage. According to our preliminary analysis, we expect a request for an increase of computing capacity in cloud environments (e.g. for ensemble simulations aimed at the uncertainty reduction, or to reduce time-to-solution in urgent computing for damage assessment) and for urgent computing demand (e.g. reducing time-to-solution for damage assessment in emergency situation: from «slow but reliable» or «rapid but simplified» to «rapid and reliable»), an increased exploitation of AI / Machine Learning for data-model integration (e.g. in digital twin applications), a support for the porting of codes to hybrid CPU/GPU architectures. The awarded solution is expected to meet all these user requirements and integrate in the supercomputing Italian infrastructure.

TeRABIT has focused its upgrade on the flexibility of the cloud to better adapt to the diversity of user workloads, while still providing high-end computing power. The tender for new equipment has been organised as a joint tender between TeRABIT, the PNRR project MEET led by the National Institute of Geology and Vulcanology. The new system, named GAIA provides to PRACE Italy an increase in computing power of about 5 times the number of nodes and 2.5 times the storage. The RI therefore offers a very complete infrastructure (HPC and cloud) that allows users to gradually exploit higher computing capacity. If the need for HPC increases, or scales beyond the cloud HPC provision, other world-class HPC systems can be integrated into the workflow to cover all

computing needs. In this context the PRACE-Italy RI can provide a scale-up access to Tier-0 ICSC systems such as Leonardo.



Figure 10: the new GAIA system

A strong focus is devoted to energy efficiency targeting very efficient CPUs (e.g. the Intel Xeon 6780e with 144 cores per 330 W) and through the continuous monitoring of the computing infrastructure carbon footprint.

Table 4 shows the projected computing capacity, in terms of cores, based on CINECA and PRACE-Italy experience of Tier-1 HPC provisioning and user utilization. From the data reported, we can observe that at least a 3x increase in terms of capacity can be obtained every 4 years, thus projecting the number of cores to around 90000-100000 in 2024 (coherent with Intermediate Objective 3.2.5 at M24 and the investment planned in the budget).

This new capacity is in fact obtained with a very competitive procurement procedure. The awarded solution offers over 130 M core-hours, providing in reality a 5x increase of resources, rather than the expected 3x. The installation of such complex system has started in 2024 but general availability to users will be provided by the end of 2025.

Year	Tier-1 System	Capacity (n.of cores)	Capacity increase	Notes
2011	PLX	3288		
2015	Galileo	8256	2.5x	
2019	Galileo-refresh	36288	4.4x	
2021	Galileo100	30288	-	68 GPUs NVIDIA V100+ Cloud OpenStack + Fat nodes
2025	TeRABIT HPC system	131.200	4x	80 GPUs A30 + 112 GPU L40s + 8 H100, Cloud OpenStack, Fat nodes (>1 TB)

Table 4 - The projected computing capacity

2.3 HPC-BD-AI

In TeRABIT, the HPC-BD-AI has been expanded to offer cloud-native access to distributed resources with HPC capabilities. These resources have been made available to TeRABIT users both with local (batch) and with Cloud interfaces (we provided IaaS approach via OpenStack and PaaS via Kubernetes instances), expanding the INFN Cloud service portfolio.

The resources are available in multiple INFN sites. Technically, they consist of:

- Clusters with 192-core CPU servers, each configured with 1.5 TB RAM, all connected via 400 Gbps network links (InfiniBand): 160 CPU nodes for a total of about 30,000 cores.
- Clusters of mixed CPU and GPU servers. The nodes of these clusters have each 4 Nvidia H100, 80 GB RAM GPUs, and are connected via 400 Gbps network links (InfiniBand): 61 nodes with 4 GPUs each, for a total of 244 GPUs.
- Clusters of mixed CPU and FPGA servers. These nodes are capable of realizing also FPGA-mesh configurations, to improve the performance of multi-FPGA applications: 10 nodes with 4 FPGA each. We have both Xilinx U55C o Terasic P0701 type of FPGA in the servers.
- All the nodes mentioned above are provided with significant local fast disks (SSDs).
- Storage solutions provisioned via a variety of access interfaces (posix and S3, for example): 118 storage nodes each equipped with 64 cores, 1 TB RAM, 384 TB HDD and 25,6 TB NVMe, corresponding to 45 PB long term storage and 3 PB high performance storage (before redundancy).

In addition, over 30 PB net disk storage has been acquired at the CNAF Tier-1, with the contribution of also the TeRABIT project, to support data processing and preservation.

These HPC-capable clusters distributed over INFN Cloud have been called in TeRABIT "HPC Bubbles". The main technical features of these bubbles are that:

- They integrate and provide access to all the existing INFN Cloud services, with consistent federated authentication and authorization methods.
- They expand the INFN Cloud product portfolio to offer HPC-capable services ranging from bare metal to higher level (PaaS to SaaS) services.
- They can be federated with each other to create Virtual HPC Bubbles, in case capacity going beyond what's available in a single bubble is required.
- They can integrate with traditional HPC centres and infrastructures, such as PRACE-Italy, Leonardo, or in general with EuroHPC centres.

The overall framework in which the HPC Bubbles expand the HPC-BD-AI infrastructure is aimed at the integration of solutions and eventually to the realization of a continuum that covers requirements from the Edge to the Cloud, to the largest HPC Centres.

Of particular interest is also that the HPC bubbles are provisioned in Cloud HPC centres that are ISO 27001, 27017 and 27018 certified. This allows not only to support full edge-to-cloud-to-HPC integration in general, but also to expand it to cover the case where ingestion, storage and processing of sensitive data is needed, such as all the cases that require management of health-

related data.

An HPC bubble at CNAF has been tuned and run the benchmarks to rank in the Top500 HPC system in the world [<https://top500.org/>]. The result allowed the Bubble to rank n.437 in the list of the fastest HPC systems as of Nov 2025, see. The result the large capabilities in terms of raw computing power of a single, simply tuned and well equipped, HPC Bubble. The Bubble is rather compact and physically occupies about two and half standard racks.

Rank	System	Cores	Rmax (PFlops/s)	Rpeak (PFlop/s)
437	Terabit HPC Bubble - ThinkSystem SR675 V3, AMD EPYC 9654 96C 2.4GHz, NVIDIA H100 80GB HBM3, NVIDIA Infiniband NDR, AlmaLinux 9, Lenovo INFN-CNAF Italy	14,400	2.91	5.36

Table 5: Nov 2025, Ranking of the CNAF HPC Bubble in the TOP500 list of the world's most powerful supercomputer

Figure 11 renders this overall framework. It is important to note that all the HPC-BD-AI services are meant to be fully integrated also with other PNRR national initiatives, such as ICSC, the national centre for HPC, Big Data and Quantum Computing, and with international initiatives, such as EuroHPC, EOSC, and others.

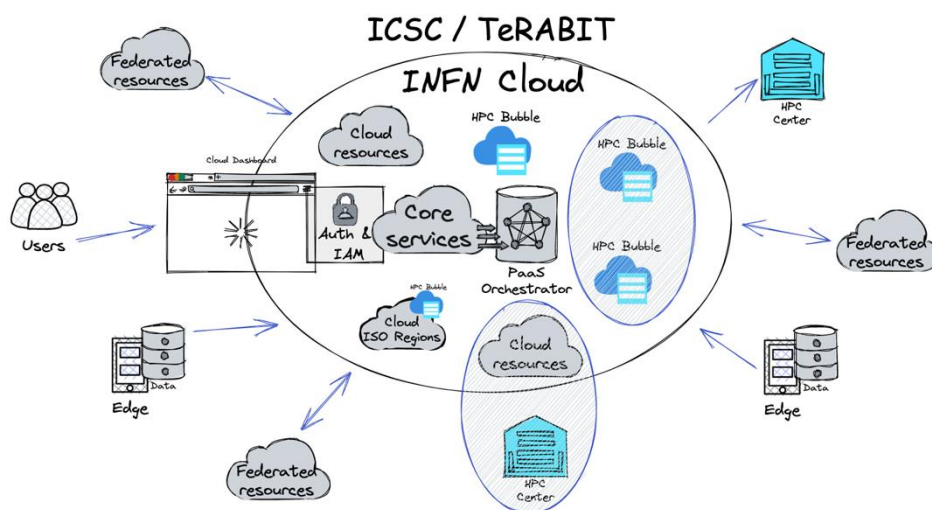


Figure 11- The INFN Cloud-based HPC-BD-AI Continuum from Edge to Cloud to HPC

3 JOINT SERVICES PLANS

At present the three infrastructures continue to be connected between them and with the rest of the world by the GARR-T network, which also allows high speed access to the infrastructures by the research community.

Each infrastructure provides a set of specific services to their users, which range from both large organizations, projects and single researcher.

The project uses the exiting services to facilitate the use of HPC, starting from the tabletop HPC

node, through the “HPC bubble”, to PRACE-Italy and possibly to the EuroHPC node Leonardo. The developments of joint services focused on the development of use cases and on an essential service to share key monitoring information.

The use cases are developed by WP5 with a set of 25 selected users as reported in D5.1.2 and D5.1.2.1 and D5.1.2.2. A full description is [1].

Among these use cases, the UrgentSHAKE case is demonstrating the capabilities of a RI federation to create a service for urgent assessment of the effect of an earthquake in North-eastern Italy. The case benefitted from the integration of the following services: fast network connectivity through GARR, HPC computing services (HPC bubbles, G100, GAIA, and Leonardo), and cloud services for workflow managing and process steering.

The specific use case is reported in D5.1.2.1 and D5.1.2.2. and in a scientific contribution [2] and paper [3].

The current developments continue to focus on sharing resources and the authentication and authorization infrastructure. See the combined deliverable of D4.3, D4.4 and D4.5 for further information.

4 REFERENCES

- [1] List of TeRABIT use cases <https://www.terabit-project.it/it/use-cases>
- [2] Zuccolo, E., Bolzon, G., Pitari, F., Rodríguez Muñoz, L., Monsalvo Franco, I. E., Scaini, C., Poggi, V., Smerzini, C., and Salon, S.: UrgentShake: An HPC System for Near Real-Time Physics-Based Ground Shaking Simulations to Support Emergency Response Efforts , EGU General Assembly 2025, Vienna, Austria, 27 Apr–2 May 2025, EGU25-9353, <https://doi.org/10.5194/egusphere-egu25-9353>, 2025.
<https://ricerca.ogs.it/handle/20.500.14083/42843>
- [3] Elisa Zuccolo, Giorgio Bolzon, Fabio Pitari, Lucía Rodríguez Muñoz, Chiara Scaini, Manuela Vanini, Valerio Poggi, Stefano Salon; Advancing Rapid Response to Earthquakes with Tiered Physics-Based Ground-Shaking Simulations: The UrgentShake System. *Seismological Research Letters* 2025;; 96 (5): 2995–3011. doi: <https://doi.org/10.1785/0220240472>