Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

terabit

# The TeRABIT project

## Terabit network for Research and Academic Big data in ITaly


Mauro Campanella Principal Investigator INFN/GARR,
Silvia Calegari Infrastructure Manager INFN


Online, 23 Sep 2024

Forum – EXPO PNRR - Borsa della Ricerca.

# AGENDA

- TeRABIT project at a glance ; Why, Who and What
- Research Infrastructures involved
- Project vision, goals and status
- The SHAKE use case example
- Challenges
- Training, collaboration
- Q&A

# Why TeRABIT: innovate and scale Infrastructures for Research

Data Centric in every discpine (STEM, humanistic, human activity) requiring

Data handling and analysis implying significant IT resources, with software and harwdare evolving very fast.

Ultra fast communication channels.
Up to Terabits per second (1 Tbps = 1.000 billions of bits per second)

MUR - General Directorate of **Internationalization and Communication**

Public notice n.3264 of 18.12.2021

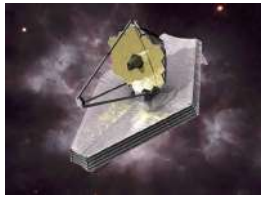
Mission 4, "**Education and Research**" -

Component 2, "**From research to business**

Investment line 3.1, "Fund for the creation of an **integrated system of research and innovation infrastructure**s",

**Action 3.1.1** Creation of new research infrastructures, **strengthening existing ones and their collaboration for scientific excellence in Horizon Europe**


apply research to production




# TeRABIT in a nutshell

| | |
|---|---|
| Applicant | : INFN (Istituto Nazionale di Fisica Nucleare) |
| Coapplicant | : OGS (Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS) |
| Other participants | : Consortium GARR, CINECA |
| Principal Investigator INFN | : Mauro Campanella (Milano) |
| Principal Investigator OGS | : Stefano Salon (Trieste) |
| Infrastructure Manager INFN | : Silvia Calegari (CNAF Bologna ) |
| Funding | : 41 ME (34.5 in assets, 4.1 personnel, 2.4 other) |
| Personnel hired with fixed-term contracts or grants | : INFN 25 persons + 1 infrastructure manager<br>: OGS 8 persons + 11 PhD + 4 master HPC |
| Unfunded effort | : 21 PM GARR, 25 PM CINECA |
| Start date | : 1 January 2023 |
| Duration | : 30 months (end date 30 June 2025)<br>extension requested to 31 Dec 2025 |

INFN Istituto Nazionale di Fisica Nucleare

OGS Istituto Nazionale di Oceanografia e di Geofisica Sperimentale

Consortium GARR

CINECA

# The project participants

INFN Istituto Nazionale di Fisica Nucleare

is the national public research body, supervised by the Ministry of University and Research (MUR), dedicated to the study of the fundamental constituents of matter and the laws that govern them. It carries out theoretical and experimental research in the fields of subnuclear, nuclear and astroparticle physics.

OGS Istituto Nazionale di Oceanografia e di Geofisica Sperimentale

is a public research body, supervised by MUR, which operates internationally in the field of physical, chemical, biological and geological oceanography, experimental and exploration geophysics, seismology and applied seismology to engineering.

Consortium GARR

is the very high capacity national network dedicated to the education, research and culture community. The GARR network is designed and managed by the GARR Consortium, a non-profit association founded under the aegis of the Ministry of University and Research by CNR, ENEA, INFN and the CRUI Foundation

CINECA

is a Consortium between Universities, amongst the largest computing centers in Italy and one of the most advanced in the world for High Performance Computing - HPC. It is a key supplier of solutions and services for universities, research centers, the Ministry of Education and Merit and the Ministry of University and Research.

# Exploit complementarity of three Research Infrastructures in operation recognized as Priorities by PNIR (2021-2027)

**status of the RIs in 2023**

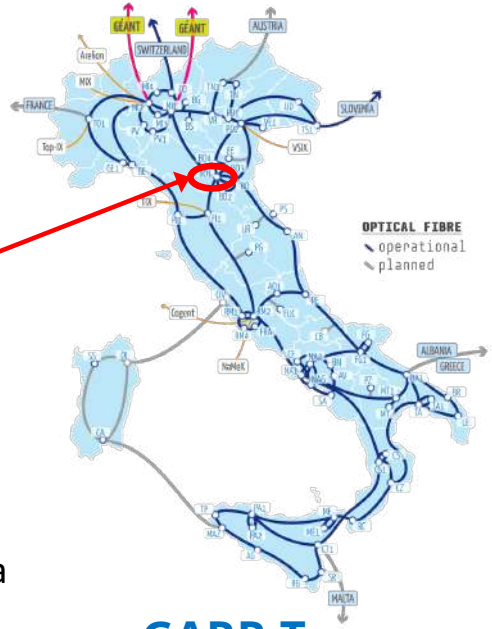
EuroHPC Joint Undertaking

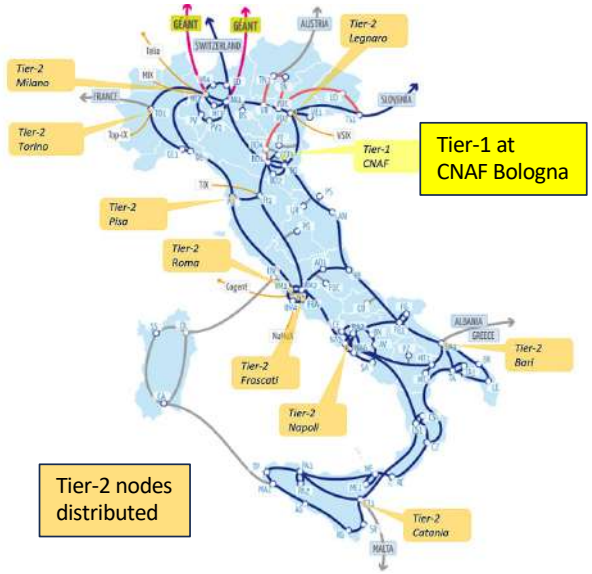
PRACE PARTNERSHIP FOR ADVANCED COMPUTING IN EUROPE

GÉANT

WLCG Worldwide LHC Computing Grid

Galileo100 - HPC, Hosted by CINECA - Bologna




**PRACE-Italy** **GARR-T** **HPC-BD-AI**

# Vision

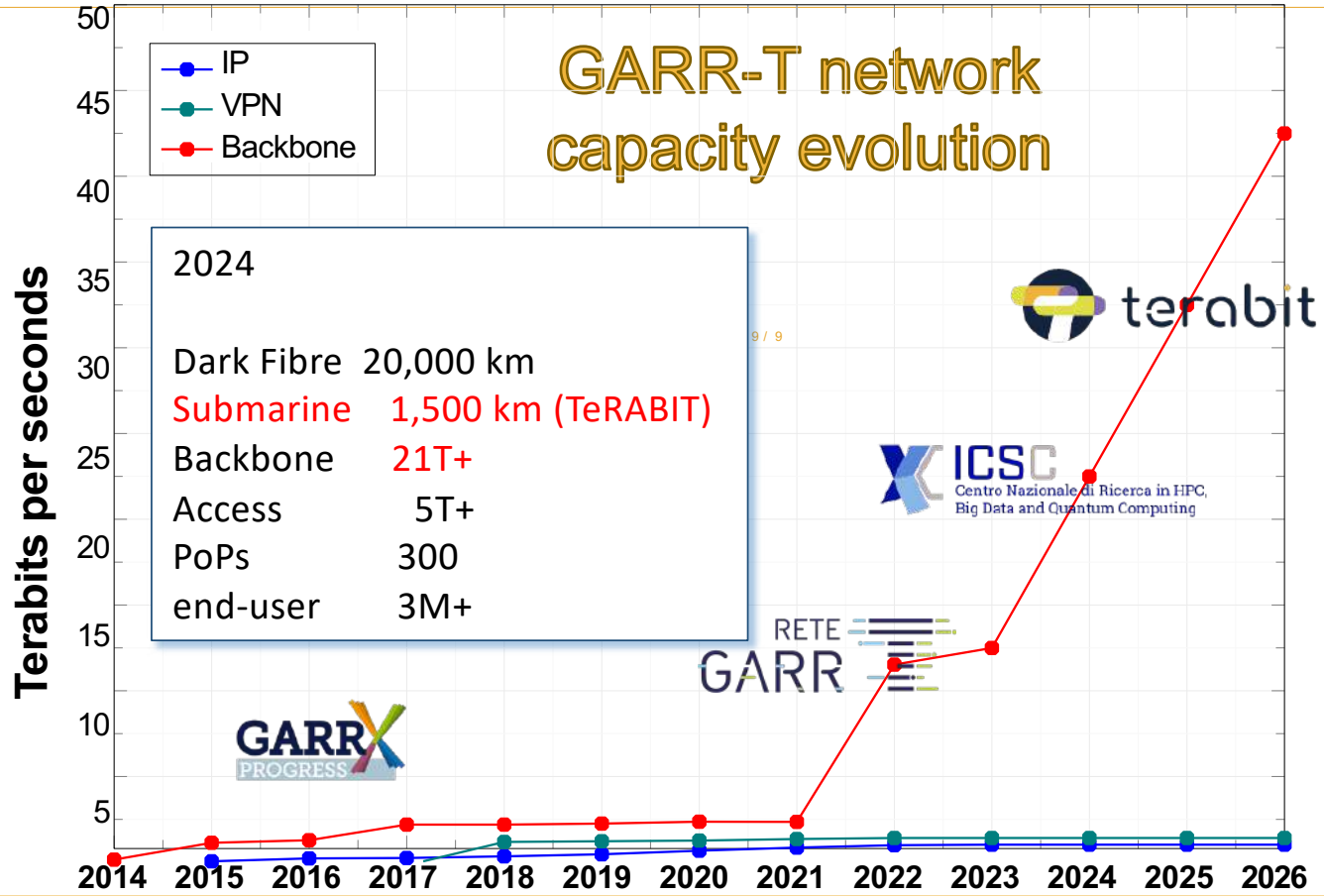
Create a distributed, hyper-connected, hybrid HPC-Cloud environment that offers services designed to meet the evolving needs of research and innovation.

The environment will federate and strengthen the three existing RIs GARR-T, PRACE-Italy and HPC-BD-AI (HPC-Big Data-Artificial Intelligence) and with ICSC national centre. Leverage their existing of connections to other national and European research infrastructures and data spaces through the GÉANT backbone.

# Main objectives

1. Enable widespread data transfer, up to Terabits per second, and services on a national scale in Italy, with particular focus on southern and island regions, all connected to Europe
2. Innovate the central HPC node of PRACE-Italy, maintaining the Tier-1 level.
3. Innovate the HPC services offered to researchers, beyond the centralized calculation model, adding distributed "HPC-Bubbles"

# GARR-T network capacity evolution




2024

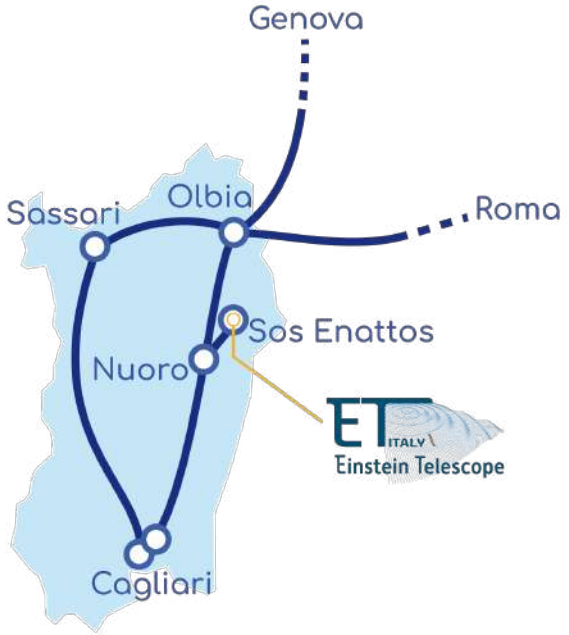
Dark Fibre 20,000 km
Submarine 1,500 km (TeRABIT)
Backbone 21T+
Access 5T+
PoPs 300
end-user 3M+

terabit

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

RETE GARR

GARR PROGRESS



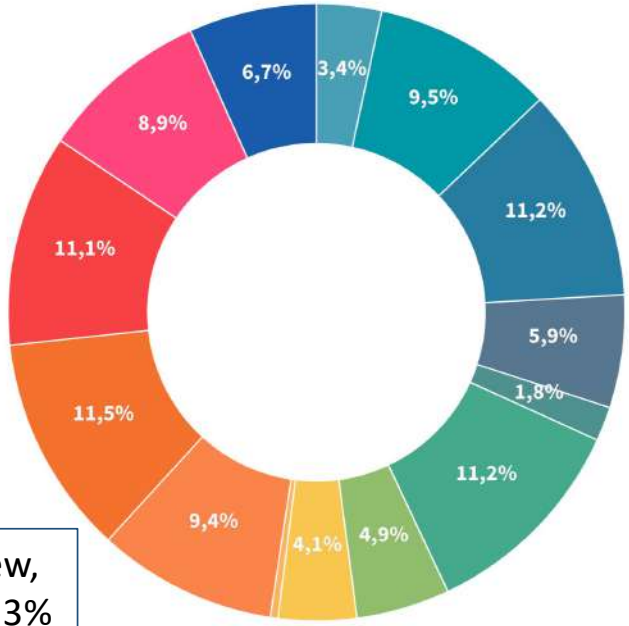

terabit

## Scientific domains

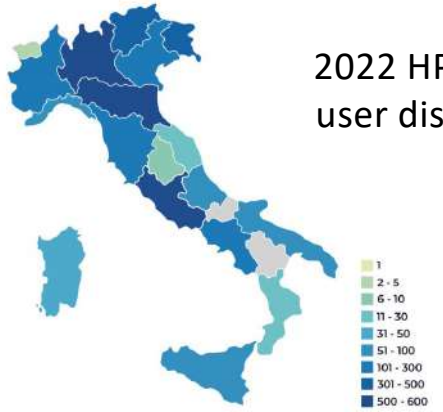
Scientists use Cineca computational resources within all scientific disciplines. The most represented three are Computational Chemistry, Condensed Matter Physics and Computational Fluid Dynamics, with about11% each, followed by Nuclear Fusion (10%), Computational Engineering, Astrophysics, and Plasma Physics with more than 9% each.

# Use and upgrade of PRACE-Italy

- Ai & Machine Learning
- Nuclear Fusion
- Computational Chemistry
- Life Science - Computational Biology
- Mathematics
- Condensed Matter
- Life Science - Bio Informatics
- Particle Physics
- Humanities & Social Sciences
- Astrophysics and Plasma
- Computational Fluid Dynamics
- Others
- Computational Engineering
- Earth & Climate Science




In 2022, after a strict peer review, the resource allocated were 109.3%




2022 HPC PRACE user distribution

| TeRABIT upgrade | NET INCREASE |
|---|---|
| Capacity (notes) | ~ x 5.0 |
| Storage | ~ x 2.5 |

# HPC-BD-AI evolution

HPC bubbles add to the existing distributed cloud infrastructure, as very compact, yet powerful, computing nodes with diverse HW

Nodes Type 1 : CPU (192 cores)
Type 2 : CPU + GPU (4x NVIDIA H100)
Type 3 : CPU + FPGA

Sites: CNAF, Bari, LNGS, Milano Bicocca, Napoli, Padova, Pisa, Roma 1, Torino

Additional Storage:
Mass storage : CNAF
High performance storage : CNAF, Bari

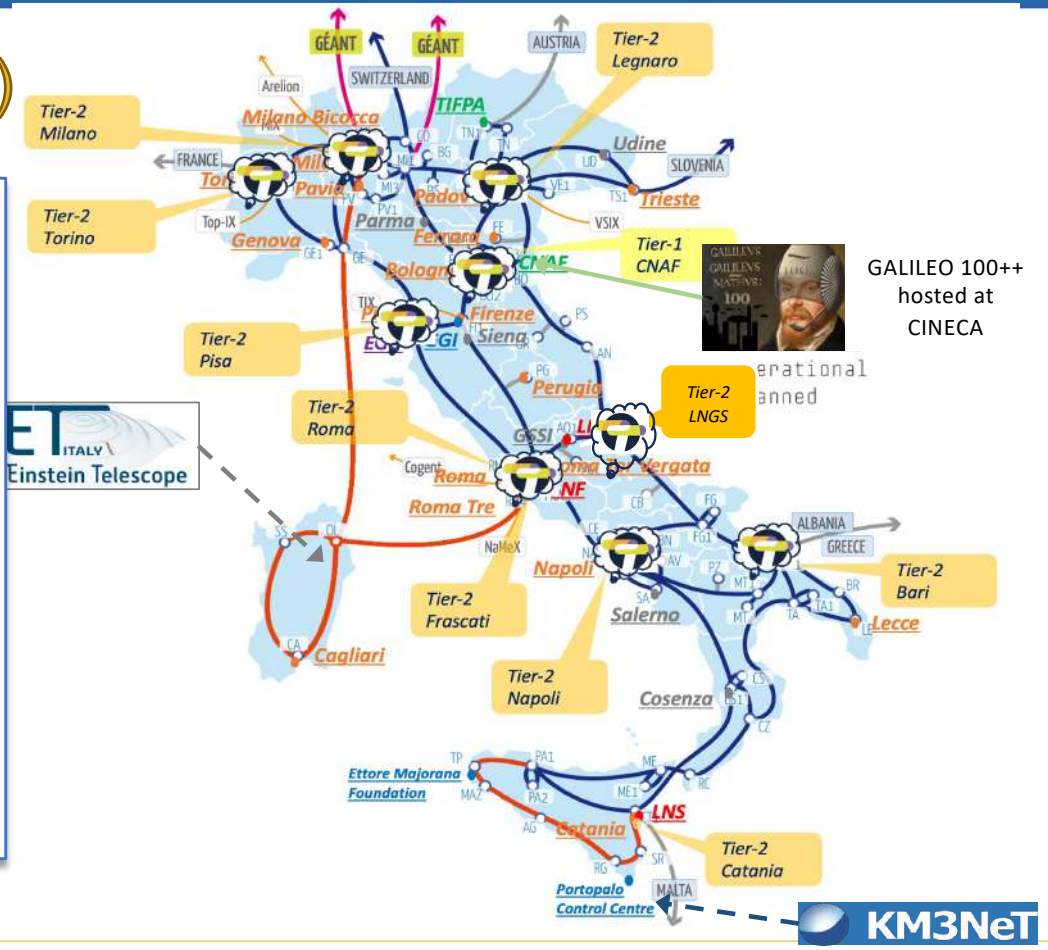
# TeRABIT final Infrastructures (2025)

The image shows the overlap of the expected combined final physical topologies of all three Research Infrastructures:

- GARR-T with (in red) the new fibres (islands)
- HPC-BD-AI with the HPC Bubbles locations
- PRACE-ITALY with the upgraded GALILEO100 hosted at CINECA

Developments are in close collaboration with ICSC

TeRABIT network extension will support other RIs (e.g. ET and KM3NeT)



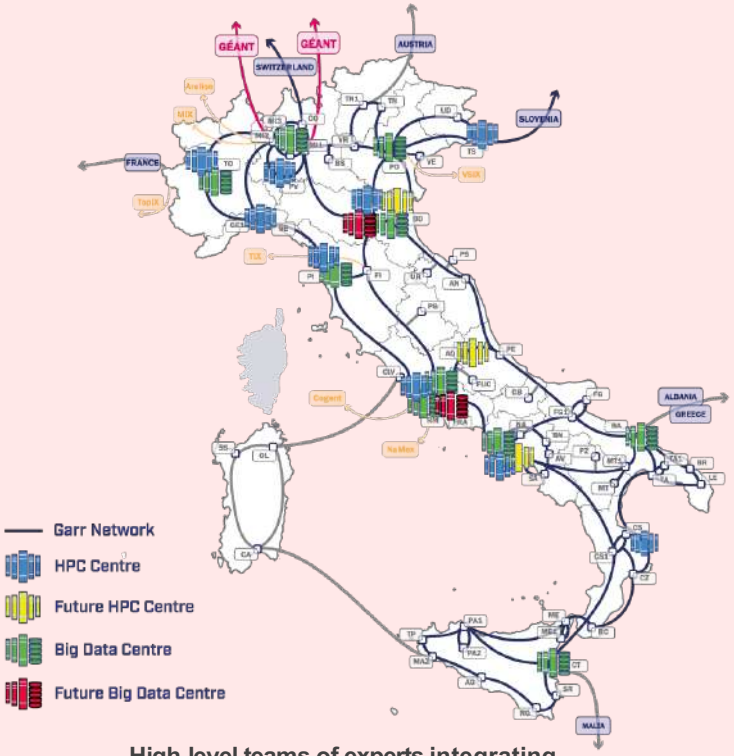
Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

terabit

## 0 SUPERCOMPUTING INFRASTRUCTURE




High-level teams of experts integrating
the Spokes working groups (mixed cross-sectional teams)

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

ICSC is made of

- 10 thematic spokes
- 1 spoke for the infrastructure

ISTRUZIONE E FORMAZIONE, IMPRENDITORIALITÀ, TRASFERIMENTO DI CONOSCENZE, POLICY, OUTREACH

1 FUTURE HPC & BIG DATA

2 FUNDAMENTAL RESEARCH & SPACE ECONOMY

3 ASTROPHYSICS & COSMOS OBSERVATIONS

4 EARTH & CLIMATE

5 ENVIRONMENT & NATURAL DISASTERS

6 MULTISCALE MODELING & ENGINEERING APPLICATIONS

7 MATERIALS & MOLECULAR SCIENCES

8 IN-SILICO MEDICINE & OMICS DATA

9 DIGITAL SOCIETY & SMART CITIES

10 QUANTUM COMPUTING

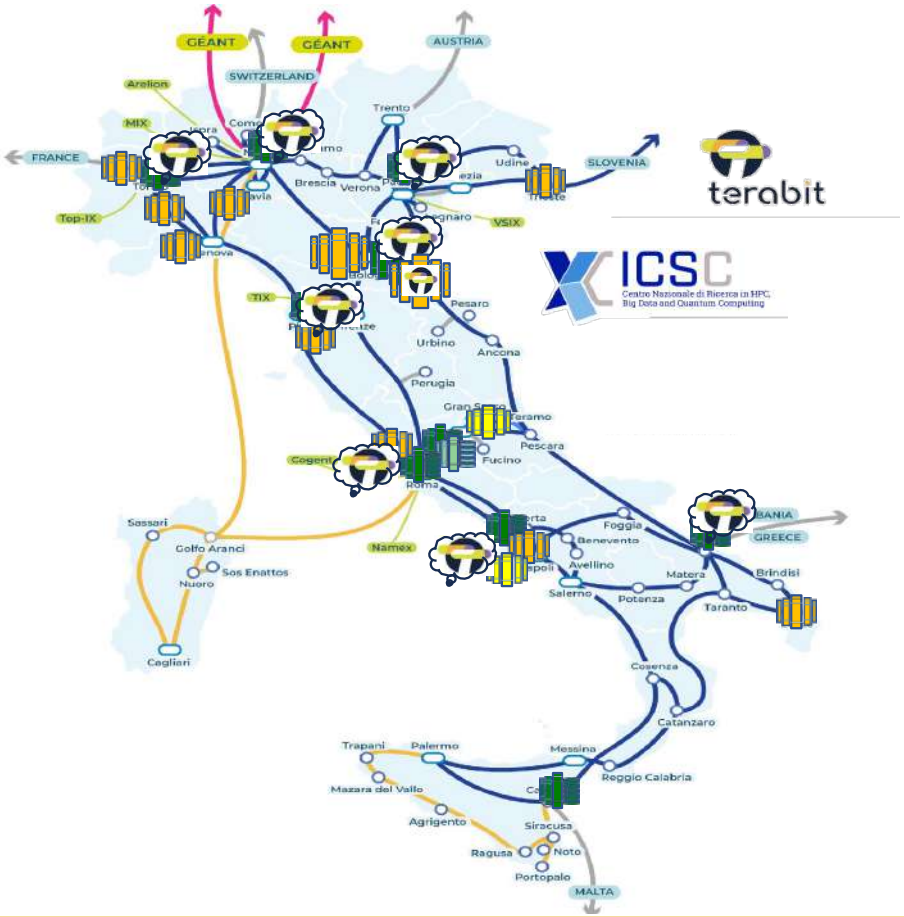
# National computing research infrastructure

ICSC and TeRABIT plan the creation of the national federated computing infrastructure for research

Access to resources should be transparent for the end-user

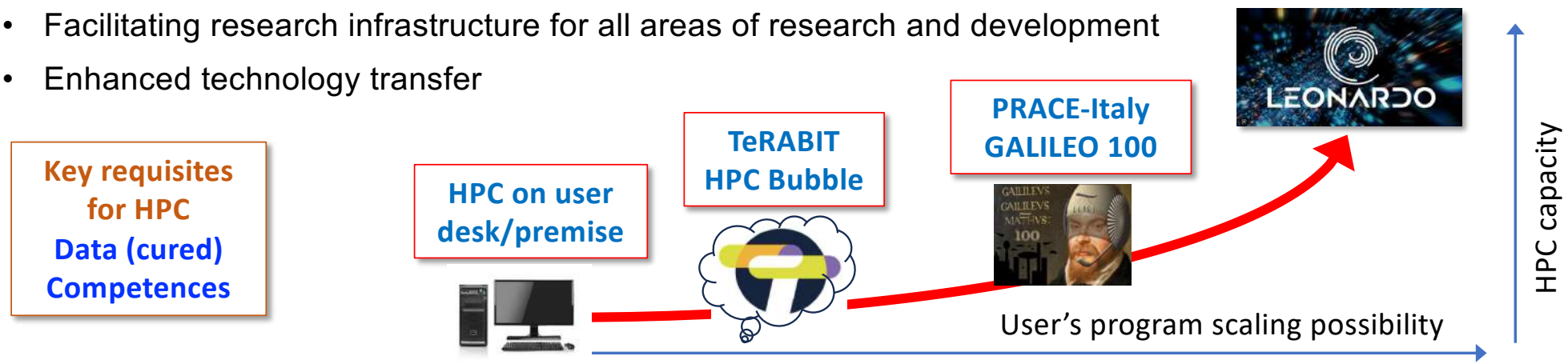
Main actors are: INFN, CINECA, GARR, OGS

And also: CMCC, ENEA, SISSA, IIT, Univ. TO, Univ. Roma Sapienza, …



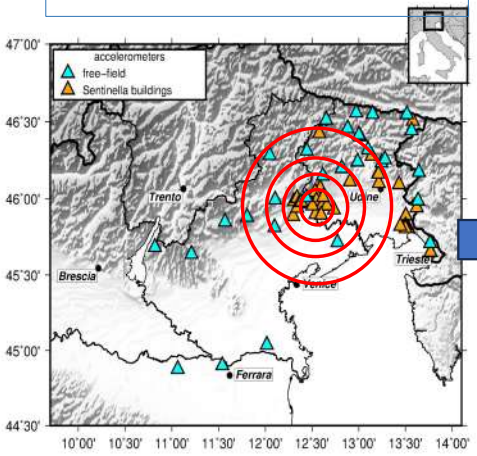
# Objectives / Innovation / Expected impact of the project

- Infrastructures strengthening to answer scaling and new requirement of research
- Tighter integration between network, data and HPC services with common, federated services
- Innovative HPC services (bubbles), modular and increasing HPC/ML capacity between the "edge", where the users and their data are, and PRACE-Italy, in synergy with ICSC (Leonardo)
- Federation and communication between HPC Infrastructures with close collaboration with the national and international HPC centres (via GÉANT) as PRACE and EuroHPC centres
- Facilitating research infrastructure for all areas of research and development
- Enhanced technology transfer



# Uses case: SHAKE (Supercomputing for eartHquAKE rapid damage assessment)

OGS SMINO network
Sensors on soils and buildings (currently 400+)




**HPC**

Physics-based code for ground motion simulation based on detailed earth models to provide

Ground shaking assessment

Impact assessment


Civil Defence
(Protezione Civile)

Challenge :
Use HPC infrastructures to provide a prompt assessment of eventual damage

# Uses case SHAKE : implication for the RIs

**REAL-TIME APPLICATION / SERVICE**

**Time-bounded results expected**

**Urgent computing service type**

| | | |
|---|---|---|
| No pre-allocated HPC resources | → | Consider Parallel executions on different infrastructures (Galileo100, HPC Bubbles) for resiliency and different parameter to produce slightly delayed, but more precise results |
| The preferred HW | → | Offer diverse HPC capabilities (CPU, GPU, FPGA) |
| Simple and automated User Access | → | Set-up a Federated AAI, considering both traditional HPC and Cloud systems and a single workflow with storage sharing |

Adequate network capabilities at all times, resiliency

# Uses case SHAKE : challenges for the user

**REAL-TIME APPLICATION / SERVICE**
**Time-bounded results expected**

OPTIMIZATION OF CODE AND BEST USE OF RESOURCES

Parallelizing code (i.e. mesh, UCSB synthetic seismograms generation, post-processing)

Choice of code language, libraries and profiling

Pre-computing input territory data assess mesh size /Green functions

Scalability tests for choice of n.of nodes, memory use , data storage size

First results: the code can be **significantly** optimized and the computation time is reduced adding nodes. The response time, *for the fastest simulations,* can now target minutes, instead of hours

# Use of HPC: challenges for TeRABIT

## NON TECHNICAL

- Digital Data-centric evolution in research (data use policies, new data analysis paradigms)
- Data preparation and maintenance (FAIR)
- Skills, permanent education, Talent retention
- Policies and Trust between RIs, Thematic infrastructures and National organization
- Cost of operation (electrical power)
- Cybersecurity
- Digital Divide

## TECHNICAL

- Federated authentication and authorization
- Rapid ICT evolution lead by commercial players, e.g. Large Language Models
- Storage (data) access and data management
- Upgrade and federate the IRs in production
- Digital divide

- **Create aTrusted Research Environment (TRE)**

# A Communities-based Project

7 HPC Master and 11 PhD Students

Education

Young technologists and researchers

Other

Case study: Earth Science

PNRR Projects

Earthquake Analysis

Developments in close collaboration with

**ICSC** National Centre for High Performance Computing, Big Data and Quantum Computing

# Users: higher education and dissemination

**Case studies**

Enhancement of HPC Training and Research for Earth Sciences and the participation of other research areas experts

**training of young technologists**

- organization of **2 workshops (1st in June 2024)**
- organization of **1 hackathon**


**HPC Master and PhD**


Doctoral and master HPC students

**Students**

- Sessions to present the TeRABIT project in High Schools distributed in Italy
  - ✓ Train young researchers and technologists of the future

# A Collaboration Plan Design Model

How to engage a collaboration?
As a function of the scope, your organization may collaborate with TeRABIT's partners or ICSC:

| | |
|---|---|
| With Partners in TeRABIT | • INFN, OGS, GARR and CINECA have well defined user access policies for research and Industry, when possible. E.g. the GARR network can be used by consortium members for research only |
| With ICSC | • Strengthening the bridge between Academia and Industry.<br>✓MoU to be agreed between ICSC and TeRABIT. |

# References


The TeRABIT project https://www.terabit-project.it/

TeRABIT conference 2024 https://www.terabit-project.it/it/eventi/conferenza-terabit-2024

GARR Italian National Research and Eduction Network http://www.garr.it

Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS https://www.ogs.it

Istituto Nazionale di Fisica Nucleare INFN https://www.infn.it/it

CINECA https://www.cineca.it/it

ICSC https://www.supercomputing-icsc.it/

HPC-TRES https://www.ogs.it/en/content/hpc-training-and-research-earth-sciences-hpc-tres

Research Infrastructures https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/european-research-infrastructures_en

Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

terabit

# Thank you. Questions ?

For information: info@terabit-project.it

## TeRABIT will be also at:

Forum-EXPO PNRR della Borsa della Ricerca, 21-23 ottobre 2024, Catania