Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INFN

terabit

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

Federating heterogeneous computing infrastructures

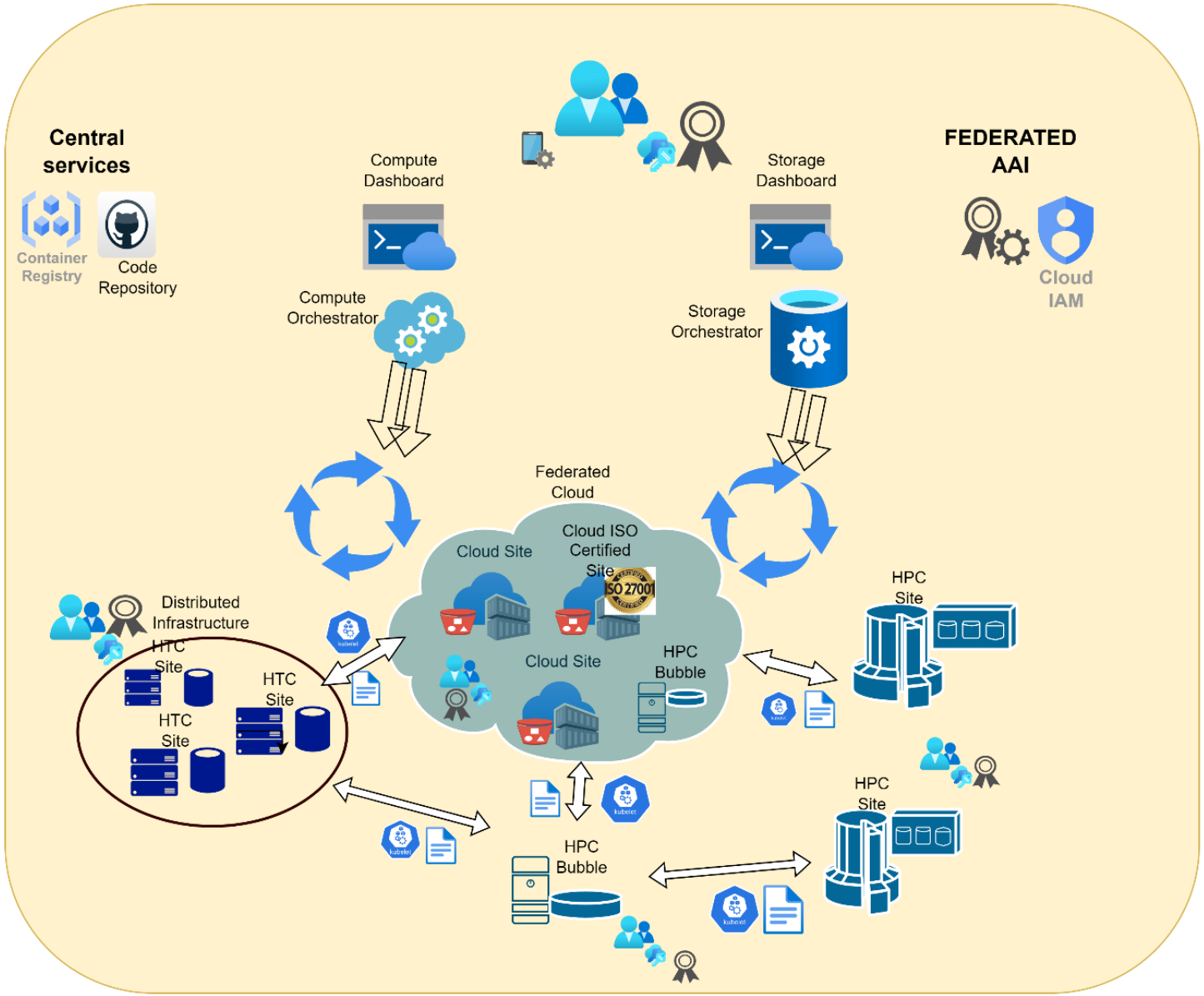
Claudio Grandi (INFN Bologna)

Heterogeneity

Enables to efficiently tackle diverse applications

- complex simulations
- high-throughput data processing
- Interactive analysis
- data storage and data access
- IOT integration
- critical data management
- …

There is no «one size fits all»!



Leonardo + Lisa High level architecture

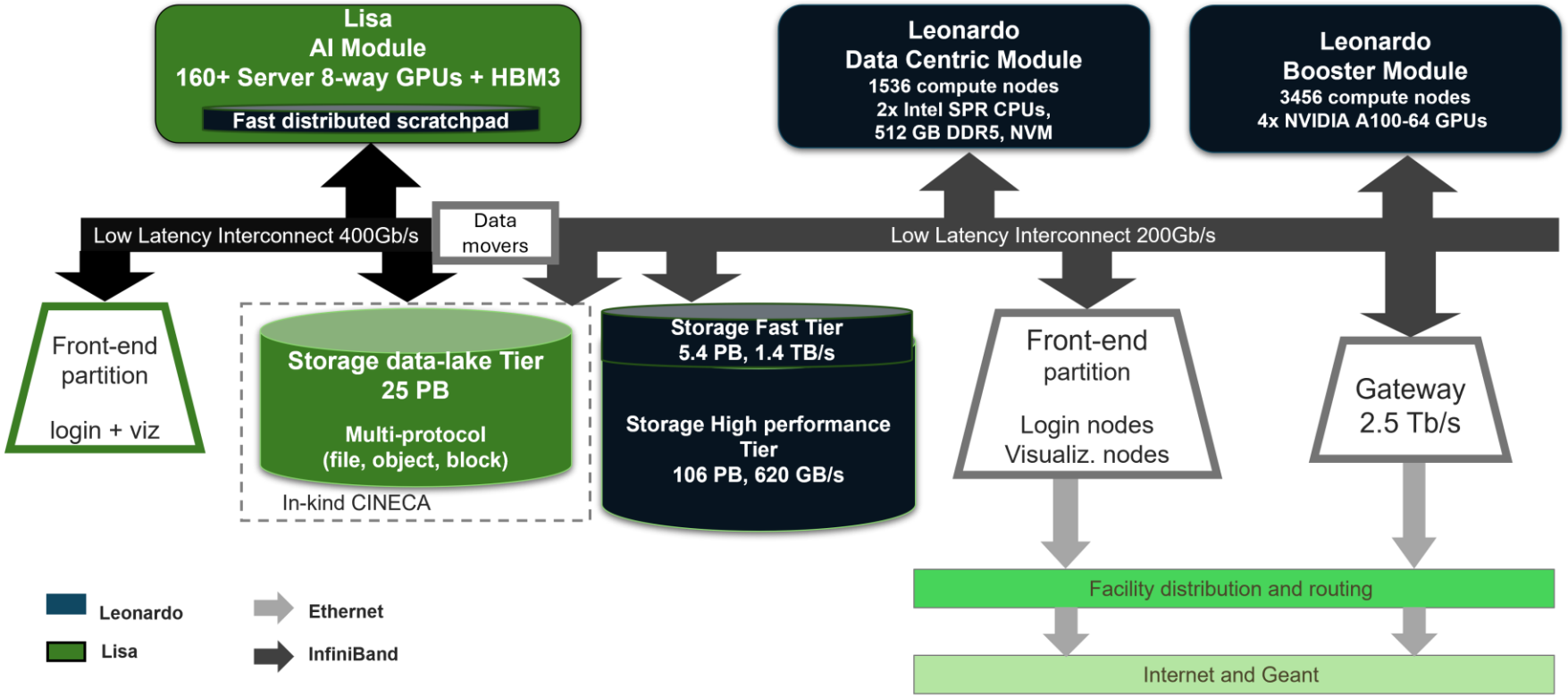
Total ~100 Pflop/s (theoretical peak)

1 compute module

Best of market Interconnect

In-kind data-lake storage

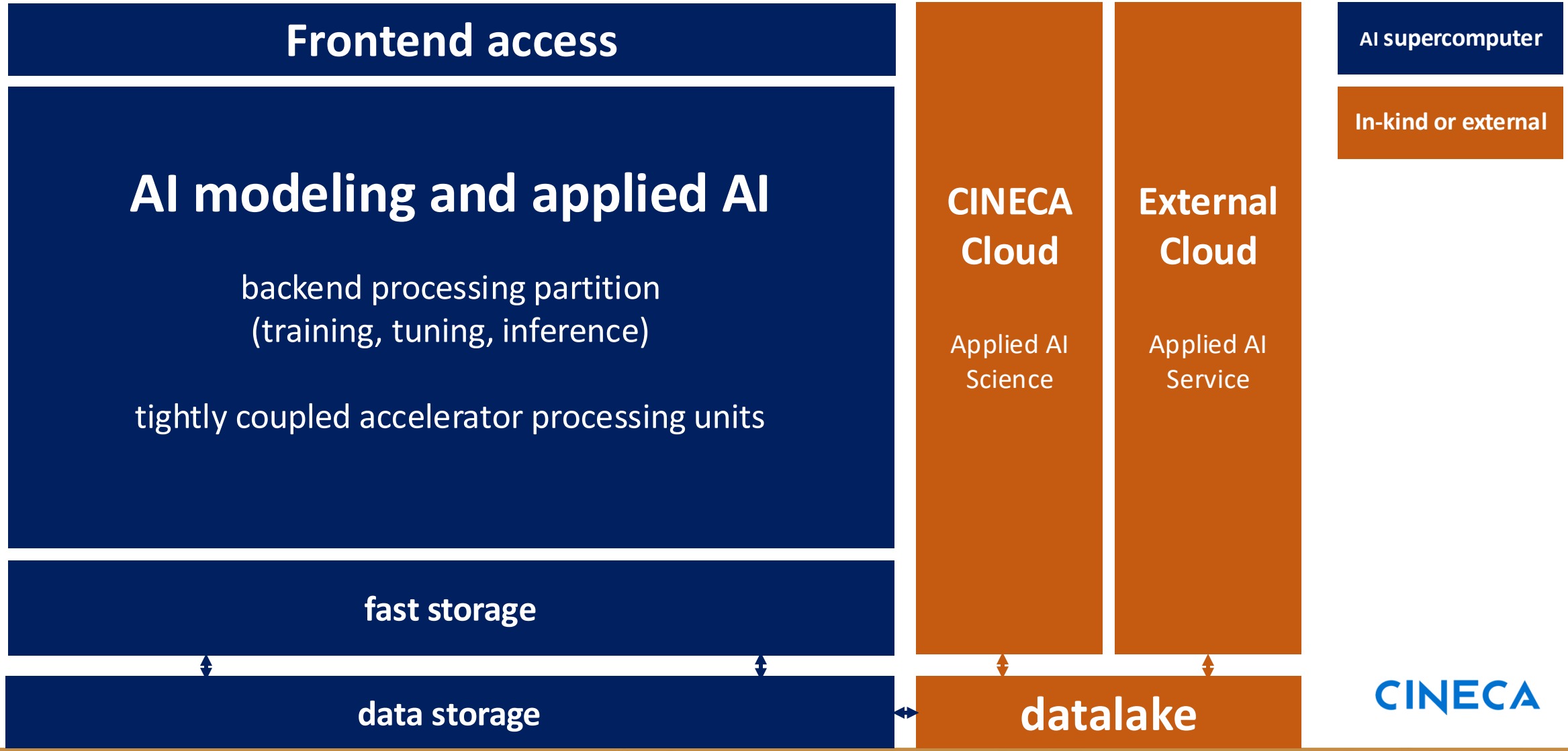
AI oriented benchmarks



CINECA

Claudio Grandi

CINECA AI Factory proposal



CINECA

CINECA AI Factory proposal

AI modeling and AI applied partition

batch processing (training, tuning, inference)

tightly coupled accelerator processing units

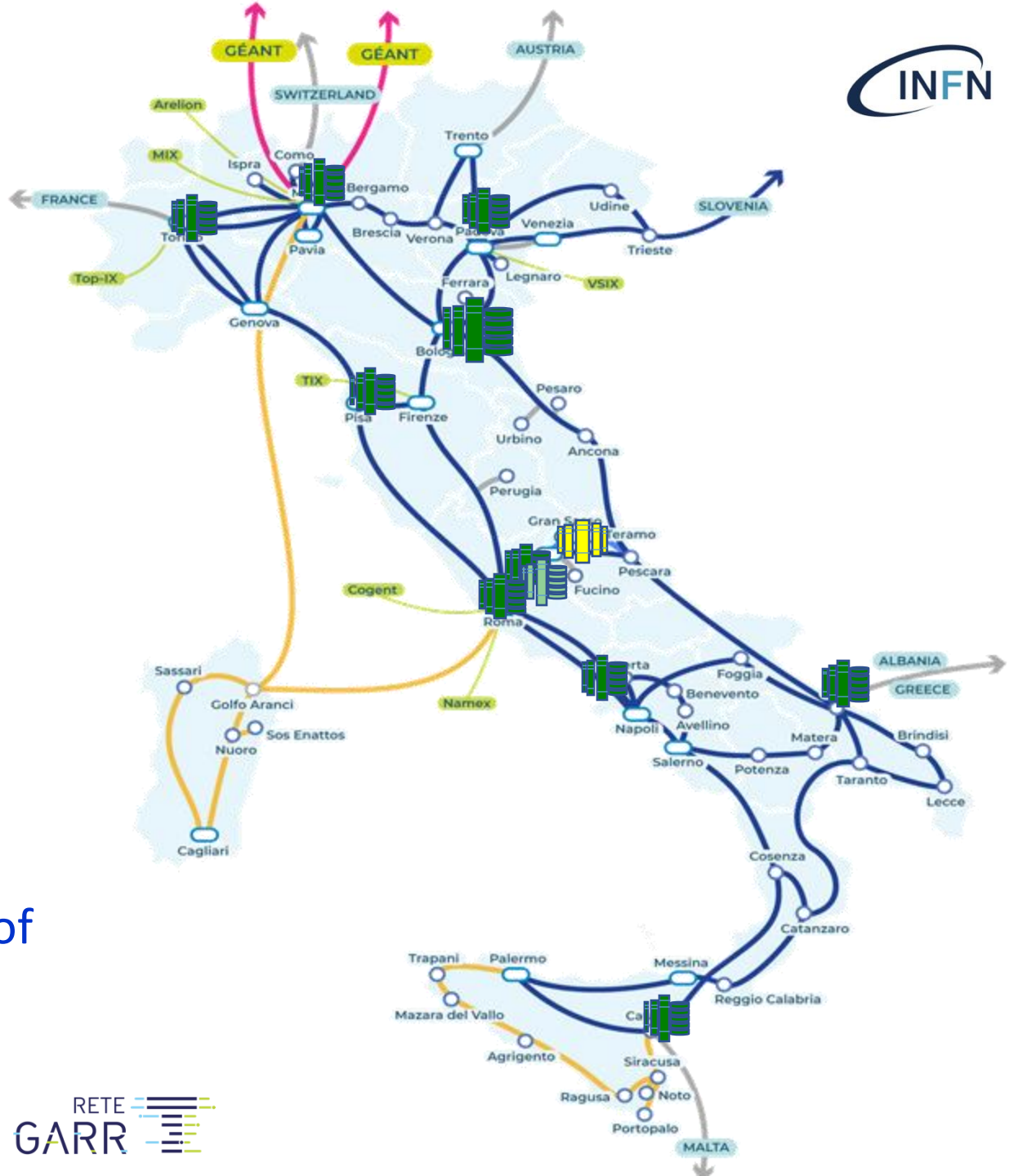
750-1000 PFLOPS
50-60 EFLOPS FP8
2000-3000 servers
120-160 racks
13-17 MW peak
8-10 MW ops
DLC/AIR: 70/30 – 90/10
32-26 °C

CINECA

INFN DataCloud Infrastructure for Scientific Computing

- Tier-1 (CNAF)
- Tier-2's (BA, CT, LNF, LNL/PD, NA, MI, PI, RM1, TO)
- INFN Cloud
  - Backbone and federated clouds
- HPC4DR (LNGS)
- (Tier-3)

DataCloud was born to address the needs of INFN research activities, but it is serving several external projects

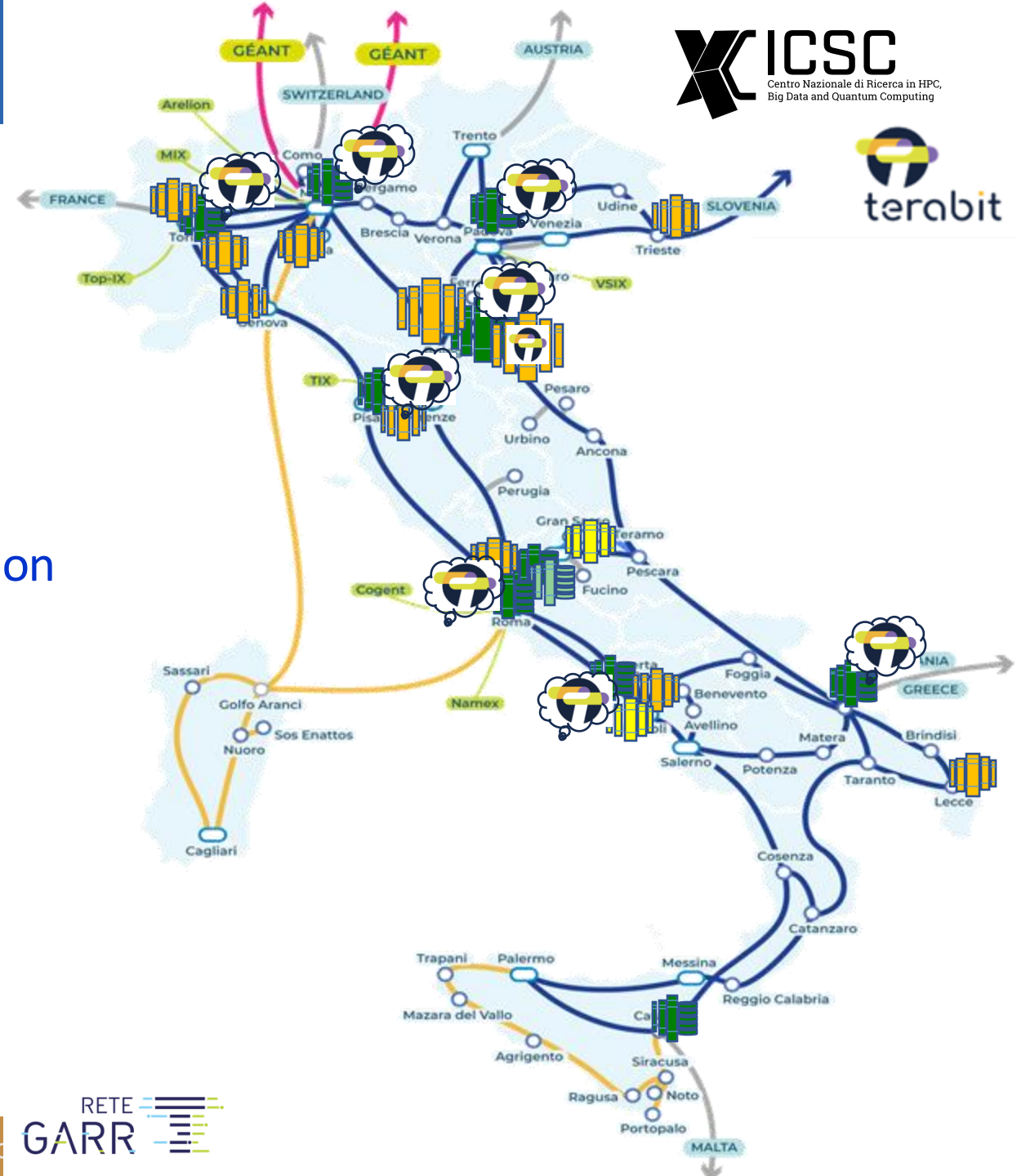


Building the Italian Cloud Federation

In the framework of the NRRP projects, in particular ICSC and TeRABIT, the INFN DataCloud model is the basis for the creation of the Italian Cloud Federation

The goal is to access all Italian scientific computing resources through uniform interfaces

Main players: CINECA, GARR, INFN



HPC Bubbles

CPU node	192 physical cores 1.5TB RAM DDR5 IB NDR 400G 20TBL (SSD) + system disks
GPU node	As CPU + 4x NVIDIA H100 SXM5 with 80GB and HBM2e memory
FPGA node	32core RAM 768GB DDR5 IB NDR 440G 4 x XILINX U55C o 4 x TerasicP0701
Storage node (CEPH Bricks)	64 physical cores 1TB RAM DDR5 384 TBL HDD + 25.6 TBL NVMe
Accessories	Switch IB, Switch ETH cables IB, cables ETH Transceivers

160 CPU nodes

61 GPU nodes

10 FPGA nodes

118 storage nodes

… in addition to HTC servers and storage

https://www.supercomputing-icsc.it/en/icsc-home/

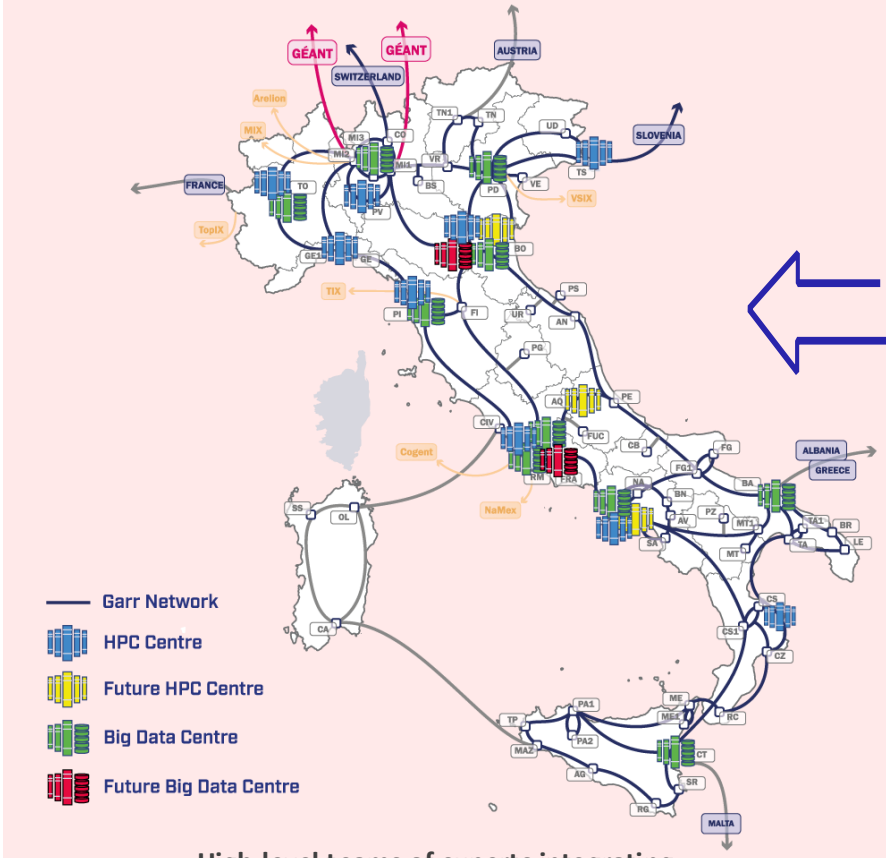
National Research Centre in HPC, Big Data and Quantum Computing

10 thematic spokes
1 infrastructure spoke (CINECA, GARR, INFN)

25 universities
12 research institutes
14 strategic private companies

320 M€ budget

0 SUPERCOMPUTING CLOUD INFRASTRUCTURE



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

EDUCATION & TRAINING, ENTREPRENEURSHIP, KNOWLEDGE TRANSFER, 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

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

TeRABIT: Terabit Network for Research and Academic Big Data in ITaly

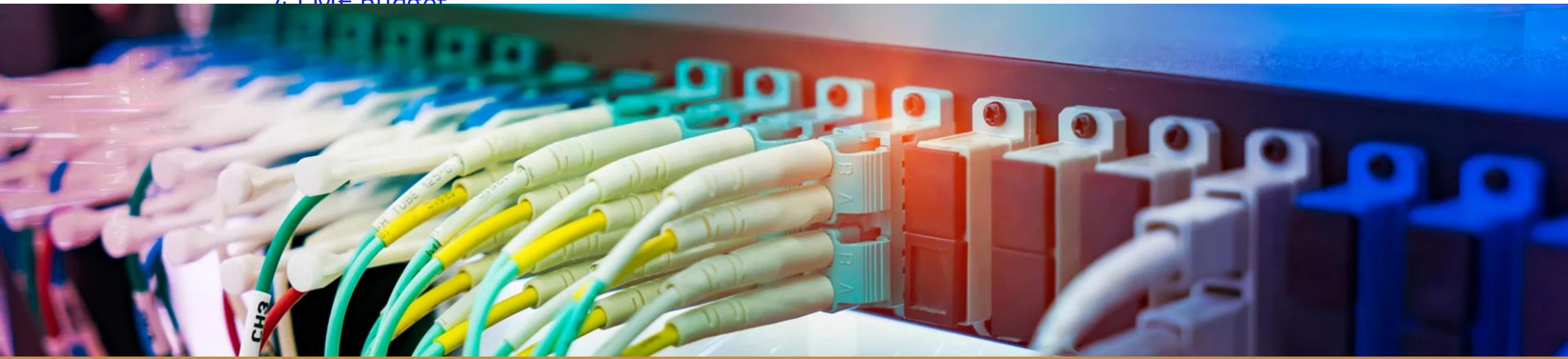
TeRABIT is a Research Infrastructure project synergic with ICSC

Partners are the same of the ICSC Spoke-0 (Supercomputing Cloud Infrastructure):

INFN, CINECA and GARR

Covers areas complementary to those of the ICSC infrastructure

~1 M€ budget



A data lake for research

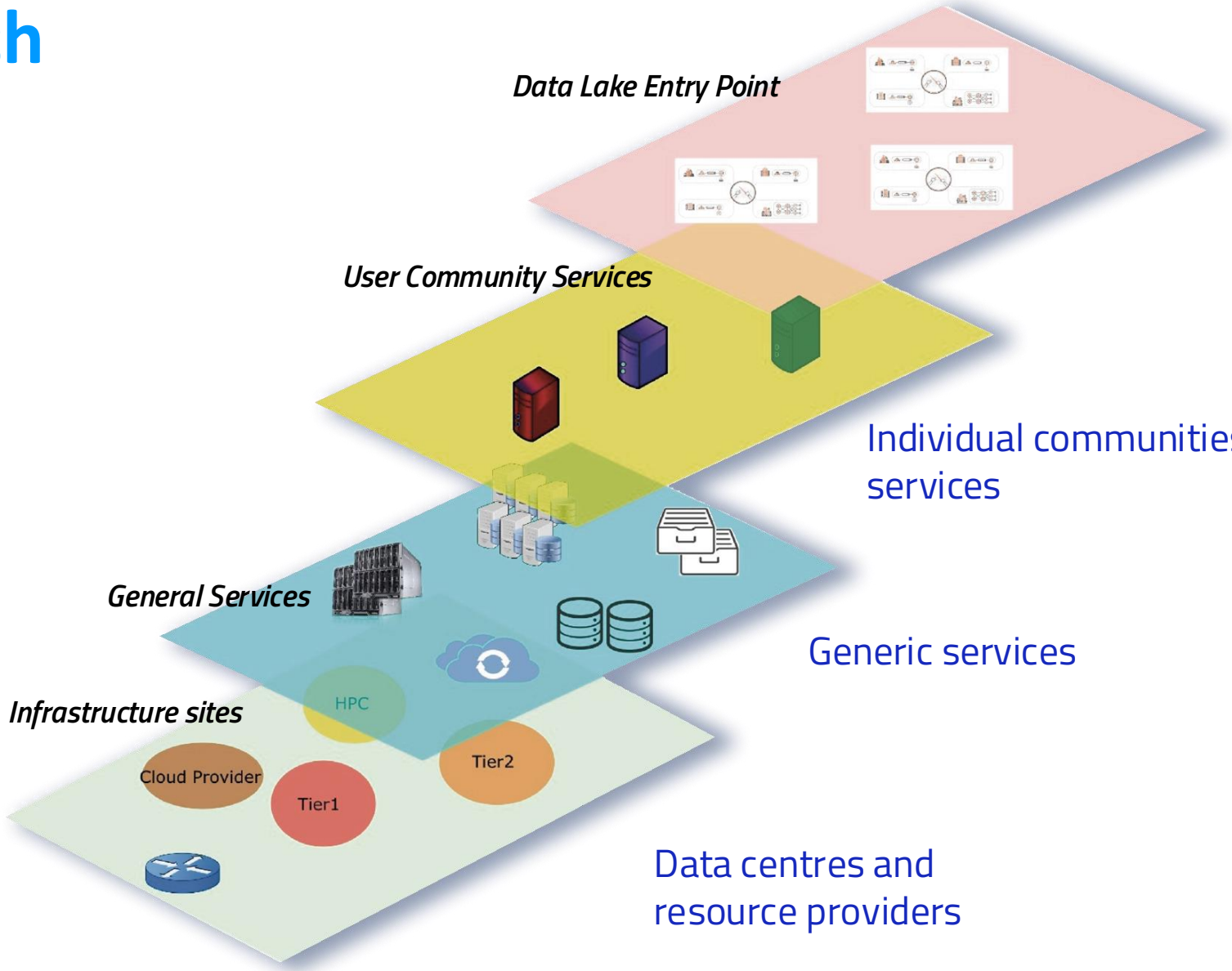
Existing infrastructures aggregation, upgraded and made available to scientific domains

A dynamic model, where infrastructures and domains can also be temporary

A clear separation between the physical and the logical levels

A high-speed network interconnection to hide the actual resource locations

A unified vision (when needed) of an Italian research data-lake

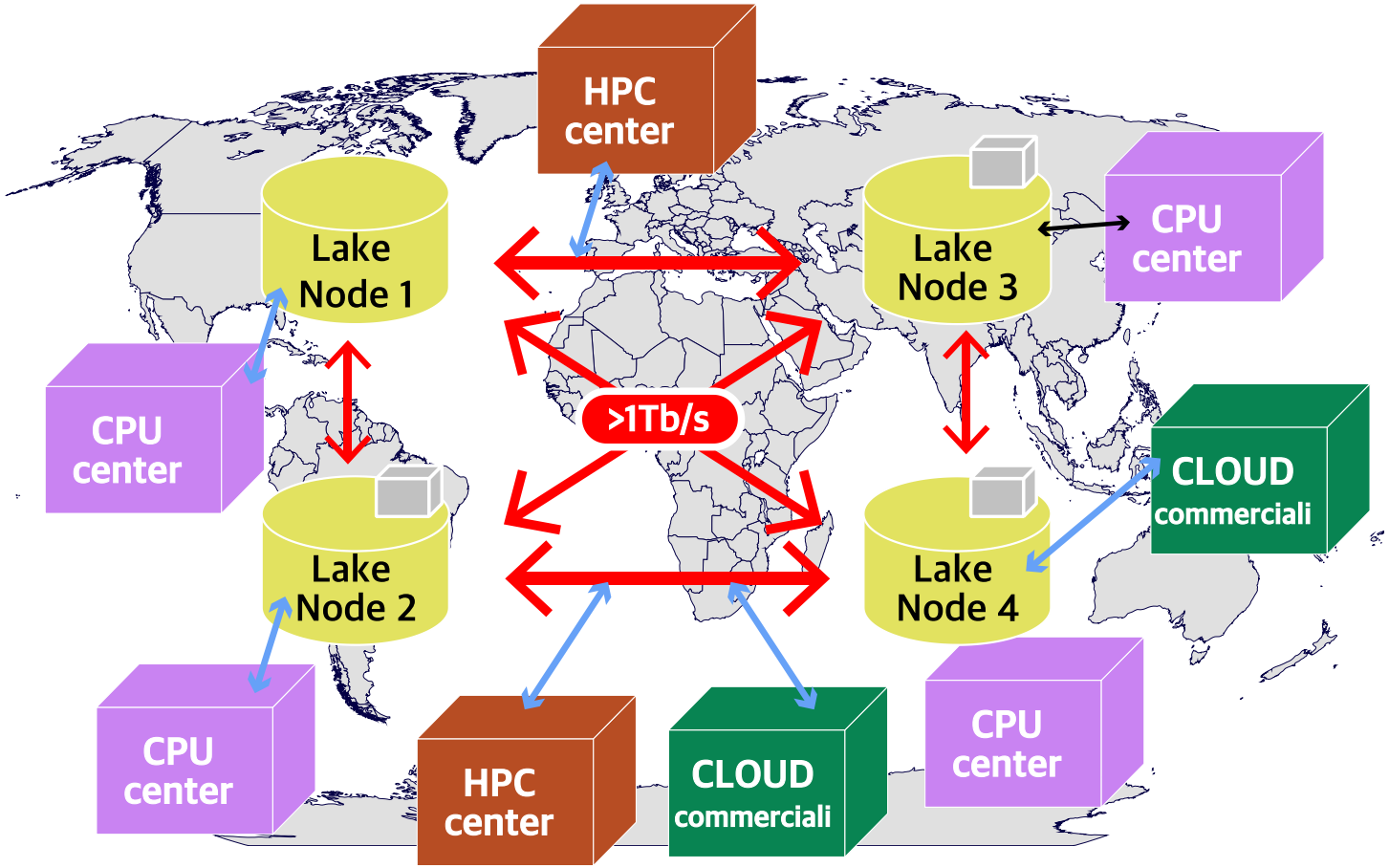


Data-centric model

Decouple storage and CPU

Storage nodes interconnected with high bandwidth network

Heterogeneous computing nodes can access data wherever they are



Inclusivity

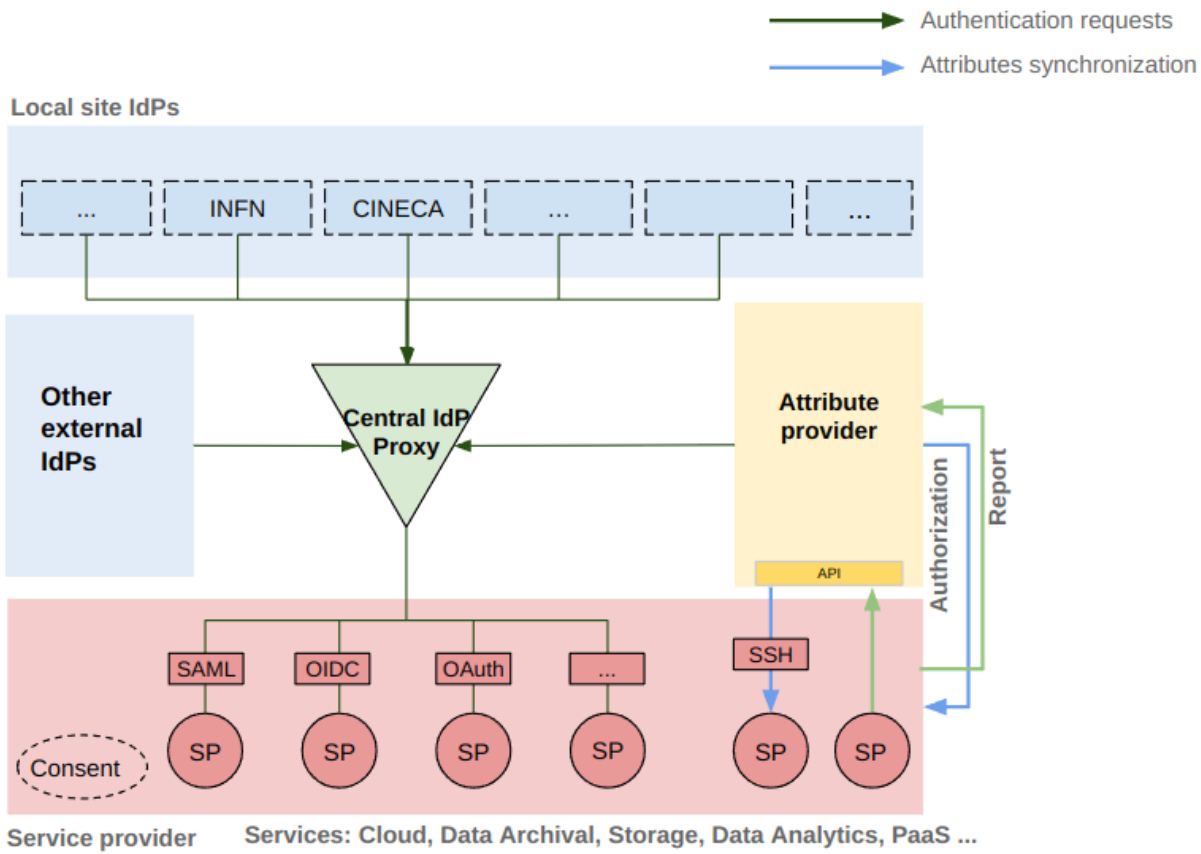
The federation will include data centres that are already in production, and part of international communities

- The procedures for joining the federation must be non-intrusive
- Standard must be used whenever possible, and developed when missing

The federation will serve users of several fields and organizations

- The procedures for user's onboarding must be as simple as possible
  - E.g.: use of Identity Federations

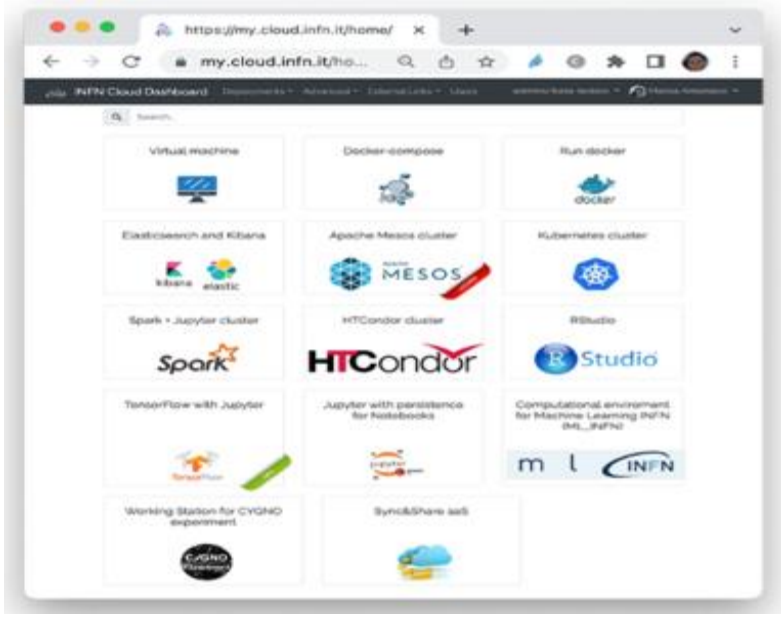
INDIGO IAM

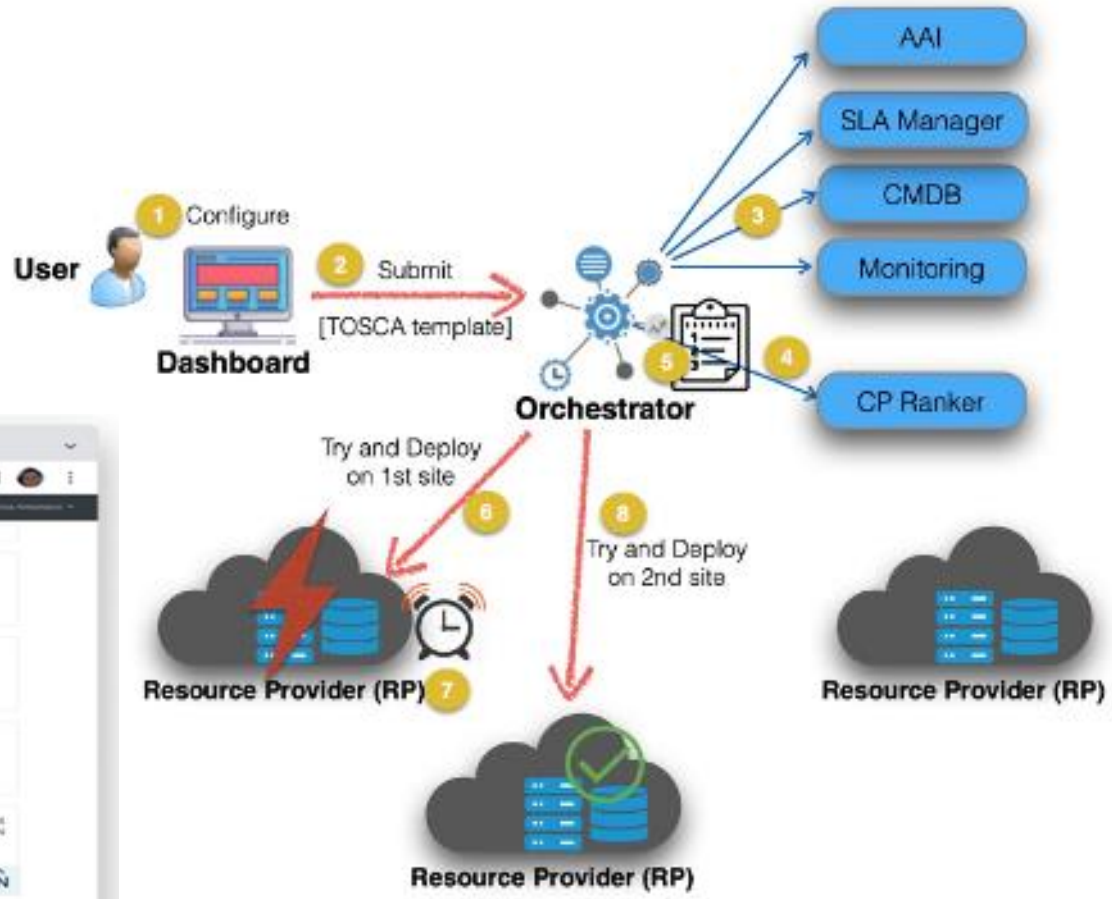


Ease of use

The federation will serve users with different computing competences

- Complexity of the underlying infrastructure hidden to the end user
- Support field experts in developing platforms that enable the effective exploitation of the infrastructure through composition of services and resources





INDIGO PaaS orchestrator

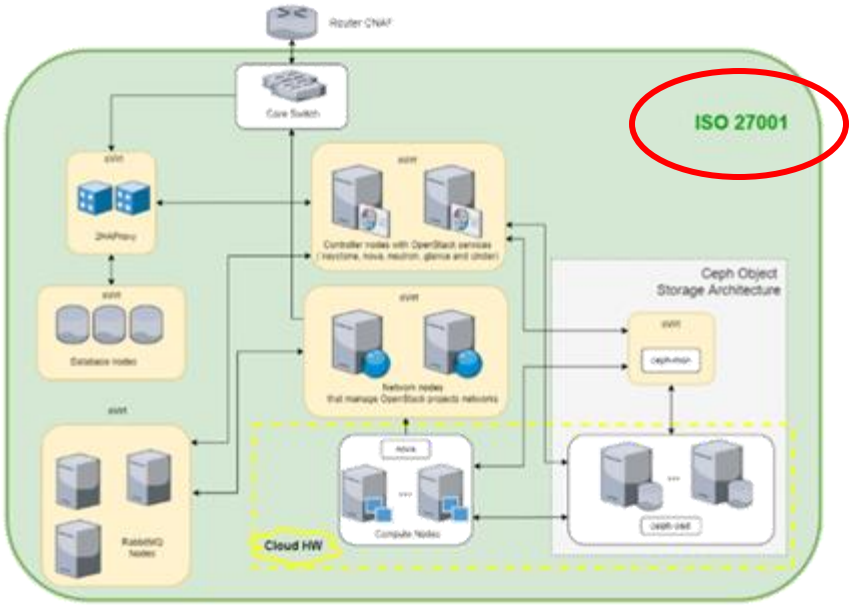
Flexibility

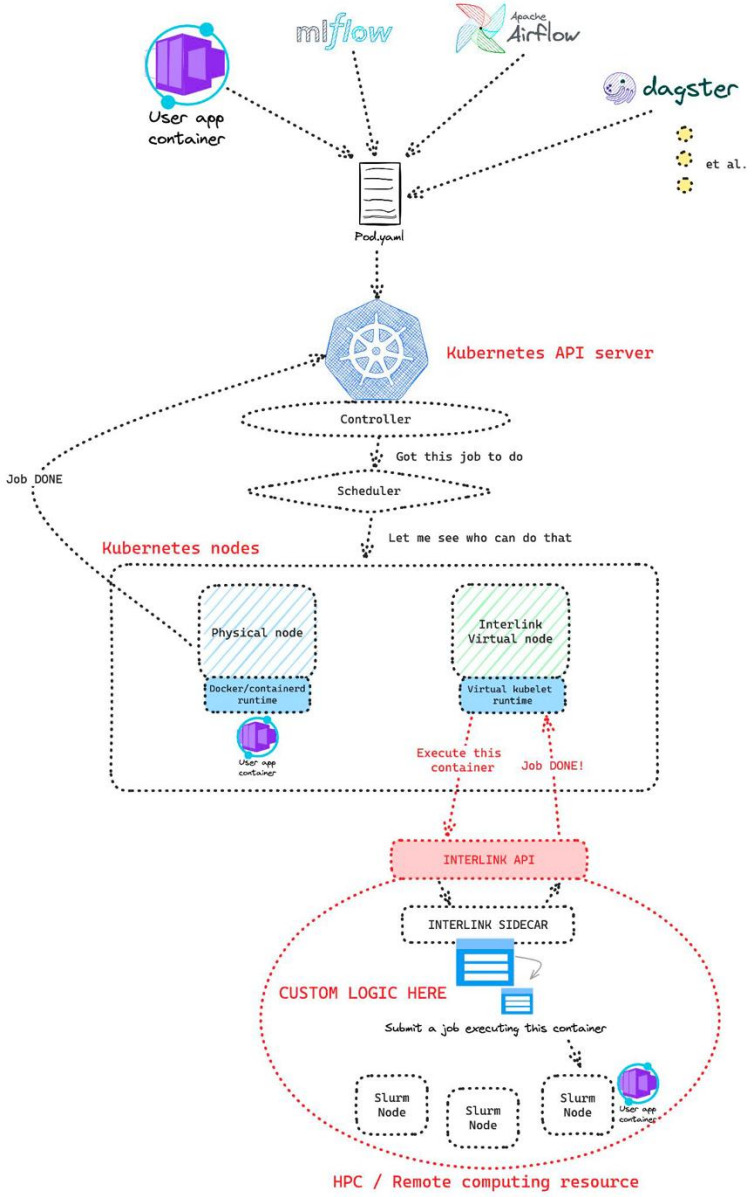
Support multiple access methods to the resources, oriented to:

a. Transparency and ease of use
b. Efficiency and effectiveness

Support application-specific requirements

E.g. enhanced privacy





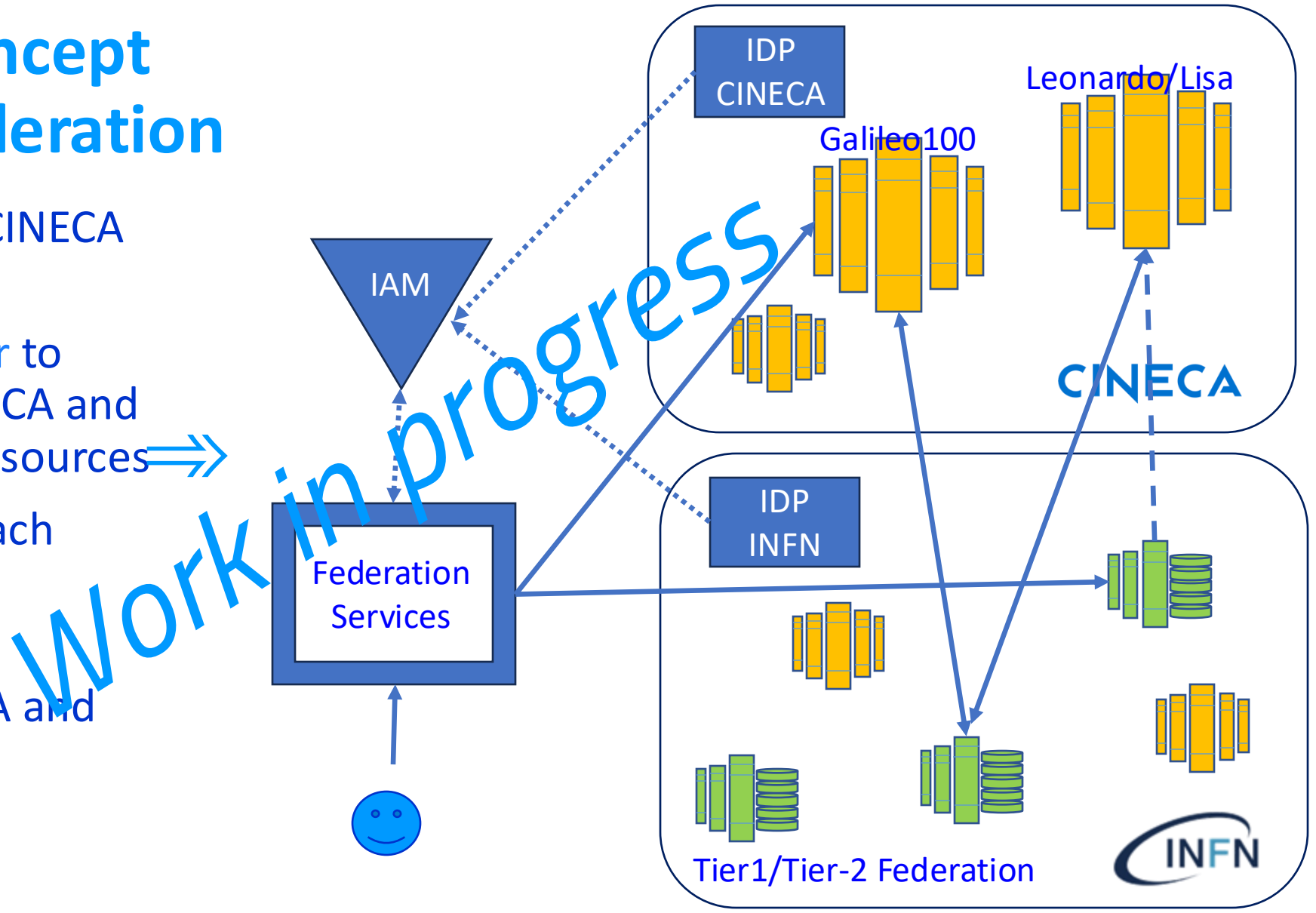
First Proof of Concept of the Italian Federation

INDIGO IAM to federate CINECA and INFN IdPs

INDIGO PaaS Orchestrator to transparently access CINECA and INFN OpenStack-based resources ⟹

InterLink offloading to reach CINECA's Leonardo Supercomputing

RUCIO to federate CINECA and INFN storage systems



Questions?

wooclap.com

Event code: **WSGARR24**

Work supported by the Italian Ministry of University and Research PNRR Mission 4, Component 2
ICSC: Investment 1.4, Project code CN00000013 - CUP I53C21000340006
TeRABIT: Investment 3.1, Project code IR0000022 - CUP I53C21000370006