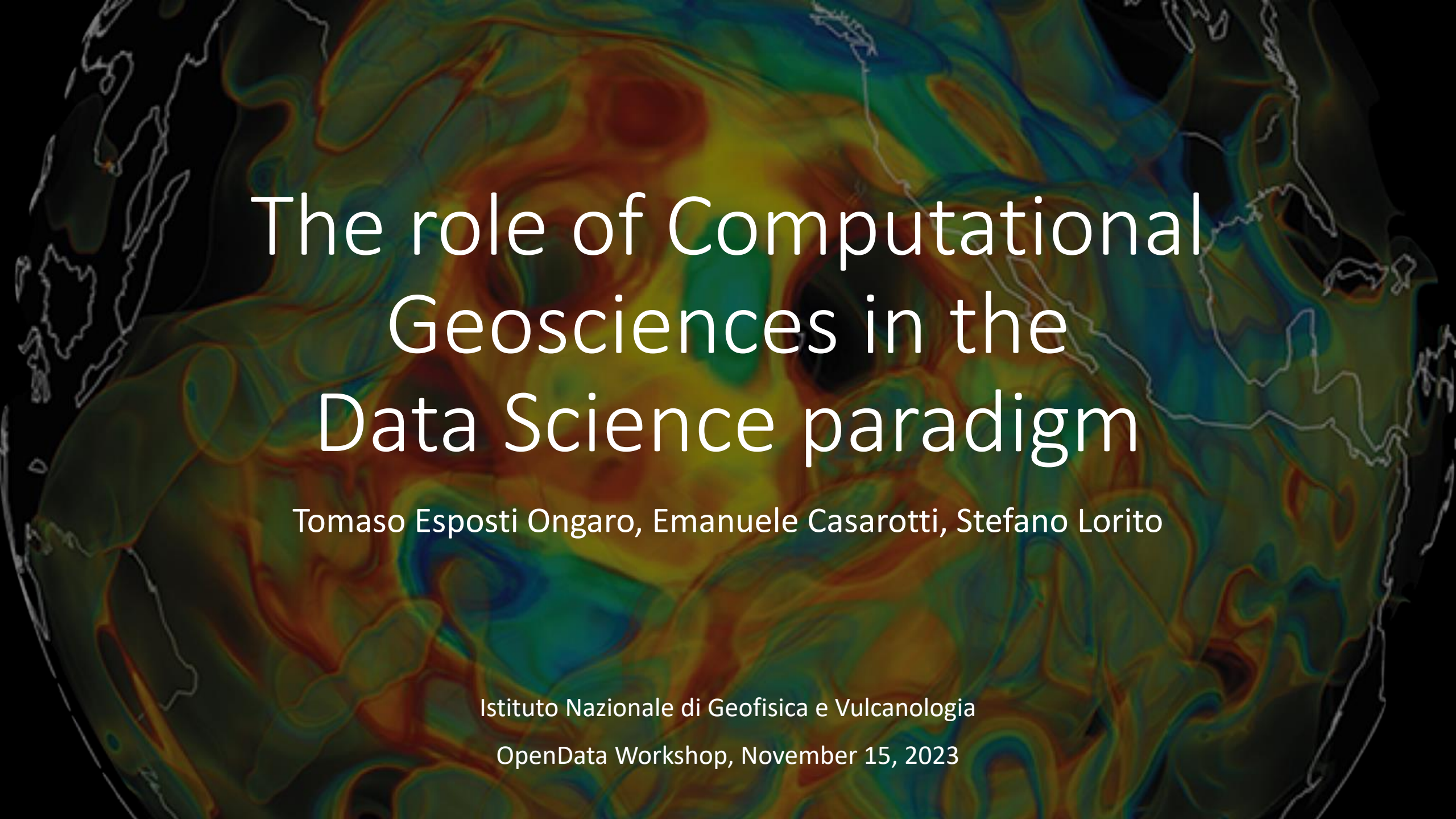
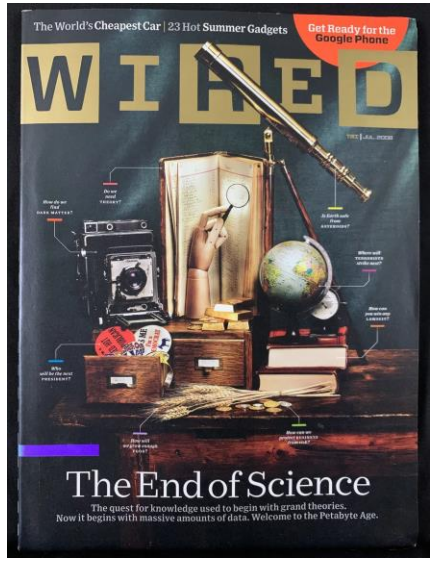


The role of Computational Geosciences in the Data Science paradigm

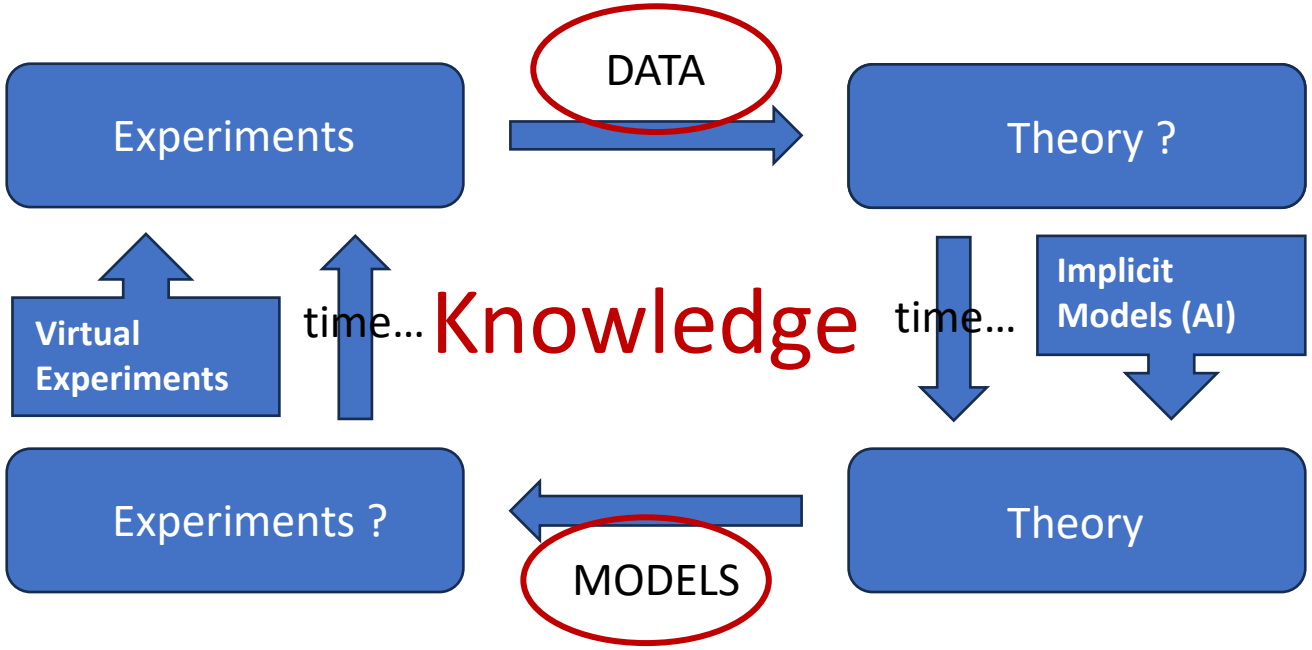
Tomaso Esposti Ongaro, Emanuele Casarotti, Stefano Lorito

Istituto Nazionale di Geofisica e Vulcanologia

OpenData Workshop, November 15, 2023



Scientific computing
- more statistics!
- *extremes*



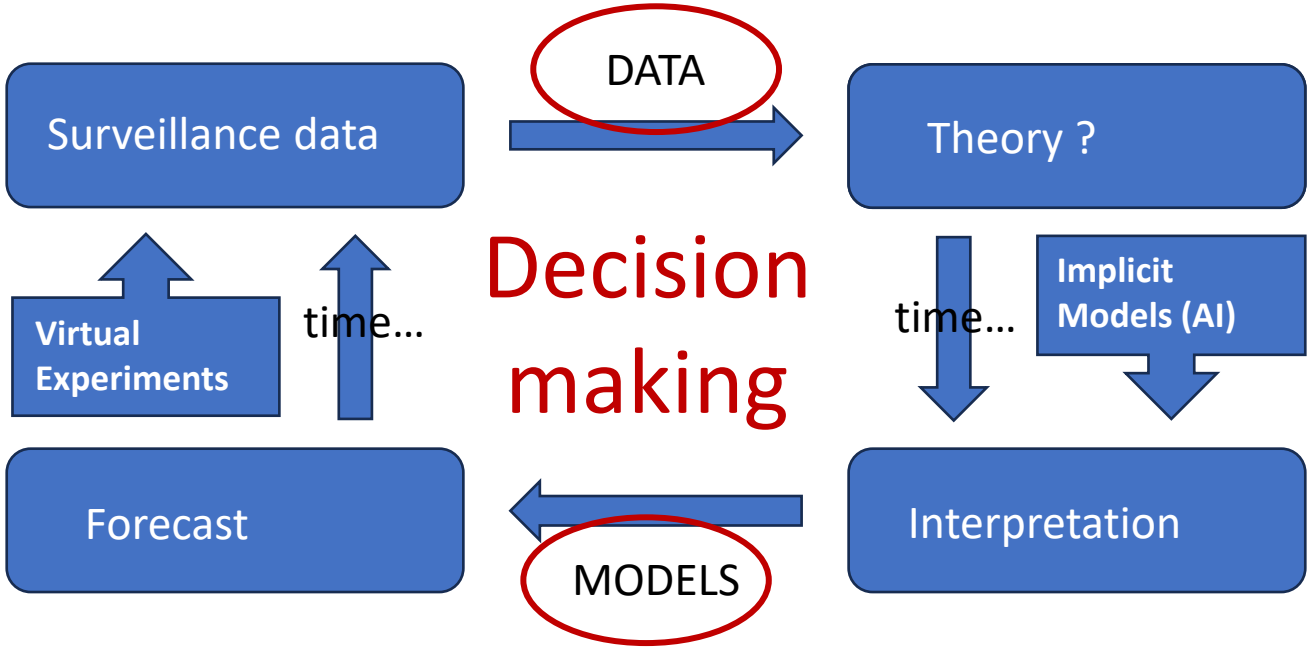
(PREDICTIONS)

Scientific computing
- reduce data

Acceleration of scientific knowledge !!

⇒Digital Twins

Scientific computing
- Scenario simulation

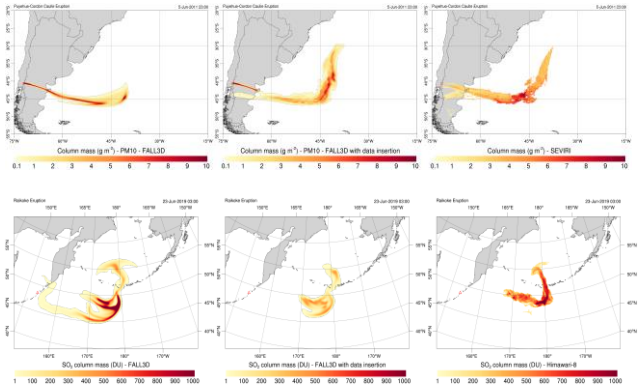


Scientific computing
- Probabilitistic forecast

Urgent Computing!!

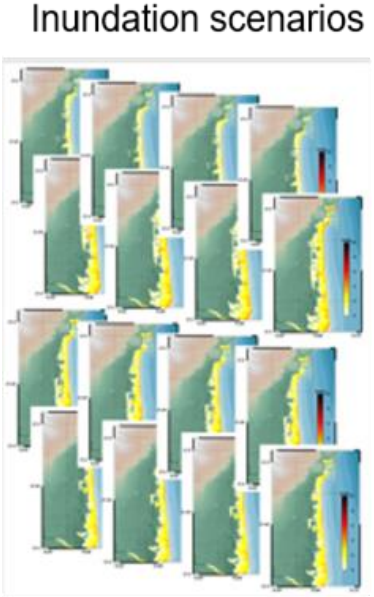
(sensitive information: Data/Model policy issues)

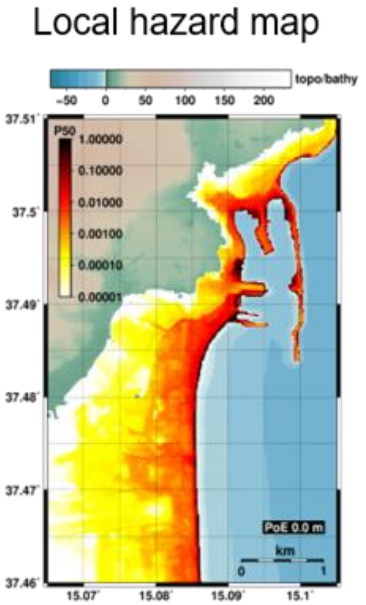
In silico experiments



More data (assimilation)

Larger Ensemble size Simulation of EXTREMES

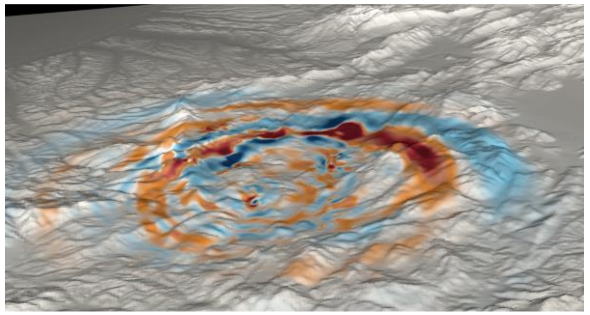




Larger spatial/temporal domains

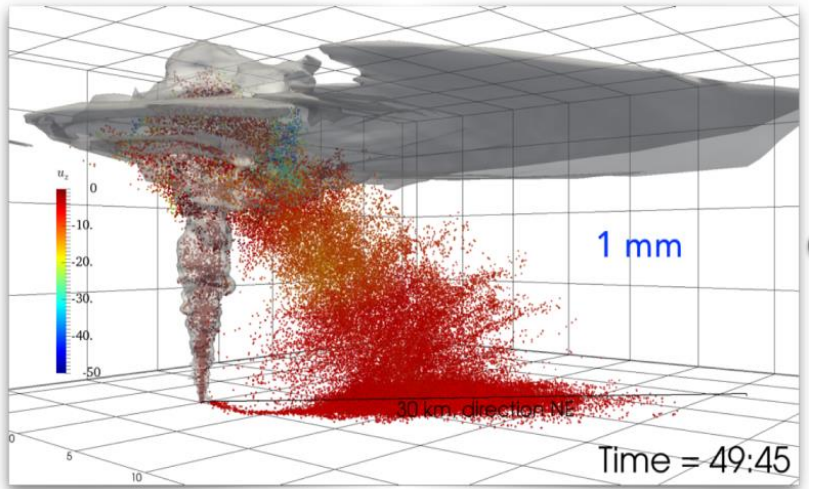
HPC

Higher spatial/temporal resolution



More physics (increase complexity)

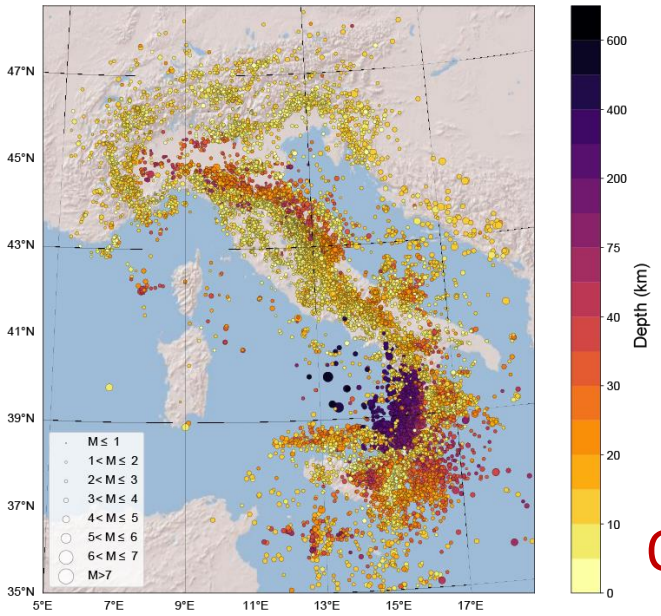
More output data (analysis)

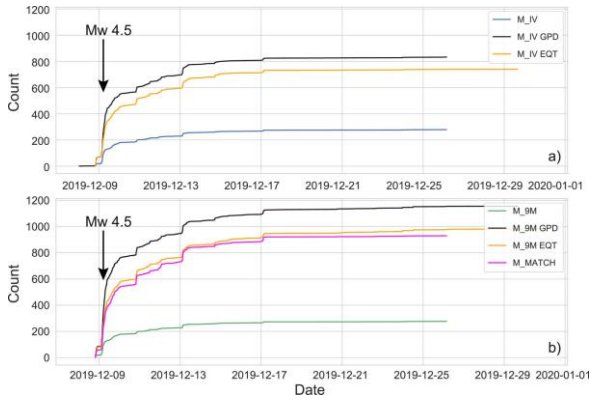


compromise between *storing* (*simulation data lakes)* and *producing on demand* simulation data yet to be found

The Big Data / AI challenge

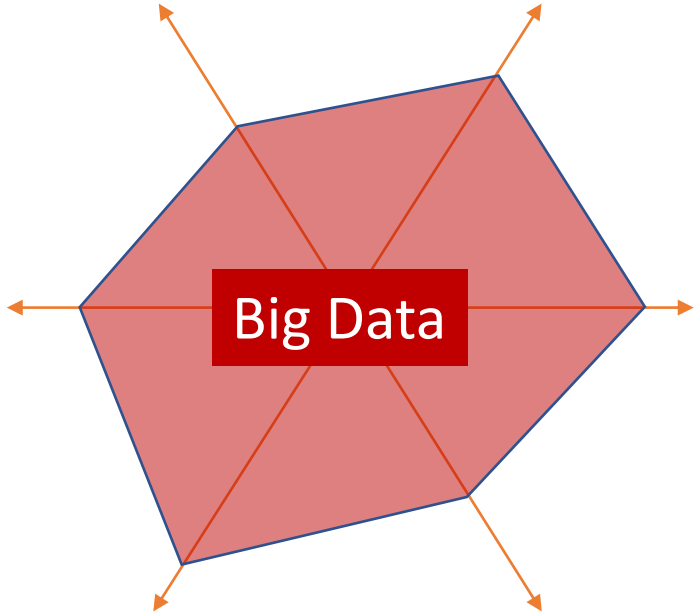
associazione BigData
from volume to value





Increased
Data volumes

Increased Data sources
(Internet of Things)



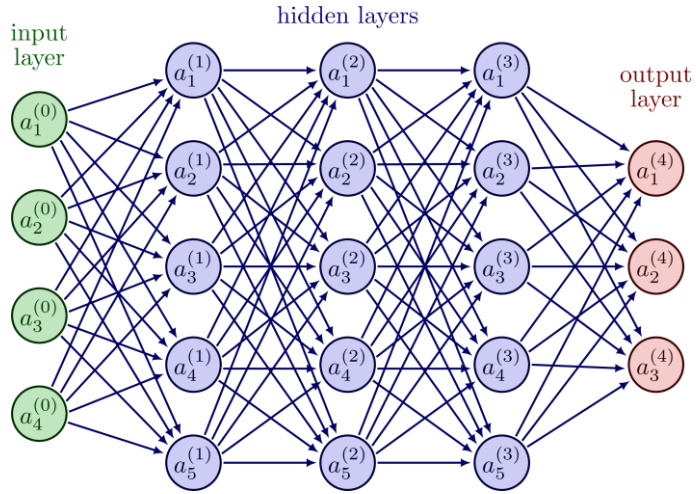
Findable Accessible Interoperable Reusable

Quality
standards

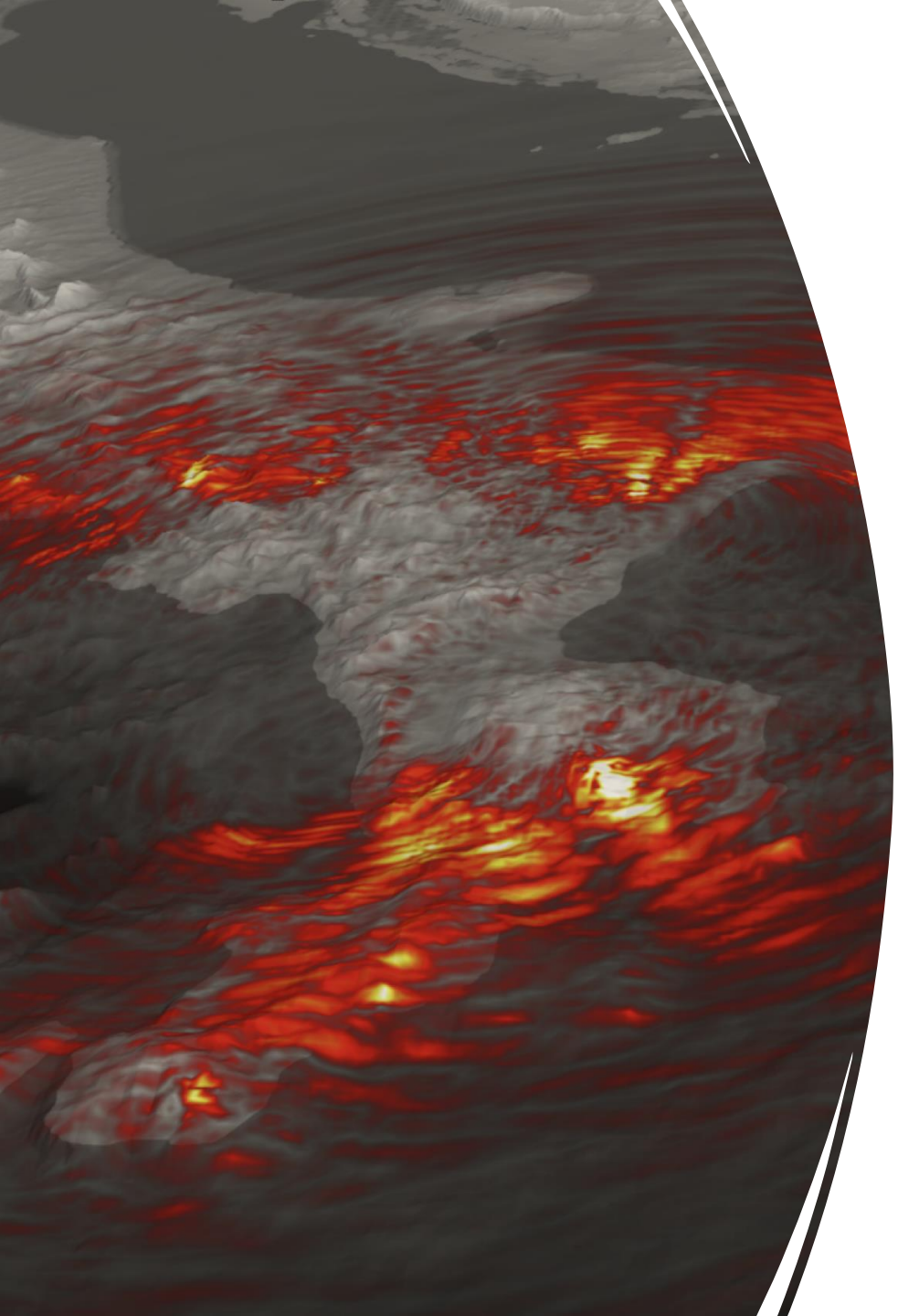
Enhanced
interoperability

Improved
Network capabilities

Evolved
ML / AI
techniques



terabit



Conclusion

- ❑ **Data Science** paradigm will accelerate scientific discovery
- ❑ **Scientific Computing** is serving to:
  - Transform the data (for the theory)
  - Reduce the data (implicit modeling)
  - Produce new data (especially for *extremes*)
- ❑ The **quality of the data** (and, thus, of the **software**) is a key issue.
- ❑ Software (as a *Digital Asset)* should also be FAIR!

- ❑ **Urgent Computing** will be a critical testbed for such issues.

Il Centro per le Geoscienze Computazionali

Pillar 1) Favorire l'accesso alle risorse di calcolo e di analisi dati, e loro armonizzazione

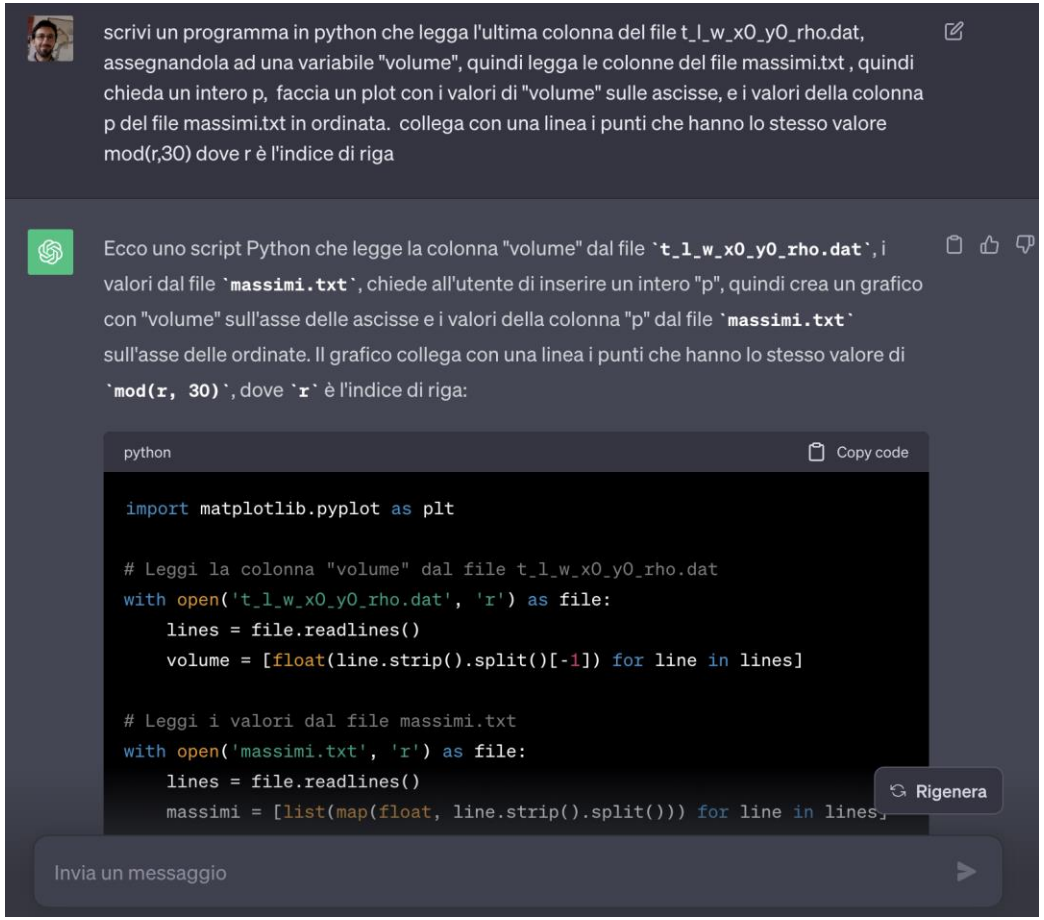
- Mantenere e gestire un sistema di calcolo di livello Tier3, per dare accesso a risorse (*hardware*) di calcolo ed analisi dati ad alte prestazioni alla comunità scientifica e tecnologica dell'Ente.
- Dare accesso a risorse *software* HPC/HPDA/AI.
- Promuovere l'ottimizzazione e l'armonizzazione delle risorse distribuite presso le Sezioni INGV, di concerto con il CSI (Centro Servizi Informatici), con i gruppi di lavoro di Istituto e le Unità Funzionali associate ai Centri di Elaborazione Dati (CED) di Sezione.
- Coordinare lo sviluppo delle risorse HPC/HPDA/AI dell'INGV in una prospettiva di scalabilità e sostenibilità economica ed ambientale.

Pillar 2) Potenziamento delle competenze (capacity building)

1. Coordinare le iniziative di formazione del personale, e condividere le conoscenze nell'ambito HPC/HPDA/AI.

2. Far crescere la comunità HPC/HPDA/AI dell'INGV, facilitando lo sviluppo tecnologico, il *porting* ed *enabling* di applicazioni ad alte prestazioni.

3. Supportare i servizi HPC/HPDA/AI dell'INGV per la valutazione della pericolosità associata a fenomeni geofisici, vulcanici ed ambientali.



(Intelligenza Artificiale: oltre ChatGPT!)

Pillar 3) Integrazione con l'ecosistema HPC Nazionale ed Europeo

1. Fornire servizi di Accesso Virtuale e Transnazionale (Virtual and Transnational Access) alle risorse HPC/HPDA/AI per gli ERIC (in particolare EPOS e EMSO), per i progetti con partecipazione INGV che lo prevedono, e per la comunità scientifica di riferimento.

2. Facilitare l'accesso alle risorse HPC/HPDA/AI di livello superiore in ambito Regionale, Nazionale ed Europeo.

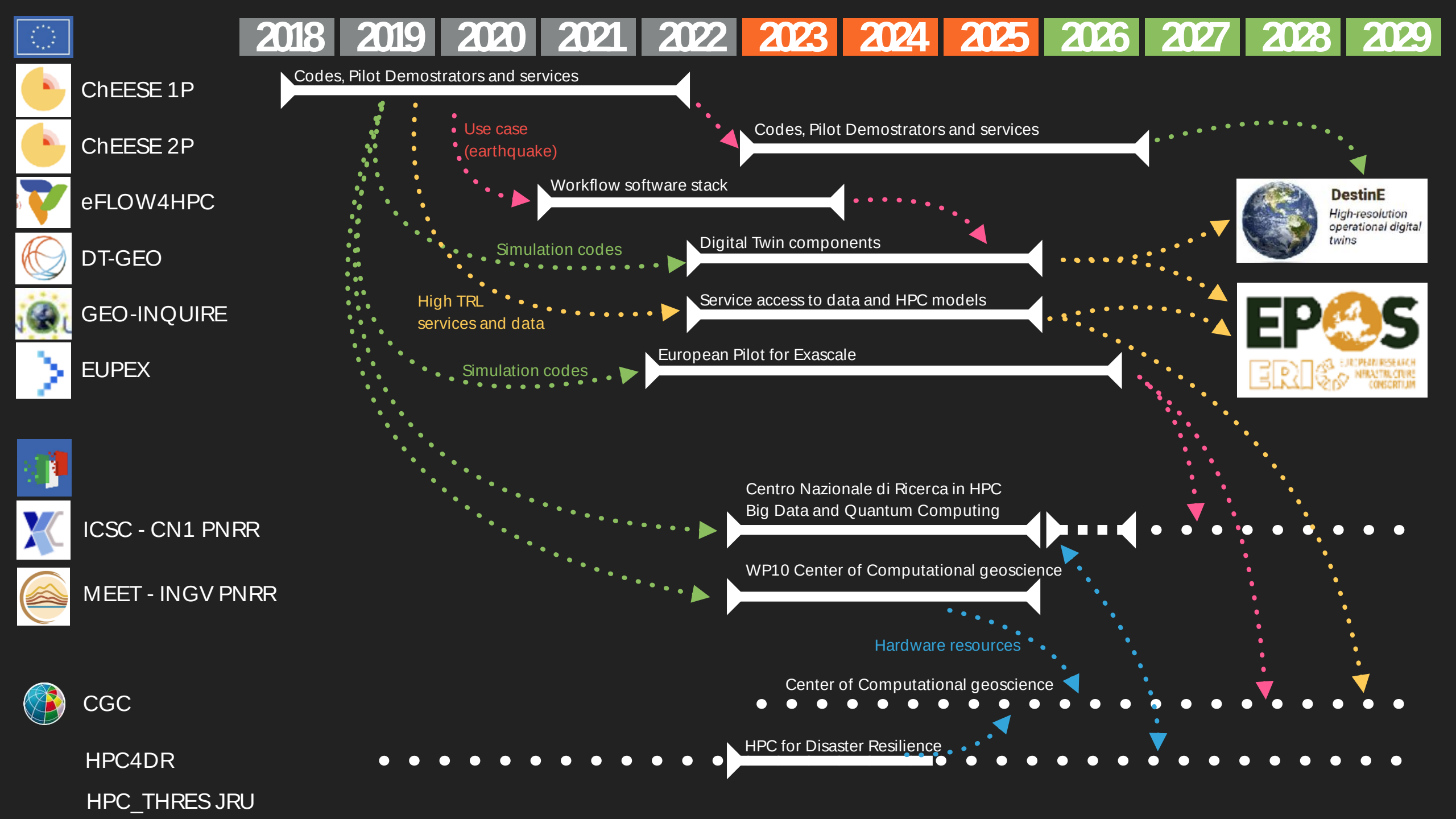
3. Integrare la strategia INGV con le iniziative della EOSC (European OpenScience Cloud) a livello Nazionale ed Europeo.

4. Supportare le attività a scala Europea (Horizon Europe) ed in particolare le attività progettuali legate ad EuroHPC e a Destination Earth.

EPOS
EUROPEANPLATEOBSERVINGSYSTEM

emso

EUROPEAN OPEN SCIENCE CLOUD


2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
ChEESE 1P
ChEESE 2P
eFLOW4HPC
DT-GEO
GEO-INQUIRE
EUPEX
ICSC - CN1 PNRR
MEET - INGV PNRR
CGC
HPC4DR
HPC_THRES JRU
Codes, Pilot Demostrators and services
Use case
(earthquake)
Codes, Pilot Demostrators and services
Workflow software stack
DestinE
High-resolution operational digital twins
Simulation codes
Digital Twin components
High TRL
services and data
Service access to data and HPC models
EPOS
ERIC
Simulation codes
European Pilot for Exascale
Centro Nazionale di Ricerca in HPC
Big Data and Quantum Computing
WP10 Center of Computational geoscience
Hardware resources
Center of Computational geoscience
HPC for Disaster Resilience

HPC@INGV vs TOP500 2021 supercomputers

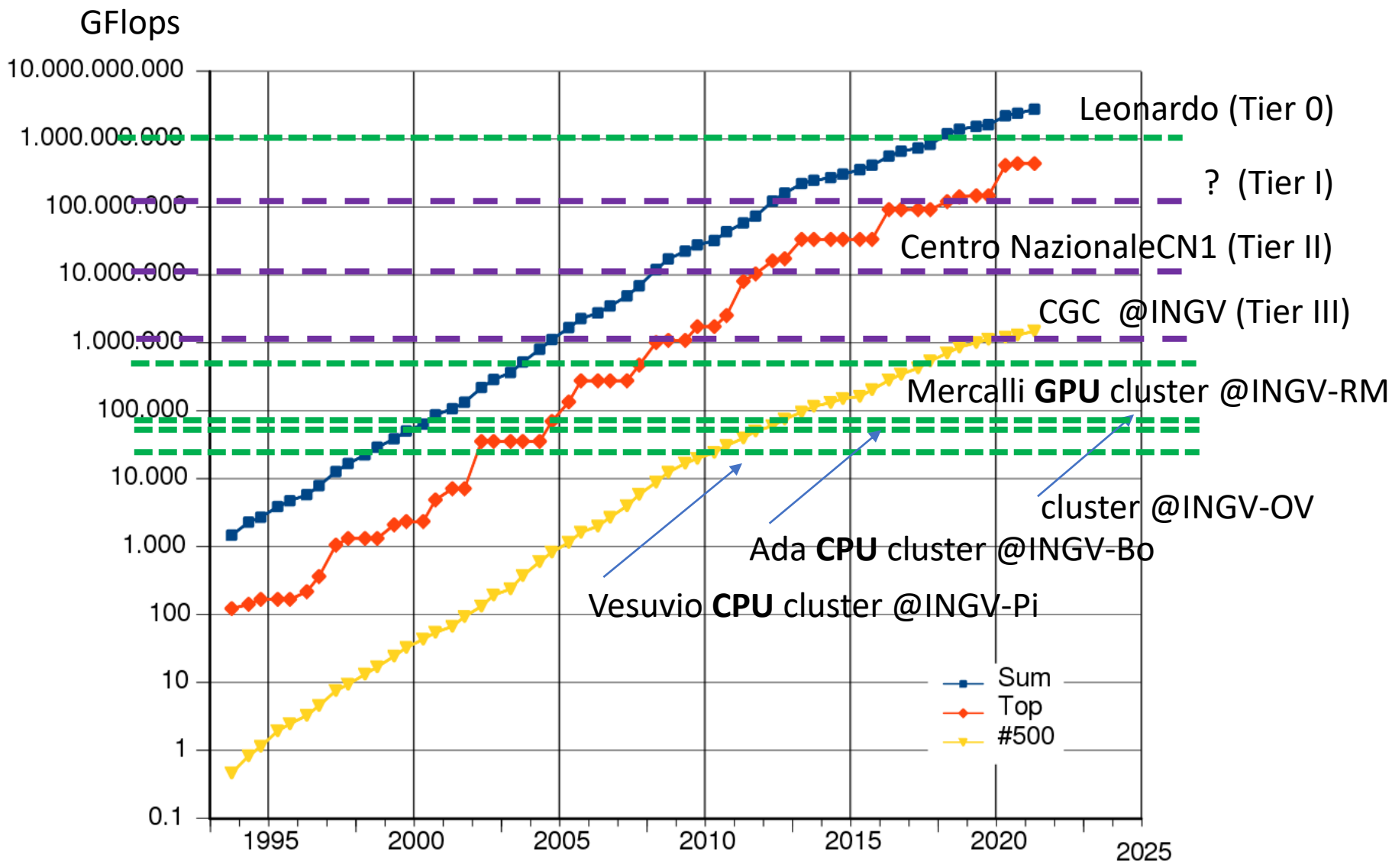
1 FLOPs
=
1 Floating-point operation / s

G = Giga (10^9)

T = Tera (10^{12})

P= Peta (10^{15})

E=Exa (10^{18})



HPC@INGV ecosystem and strategy

Target	Risorse di calcolo	Target	Sostenibilità
~ exaFLOPS	• Centres of Excellence • PRACE	• Scientific excellence (grand challenges) • Urgent computing • Probabilistic hazard assessment / forecast /early warning	European Commission (Long-term?)
~10 petaFLOPS	• Centro Nazionale HPC	• Hazard and risk assessment for natural catastrophes (**High TRL**) • Service to the society (**High TRL**)	• PNRR • ?
~1 petaFLOPS	• Centro Geofisica Computazionale INGV	• Fundamental research (**Low TRL**) • Service to the scientific community (**Intermediate TRL**)	• PNRR • Progetti Infrastrutturali • Servizi HPC
~ 10-100 teraFLOPS	• Sezioni INGV (**federazione**)	• Fundamental research and development (Low TRL)	• Progetti PON • EOSC • Fondi ordinari