Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

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

Highlights of ICSC research activities

Alessia D'Orazio

Spoke 3 Annual Meeting – La Biodola – 5-9 May 2024

Disclaimer

Not covering general description of Spokes' research programmes

Not comprehensive overview of Spokes' relevant activities and results: just hints of 1 or 2

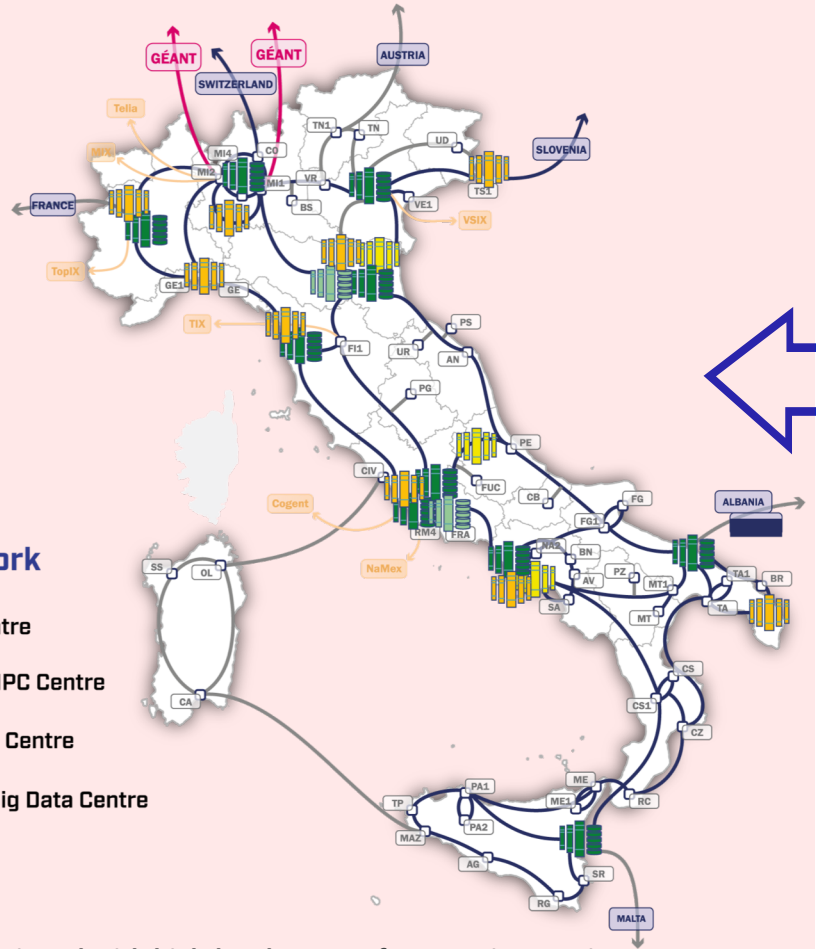
KPI: Already >> 1K publications tagged «ICSC»

Innovation Fund projects covered in D. Salomoni talk on Wed.

The ICSC will include

10 thematic Spokes and **1 Infrastructure spoke**

0 SUPERCOMPUTING CLOUD INFRASTRUCTURE



equipped with 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 APPLICATI

7 MATERIALS & MOLECULAR SCIENCES

8 IN-SILICO MEDICINE & OMICS DATA

9 DIGITAL SOCIETY & SMART CITIES

10 QUANTUM COMPUTING

Spoke 0 - Supercomputing Cloud Infrastructures

PoC of Cloud Federation (due 30/6/2024)

- ❑ Extend the INFN-Cloud approach to ICSC (and TeRABIT): demonstrate **interoperability** with CINECA
  - ✓ Identity federation, accepting CINECA and INFN registered users (Indigo IAM)
  - ✓ Computing federation with Galileo100 (PaaS orchestrator and dashboard), and «offloading» to Leonardo
  - ✓ Data management (Rucio/FTS)
- ❑ **Development needed** to bring the approach to production with multiple identity providers

Actual Access to existing resources for ICSC users

Differentiated for CINECA and INFN

- ❑ CINECA: usual user registration procedure
- ❑ INFN: adapting INFN AAI to register ICSC users: *release end of May*
  - ✓ ***For all cloud users***: Employer's attestation of user's training as an IT user and/or IT sysadmin required

Resources Procurements

- GARR: procedures for network extension almost complete
- CINECA: RFP for Leonardo upgrade (LISA) in preparation
- INFN: First tranche of HTC resources being deployed, more to come in 2025

Quantum educational system

NEW

- Short-term purchase of a general-purpose quantum computer for scientific use.
- Operating using a "gate model" system and qubits technology based on superconductors
  - ✓ Small QPU (no more than 20 qubits)
- Integrated with LEONARDO supercomputing.
- **Used as educational machine for users**.
- In addition to Quantum machine with neutral atoms technology → from EuroHPC tender (ICSC co-funding)

PLAN: making Leonardo's HPC environment prepared to accomodate almost all currently existing qubits technologies

Spoke 1 – Future HPC & Big Data

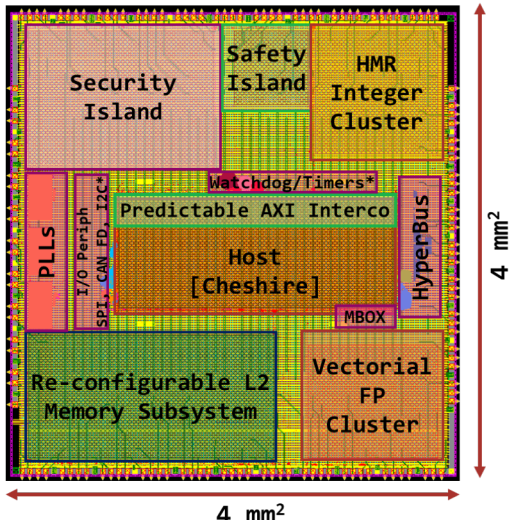
Living Lab: Hardware and Systems (HWS)

Monte Cimone (MC) series of RISC-V HPC prototypes

Upgrade MC1 → MC2: 40x BW, 64x FLOPS (DP)

x2 Sophgo SG2042
- 64-cores T-head C920 RISC-V CPU
- 2GHz
- 64MB cache
- PCIe Gen4x16
- RV vector extension 0.71

Delivered the 1st RISC-V based system for Automotive (CARFIELD platform)
- Tape-out in Intel16 Technology 22 November, 2023



Next Step: Design of open-source RISC-V System on Chip for Space (Astral)

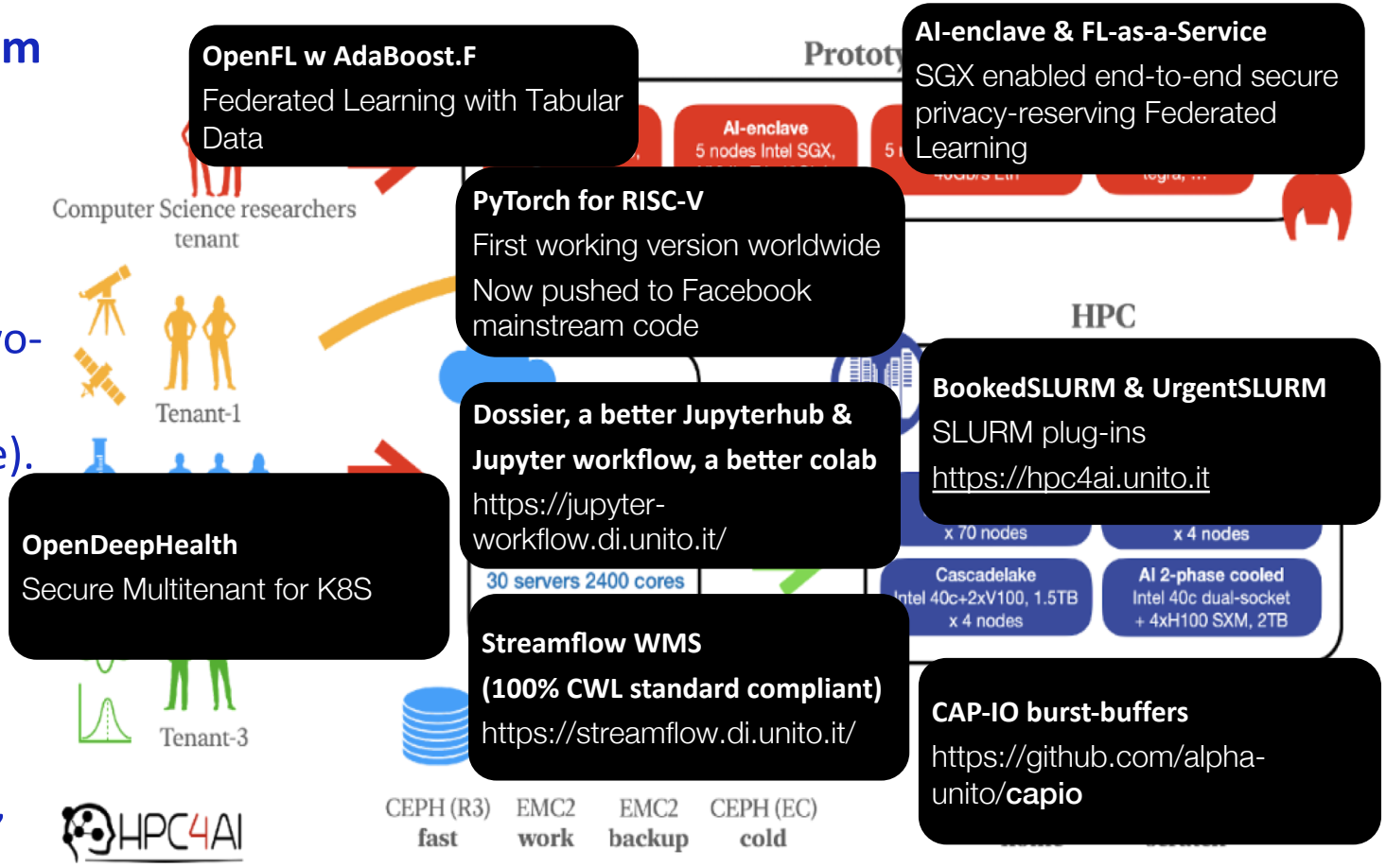
Spoke 1 – Future HPC & Big Data

Living Lab: Software & Integration (SWI)

HPC4AI cloud-HPC system of systems

Most of the middleware of the cloud-HPC system is «made-in-ICSC»

- ❑ OpenStack cloud coupled with modular HPC system with UNITO software
- ❑ Two-phase cooling prototype Intel 2 sockets + 4xH100 - first commercial GPU machine with two-phase cooling (fully user-programmable
- ❑ 4 x GraceHopper GH200 (150 servers worldwide). 4 x NVidia DevKit Arm+2xA100+2xBF2.
- ❑ RISC-V: SiFive, Esperanto, Canaan 908 (vec 1.0)
- ❑ NVidia AI lighthouse university: 300 NVidia Enterprise free licenses available
- ❑ **Several in-house middleware tools used in Innovation Grants** : StreamFlow WMS, CAPIO burstbuffers, JupyterWorkflow, UrgentSLURM, AI-enclave Federated Learning



https://hpc4ai.unito.it/livinglab-swi-icsc/

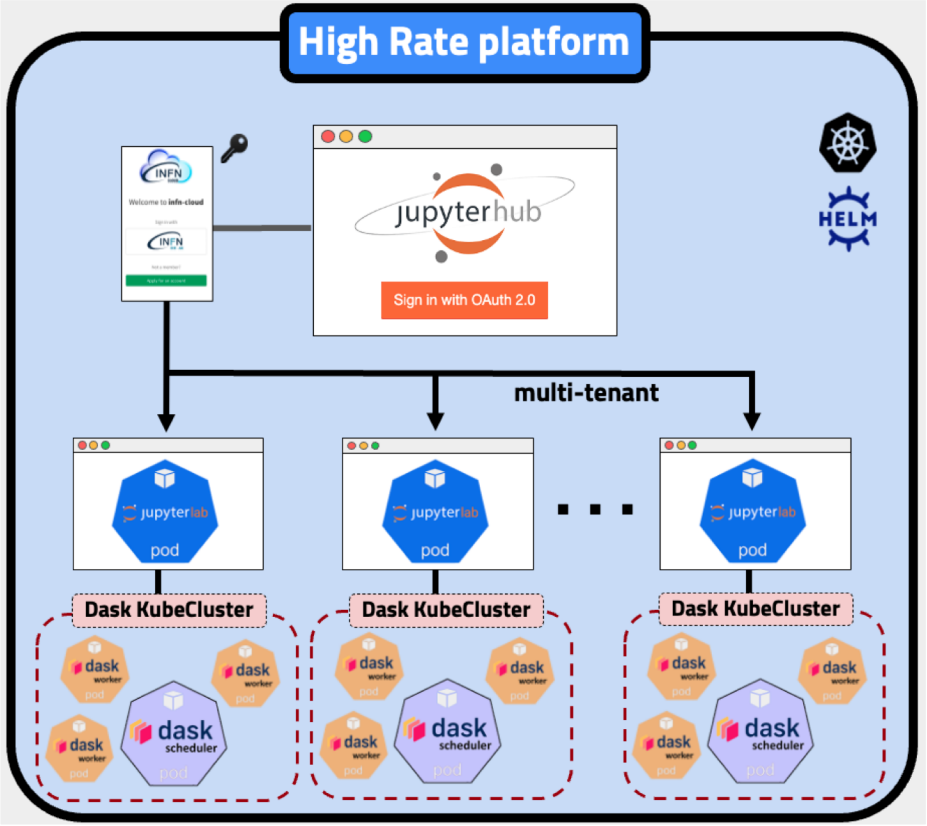
Spoke 2 – Fundamental Research & Space Economy

Quasi interactive analysis of High Energy Physics big data with high throughput

Tommaso Diotalevi[1,3], Francesco Giuseppe Gravili[1,7], Muhammad Anwar[2], Matteo Bartolini[1,5], Antimo Cagnotta[1,6], Adelina D'Onofrio[1], Paolo Mastrandrea[1], Gianluca Sabella[1,6], Federica Maria Simone[1,2], Bernardino Spisso[1], Alessandro Tarasio[4], Tommaso Tedeschi[1]

Flagship use case

https://shorturl.at/hoFIX



- Demonstrate the capability to process large amounts of data (GB-TB-PB) via a **Cloud infrastructure** extendable to 1-100-10000s nodes.
- Tested on HEP Samples, but data-agnostic – even for sensitive data (using an industry-grade Authorization and Authentication Infrastructure)

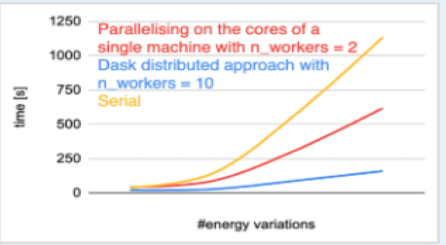
Z*ee* simulations at Future Colliders (FCCee)

Feasibility study: Mimic systematic variations applying a gaussian smearing to e^+e^- energies many times.

Preliminary results

Events selection and histogramming: interactively with **ROOT RDataFrame** and **Jupyterlab** [4]

Dask [5] used as backend.



- Considering the overall execution time as metric and running the same workflow, there is a performance improvement in the distributed approach wrt the standard/serial approach;
- Moreover, it was tested that scaling resources, the performance further improves.

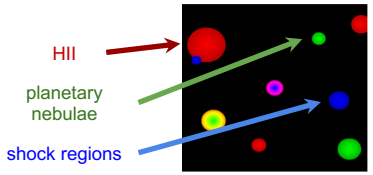
Will be eventually provided as a service in the ICSC Datalake.

Spoke 2 – Fundamental Research & Space Economy

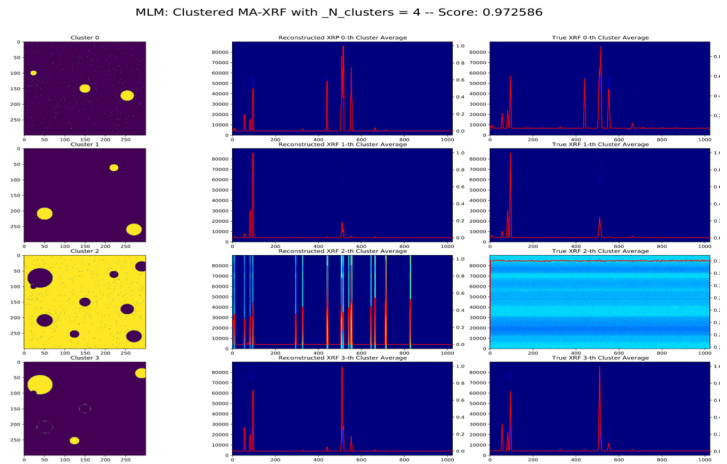
From stars to frescos: extended spectral computer vision

- Datacube segmentation via Deep Spectral Clustering: cutting-edge technique aiming **adding a 3rd dimension to images**.
- First application to Synthetic Astrophysical Datacubes, obtained from images with HII, planetary nebulae and Shock regions, including synthetical nebular optical spectra.
- Extended to pairings and frescos with embedded spectral measurements

Synthetic Astronomical Image

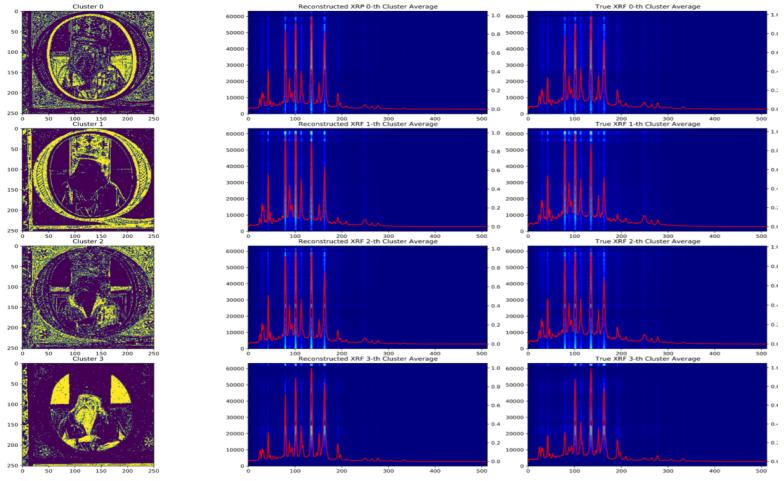


Clustering through embedded spectra

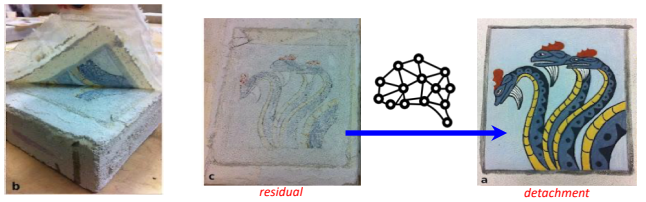


Miniature contained in the Codex Heroica by Phoilostratus, illuminated by the Florentine painter Attavante degli Attavanti

Original Decoded

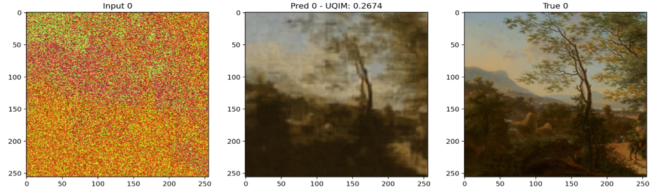


Digital restoration of frescos and paintings (in progress)



Test on several DNN Train on multiple synthetic dataset. architectures

Input Predicted Original



https://arxiv.org/pdf/2401.17695

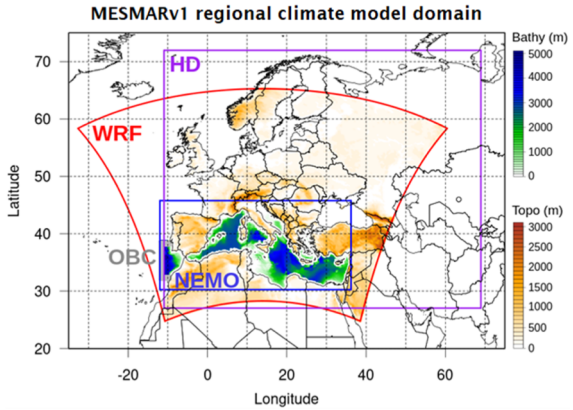
Spoke 4 – Earth & Climate

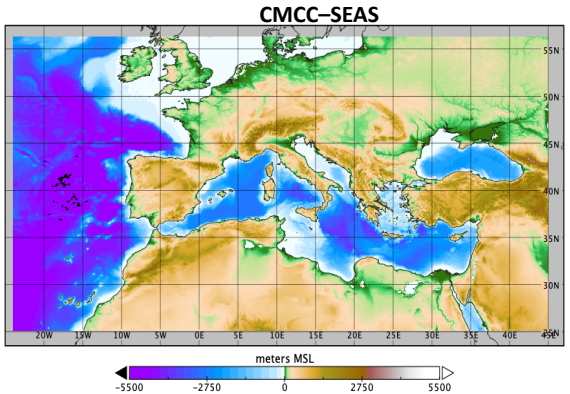
Development of a set of climate models to produce new high-resolution climate change scenarios for the Mediterranean basin and the Italian Peninsula

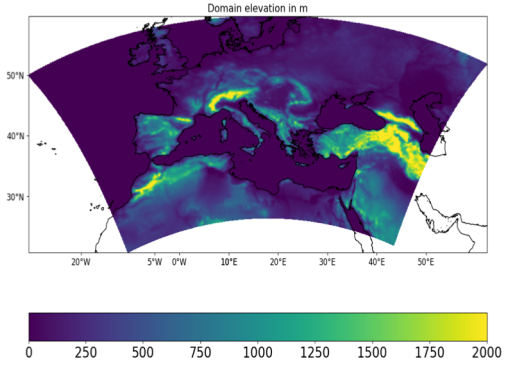
Coupled models	Components	Coupler	Test/evaluation	Configuration	Partner
CMCC-SEAS	WRF, NEMO	OASIS3	Test 2021 JFM & JAS	South European Seas	CMCC
RegCM-ES	RegCM5, MITgcm, BFM, ChyM	ESMF	Test 79-86	Mediterranean	OGS
RegCM-ES	RegCM5, MITgcm, BFM, ChyM	ESMF	Test Jan 2006-Dec 2011	North Adriatic	OGS
MESMAR v1	WRF, NEMO, HD	OASIS3	Sens. Data ass tests 1998-2022	Mediterranean	CNR-ISMAR
ENEA-Reg	WRF, MITgcm, HD	ESMF	ERA5/1980-2014 CMIP6 SSPs - SLR	Mediterranean	ENEA

Interesting for scientists on climate change impact on several sectors: agriculture, water management, energy production, insurance, etc..

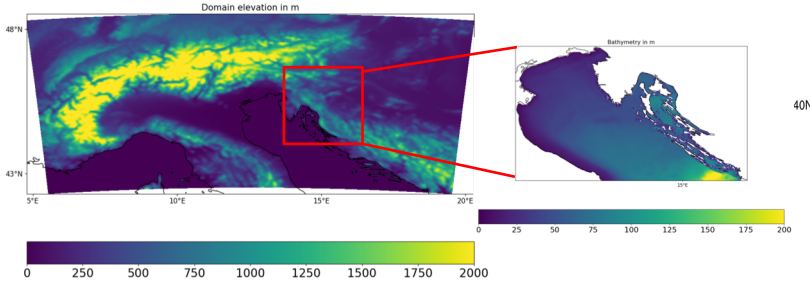
Reg-ESMs – coupled configurations and simulations



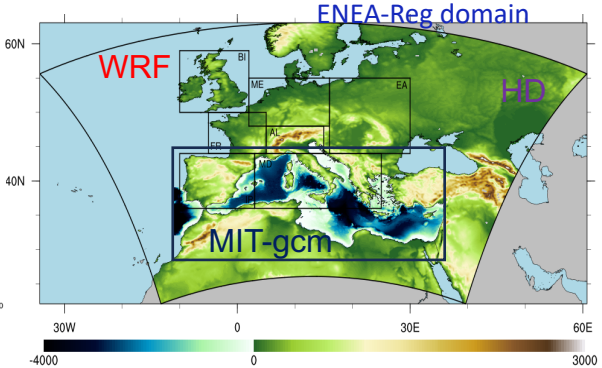




RegCM-ES domains:
Med (12km;1/12°)
North Adriatic (3km CPM;1/128°)



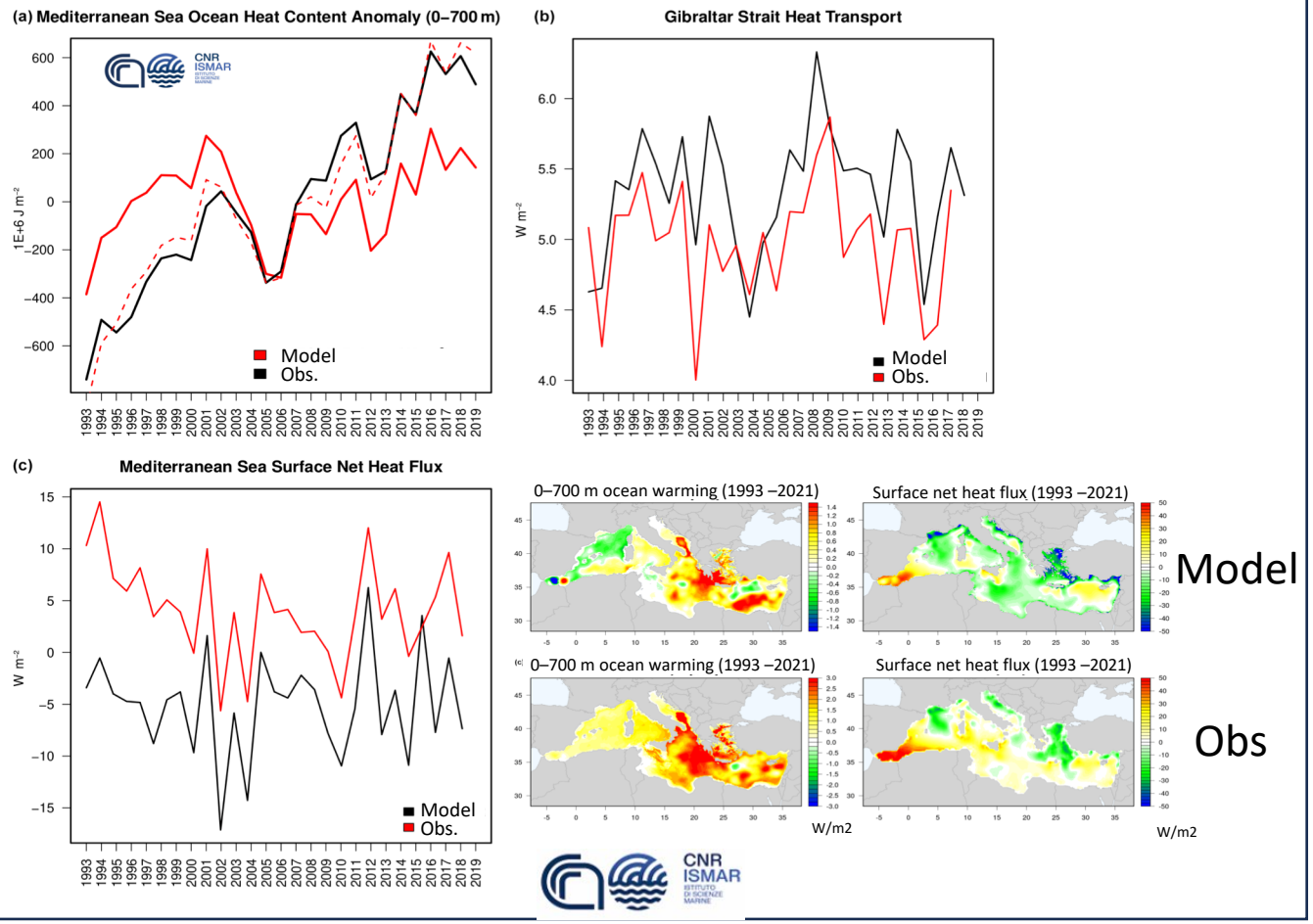
ENEA-Reg domain



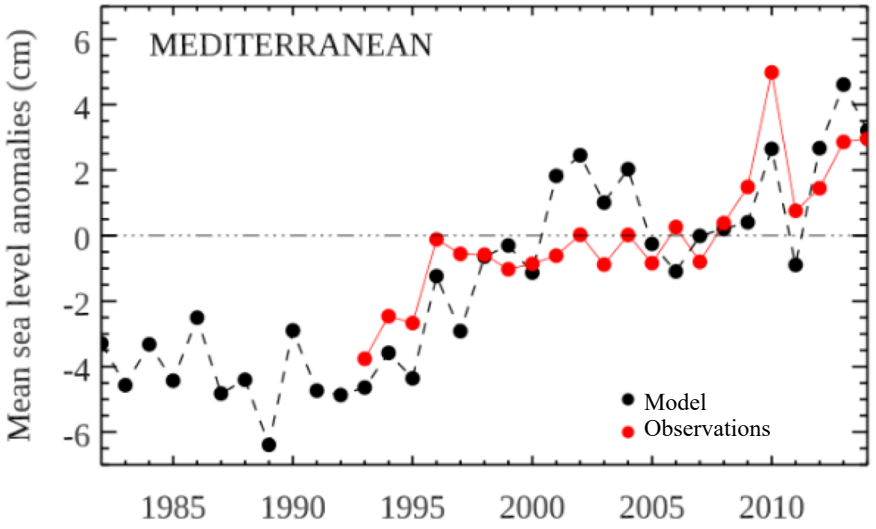
Spoke 4 – Earth & Climate

Reg-ESMs – coupled configurations and simulations – Preliminary results

MESMAR- Very good variability and consistency of year-to-year OHC and Heat Budget terms versus independent reanalyses (although offset exists)



Simulated average sea surface elevation (red dots = satellite observations), reference period 1983–2014



ENEA

Spoke 5 – Environment & Natural Disasters

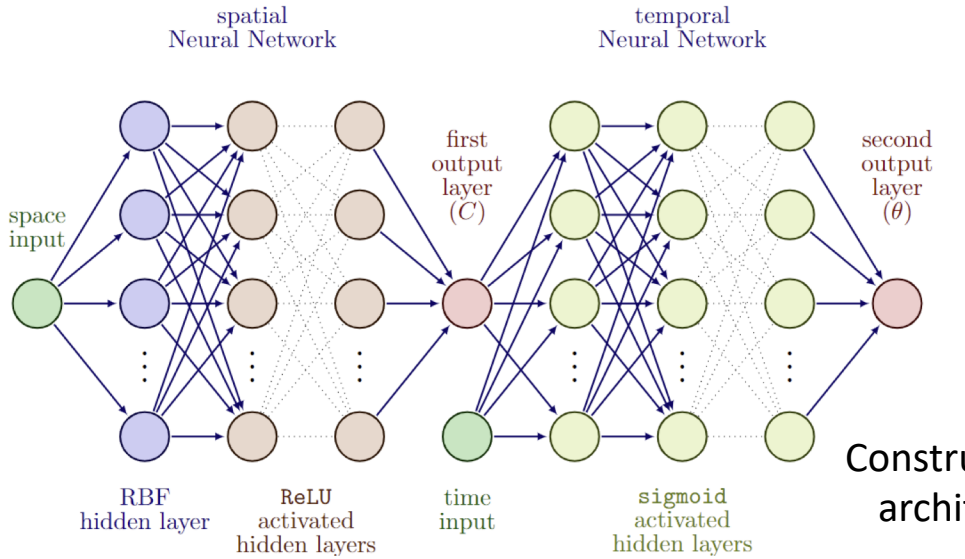
Multi-hazard modelling and analysis of damages and losses (WP5)

Aims at delivering a **framework for models and simulations to deal with multi-hazard risks**

- **Primary objective**: **identification of regions** where multi-hazard are present with non-negligible effects
  - ✓ Development of models and simulations for selected sites

Data-driven approach for calibration of nonlocal models kerne in unsaturated soils

- Radial basis functions (RBFs) are used as activation functions in Physics-Informed Neural Networks (PINNs).
- Model: data-driven method for solving an inverse problem in peridynamics kernel computation



Constructed NN architecture

Approach interesting for:

- Solving fluid dynamics or advanced materials physics problems (Spoke6)
- Speeding up numerical solutions of meteorological equation systems (Spoke4)

Models useful for:

- Decision-makers: can assess overall vulnerability and prioritize mitigation efforts effectively

F.V. Difonzo, L. Lopez, S.F. Pellegrino. Physics Informed Neural Networks for an Inverse Problem in Peridynamic Models. Engineering with Computers. To appear (2024).

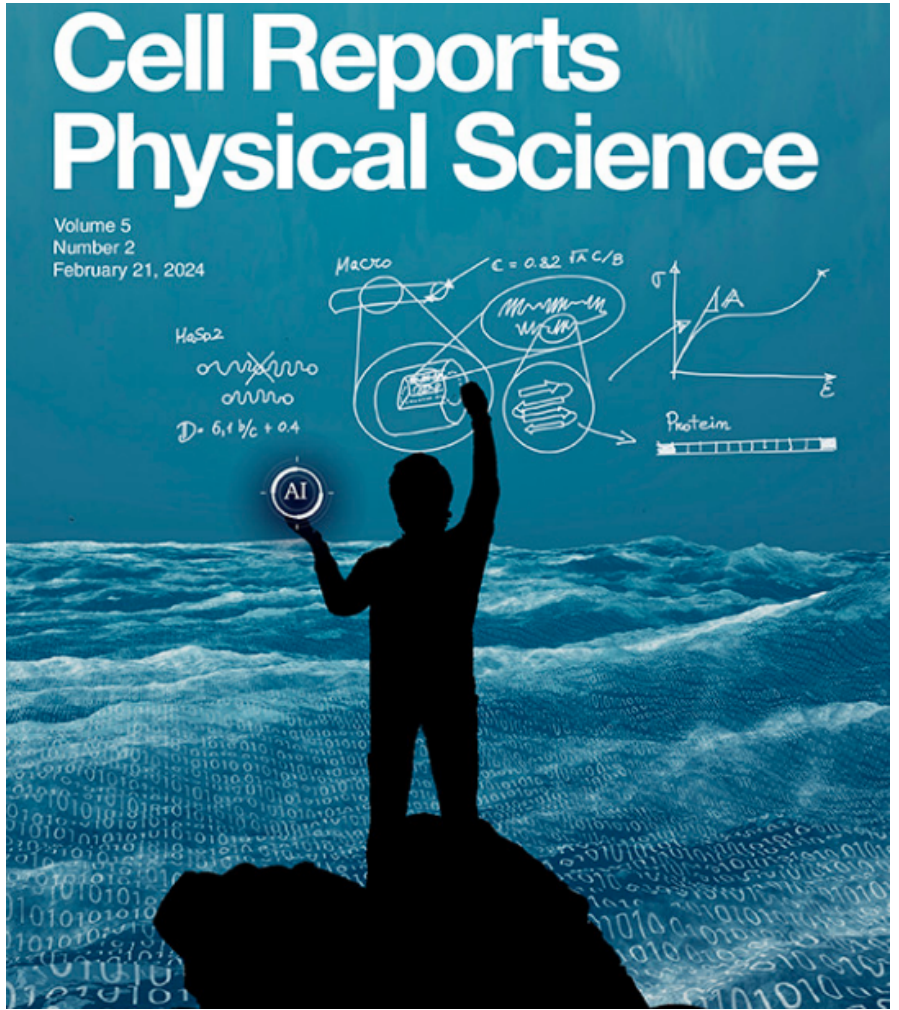
Spoke 5 – Environment & Natural Disasters

Definition of a Machine Learning, Quantum Computing and AI platform to design and exploit Digital Twins (WP7)

Identification of new frameworks for deduction of random laws in hierarchical systems (soft matter)

Evolutionary Polynomial Regression (EPR) Data Modeling Technique:

- ❑ Evaluates multi-scale correlations in hierarchical materials
- ❑ Tested with biological material data and engineering optimization problems; it will be extended to other materials relevant for Spoke 5
- ❑ Allows deduction of causal relationships between measurable quantities at different scales



V. Fazio, N. M. Pugno, O. Giustolisi, G. Puglisi. Physically-based machine learning for hierarchical materials. Cell Reports Physical Science, Volume 5, Issue 2, 2024, https://doi.org/10.1016/j.xcrp.2024.101790.

Approach interesting for:

- ❑ Solving fadvanced materials physics problems (Spoke6)

Spoke 6 – Multiscale Modelling & Engineering Applications

8 Flagship Areas

FP1 - Advanced Materials and New Devices
FP2 - Big Data Analytics & AI
FP3 - Computational Fluid Dynamics
FP4 - Green Energy
FP5 - Life Science
FP6 - HPC for Complex Systems
FP7 - Electro magnetics Systems
FP8 - Innovation Projects

WP1: Simulations of complex multiscale and multiphysics models

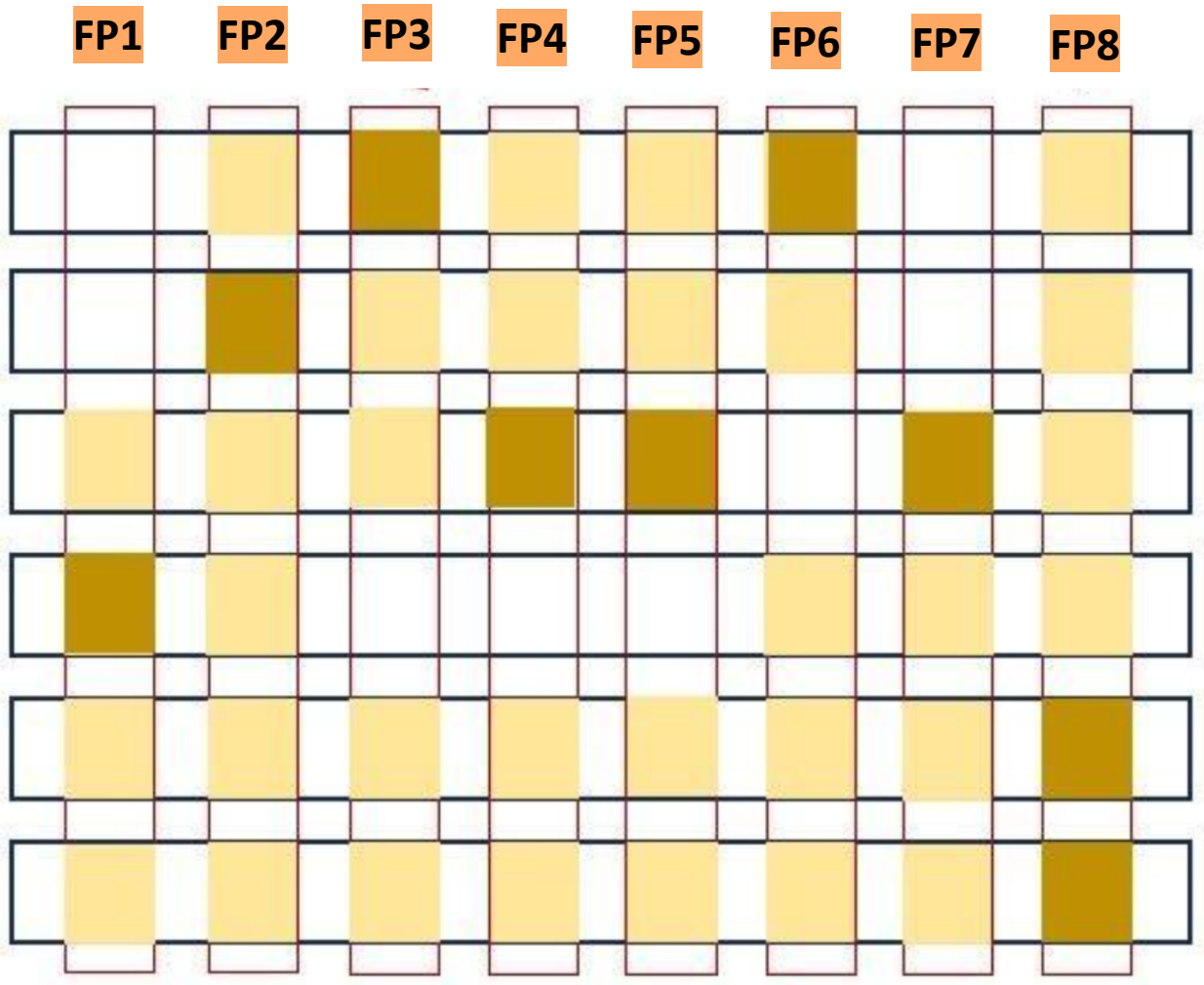
WP2: Simulation of large systems via multiscale and AI methods

WP3: Digital twin modelling and tools for engineering disciplines

WP4: Engineering applications: verticals and method/tools integration

WP5: Proof-of-concept PoC demonstrators

WP6: Training, Technology transfer activities and Outreach



Spoke 6 – Multiscale Modelling & Engineering Applications

HPC for complex Systems(FP6)

- ❑ Quantum annealing promises to solve hard combinatorial problems (e.g., the traveling salesman problem) much faster than any existing method
- ❑ Actually, so far there was not a full consensus about the feasibility of QA
- ❑ To shed some light on the problem, we developed highly tuned codes for Monte Carlo simulations and Transfer Matrix diagonalization (The QISG suite: High-performance codes for studying quantum Ising spin glasses, DOI: 10.17632/g97sn2t8z2.1)
- ❑ Used 1024 GPU for 10 months for a total of ~ 7 millions GPU hours to measure with unprecedented precision crucial properties of the system.
- ✓ Massimo Bernaschi Isidoro Gonzalez, Victor Martin-Mayor, Giorgio Parisi
  The Quantum Transition of the Two-Dimensional Ising Spin Glass arXiv:2310.07486
- ❑ Results confirm that a quantum annealer may work (there is no fundamental obstacle as some authors suspected). In March 2024, a D-Wave team confirmed experimentally their results! **arXiv:2403.00910**

  This does not mean that is easy to build a real QA!
- ❑ Work is in progress to extend our HPC approach to the study of other complex systems as part of one of the Spoke6's flagship projects.

Spoke 7 – Materials & Molecular Sciences

Flagship Code (WP1)

extension, optimization, maintenance, and distribution of world-leading flagship codes for the high-performance numerical simulation of materials and molecular systems.

new developments have been included in at least one “interim release” (not public, only meant for benchmarking) for each code.

Deployment of codes on available HPC architectures

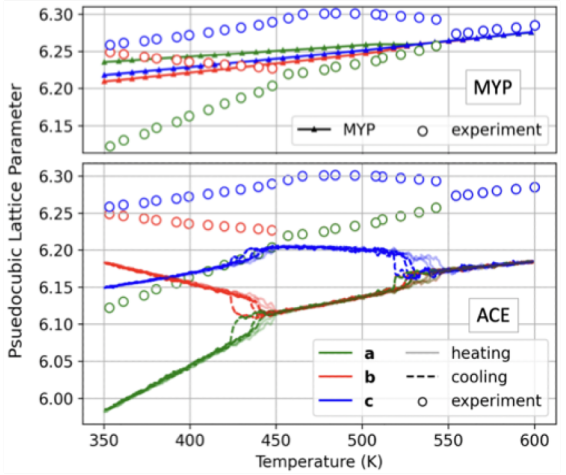
All flagship codes installed on the Leonardo machine

documented in the official website for chemistry and physics

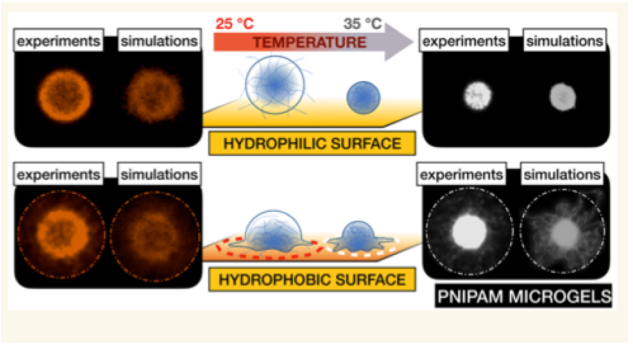
Spoke 7 – Materials & Molecular Sciences

Data-driven methods for materials and molecules

- **Advanced machine-trained interatomic potentials**
  - ✓ development of ML interatomic potentials for perovskites obtained from Bayesian linear regression of atomic cluster expansion coefficients
  - ✓ different interpolation/ML techniques evaluated
- **Novel machine learning methods** working in progress
- **Artificial intelligence for macro-scale simulations:**
  - ✓ elastic properties of microgels at a liquid-liquid interface as a function of increasing packing fraction
  - ✓ implementation of a phase-field (PF) code to simulate the dynamics of spinodal decomposition
  - ✓ development of neural-network based methods to learn large-scale continuum temporal evolution from phase field



W. Baldwin et al.,(2023), under revision
https://doi.org/10.48550/arXiv.2304.04714



X. Shaulli et al. ACS nano, 17 (2023)
https://doi.org/10.1021/acsnano.2c07569

Spoke 8 – In Silico Medicine & Omics Data

Development of vertical solutions for In silico Trials and Digital Twins in healthcare

- ❑ Beyond the state of the art of reverse vaccinology: predicting vaccine efficacy with the universal immune system simulator for influenza. *Giulia Russo, Elena Crispino, Avisa Maleki, Valentina Di Salvatore, Filippo Stanco and Francesco Pappalardo, BMC Bioinformatics (2023) 24:231* https://doi.org/10.1186/s12859-023-05374-1
  - ✓ reverse vaccinology is an in silico approach to identify potential proteins for vaccine production
    - ➢ it includes bioinformatics tools and processes like subtractive proteomics, computational vaccinology, immunoinformatics, and in silico procedures.
  - ✓ current state of reverse vaccinology lacks the ability to predict vaccine efficacy: advanced immune system simulator proposed to test efficacy of a vaccine against the designed disease.

- ❑ A web-based platform for extracting and modeling knowledge from biomedical literature as a labeled graph. *Antonio Di Maria, Lorenzo Bellomo, Fabrizio Billeci, Alfio Cardillo, Salvatore Alaimo, Paolo Ferragina, Alfredo Ferro, Alfredo Pulvirenti, Bioinformatics, btae194,* https://doi.org/10.1093/bioinformatics/btae194

- ❑ NetMe 2.0 is a web-based platform that extracts biomedical entities and their relations from input texts and models these entities as a BioMedical Knowledge Graph (BKG)

tailored for biomedical sector but its underlying technology and approach to entity extraction and relationship modeling can potentially be adapted for use in other scientific domains: ex. environmental science to track research on climate change or in chemistry to map out chemical compound interactions from research papers

Spoke 9 – Digital Society & Smart Cities

Digital Twin Space: The Integration of Digital Twins and Data Spaces https://ieeexplore.ieee.org/document/10386737

- Urban Digital Twins (UDTs) are complex examples of Digital Twins, virtualizing urban assets and providing advanced analysis and prediction services.
- The Digital Twin Space architectural proposal aims to convert Data Spaces into DTs, using the **UDT as a running example.**
- **The paper presents a model to bridge the gap between the Digital Twin paradigm and Data Space concept.**
- The model was validated on a smart photovoltaic (PV) prototype, demonstrating all components fit.
- DTS is still not fully compliant with data sovereignty and FAIRness principles, therefore next steps will include the complete mapping of the DTS on the Data Space definition provided by the International Data Space Association
- Further validation of the model in the case of smart buildings and a suburban area.

Spoke 9: many preliminary results (prototypes) all somewhat on the same level: difficult at the moment to extrapolate only some that are more relevant or promising

Spoke 10 – Quantum Computing

WP1. Software	WP2. Mapping, compilation and quantum computing emulation	WP3. Firmware and hardware platforms

Quantum Applications in Science and Technology (WP1)
Application of **noise-robust** quantum computing algorithms in scientific and technological domains such as fluid dynamics, material science, electronic structure, bioinformatics, and image compression

Quantum Simulation and Modeling (WP1)
Quantum algorithms for noise shaping in quantum hardware, simulation of quantum physics processes, real-time evolution of quantum systems, for modelling of quantum noise in NISQ circuits, quantum gauge systems, solid-state physics modeling

Quantum Simulations and Emulation (WP2)
Investigation into the scalability of quantum algorithms. This also involves performance analysis and tuning of quantum emulators with specific backends

Quantum Benchmarking (WP2)
Implementation of an evaluation infrastructure for the comparative benchmarking of quantum algorithms against classical ones

A lot of publications and preprints

SSI – Transversal Research Group on Societal Implications and Impact

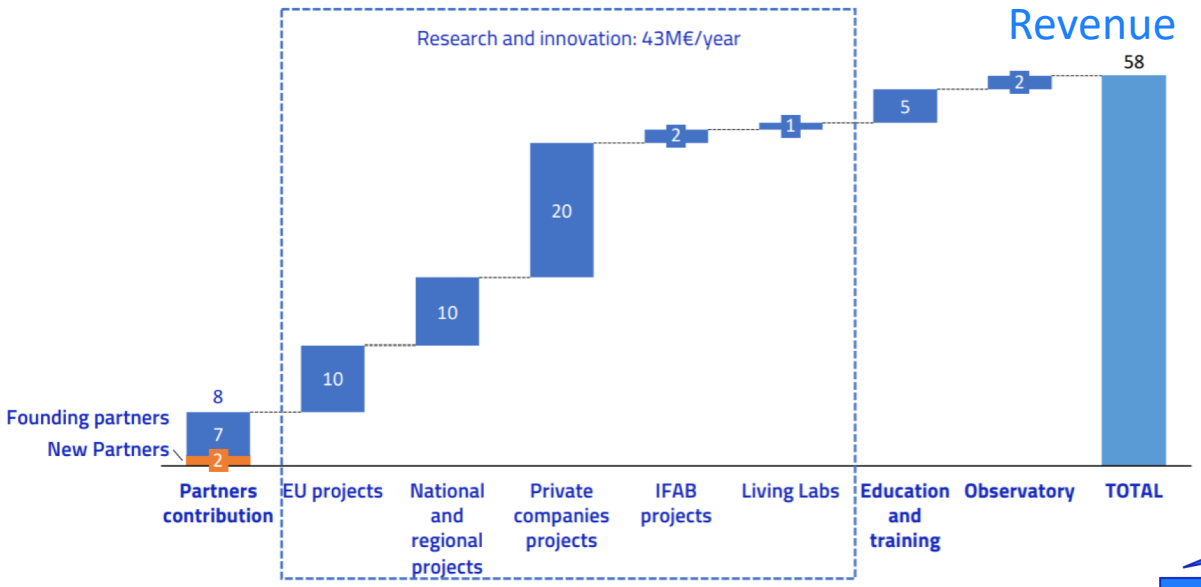
SSI research group of Big Data and HPC is an inter-spoke group that will operate within the Spokes' Board.

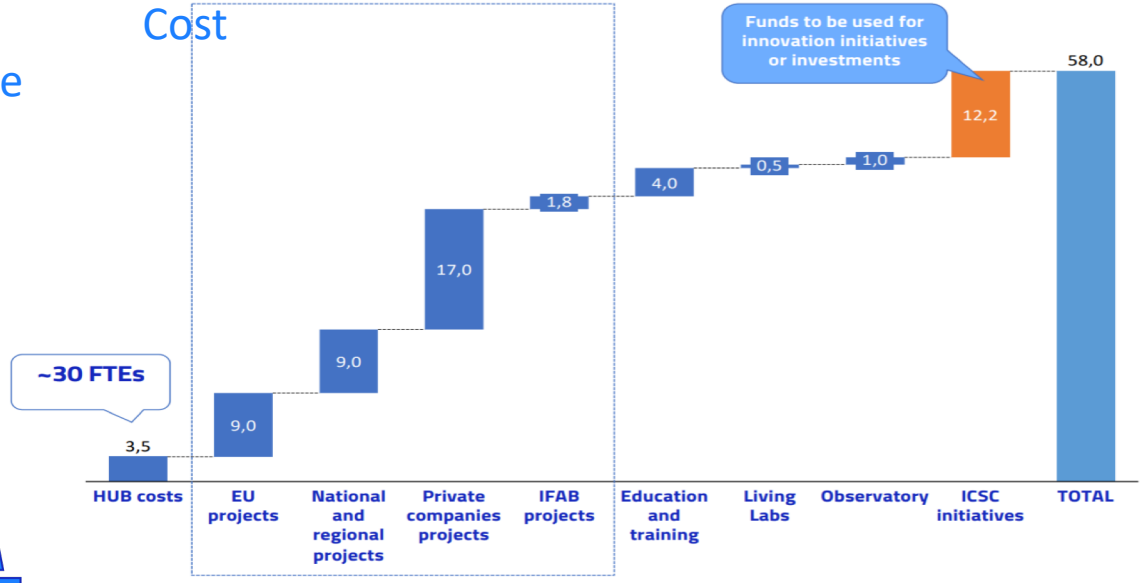
- Carry out research activities contributing to assessing the impact of advancements in High-Performance Computing (HPC), Artificial Intelligence (AI), and Big Data
  - ✓ Include, for instance, complex socio-technical dimensions encompassing scientific, social, ethical, legal, and political implications of these technologies.
- Aims to provide insights that inform policy, drive innovation, and contribute to the responsible development and deployment of advanced technologies in society.
- specifically address the diverse dimensions of impact and risk associated with AI (as delineated e.g. in the AI Act).

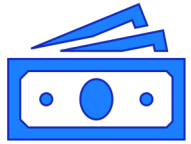
Research Group composition

- Luciano Floridi (Leader) –UniBO
- Antonino Rotolo (co-leader)- UniBO
- Nicola Barban - UniBO
- Guido Boella – UniTO
- Aldo Gangemi – CNR
- Paola Inverardi - UniAQ
- Biagio Aragona - UniNa

ICSC Long – Term Sustainability (beyond 2026)







- Current analysis suggests a cost profile of around 58 M€/year
  - ✓ Includes revenues and costs from partners, EU and national projects, living labs, training, industrial projects, Observatory outcomes exploitation, and Foundation costs.
  - ✓ The number is a "raw number" and should be continually re-evaluated based on market conditions, opportunities, and feedback.

But well before coming to the “money”, **we must HAVE a CLEAR VISION and purposes**

ICSC – Vision for sustainability beyond NRRP

Start from the beginning
What we claimed in the ICSC proposal

Vision

In the aim of the call, the Centre has the role of federating and coordinating persistent development process for the country's most significant assets in the HPC, Big data and Quantum Computing domain, in order to unlock the discovery and innovation potential, build a globally attractive ecosystem, strengthen Italian competitiveness and contribute to Europe leadership.

The Centre intends to fulfill this strategic role by multiplying the opportunities for partners and stakeholders, by building capacity and maximizing the socio-economic impact in a competitive environment, by implementing something that does not exist today, without overlapping or competing with individual institutes and research organizations already present, instead creating added value for its partners and for society. The key to the success of the initiative relies precisely in the ability to create these conditions and the appropriate ramp up in the start-up phase (PNRR phase: 2022-2025): a unique ecosystem that is strategic for society and for the country and capable of creating value for their partners by playing a role that individually they could not afford or sustain.

How to implement this vision? Where ae we going?

ICSC – Vision for sustainability beyond NRRP

Some questions that require an clear and shared answers

What is the **added value** that ICSC brings to academia, industry and society?
What, specifically, about our value proposition makes us "**unique**"?

What are our **customers**' fundamental, long-term **needs**?
Who are these prospective customers/clients?

What part will **the ICSC Foundation play**?
We shall be considering its evolution, for instance, in terms of its relation to stakeholders, competence, assets, services, toward Italy and beyond.

Current **hub and spoke model**, required by the PNRR action. But, neither the current configuration, nor the current ICSC composition should be considered a dogma.
Is the actual HUB and thematic Spoke configuration sustainable beyond the PNRR?

a *future governance structure and model* must take into proper account inputs from both academia and industry.

Thanks for your attention