Blueprint Workshop: Towards a National-Scale AI Collaboration in HEP

Position of **Italy/INFN**

Presenter: Prof. Daniele Bonacorsi (INFN, University of Bologna)

18–19 May 2026
SPIA, Washington DC

(AI-oriented) infrastructure in Italy/INFN [1/2]

National-scale AI and computing ecosystem coordinated by INFN

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

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

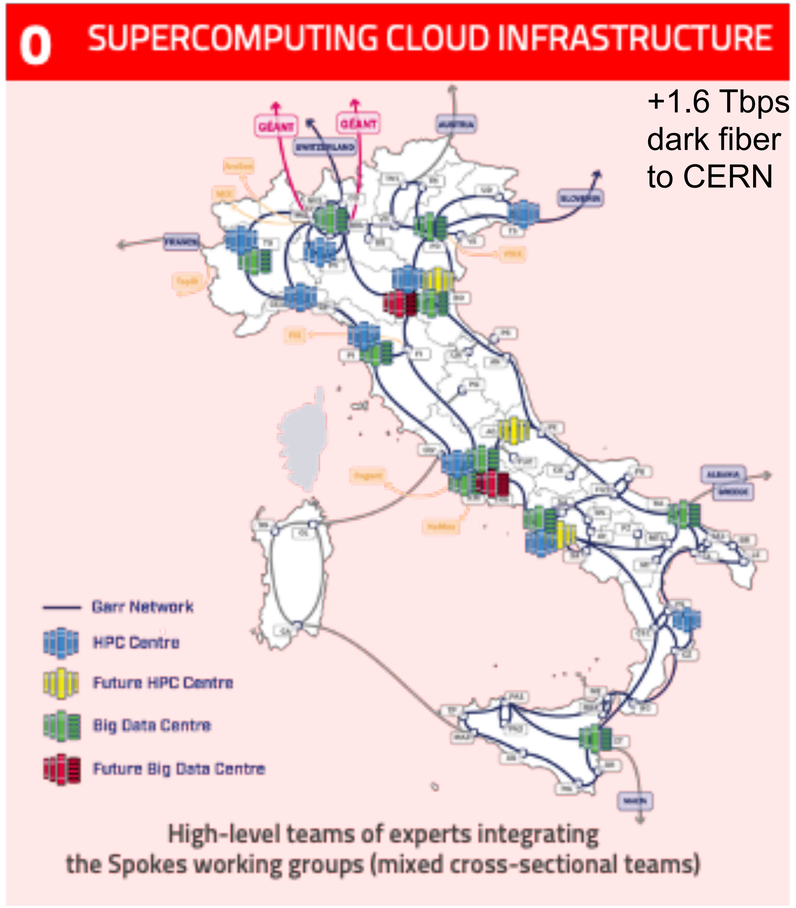
Overall, a contry-wide investment of >1B€ across advanced computing infrastructures, data platforms, and next-generation digital services

Main nodes:

- **Bologna DAMA Technopole** as flagship hub (integrating HTC, HPC, Cloud, Quantum)
- **Federated nationwide infrastructure** composed of (10) INFN WLCG centers (HTC, HPC, Cloud)

All major sites interconnected at ~Tbps scale to enable seamless large-scale AI training, distributed data analysis, and real-time scientific workflows. Soon: common Single-Sign-On (SSO), unified Data Management (DM)

AI-view → **designed as a platform for AI-driven science**, capable of supporting foundation-model training, large-scale simulation, real-time inference, distributed agentic workflows for data-intensive research communities



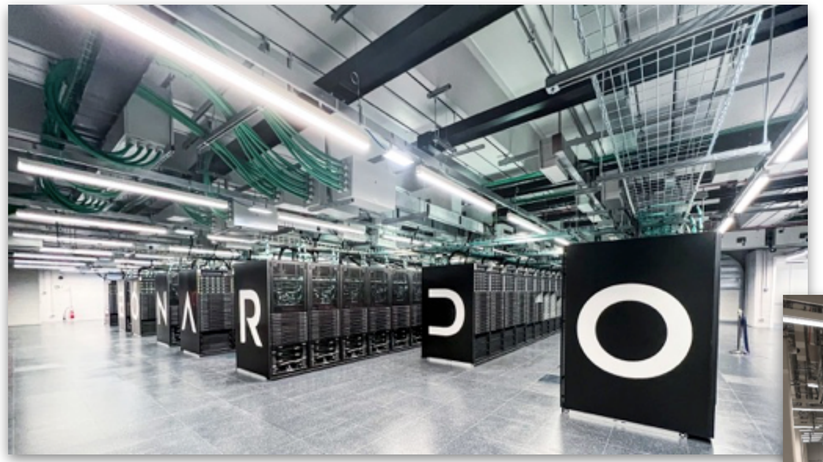
(AI-oriented) infrastructure in Italy/INFN [2/2]

Leonardo (CINECA EuroHPC flagship system): ~14,000 Nvidia A100 GPUs, ~6,000 compute nodes, ~250 PetaFlops

Tightly integrates HPC, AI, and data-intensive computing workflows. Hosts an AI-optimized accelerated partition for foundation-model training and large-scale inference

LISA AI: next-generation AI-focused platform: ~1,300 Nvidia H100 GPUs, ~1,300 PetaFlops BF16 AI perf

Designed for large-scale AI training and inference, supporting frontier scientific, multimodal, and agentic AI workloads at scale.



Leonardo

Tier-1

WLCG Tier-1 @ INFN-CNAF: colocated, ~200 PB federated storage, ~2M HS23 computing capacity

Historical large-scale international support for distributed data analysis, simulation. Additionally, all Tier-2's collectively are comparable in scale to the Tier-1

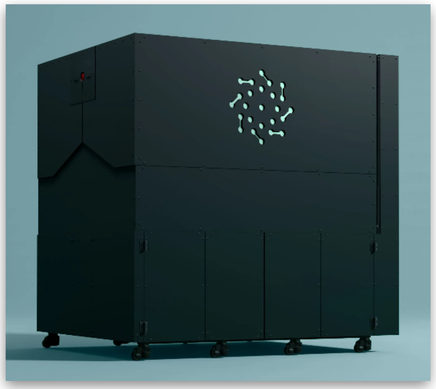
IT4LIA AI FACTORY

IT4LIA AI Factory (2027): expected ~160 ExaFlops AI inference perf, ~8,200 Nvidia Blackwell B200 GPUs

Next-generation EuroHPC AI infrastructure hosted at the Bologna DAMA Technopole. Designed as a EU-scale AI platform for foundation models, genAI, extremely-large-scale scientific and industrial AI applications (support research + startups, SMEs, PA)

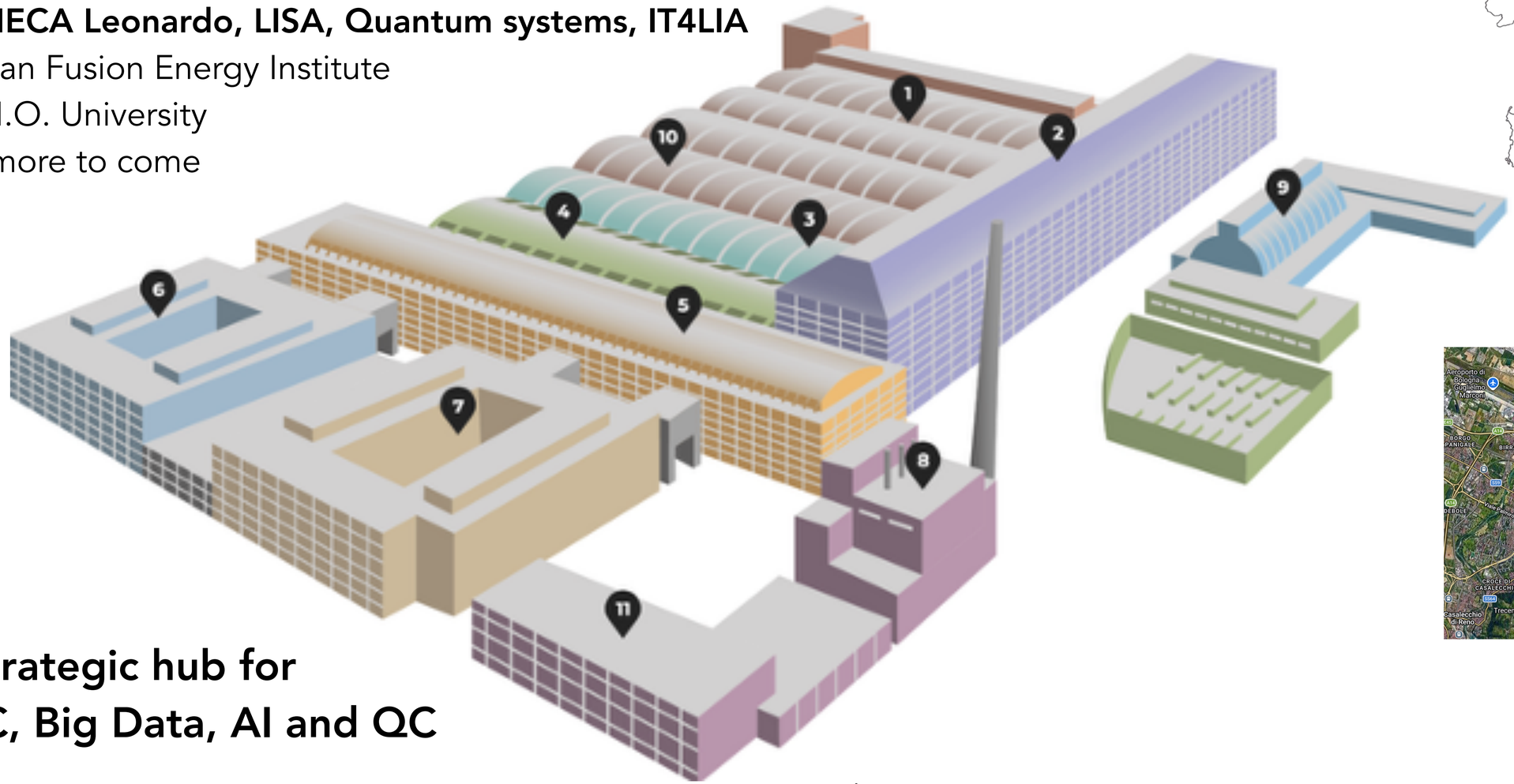
Quantum computing: IQM 54 qbits (superconducting), Pasqal 200 qbits (neutral atoms)

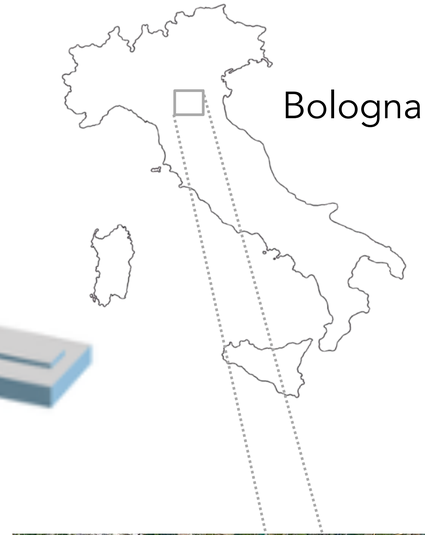
Investment in colocated (Bologna Technopole) real quantum hardware, supported by other machine in Italy, plus a theo+exp strong quantum community



The DAMA Technophole in Bologna

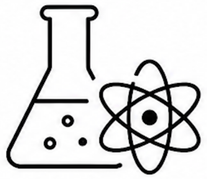
❶ ECMWF (European Center for Medium-term Weather Forecast)
❷ Hub for industrial research
❸ **INFN-CNAF WLCG Tier-1**
❹ **CINECA Leonardo, LISA, Quantum systems, IT4LIA**
❺ Italian Fusion Energy Institute
❻ U.N.O. University
❼ … more to come



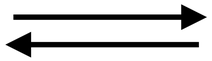


A strategic hub for
HPC, Big Data, AI and QC

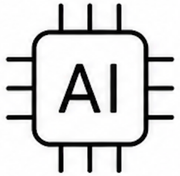
"**AI** for **Science**" and "**Science** for **AI**" in Italy/INFN



AI 4 Science:
a multi-domain research ecosystem



Science 4 AI:
foundational research



(focus only on HEP in this context, of course)

FAIR initiative (~116 M€), including INFN

2022 (another era!): a survey listed 150 AI applications being developed inside INFN

2024: INFN published a concise institute-wide AI strategy paper (more here)

Future Artificial Intelligence Research
https://fondazione-fair.it/en/

AI_INFN
Artificial Intelligence technologies for INFN research

The **AI_INFN** initiative (since 2020):

focus on **medium-scale infrastructures** (give access to researchers to quality hw/sw environments, seamless integration of Cloud/ HPC/Quantum, etc), **training and support** (schools, hackathons, support lists, thoroughly explained realistic use-cases on github, etc), collection of knowledge and solutions for INFN use-cases (more here)

Strong tradition in Italy on physics of complex systems - also for AI **(Prof. Giorgio Parisi, 2021 Physics Nobel Laureate)**

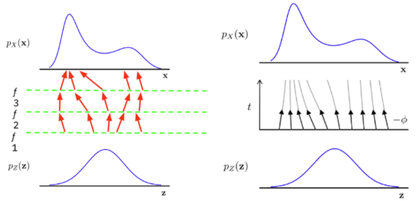
Just some examples

Flash Simulation (*e.g.* in CMS)

CMS CR -2025/027

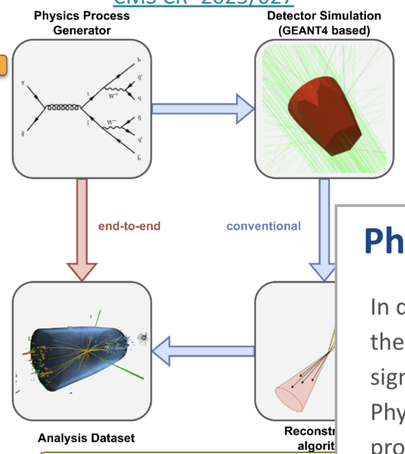
Thesis Vaselli

Continuous Flow Matching algorithms are used to predict vector fields that can be solved at inference time to connect noise distributions to physics distributions.



Similar experience in LHCb with GANs (Lamarr)

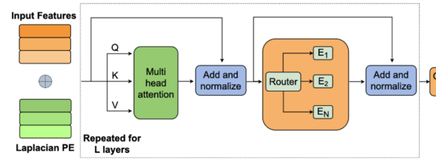
Thesis Barbetti

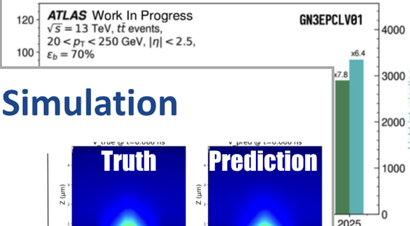


AI_INFN
Artificial Intelligence technologies for INFN research

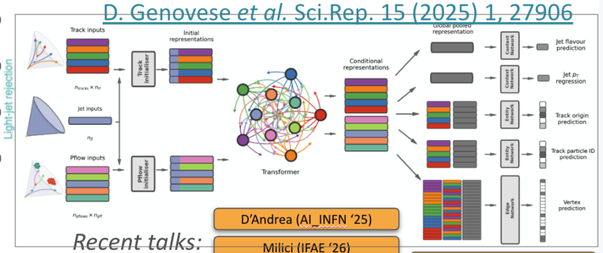
Jet tagging and jet labelling (*e.g.* in ATLAS)

Advanced algorithms (GNN, Transformers, MoEs) can greatly improve event interpretation.





D. Genovese *et al.* Sci.Rep. 15 (2025) 1, 27906



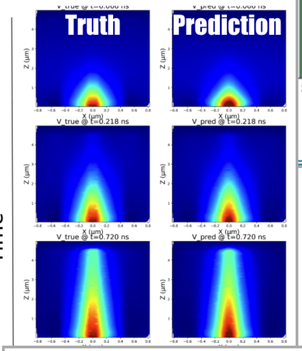
Recent talks:

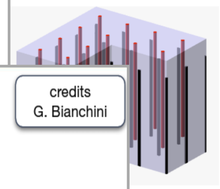
D'Andrea (AI_INFN '25)
Milici (IFAE '26)
Giagu (AI_INFN '25)
Cossutti (IFAE '26)

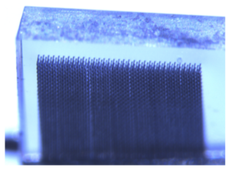
JB-2026-001

Physics-Informed Neural Networks for Simulation

In diamond detectors modelling signal propagation through the impedance structure of the sensor is crucial to predict signal shape.
Physics-Informed Neural Network were investigated to provide meshless solution to third-order PDE defined by quasi-static Maxwell Equations.







arXiv: 2509.21123
Poster @ INFIERI
Talk @ TREDI

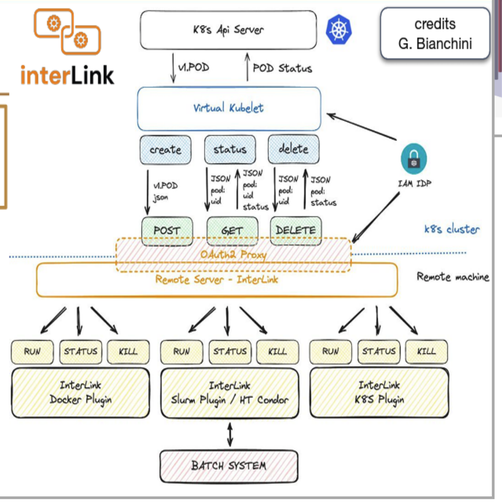
Offloading:

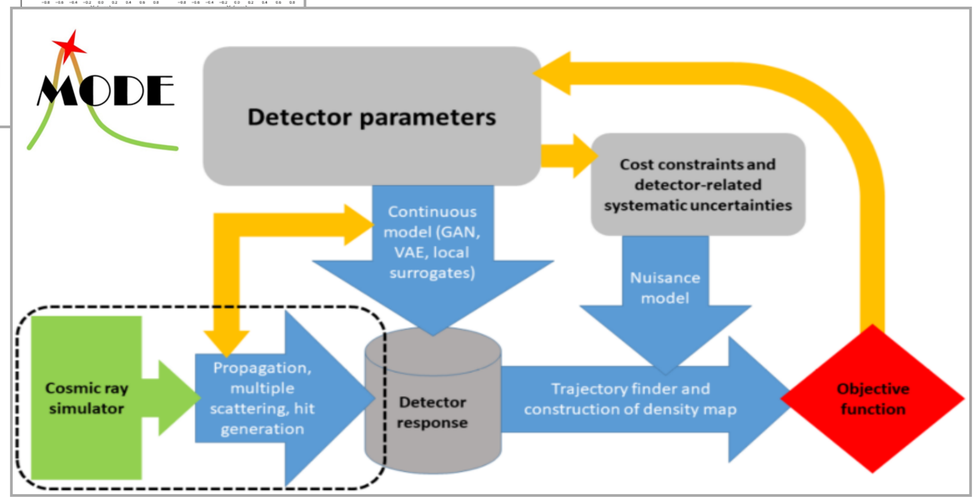
InterLink & Virtual Kubelet

Provide an abstraction for the execution of a Kubernetes pod on any remote resource capable of managing a Container execution lifecycle.

The project consists of 3 main components:

- **Kubernetes Virtual Node:** based on the VirtualKubelet technology. Translating request for a kubernetes pod execution into a remote call to the interLink API server.
- **InterLink API server**: handle requests coming from the VK and forwarding them to the sidecar;
- **Sidecar**: runs the containers on the infrastructure and returns the result. It can also communicates with the InterLink server.





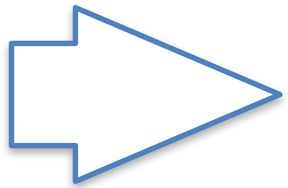
INFN interests

Broadly on this initiative, INFN interests match well the “**Grand Challenges**" outlined in the document

In light of that, **INFN would be interested in establishing collaborations with Genesis**

- in general, and in particular about the HEP domain which is our core mission

More specifically (albeit still at a broad-brush level) → *next pages*



INFN interests - in connection with..

Grand Challenge 1

Accelerated Experimental Design

(selected) **ongoing** activities:

- AI-driven co-design of facilities and detectors [in connection with the European coalition for AI in Particle Physics (EuCAIF)]

INFN interested to collaborate (primarily) on:

- Digital twins for AI-native facilities / detectors simulation
- AI-assisted cost and risk modeling

Grand Challenge 2

Intelligent Sensing and Instrumentation

(selected) **ongoing** activities:

- AI-accelerated simulations (e.g. FlashSim)
- More and more pervasive AI exploitation as baseline in most analysis workflows
- Exploration of agentic AI (also) in analysis workflows

INFN interested to collaborate (primarily) on:

- AI-driven reconstruction
- Multimodal foundation models
- AI agents as orchestrators of end-to-end analyses

INFN interests - in connection with..

Grand Challenge 3

Autonomous Experiments

(selected) **ongoing** activities:

- Work-in-progress in AI-assisted predictive maintenance on INFN data centers (e.g. INFN-CNAF)

INFN interested to collaborate (primarily) on:

- LLM-based operational assistance to support ops and reduce export intervention
- Integrated multimodal monitoring, e.g. incl. emerging Vision-Language-Action (VLA) paradigms

Grand Challenge 4

From Data to Discovery

(selected) **ongoing** activities:

- GenAI for (flash) simulation (normalising flows, transformers, …)
- AI for real-time (DL model distillation on FPGA, neuromorphic chips, spiking networks (SNNs))
- Algorithms for offline reco and analysis (graphs, LLM/LRM, …)

INFN interested to collaborate (primarily) on:

- AI-native trigger systems
- Extended use-cases of FPGA deployment of AI models
- Projects to invest on AI-enhanced triggers for higher sensitivity to rare and unexpected signals
- End-to-end generative systems to lower computing costs for large experiments
- Physics-oriented LLM models (operations, discovery, literature distillation)

INFN interests - in connection with..

"**Cross-cutting Themes and Emerging Opportunities**"

(selected) **ongoing** activities:

- Quantum AI

INFN interest to collaborate (primarily) on:

- AI on simulated / real quantum systems

Your view on "**Collaboration - Building a national effort**"

Ongoing in INFN:

- Italian Research Datalake (multi institute, including industries)
- Integration HPC/Cloud/Grid/Quantum
  - e.g "how to include HPC resources in WLCG-like processing" or "how to use WLCG-like AAI to access HPCs")

INFN interest to collaborate (primarily) on:

- Middleware to transparently access heterogeneous resources (data management, workload management, federated access, …)
- Integration with continental level / global initiatives (EOSC, EuroHPC, WLCG, …)

Conclusions

INFN stands as a reference institute in Italy for such initiatives in HEP.

INFN would be interested in establishing collaborations with Genesis.

Bilateral contacts are key. Thank you for the invitation to INFN, and glad to discuss further.

Acknowledgements: A. Zoccoli, O. Adriani, T. Boccali, L. Anderlini