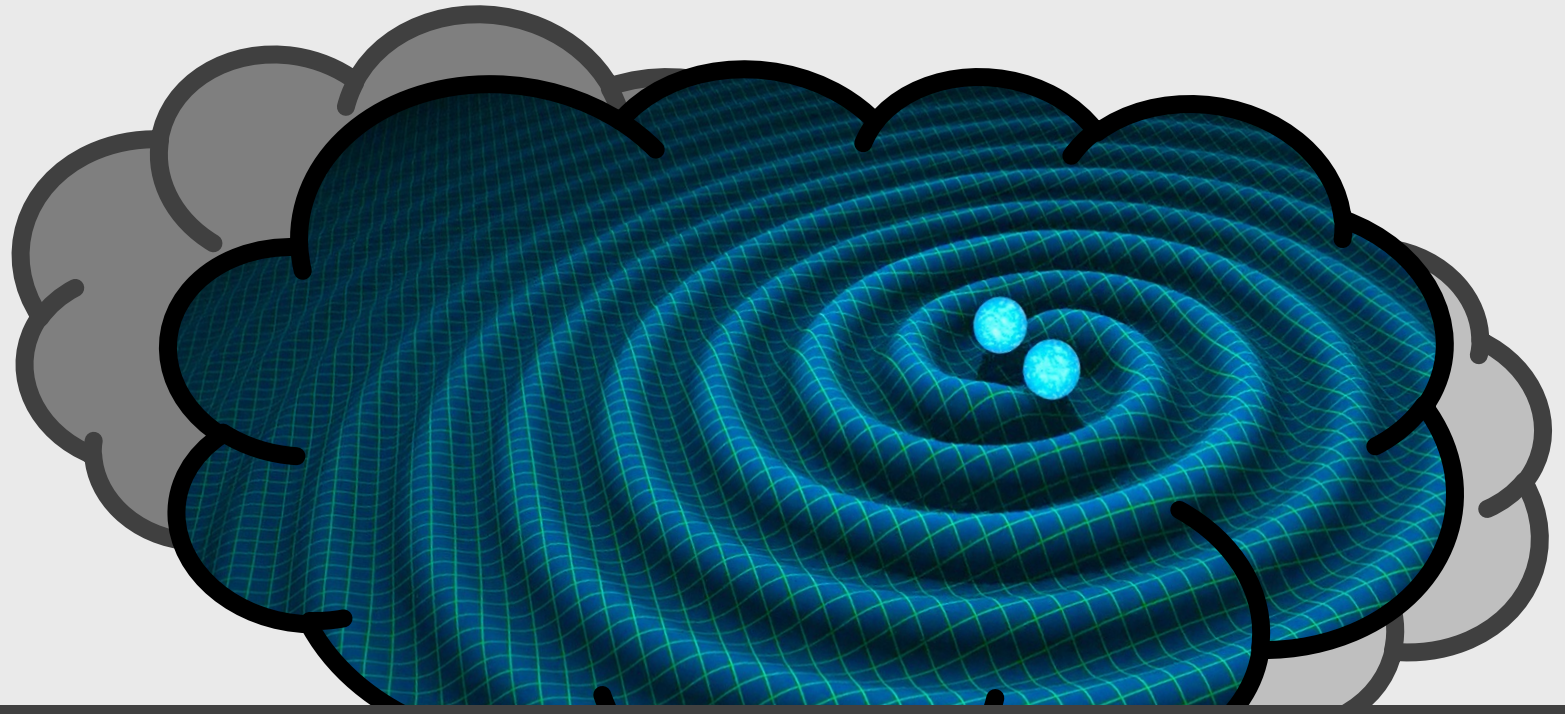


E-INFRASTRUCTURE BOARD

Stefano Bagnasco, INFN

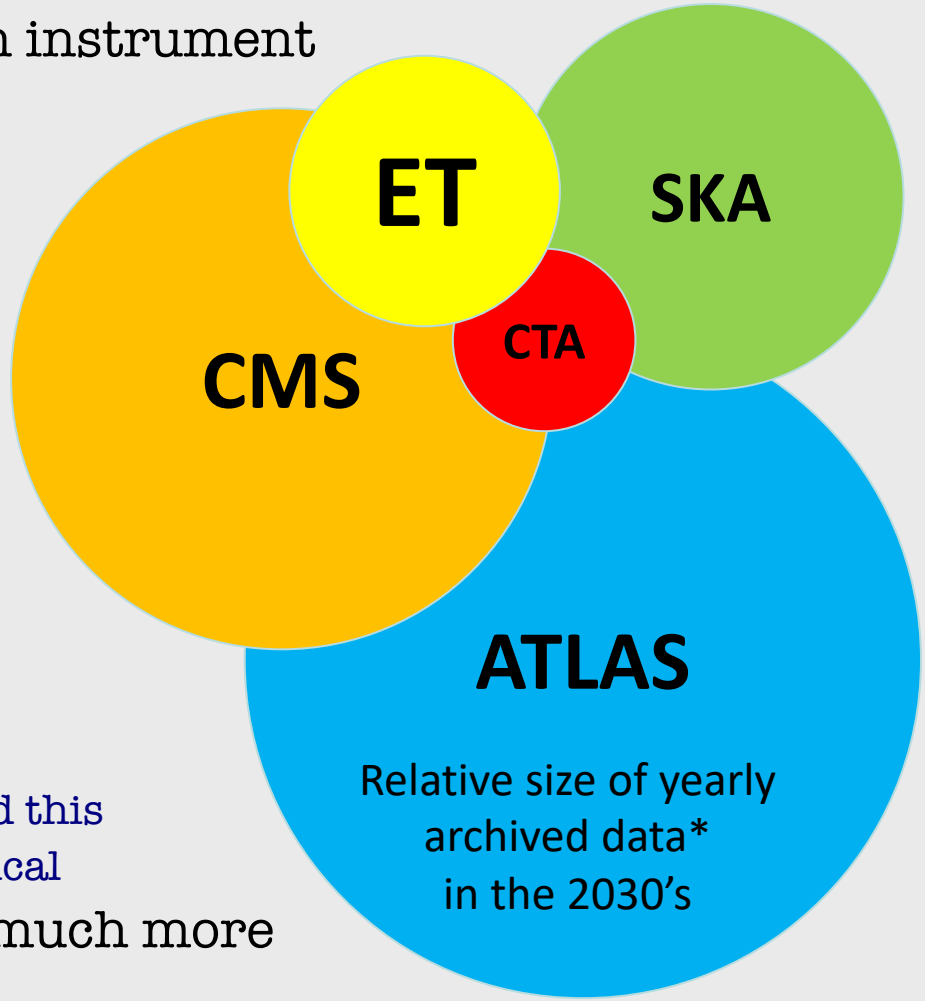
For the e-Infrastructure Board

ET in Italia:
scienza e tecnologia per la candidatura

Mar 18-19, 2025

CHALLENGES SUMMARY

- Luckily, raw interferometer data don't grow much with instrument sensitivity
  - We're not exploding like HL-LHC!
  - We expect about few tens of PB of raw data per year
  - No big deal today, piece of cake by 2030s
- The amount of scientific information encoded in the data does grow (a lot)
  - And the computing power needed to wring it out
  - It's already a task to estimate the computing power needs!
- Current computing algorithms and strategies won't trivially scale
  - High rates, long and overlapping signals,...
  - Need to rethink some of them
  - We need to support and run Mock Data Challenges to understand this
  - LL searches (including pre-merger) and LL PE are the most critical
- Low-latency alert generation and management will be much more complex
  - Design, planning and implementation already started outside of GW
  - Need to have our say now (see also dedicated session on Wednesday)



*to say nothing of weather forecast, genomics, Earth observation, oil industry, GAFAM and everybody else

GENERAL STRATEGY

- Use Mock Data Challenge support as “milestones”
  - More about this later
- Use IGWN infrastructure as baseline
  - IGWN uses the European computing centres as an extension of the OSG (which is nonoptimal…)
- Provide “workflow evaluation kits”…
  - Partial functionalities to evaluate tools and architectures
- …using ESCAPE as the first toolbox
  - But not the only one!
- Develop a. long-term R&D program (together with Virgo)
  - Together with Virgo “post-O5”: both can benefit from synergies

DRAMATIS PERSONAE

EIB Chairs: Stefano Bagnasco (INFN), Patrice Verdier (IP2I Lyon - IN2P3)

ET-PP WP8 leaders: Achim Stahl (U. Aachen), Nadia Tonello (BSC)

Division 1: Software, frameworks, and data challenge support
Andres Tanasijczuk (UC Louvain)

Division 2: Services and Collaboration Support
Antonella Bozzi (EGO)

Division 3: Computing and data model, Resource Estimation
Gonzalo Merino (PIC)

Division 4: Multimessenger alerts infrastructure
Steven Schramm (Université de Genève)

TTG: Technology Tracking working Group
Sara Vallero (INFN Torino)

Task 8.1: T0 data center
Leader: Patrice Verdier (IP2I-IN2P3)

Task 8.2: Computing and Data Model
Leader: Paul Laycock (Geneva)

Task 8.3: Resources
Leader: Silvio Pardi (INFN Napoli)

Task 8.4: Data Access Implementation
Leader: Nadia Tonello (BSC)

Liaison with OSB Div. 10: John Veitch (University of Glasgow), Elena Cuoco (Bologna)

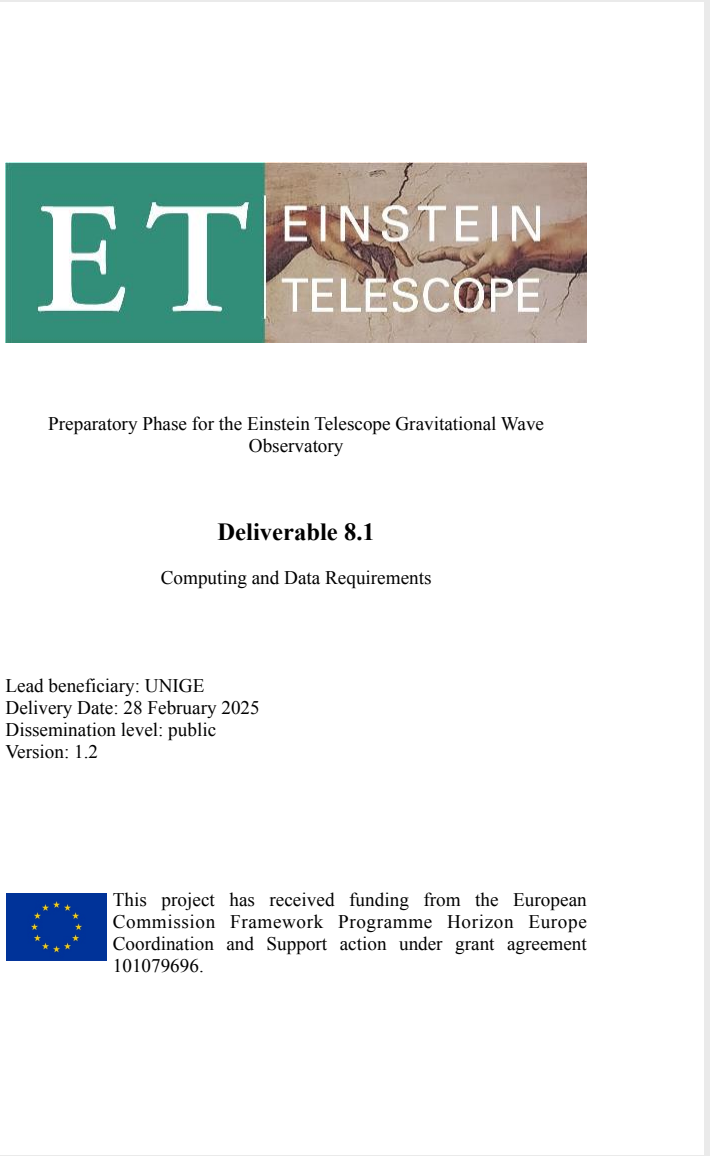
Joint ET-PP WP8 & ETC-EIB management (e.g., weekly call for coordination)

ET-EIB + ET-PP WP8 WORKSHOPS

- October 2023, Observatoire de Genève: requirements gathering
- July 2024, INFN-Napoli: Computing Infrastructures availability
- February 2025, INFN-Bologna: Joint EIB/OSB on Mock Data Challenges
- May 2025, Université de Genève (TBC): MM/LL challenges
- ...

- “Computing and Data Requirements”
  - https://apps.et-gw.eu/tds/?r=19444

- Will also need to define:
  - Data Management Plan
  - ET Computing Cloud MoU
  - ET Open Source Policy
  - ET Code best practices
  - …



ET EINSTEIN TELESCOPE

Preparatory Phase for the Einstein Telescope Gravitational Wave Observatory

Deliverable 8.1

Computing and Data Requirements

Lead beneficiary: UNIGE
Delivery Date: 28 February 2025
Dissemination level: public
Version: 1.2

This project has received funding from the European Commission Framework Programme Horizon Europe Coordination and Support action under grant agreement 101079696.

MOCK DATA CHALLENGES

- MDC as multipurpose tools
  - Develop and exercise analysis code and strategies
  - Collect data about analyses' requirements and performance
  - Build the data analysis community and bootstrap new groups
  - Educate the community in the use of common distributed computing tools and best practices for sustainability and manageability
  - Iteratively test the distributed computing infrastructure

- Mock Data Challenge support plans
  - MDC1: provide data distribution layer (OSDF: CVMFS + cache) and survey the activities
  - MDC2: provide (possibly a set of) prototype tools for workload management etc.
  - MDC3-n: iterate

MOCK DATA CHALLENGES

- MDC1 still underway
  - This round: only data distribution through the OSDF-powered IGWN infrastructure provided
  - All-comers, survey “by hand” to get information
- Tentative timeline:
  - February 2025: MDC workshop
  - Results from MDC1 - ET symposium
  - Presentation of MDC2 - ET annual meeting
- Regular OSB/EIB meetings restarted
  - For coordination and support

TODO LIST

- Define a (set of) standard software distribution(s) [Elena]
- Provide a CVMFS repository for distribution (and policies for inclusion) [Stefano]
- Call for computing resources requests (also for MDC datasets production) [Stefano]
- ET Computing Cloud MoU [Stefano, Elena]
- Data/metadata catalogue (annotations) [Paul]
  - or at least (for the time being) clear and permanent association between config file(s) and dataset
  - Define the requirements for metadata {Oscar]
- New MDC production code review for modularity [Anuradha]
- Prioritized list of requested datasets [Tania & Gianluca]
  - Resources (CPU, storage, personpoiwer for development & running)
  - Strategies for reuse and ”modular” generation
- Set up MDC analysis Code working group [Paul]
- Start definition of best practices for code [Andres]
- Organize MDC timeline definition meetings (or whatever) [Stefano]
- Define a plan for user-facing portal (docs, tutorials, data and code repos, à la GWOSC) [Alberto]
- ETAP/MADDEN/ETIC/whatever coordination (workshop?) [Federica]
- “Baseline” technology Q&A-based seminars [Sara]

Computing Technology Laboratory for ET

The Objectives

- Porting of code on architectures else than x86 (GPU, Arm…)
- Test on "unusual" configurations (large-RAM to hold large templates…)
- Benchmarking and architecture evaluation for sustainability (carbon footprint…)
- Parallel computing and HPC in a fully controlled and customizable environment
- Support to development of middleware and framework for ET
- Support to computing needs for site characterization

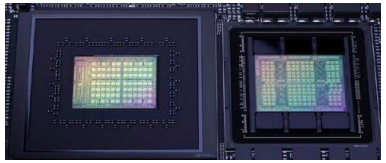
… and obviously for "standard" computing activities (MDC, waveform, analysis)

Lia.Lavezzi@to.infn.it

Hardware Resources

- 4x server Dell PowerEdge R7525, 64 CPUs
  with various GPU types:
  - 2x AMD MI100, 32GB
  - 1x NVIDIA A16, 4x16GB
  - 2x NVIDIA A5000, 24GB
  - 10 free slots

- 1x **nVidia GraceHopper** evaluation system by E4
  - CPU+GPU coherent memory model
  - 900 GB/s coherent interface NVLink–C2C

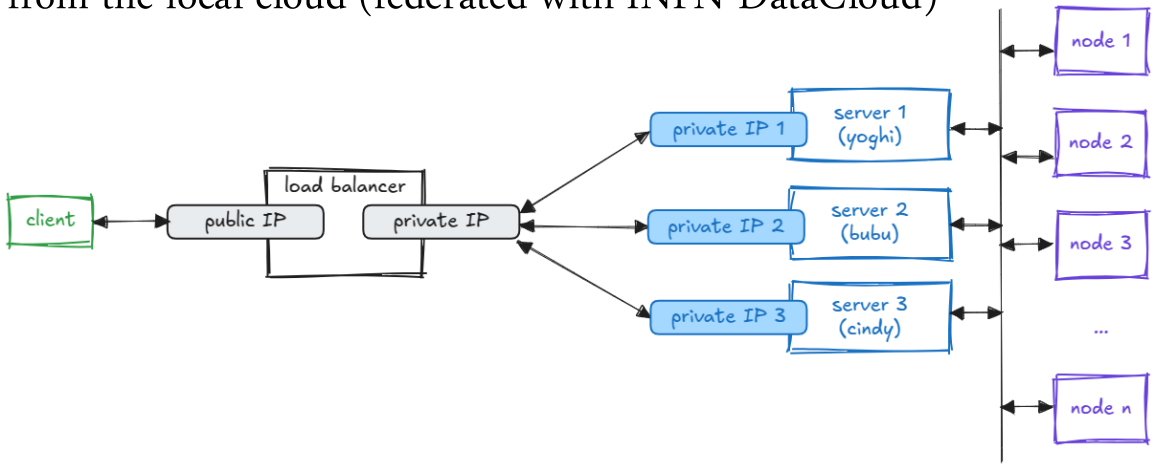


- 3x server Lenovo SR675v3, 64 CPUs
  - 1TB RAM, expandable
  - 2x NVIDIA L40S 48GB
  - free GPU slots
- Connectivity upgrade (**InfiniBand** + NVIDIA ConnectX–7 NDR 400Gb/s)

Lia.Lavezzi@to.infn.it

Design

- **Kubernetes** in High Availability (three master nodes)
- **Storage**
  - dedicated storage for data, provided by site
  - /home on glusterfs from machines disks
- **Authentication** via ET AAI (ongoing @CNAF)
- User **access** via Jupyter Notebooks (as a first solution)
- Possibility to use resources from the local cloud (federated with INFN DataCloud)

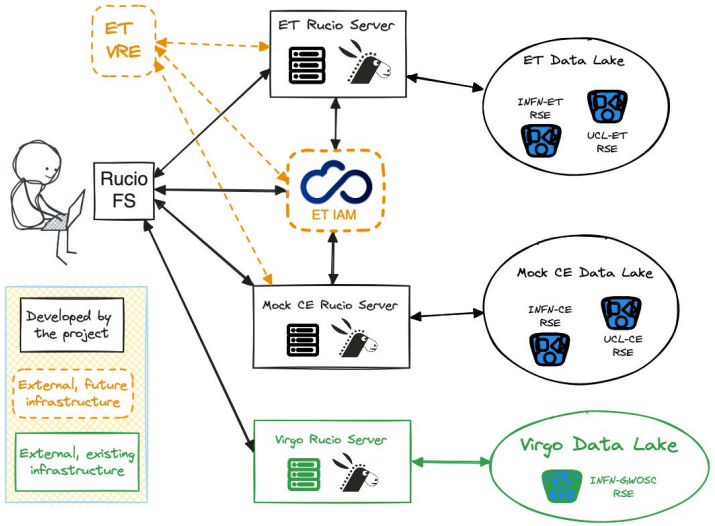


Lia.Lavezzi@to.infn.it

CTLAB TORINO

- Personnel
  - Lia Lavezzi (Tecnologa ETIC) – Software architecture
  - Luca Tabasso (CTER ETIC) – Infrastructure support
  - SB (DT, eIB cochair) - General worrying
  - Federica Legger (Tecnologa) Madden PI
  - Sara Vallero (Tecnologa) TTG chair
  - MADDEN PostDoc
- Synergies & such
  - Tier-2 di Torino (infrastruttura)
  - BETIF: complementary hardware and scope, plans to converge on similar access and interface architectures (and possibly AI topics, through Daniele Bonacorsi, Elena Cuoco)
  - TeRABIT: hardware integration with the Torino “HPC Bubble”
  - ICSC: partial integration into INFN DataCloud infrastructure
  - InterTwin: EU project on Generative AI for noise simulation (Vallero, Legger)

MADDEN - Multi-RI (Research Infrastructure) Access and Discovery of Data for Experiment Networking

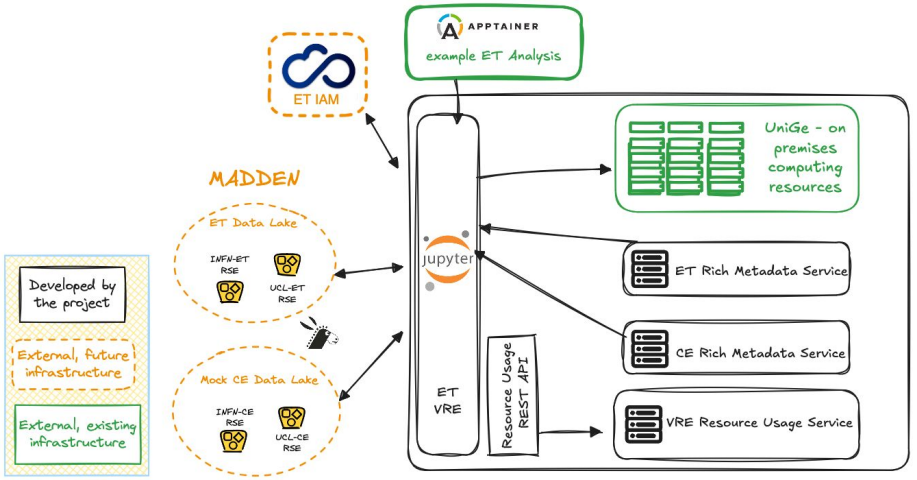


- Funded by 1st call OSCARS cascading grants
  - Participating organizations:
    - INFN Torino (PI: F. Legger)
    - UC Louvain (coord.: A. Tanasijczuk)
  - Feb. 1st, 2025 - Jan. 31st, 2027
  - Overall funding: 210 K€
- Status:
  - 2 post-docs being hired
  - currently setting up ET Data Lake with MDC data

- Main objectives:
  - Build a multi-RI **Data Lake** managed with **Rucio** and interface it with the VRE (Virtual Research Environment) developed by companion project ETAP
  - Develop and test **RucioFS**, a tool to provide a POSIX-like view of the Rucio catalogue in a multi-RI environment
  - Extend RucioFS to support advanced querying capabilities using metadata

Federica.Legger@to.infn.it

ETAP - Einstein Telescope Analysis Portal



- Funded by 1st call OSCARS cascading grants
  - Participating organizations:
    - University of Geneva (PI: Paul Laycock)
  - Jan. 1st, 2025 - Dec. 31st, 2026
  - Overall funding: 250 K€
- Status:
  - kicking off
  - deploying VRE components in Kubernetes
  - discussion started with relevant VRE and REANA experts

- Main objectives:
  - Deploy the CERN ESCAPE **VRE (Virtual Research Environment)** at University of Geneva
  - Connect to **multi-RI Data Lakes** managed by Rucio (MADDEN)
  - Deploy multi-RI **Metadata services** from the HEP Software Foundation (HSF)
  - Design a flexible computing resource monitoring service

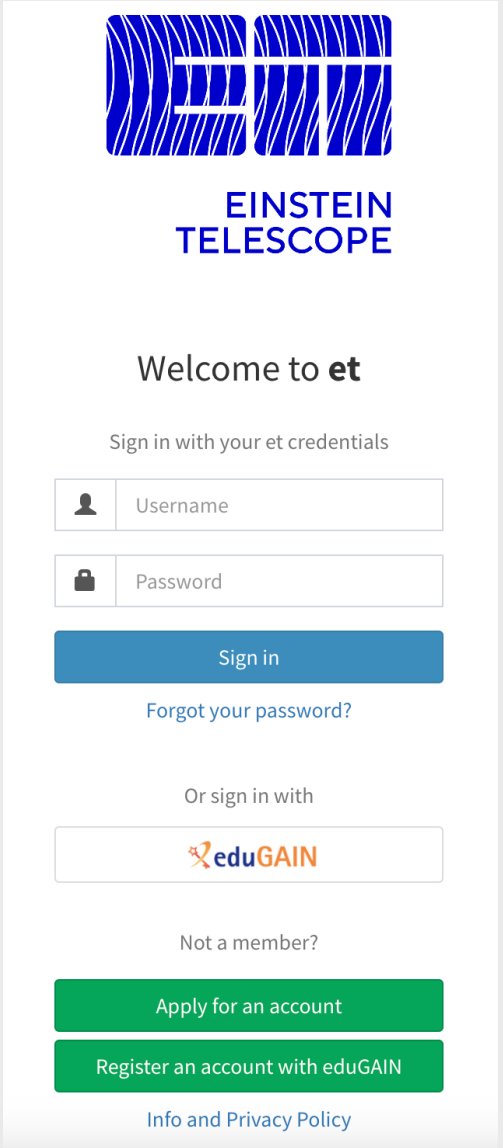
Federica.Legger@to.infn.it

THE WIDER COMPUTING COMMUNITY

- Representation or interactions with
  - WLCG Collaboration (SB)
  - ESCAPE Open Collaboration (SB, Patrice Verdier, Elena Cuoco)
  - IGWN (SB, with caveats)
  - JENA Computing Initiative (SB, Paul Laycock, Patrice Verdier)
  - HEP Software Foundation (Paul Laycock)
  - SPECTRUM Project (SB)
  - Major computing conferences (e.g., CHEP)
  - ...

IAM DEPLOYMENT AT CNAF

- Installed and configured!
  - Thanks to Michel Jouvin
  - Missing: interface to ETMD
- Administrative problem: managers need to be formally appointed by INFN Director General
  - GDPR-related issue
  - Impossible for non-INFN people
- Signed a specific MoU with CNAF



THANKS!

- Questions?