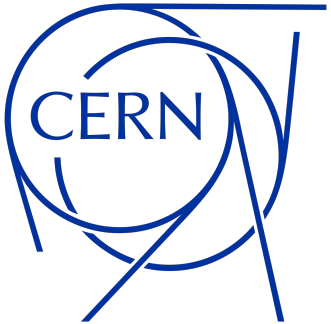


The CERN Virtual Research Environment within the ESCAPE Collaboration

Giovanni Guerrieri (CERN), Xavier Espinal (CERN) on behalf of the ESCAPE Collaboration

October 21th 2024

CERN: Conseil (organisation) Européen(e) pour la Recherche Nucléaire

International Research Organization focused on fundamental physics research, specifically in the field of **particle physics**

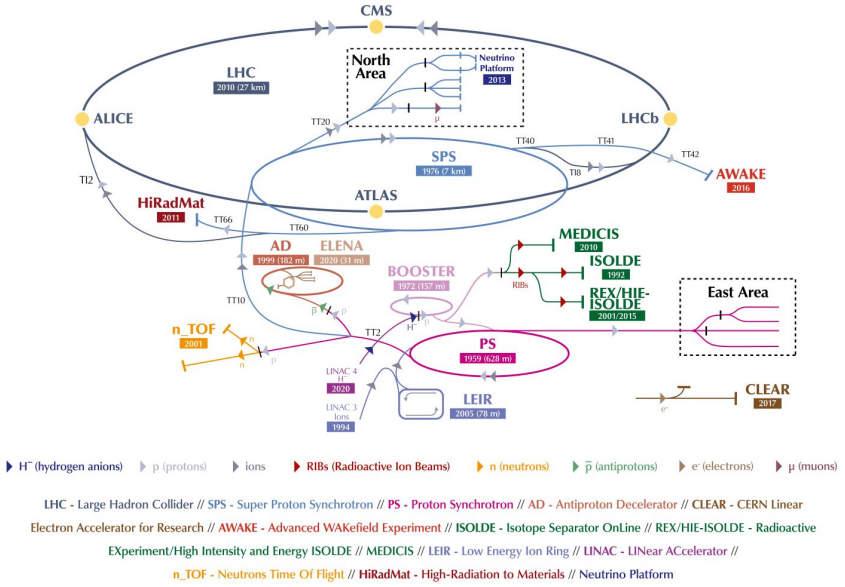
Established in 1954 and located in Geneva, Switzerland

24 European member states, **10** associate members, **4** observers and more than **60** non-member states

Groundbreaking advancements to science and technology

- Discovery of the Higgs boson in 2012, confirming the mechanism that gives particles mass
- Development of the World Wide Web in 1989 by Tim Berners-Lee, revolutionizing global communication
- Advances in accelerator and detector technology with applications in medicine, energy, and industry
- Significant contributions to Open Science, e.g. Zenodo

Budget: **~1255 MCHF** (2024)



ESCAPE: European Science Cluster of Astronomy and Particle Physics

Consortium of 31 members, including:

- 10 ESFRI projects & landmarks: **CTA, EST, FAIR, HL-LHC, KM3NeT, SKA, LSST, VIRGO, ESO, JIVE**
- 2 pan-European International Organizations: **CERN and ESO**
- 2 European Research Infrastructures: **EGO and JIV-ERIC**
- 4 supporting European consortia: **APPEC, ASTRONET, ECFA and NuPECC**

Budget: **15.98 M€**

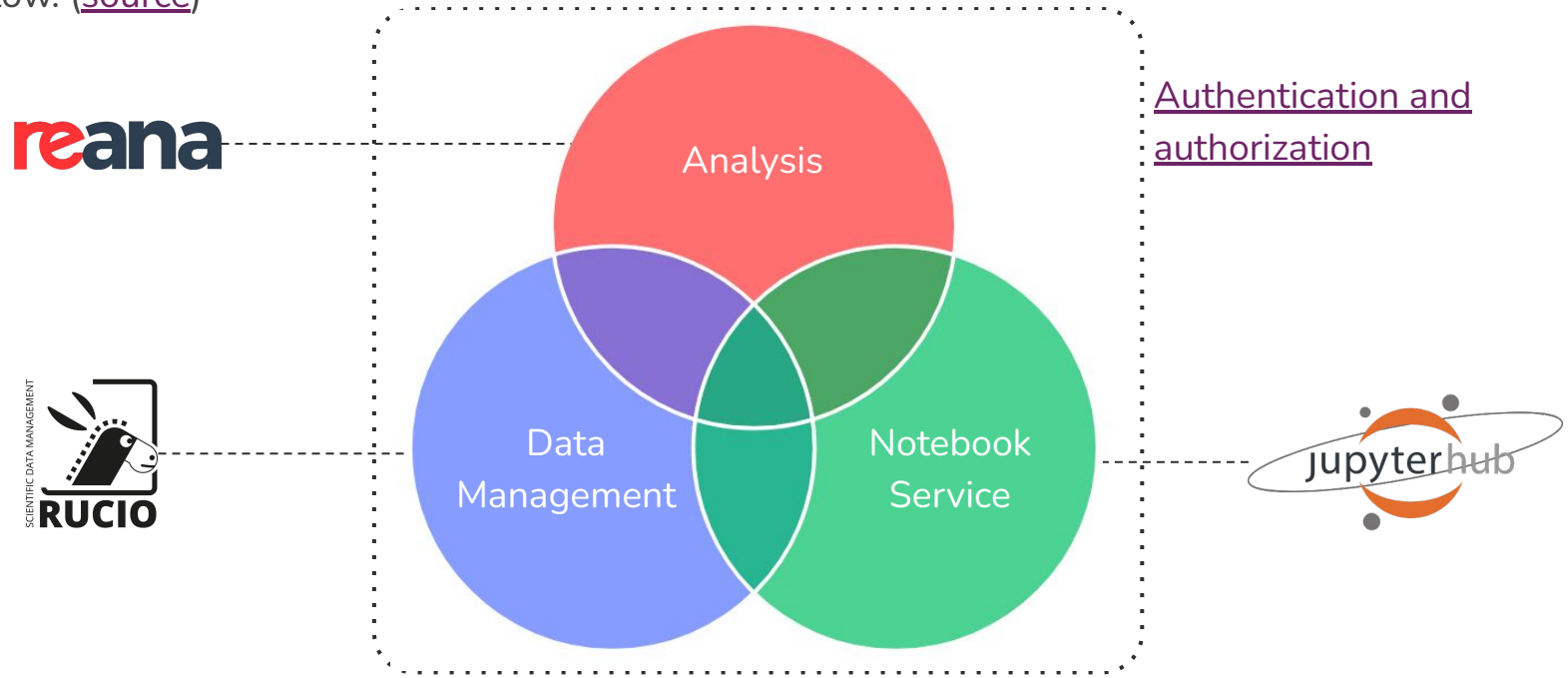
Duration: **48 months** (1/2/2019 -31/1/2023)

Key high-level ESCAPE project results

- Provided a full **Exabyte-scale Data Management** prototype (Storage, Transfer and Access) service with **common AAI** framework for distributed scientific computing
  - Validated through Data Challenges
- Developed a **catalogue** to publish digital scientific products of ESCAPE communities
  - ESCAPE research infrastructures onboarded
- Developed **interoperability standards** for astronomical data services
  - Virtual Observatory (VO) services prepared to be integrated with EOSC
- Produced a **reusable analysis toolkit** for integrating diverse service offerings, and a **Virtual Research Environment** prototype that integrates ESCAPE services
  - Adopted and replicated by a number of research infrastructures
- Several **Citizen Science** projects built upon ESCAPE developments
  - Attracted significant interest/involvement
- Dark Matter and Extreme Universe Science Projects as demonstrations of Open Science in ESCAPE
  - Integrated with EOSC platform through EOSC-Future

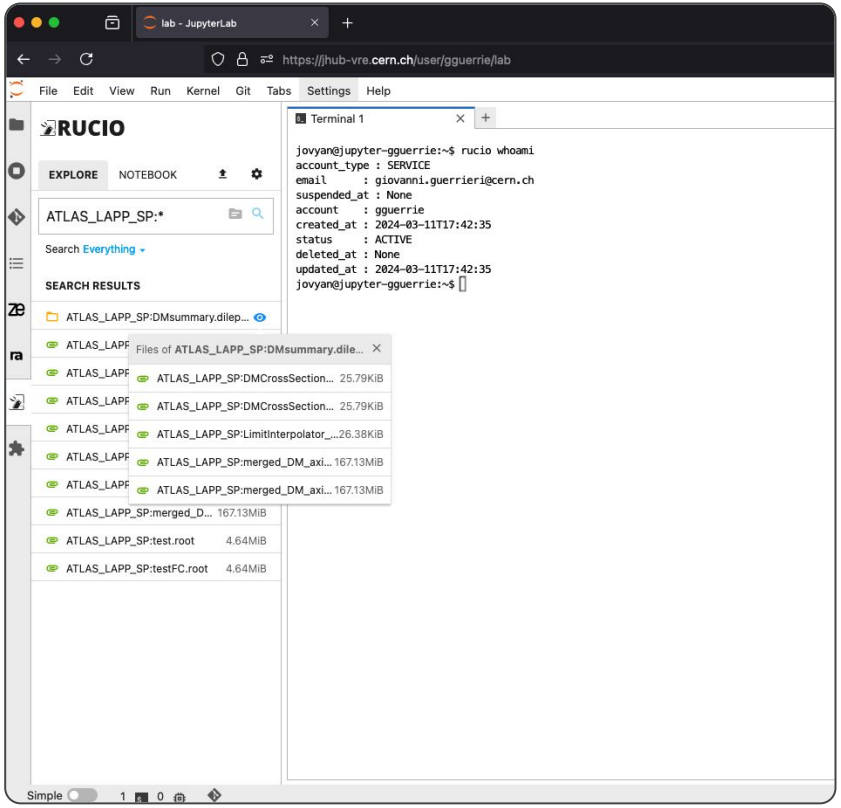
The CERN Virtual Research Environment

Virtual Research Environment (VRE): infrastructure and services that provide ***integrated*** data access, experiment's software and computational resources to execute one or more elements of an analysis workflow. (source)



The CERN Virtual Research Environment

- Modular
  - Integrates software, tools and packages
  - Can connect to remote storage and computing resources
- Flexible
  - *Ad-hoc* workflows can be created via easily editable declarative files
  - Can be installed on different machines independently of CERN requirements
- Reproducible
  - Deployment is kept accessible and documented to be used as a blueprint for other infrastructures.
  - Allows analysis preservation.



Example: from experiment's raw data to final results

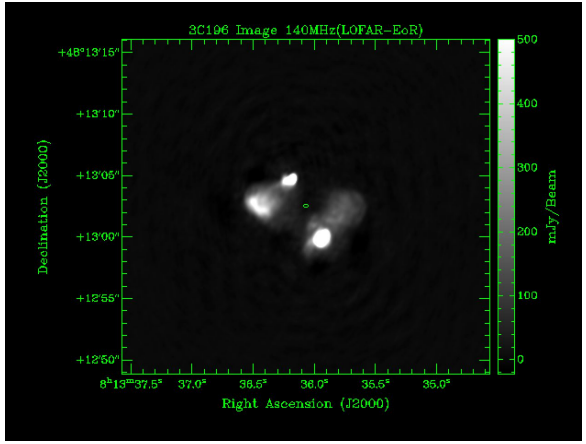
A multi-wavelength analysis in the VRE (3C196 quasar)

1. Data **injected** to the Data Lake from different radio sources from **external** locations

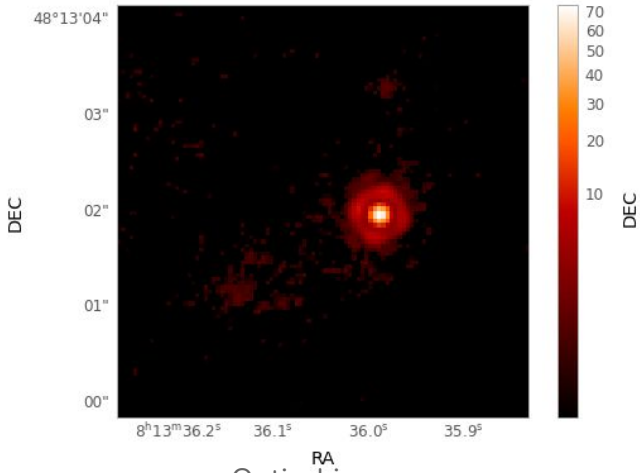
2. Download of data, **process** and **storage** of results back to the Data Lake

3. Combined **optical** data located via the VO plugin available in the **VRE**

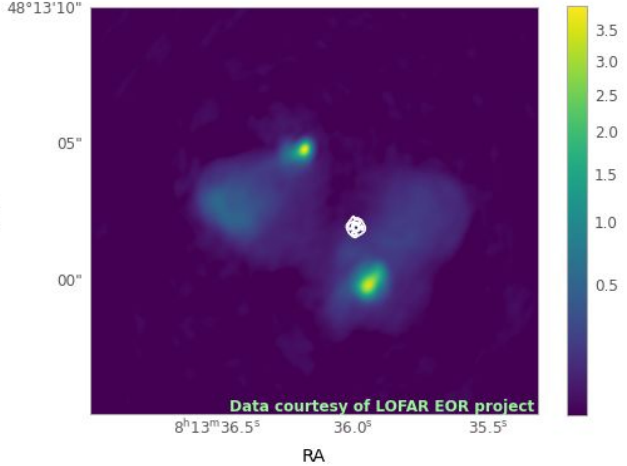
4. Optical and **radio** data combined.
Results uploaded back to the Data Lake.



Radio image



Optical image



Combined image

The CERN / ESCAPE use cases

- Landmark experiments adopting CERN / ESCAPE data management technologies :
  - Square Kilometre Array Observatory (SKAO): link to presentation
  - Cubic Kilometre Neutrino Telescope (KM3NeT): link to presentation
  - Cherenkov Telescope Array Observatory (CTAO): link to presentation

- Other ESCAPE stakeholders onboarding leveraging the VRE:
  - Einstein Telescope (ET): link to presentation

Efficiency gains and Economies of Scale in Scientific Computing as a direct result of Open Science initiatives

Courtesy of xavier.espinal@cern.ch

- **Common software and computing tools and framework:**
  - Well established systems easily adoptable by new RI's/experiments, largely based on technology and suited for heavy duty computing activities, eg. data management , data lifecycles, data processing and analysis. **Limit re-inventing systems to address similar challenges**.
  - Coherent approach to access and use scientific resources on grid sites, clouds and HPC centers. With similar tools, protocols, authentication mechanisms, etc. **Limit duplications of infrastructures/support systems to achieve common goals**.
- **Shared practices on data analysis based on open data and open science**
  - New paradigms on scientific data analysis. Virtual research environments based on Open Science being successfully prototyped. **Coherent data processing frameworks**, all-you-need in a common place, eg. software, data and code, and maintained/validated by the experiments.
  - Minimising probability of mistakes in the data processing chain leading to more **effective usage of computing resources** in the centres around the world, eg. visual real time preview of analysis on micro-data samples before offloading large scale processing to the grids, clouds or HPCs.
- **Alignment in coding strategies and hardware approaches**
  - Favouring **common practices and R&D on code frameworks/algorithms for G/T/CPU and Accelerators** could maximise the outcomes of ML/AI models and NN training methods with an impact in the way the hardware is used (accelerators are extremely power hungry) and promote code/models reusability between disciplines.
  - Common voice to have **critical mass to influentiate hardware designs** tailored also for scientific activities. eg. GPU industry seems disfavouring double precision floating point (not fundamental for gaming)
- **Common Open Science practices and its impact on Society**
  - Commonalities in Data Management and Data Access framework will **favour Open Science initiatives massively**. Easy way to tag/publish open data from the RIs and access to Experiment's validated Virtual Research portals.
  - Lower the barrier to access scientific data and boost **Citizen Science campaigns** (galaxy identification, particle identification, star classification, training and feeding of ML/NN, etc.

Summary

- ESCAPE demonstrated Scientific Collaboration across different disciplines is leading towards common Scientific Computing models and tools
  - SKA, CTA/MAGIC, KM3Net, Einstein Telescope, etc. adopting ESCAPE Data Lake, Virtual Research Environments, common Software Repositories and the common Authentication, Authorization and Identity Management framework.
  - In good position to provide input and support in defining the EU Node's roles and capabilities both in terms of thematic approaches (per cluster/science) and facility specific activities (HTC/HPC)

- Further consolidation ongoing with the implementation of ESCAPE "community-based competence centre" (CCCs) via the OSCARS project to expand synergies with all European Science Clusters
  - *Collaborative **network of people** in the context of the Science Clusters **providing expertise, best practices and services** in relation to Open Science, and the **promotion of cross-disciplinary collaboration***

- Although still in the initial phase of evaluation, the impact of fostering Open Science activities on efficiency gains and economies of scale in scientific computing is significant.

Thank you!

home.cern