OSCARS
Open Science Clusters' Action for Research & Society

Funded Project

MADDEN - Multi-RI Access and Discovery of Data for Experiment Networking

Presenter: GIOVANNI GUERRIERI, CERN, ORCID: 0000-0002-3403-1177
On behalf of Project PI: FEDERICA LEGGER, INFN, ORCID: 0000-0003-1400-0709

Implemented by

INFN UCLouvain

Funded by the European Union

CHALLENGE ADDRESSED

OSCARS

What problem(s) are you going to solve?

- Scientific collaborations often **need to share data**, especially in fields such as Gravitational Wave (GW) research, where the same GW event is detected by several interferometers
- Each Research Infrastructure (RI) typically manages its own data in **isolated** Data Lakes
- We need to establish **a common data infrastructure** that allows scientists to securely access data across multiple RIs
  - Current interferometers: LIGO (US), Virgo (Europe), KAGRA (Japan)
  - Future interferometers: Einstein Telescope (ET) in Europe, Cosmic Explorer (CE) in US

What are you planning to do to solve the problem?

- We will use **Rucio** to build a multi-RI Data Lake for the GW community
  - Rucio is an open-source framework for data management, distribution and access, initially developed for the ATLAS experiment at CERN and now used by communities across several domains (HEP, Nuclear, Astro)
  - LIGO and Virgo are already using Rucio in production, and ET is currently evaluating its use in the design of their computing model
  - We will extend Rucio to **seamlessly support a Multi-RI setup**
- To increase Rucio usability for scientists we will further develop and evaluate **RucioFS**, a tool that provides a POSIX-like view of the Rucio catalogue

What will be the results and how do you plan to make them available to the broader community?

- A **multi-RI Data Lake** for ET and CE will be built using Rucio
  - Two independent Rucio instances for ET and CE will be deployed
  - The ET IAM (Identity and Access Management) will be used for authz
  - ET users will seamlessly access data from both Rucio instances
- The Data Lake will be interfaced with the **ET analysis facility** based on the ESCAPE VRE (Virtual Research Environment) implemented by the companion ETAP project
- **RucioFS** will be tested at scale and extended to use metadata in queries
- A technology demonstrator for the Virgo collaboration will be set up

What risks could limit the success of the project, and how can they be mitigated

- Integration of the multi-RI Data Lake with the ESCAPE VRE relies crucially on ETAP; **close cross-coordination** has been established between the two teams
- Authorization with **IAM and metadata integration are external dependencies**: project members and collaborators are members of the relevant collaborations (LIGO/Virgo and Einstein Telescope) and were involved in former ESCAPE/INDIGO projects
- **RucioFS might not work at scale** (or not worth the effort, we do not want to reinvent the wheel) and we are ready to look for alternatives

TEAM

- PI: Federica Legger, *INFN* (ORCID – 0000-0003-1400-0709)
- Lia Lavezzi, *INFN* (ORCID – 0000-0002-4928-8151)
- Andres Tanasijczuk, *UCLouvain*
- ???
  - Two post-doc positions both at INFN and UCLouvain have been opened and are being filled right now