OSCARS
Open Science Clusters' Action for Research & Society

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

Principal Investigator: **F. Legger,** INFN
Project team members: **Lia Lavezzi** (INFN), **Andres Tanasijczuk** (UCLouvain)

OSCARS

- Funded by 1st call OSCARS cascading grants

- Start date: **February 1st 2025**

- Duration: **18 months**

- Funding: **210 K€**

- Institutes: **INFN Turin, UCLouvain**

- Personnel: hiring 2 postdocs (1 in Torino, 1 in Louvain)
  - selections on-going

- Showcase **Rucio** for Distributed Data Management for ET (and possibly Cosmic Explorer) community -> similar to current LVK use case

- Build a multi-RI Rucio Data Lake integrated with:
  - **ET AAI** (Indigo IAM)
  - **VRE** from partner project **ETAP**

- Further develop **RucioFS** and test it at scale
  - Extend RucioFS to support Rucio metadata querying

- Build a mini Rucio Data Lake for Virgo to be used as a technology demonstrator for LVK

- The **Multi-RI data lake** for ET and CE is the innovative development
  - **OIDC** for auth
  - CE Rucio to trust ET IAM as an IdP
  - allow Rucio client to connect to more than one Rucio server
- POSIX like view of the data with **RucioFS**

OSCARS
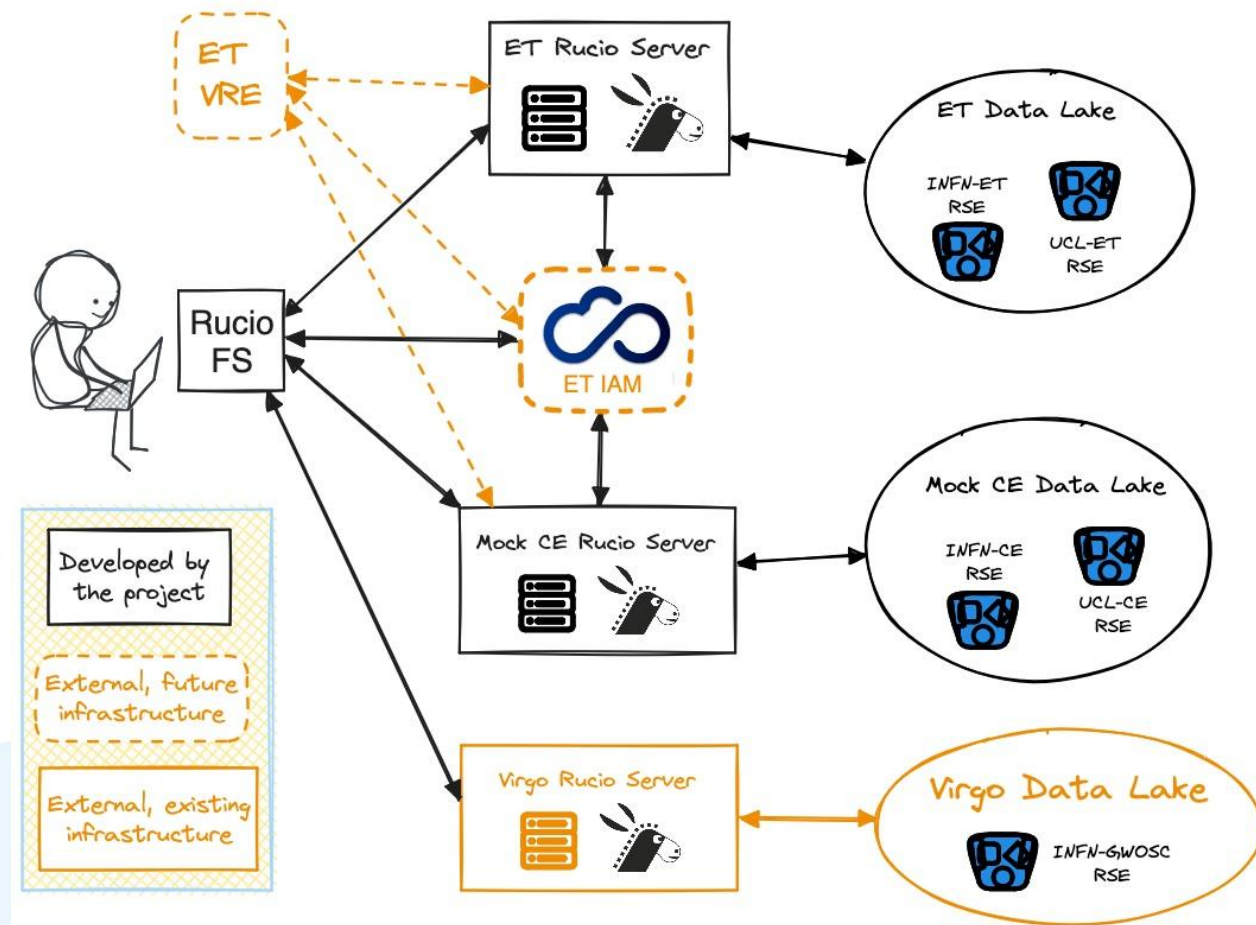
- (External) Authz: ET AAI service (Indigo IAM)
  - Service being installed at CNAF (admin. Michel Jouvin)
- ET Data Lake
  - ET Rucio server already installed in INFN-Torino ✅
  - 2 RSEs (the one in UCLouvain used as OSDF origin for MDC1 data distribution, one in INFN-Torino) already connected to ET Rucio server ✅
  - Need to connect to ET AAI
- (External) Integrate with ETAP VRE
- Distribute ET Rucio client in docker image and CVMFS (configured to use ET AAI as auth server and OIDC)

OSCARS

- Mock CE Data Lake
  - Mock CE Rucio server will be installed in INFN-Torino
  - 2 RSEs will be installed (one in UCLouvain, one in INFN-Torino) physically in same ET Rucio storages, but logically independent
  - Can use user/passwd for CE mock user auth in this Rucio instance (no need for a mock CE AAI)
  - Create generic "ET reader" account in the CE Rucio instance, with read-only permissions, and map ET ID tokens to that account
- Replicate setup to create a mini Data Lake for LVK community
  - Use Virgo open data
 - Connect to existing Virgo Rucio server in Turin

- GW community used to browse data in POSIX-like filesystem (CVMFS)
  - Caveat: LVK will retire CVMFS for data in summer 2025
- RucioFS is a POSIX-like FUSE mount filesystem (read-only)
  - Transforms the scope/container/dataset/files Rucio schema into a hierarchical structure of directories
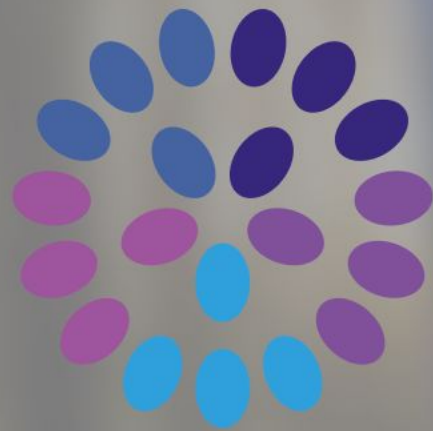 - Supports multi-RI (separate mount points under /ruciofs for each Rucio server)
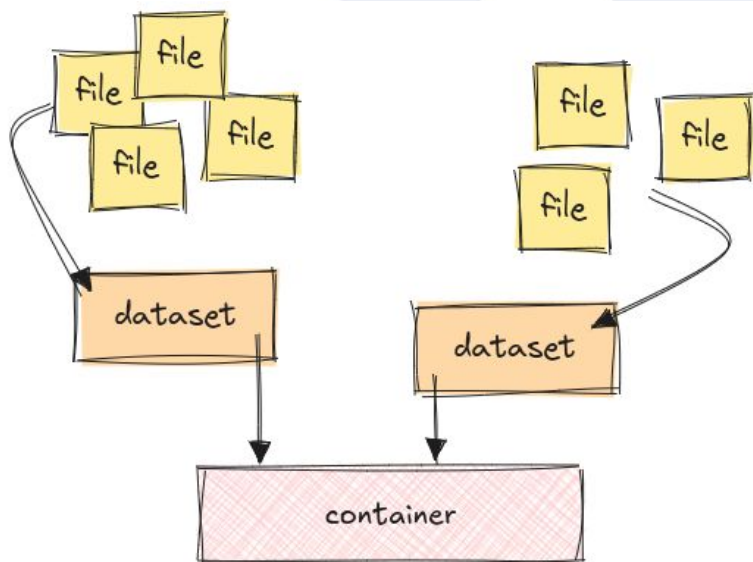
C. Fronzé (RucioFS original author)

RucioFS: planned developments

- Already part of Rucio code base, but not production-ready
  - File download at first access, with caching system to speed up query execution, but ... cache is not updated unless remount
  - Need to implement some automatic periodic cache refresh
  - Verify its scalability
    - Thousands of simultaneous queries
    - Billions of files in the catalogue(s)
- Add metadata support in RucioFS queries
  - First for Rucio metadata
  - Maybe also for HSF's rich metadata DB (OSCARS ETAP project)
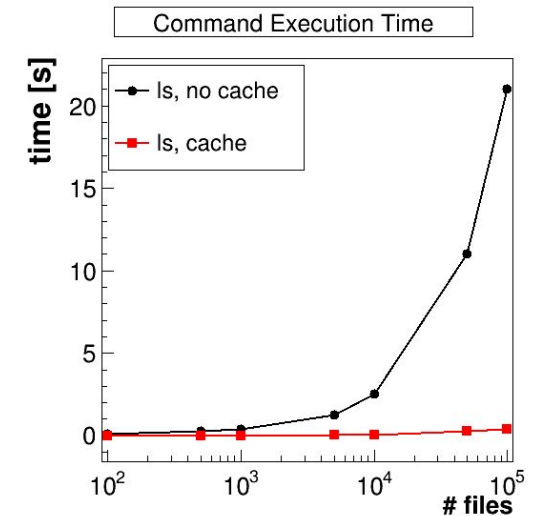
First scalability tests
[L. Lavezzi]

OSCARS

Thank you