

MADDEN Planning

Meeting with Rucio - 12.6.2025

Minutes:

- **multi-RI authentication:**
 - assuming two Data Lakes: ET and CE
 - connect CE data lake so that ET users can access read-only data from CE Data Lake. CE Rucio server should have a “ET_reader” account
 - Martin confirms this is the way to go
 - Changes needed:
 - in the Rucio authentication layer
 - in the client (API and CLI), how to display the results to the users
 - Discussion:
 - do we need finer granularity (such as only a few people from ET being allowed): not on first instance
 - allowed to read only some data (to remove for example embargoed data): might be really complex, not there yet. Interest also from astronomer communities
 - access to other data lake should be read-only: this should be fine
- **Action items:**
 - Start with a design plan (in the form of a google doc) of the rucio client
 - rucio dev meetings: every Thursday at 3pm, Matti/Nikita should join
 - Martin will send on-boarding guide for developers
- **POSIX-like view of Rucio file catalogue** (rucio-FS or plugin to CVMFS?)
 - no summer student
 - no effort from Rucio team on this. We could provide it within MADDEN, if developed tool becomes mainstream then there will be community effort to support it
 - **Action items:**
 - do some brainstorming among us (at least become familiar with current rucioFS prototype), then start discussing with CVMFS guys
 - Nikita/Matti to have a look at the current code:
<https://github.com/rucio/fuse-posix>
 - Lia can help to set it up

Project implementation

WP1: Multi-RI data management (INFN) 1/5/2025-31/12/2025

1.A) Provide ET data management (DM) system

An ET Rucio instance will be made available with two Rucio Storage Elements (RSEs). We will distribute a ready-to-use Rucio client via Docker image and CVMFS. We will provide a Rucio WebUI.

Status: On-going

Tasks:

- ET rucio server available at:
 - et-rucio-virgo.to.infn.it:
 - Alma9.5 - OK
 - move to a faster machine? Not now
 - Put everything running under rucio-admin account rather than root
 - provide a new machine or do it on the old one?
 - RSE
 - UCLouvain: attached? how much space?
 - Torino:
 - storage currently on same et-rucio-server machine (130 GB x2 disks)
 - get a new storage?
 - PIC: ?
 - authentication via user/passwd or x509
- Distribute Rucio client via:
 - github/ET gitlab + dockerfile, config file, and instructions how to use it
 - rucio client can also be installed with pip for people not using docker
 - cvmfs?

1.B) Setup the authentication of ET users to the ET DM system

We will integrate the identity and access management (IAM) ET AAI, based on the ESCAPE chosen IAM solution (Indigo-IAM).

Status: to be started

Tasks:

- understand how rucio needs to be configured to use tokens
- understand what should be allowed by IAM - ask Stefano

1.C) Install Cosmic Explorer (CE) data management instance

To mimic the multi-RI case, an independent CE Rucio instance will also be installed at INFN-Torino, and RSEs with mock CE data will be attached.

Status: to be started

Tasks:

- identify machine in Turin where we can install ce-rucio-server
 - will use the current et-rucio-server machine once we move ET rucio server to new machine
 - install and operate ce-rucio-server. Start from this as example:

<https://baltig.infn.it/INFNTO-Calcolo/et-rucio-setup>

- setup small RSE for CE mock data

1.D) Provide access to the CE Data Lake for ET users -> UCLouvain

We will make the CE Data Lake accessible to ET users via the OIDC protocol, configuring the CE Rucio instance to trust the ET IAM service as an external Identity provider (IdP). We will create a generic ET account in the CE Rucio and map all ET identities to it, with read-only permissions.

Status: to be started

Tasks:

1.E) Deploy an additional RSE in the production Virgo Data Lake

We will attach to the production Virgo Rucio server an additional RSE deployed in Torino. This setup will enable Virgo users to evaluate the developed technologies.

Status: to be started

Tasks:

- identify where to setup an additional RSE for Virgo

WP2: RucioFS (INFN) 1/8/2025 - 30/4/2026

2.A) Development of RucioFS

The current implementation of RucioFS (C/C++) lacks some features to be production ready:

- a) authentication via X.509 and OIDC tokens, using as examples the existing codes in the Rucio client (Python) and translating them to C++;
- b) update the POSIX view when updates to the Rucio DID namespace are made, by adding a system of caching of the POSIX view to speed up the access to the database (DB).

Status: to be started

Tasks:

- check if RucioFS is really what we want
- possible plugin for Rucio-CVMFS interaction
 - Federica to check with Martin first

2.B) Scale and stress tests of RucioFS

We will test and verify the scalability of RucioFS under multiple concurrent clients and up to billions of files in the catalogue. We will fill the ET and CE Data Lakes with billions of test files, creating millions of Rucio datasets. We will then deploy hundreds of RucioFS clients on the INFN Cloud to test thousands of simultaneous queries.

Status: to be started

Tasks:

WP3: Multi-RI metadata queries (UCL) 1/11/2025-30/12/2026

3.A) Add support to RucioFS for Rucio metadata

To leverage Rucio metadata capabilities for data discovery we will extend RucioFS to run parallel metadata queries across multiple Rucio instances and retrieve the corresponding data. This is a coding development to add missing features in RucioFS, i.e. metadata queries and filtering, and to distribute the queries among multiple Rucio servers.

Status: to be started

Tasks:

3.B) Implementation of Rucio metadata for the GW use case

We will define the type and amount of needed metadata and the queries for benchmarking the DB performance. Metadata is saved in the Rucio DB in JSON format without indexing. We will investigate the scalability of Rucio metadata support for GW science use cases. If the DB performance under the queries is poor, we can introduce indexing on the metadata column and investigate other optimisation possibilities. Integration with the HSF MDB adopted by the ETAP project will be implemented as needed.

Status: to be started

Tasks:

- understand which metadata should be implemented: files metadata should go in Rucio, other metadata in HFS metadata
- Paul is talking with Stefano et al. to work it out

WP4: Multi-RI testbeds (UCL) 1/5/2026-31/10/2026

4.A) Technology demonstrator for Virgo

We will build a fully functional technology demonstrator for Virgo using RucioFS to browse the Virgo Data Lake. RucioFS will be packaged and made available to Virgo users, and properly documented.

Status: to be started

Tasks:

4.B) Support for ET Mock Data Challenges

We will make available RucioFS to the ET community, and provide support to use the developed technology in upcoming Mock Data Challenges.

Status: to be started

Tasks:

GANTT chart

WP / Task	M1-3	M4-6	M7-9	M10-12	M13-15	M16-18	M19-21	M22-24
	1/2/25 - 30/4/25	1/5/25 - 31/7/25	1/8/25 - 31/10/25	1/11/25 - 31/1/26	1/2/26 - 30/4/26	1/5/26 - 31/7/26	1/8/26 - 31/10/26	1/11/26 - 31/1/27
WP1 (INFN)				8				
1.a		2						
1.b		1						
1.c			1					
1.d			3					
1.e				1				
WP2 (INFN)					9			
2.a					5			
2.b					4			
WP3 (UCL)							12	
3.a					6			
3.b							6	
WP4 (UCL)							5	
4.a						2		
4.b							3	