Data Lake as a Service for Open Science

Riccardo Di Maria (CERN)

Muhammad Aditya Hilmy (CERN IT/OpenLab)

March 25th, 2022 - Data Management & Big Data, Session II, ISGC 2022

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

Science Projects

HiLumi HL-LHC PROJECT
SKA SQUARE KILOMETRE ARRAY
VIRGO
FAIR
KM3NeT
JIVE Joint Institute for VLBI ERIC
cta cherenkov telescope array
LSST
EST european solar telescope
ESO
CERN

ESCAPE

European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

EUROPEAN OPEN SCIENCE CLOUD

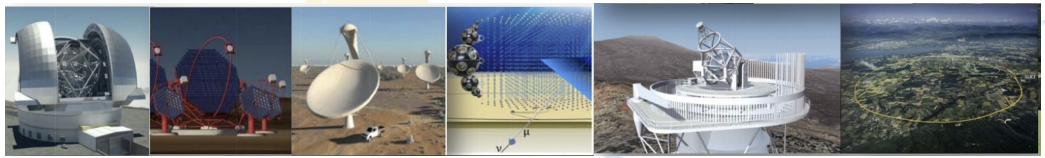
Horizon2020
European Union Funding
for Research & Innovation

Partners

CERN
INFN
GSI
DESY
Nikhef
SURF SARA
rijksuniversiteit groningen
CCIN2P3
cnrs
PIC port d'informació científica
LAPP Laboratoire d'Annecy de Physique des Particules
INAF ISTITUTO NAZIONALE DI ASTROFISICA NATIONAL INSTITUTE FOR ASTROPHYSICS

Project Goals

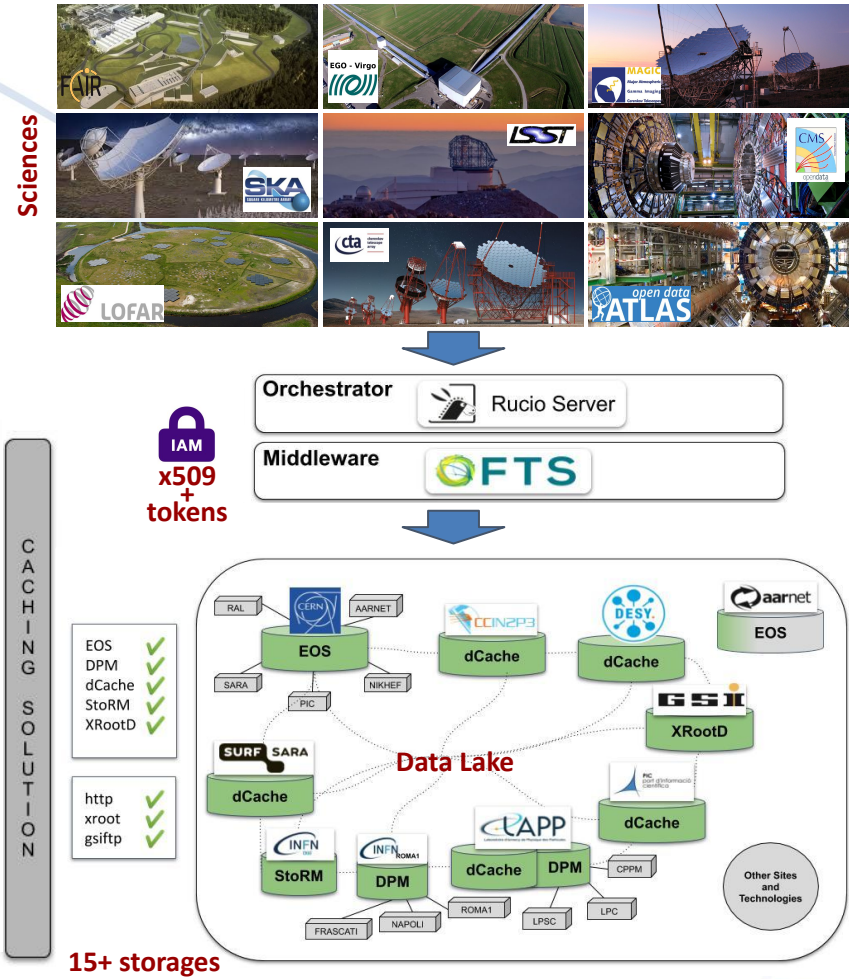
- Prototype an infrastructure adapted to exabyte-scale **future needs** of large science projects
- Ensure sciences drive the development of EOSC
- Address FAIR data management principles



Funded by the European Union's Horizon 2020 - Grant N° 824064

The Data Lake

- Data Lake as modular ecosystem of services and tools shaped around the ESCAPE scientific communities
  - federated data management and access solution
  - heterogeneous resources
    - e.g. integration of HPC and **commercial Clouds**
- Hiding complexity and providing transparent access to data
  - layer for orchestration of resources as entry point for sciences
    - define data policies and rules
  - content delivery and caching layer
    - HTTP data access and Tokens awareness for future sustainability
    - latency hiding and file re-usability
    - facilitate ingress/egress with Clouds and HPC
- Storage and compute resources not necessarily colocated

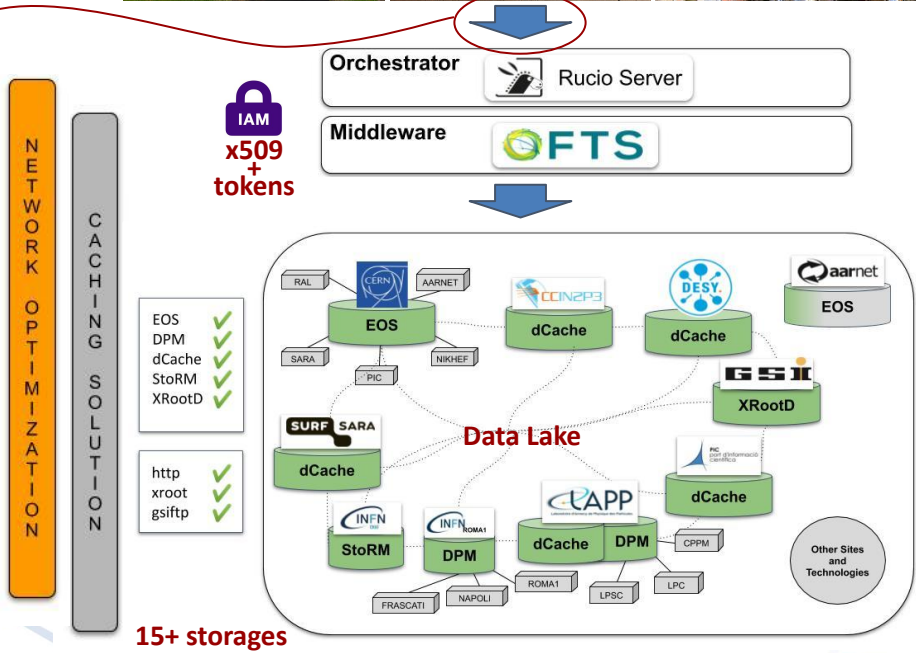


Further info: https://wiki.escape2020.de/index.php/WP2_-_DIOS#Datalake_Status

Riccardo.Di.Maria@cern.ch
Funded by the European Union's Horizon 2020 - Grant N° 824064

DLaaS for Open Science

- Goal: make **end-user comfortable** in embarking on a Data Lake experience
  - abstract the complexities of the Data Lake from the scientists
    → focus on doing science instead of data procurement
- An ever-increasing number of experiments are looking at Rucio Data Management system
  - **DLaaS** potentially interesting for both **aficionados** and **newcomers**



Riccardo.Di.Maria@cern.ch
Funded by the European Union's Horizon 2020 - Grant N° 824064

As It All Started

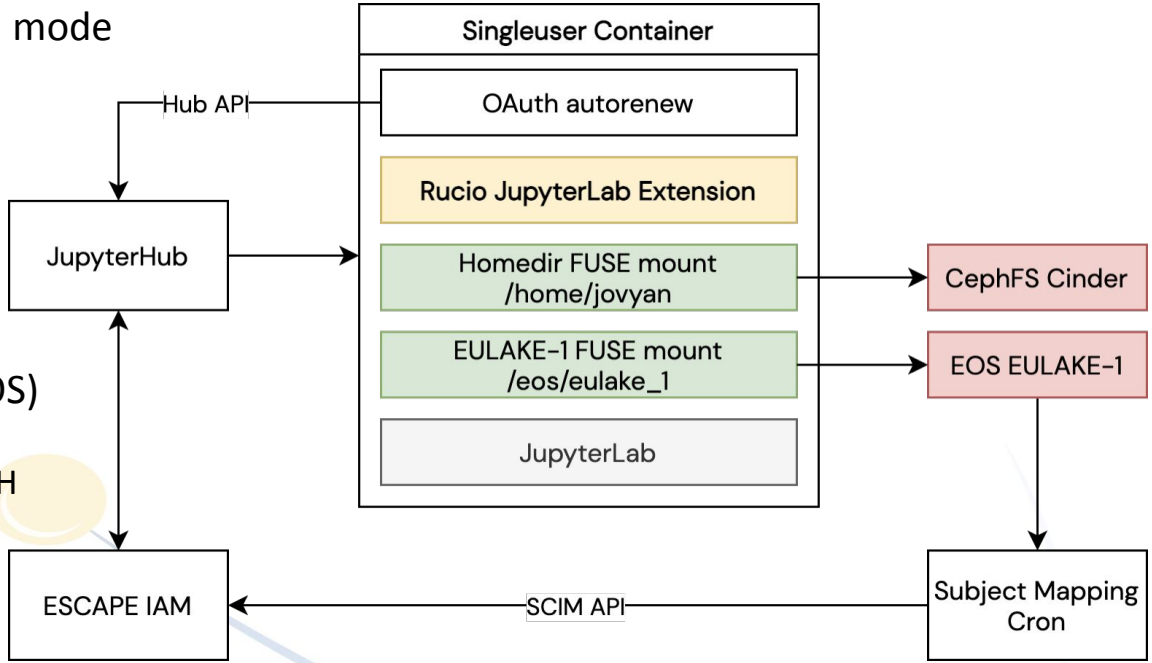
- An idea presented at CS3 2020 by the Rucio team [1]
- Development of a “Rucio JupyterLab Extension” as part of GSoC 2020 [2,3]
- A long time has passed, many things have happened…
  - CERN OpenLab Summer Student to concretise the effort in 2021
    - deployment, docker-images, documentation
  - DataLake-as-a-Service (DLaaS) in production-like phase
    - extensively exploited during ESCAPE “Data and Analysis Challenge” in November 2021 by SKA, MAGIC, CTA, ATLAS, KM3NET, LOFAR, FAIR
      → 3rd ESCAPE DIOS Workshop
    - EU projects e.g. EOSC-Future and CS3Mesh4EOSC/ScienceMesh
    - other communities e.g. EGI

Funded by the European Union’s Horizon 2020 - Grant N° 824064

DLaaS Implementation

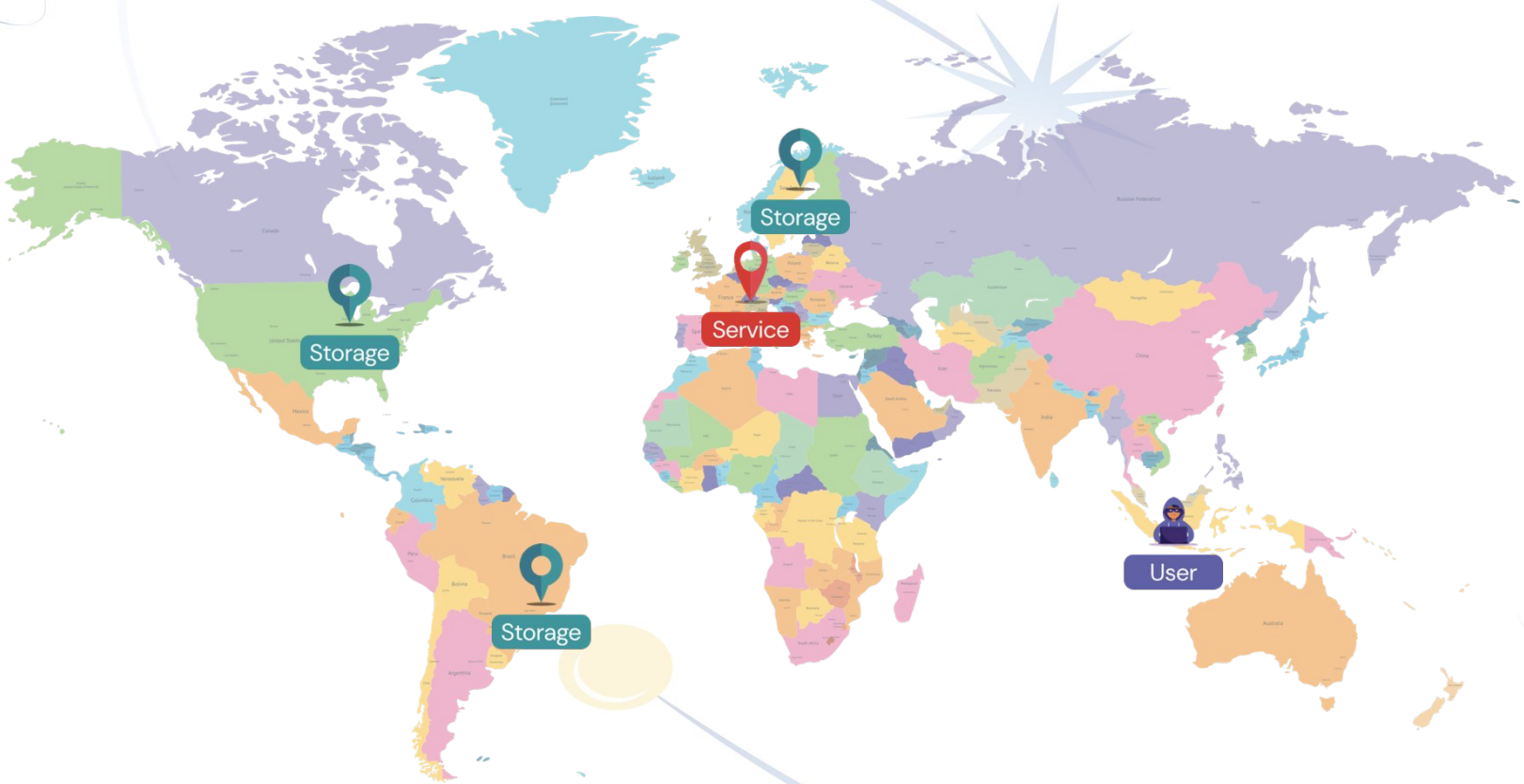
RSE = Rucio Storage Element

- Deployed in Kubernetes, using Zero-to-JupyterHub Helm chart → https://escape-notebook.cern.ch
- OAuth authentication using ESCAPE IAM (X509 still supported)
- Rucio JupyterLab Extension in Replica mode (i.e. TPC to local storage) used
  - download mode still possible (if configured)
  - connected to ESCAPE Data Lake
  - automatically pre-configured to use OpenID Connect
  - 2 FUSE mounts to EULAKE-1 (EOS)
    - ESCAPE RSE in r-only
    - additional RSE in r/w: SCRATCH
  - making files available *aka* creating a replication rule to move files to EULAKE-1



Funded by the European Union's Horizon 2020 - Grant N° 824064

DataLake-as-a-Service for Open Science



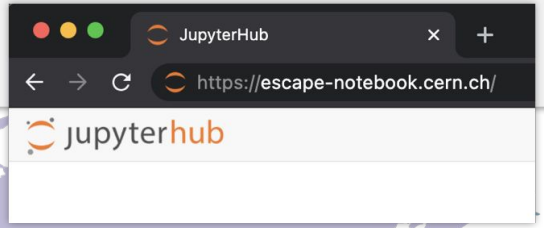
Riccardo.Di.Maria@cern.ch

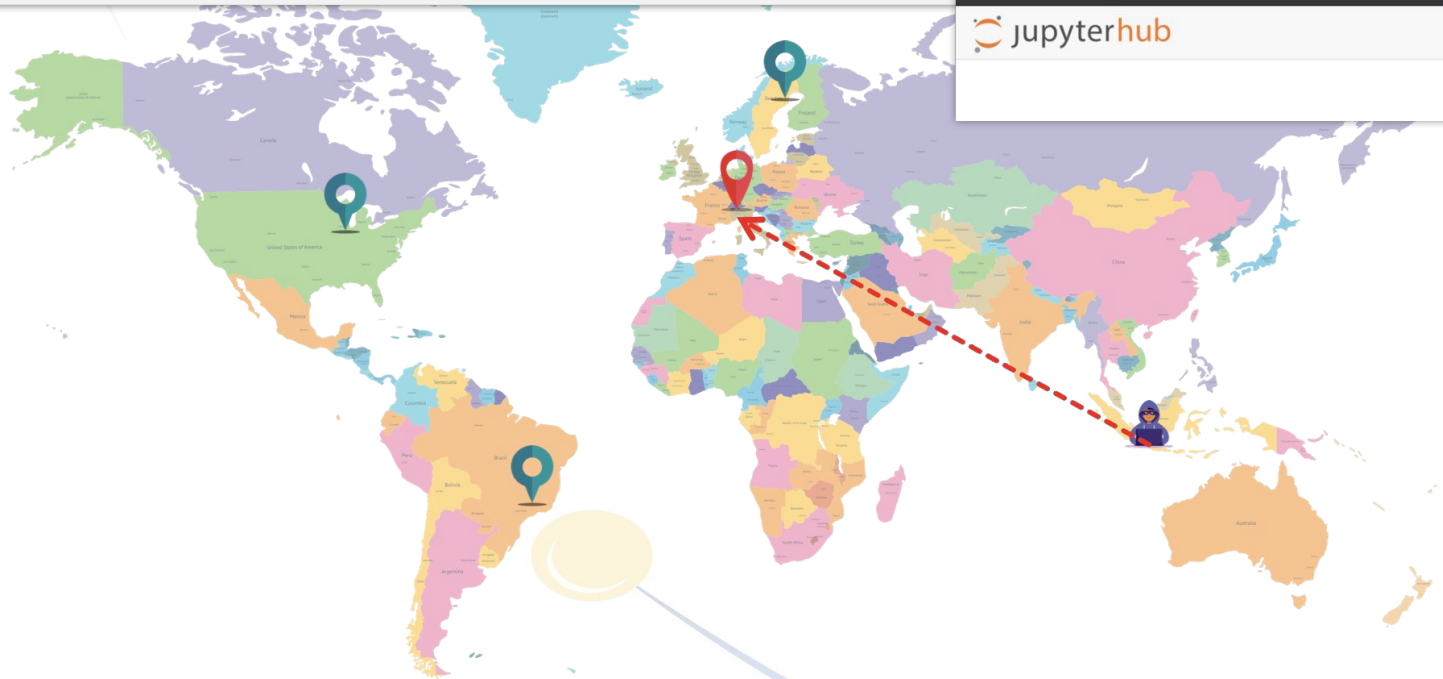
Funded by the European Union's Horizon 2020 - Grant N° 824064

Contact DLaaS

→ Scientists Contact DataLake-as-a-Service

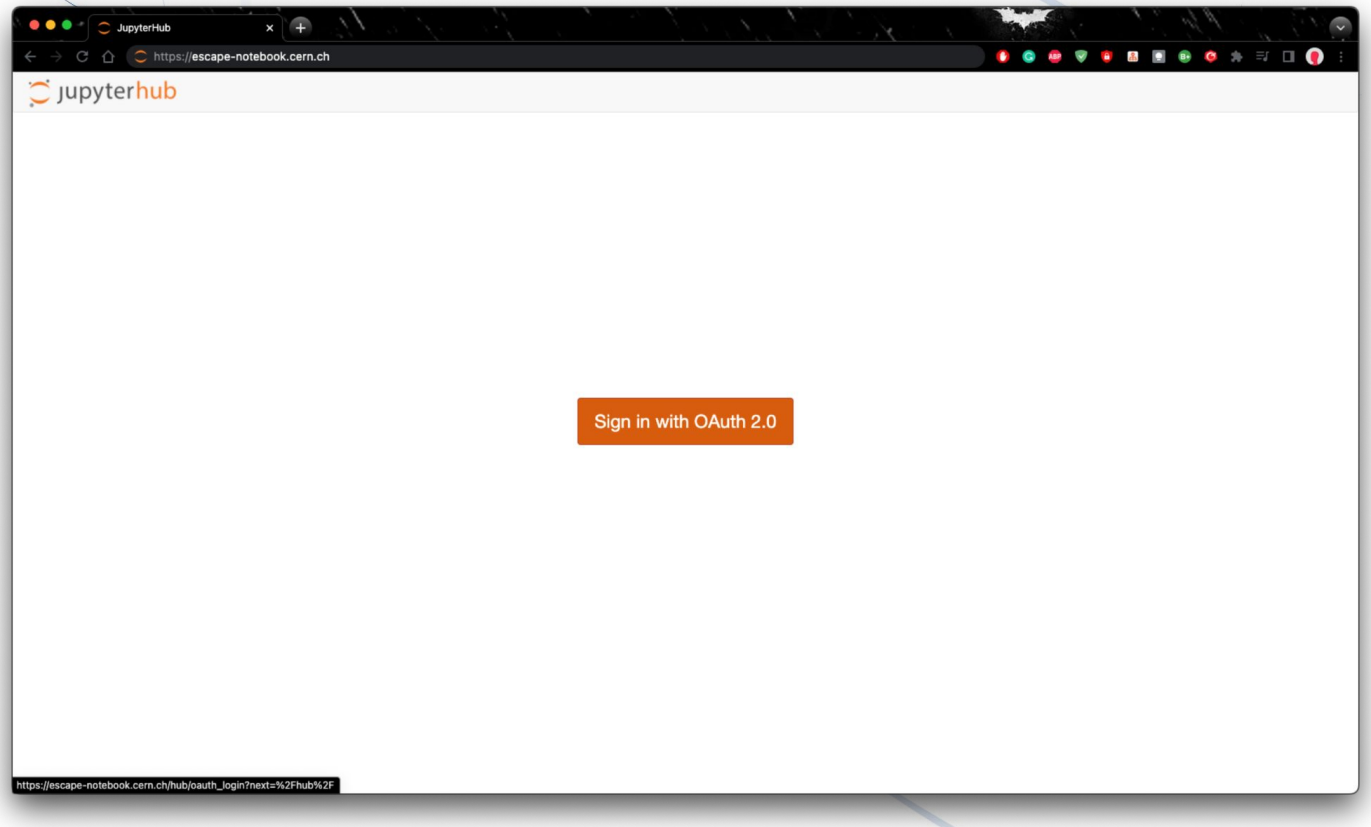
Requests are handled by Jupyter servers at CERN, Geneva





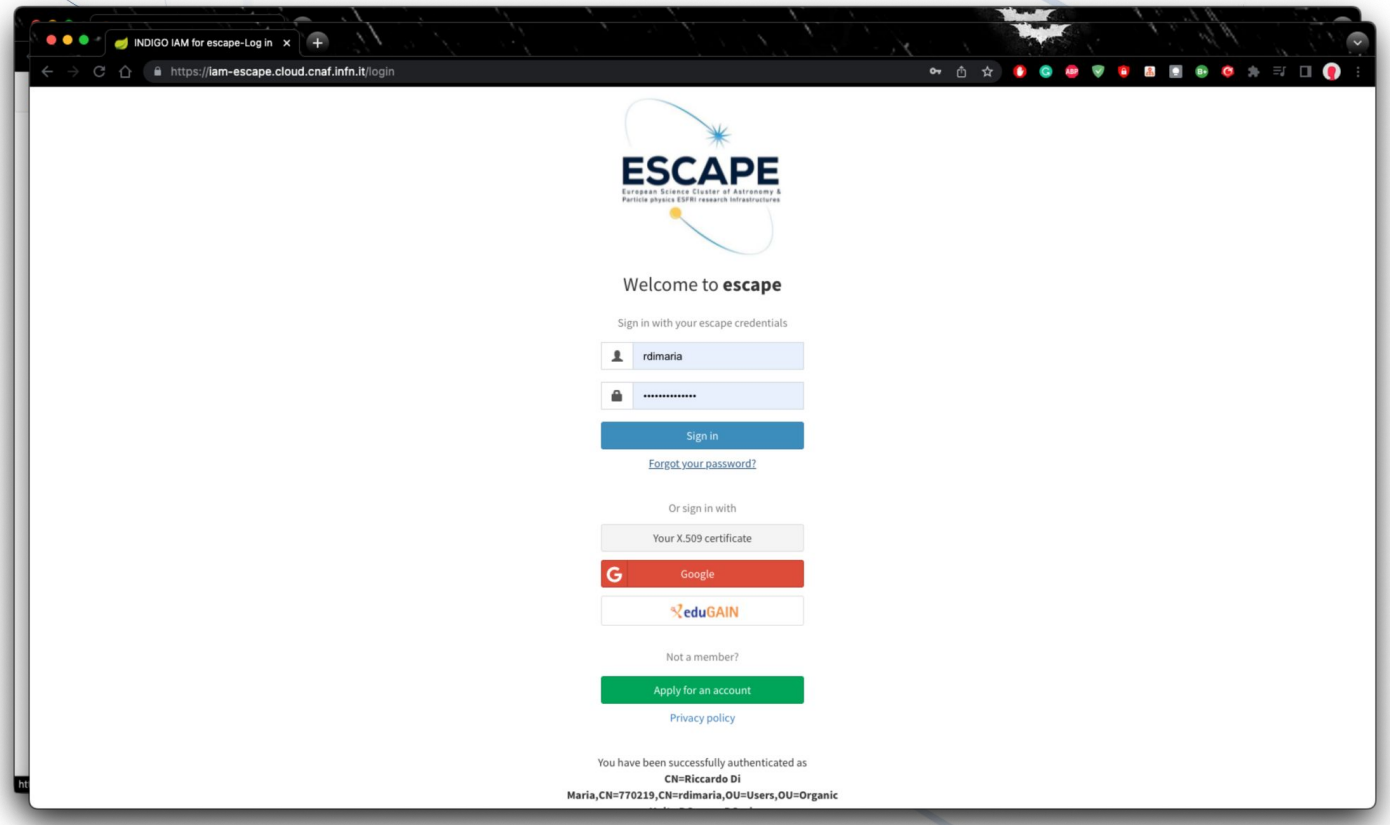
Funded by the European Union's Horizon 2020 - Grant N° 824064

Sign in with OAuth 2.0



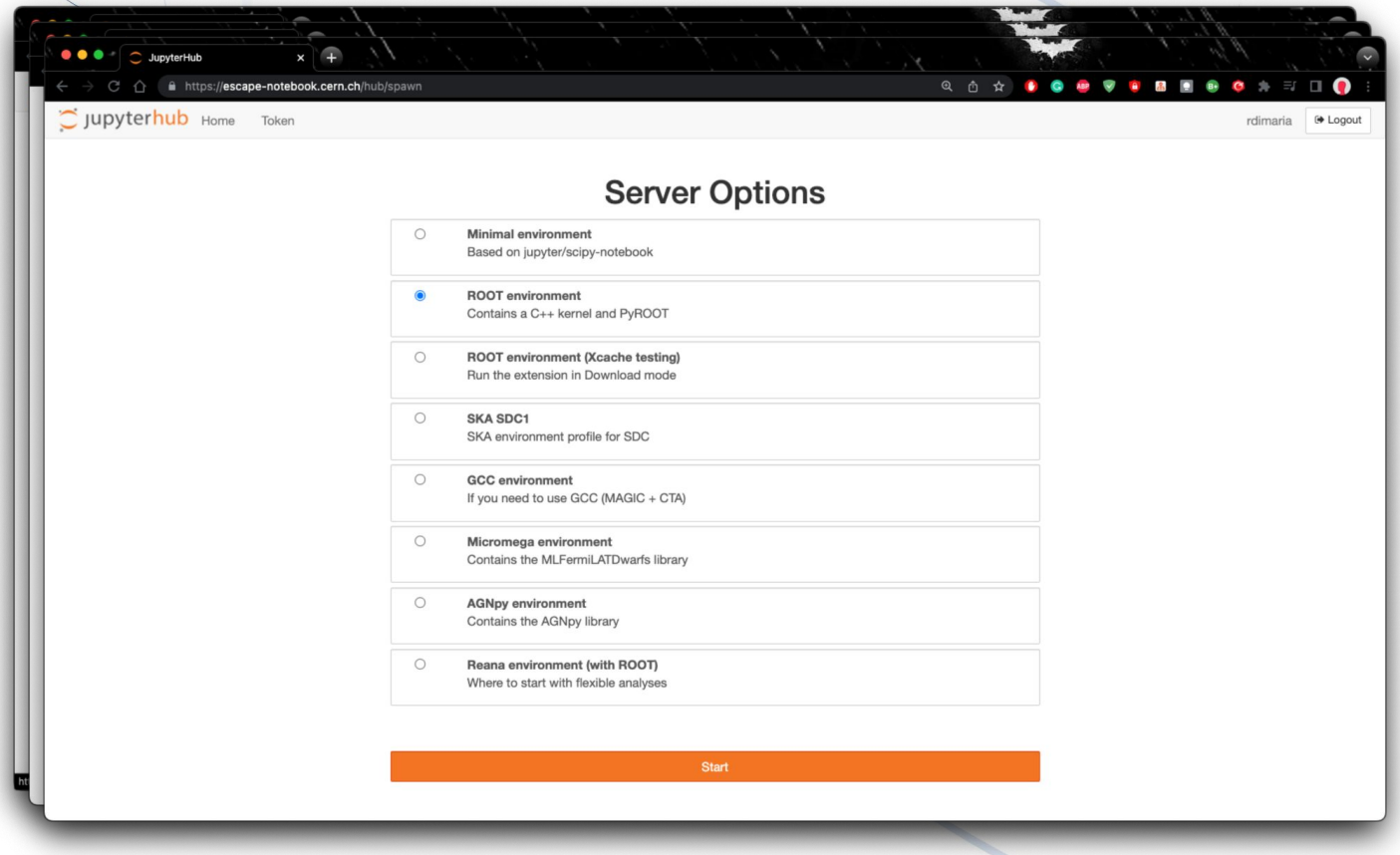
Funded by the European Union's Horizon 2020 - Grant N° 824064

ESCAPE Indigo IAM for Auth



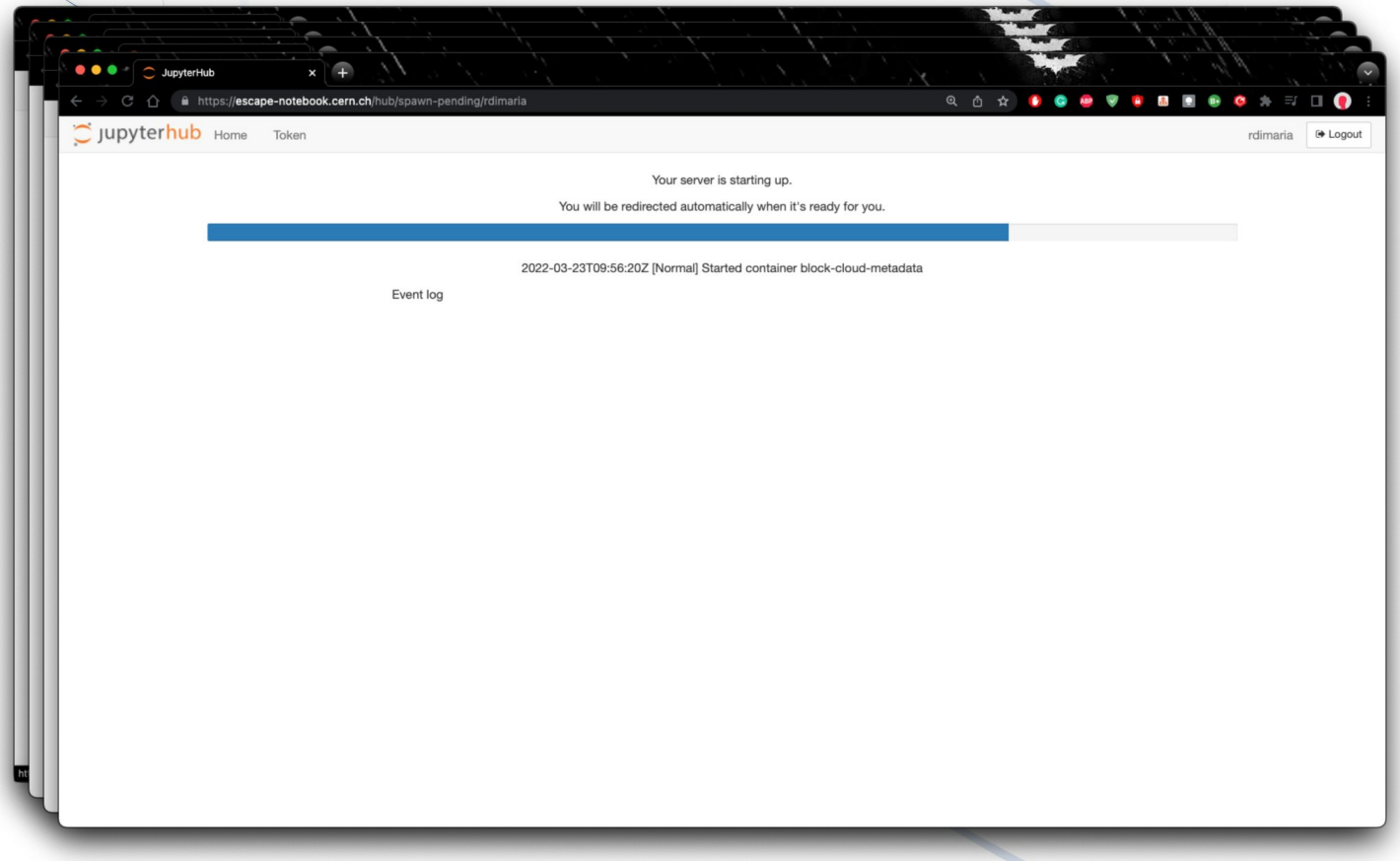
Funded by the European Union's Horizon 2020 - Grant N° 824064

Multiple Server Options for Sciences



Funded by the European Union's Horizon 2020 - Grant N° 824064

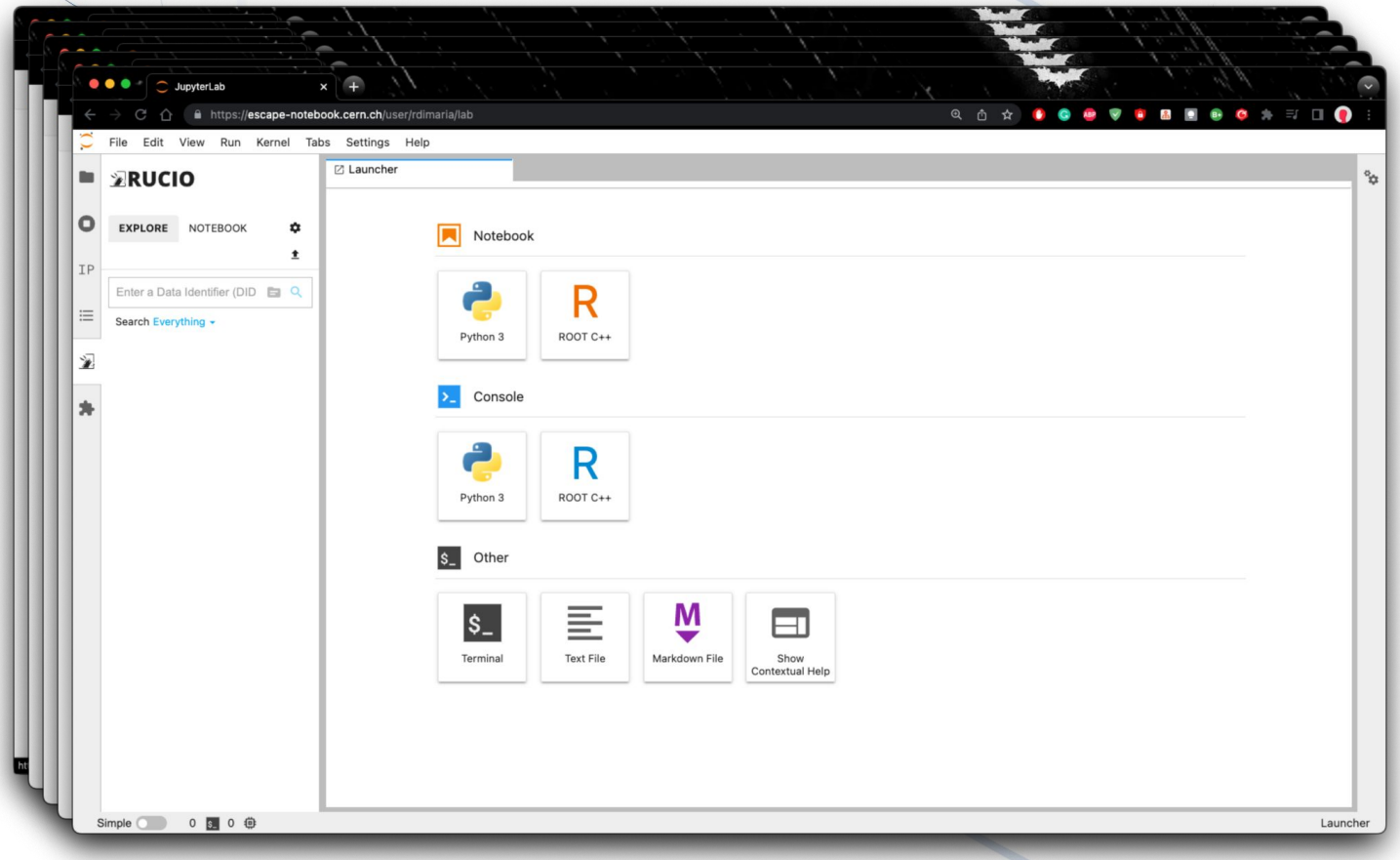
K8s-Pod per User



Riccardo.Di.Maria@cern.ch

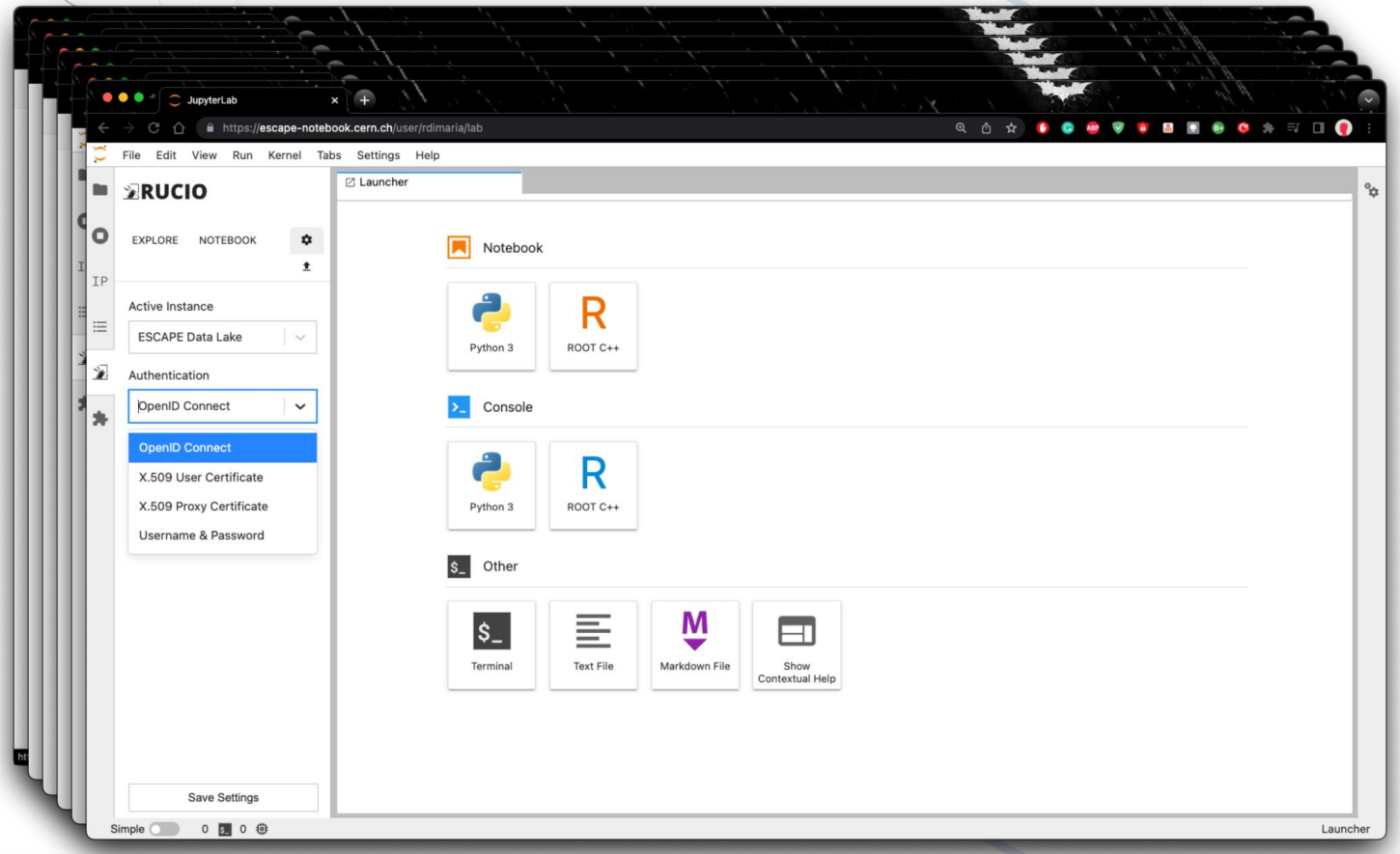
Funded by the European Union's Horizon 2020 - Grant N° 824064

Jupyter Notebook with Rucio Extension



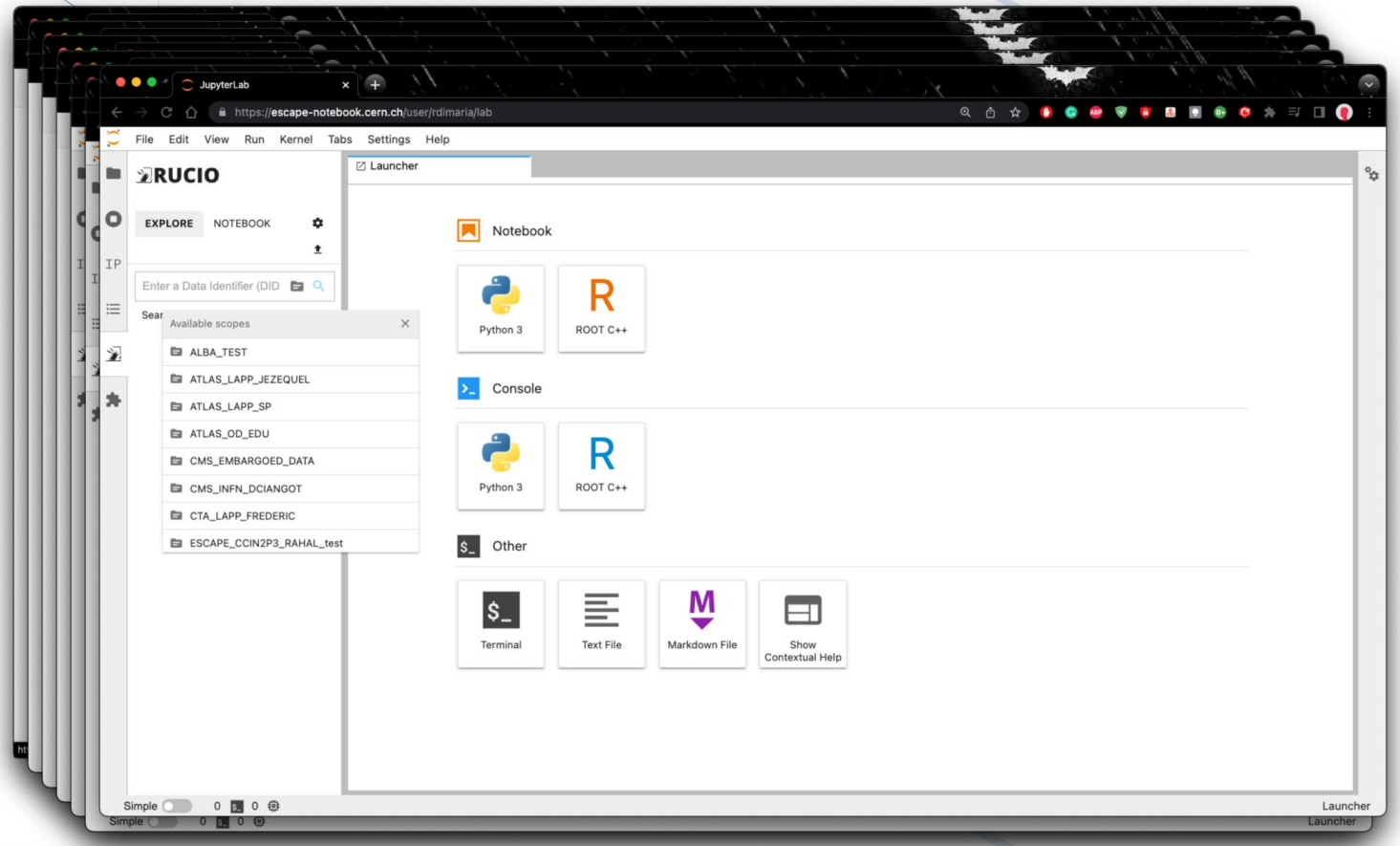
Funded by the European Union's Horizon 2020 - Grant N° 824064

ESCAPE Rucio Instance and OpenID Connect



Funded by the European Union's Horizon 2020 - Grant N° 824064

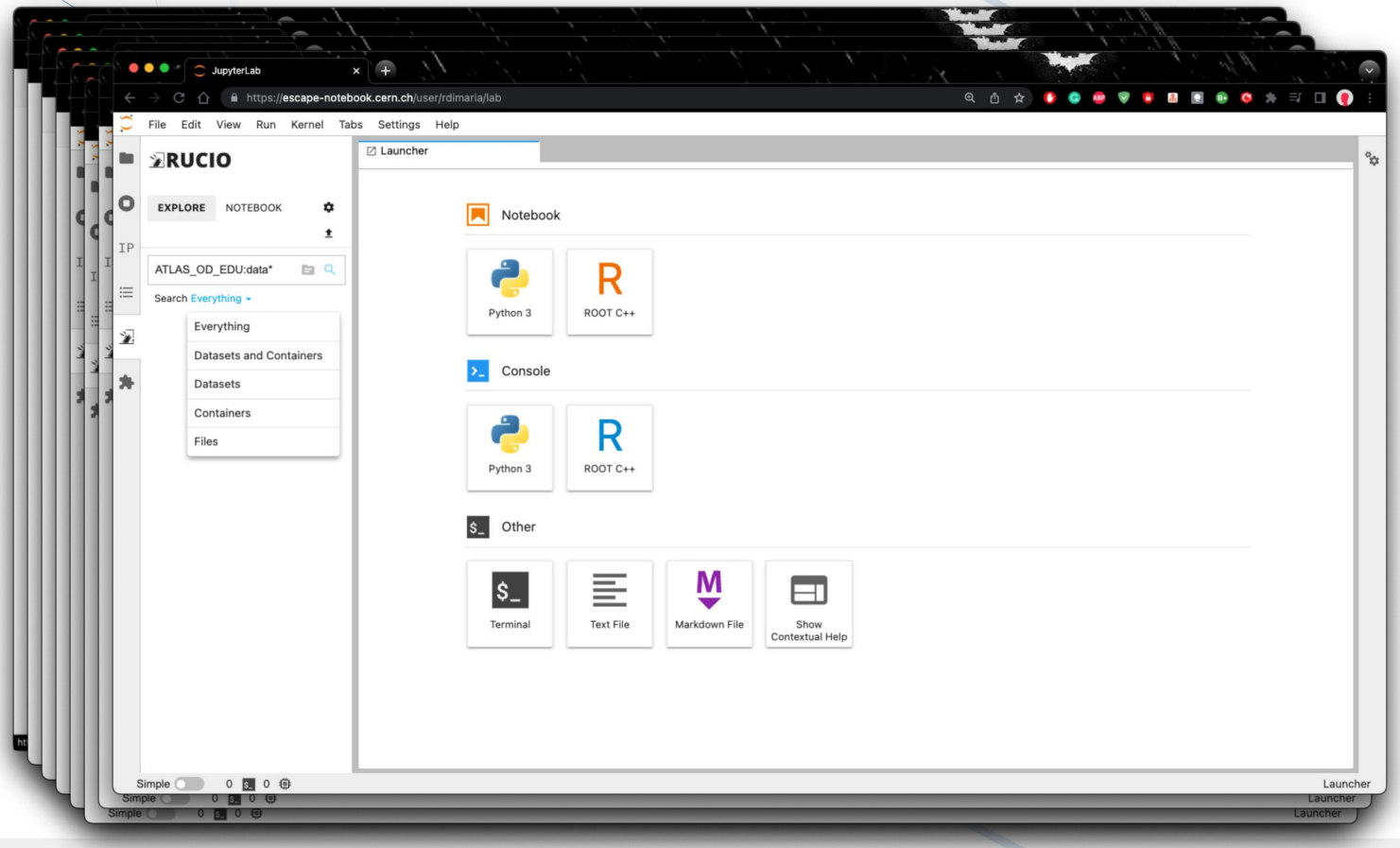
Data Browsing



Riccardo.Di.Maria@cern.ch

Funded by the European Union's Horizon 2020 - Grant N° 824064

Data Browsing

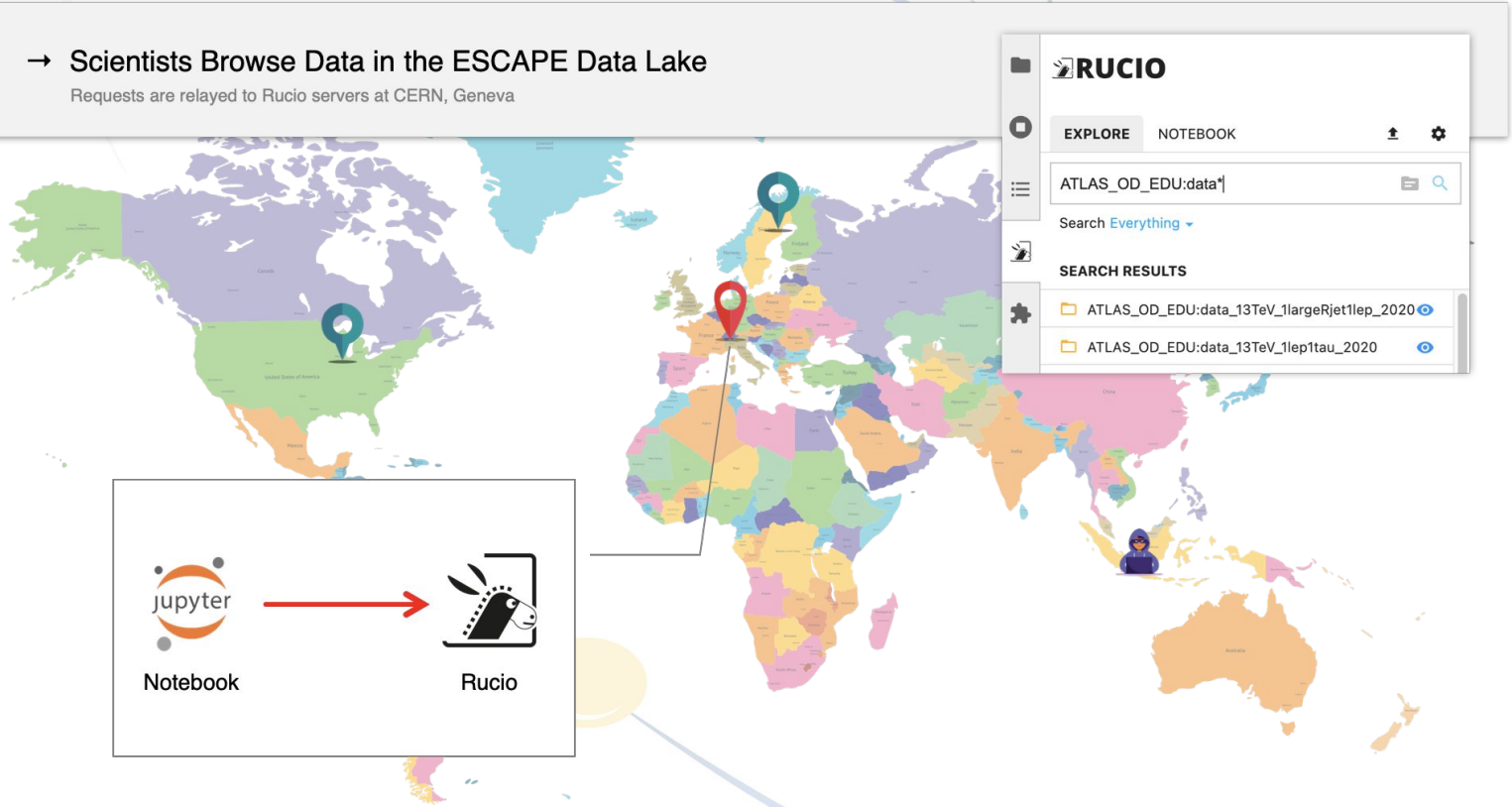


Funded by the European Union's Horizon 2020 - Grant N° 824064

Data Browsing

→ Scientists Browse Data in the ESCAPE Data Lake

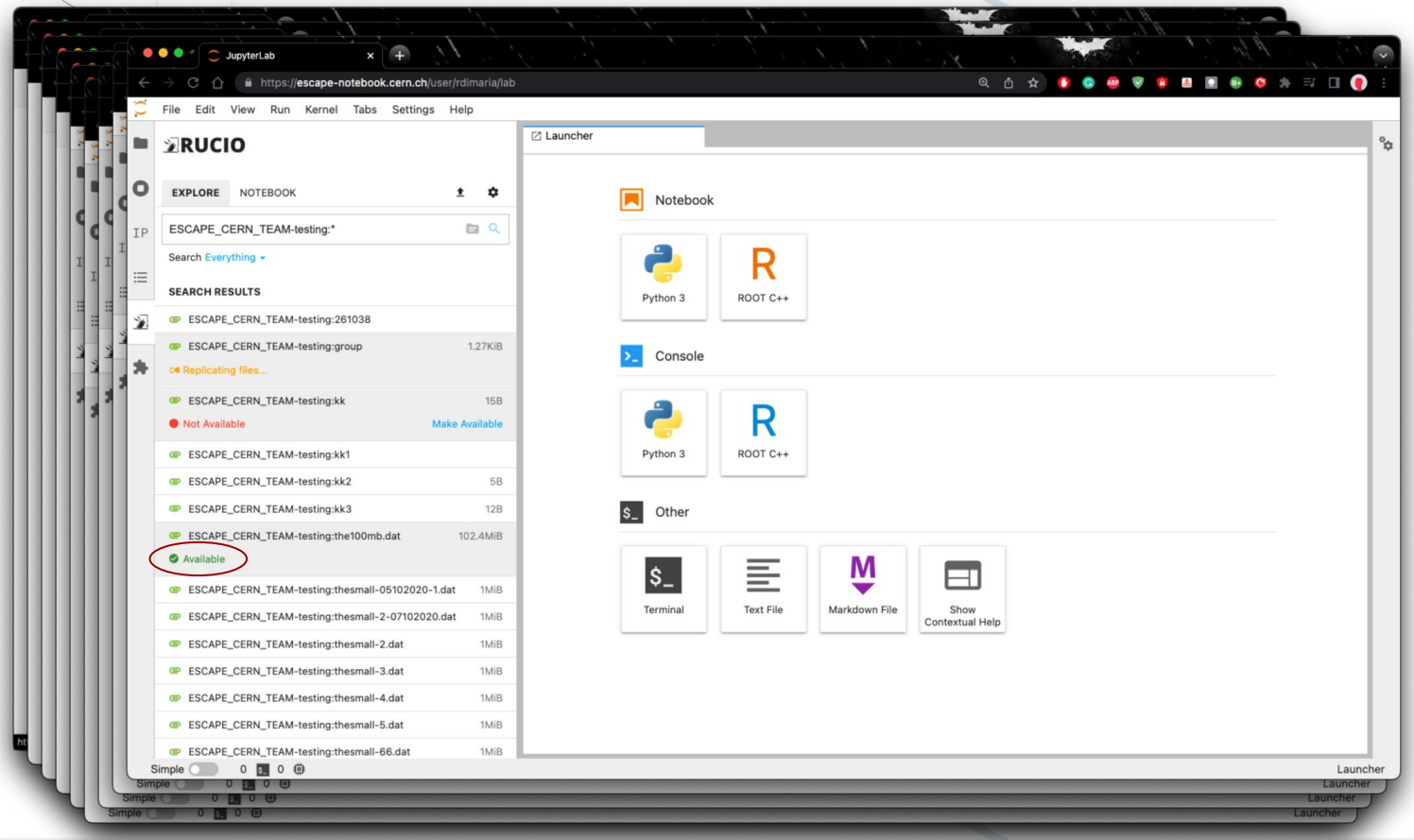
Requests are relayed to Rucio servers at CERN, Geneva



Riccardo.Di.Maria@cern.ch

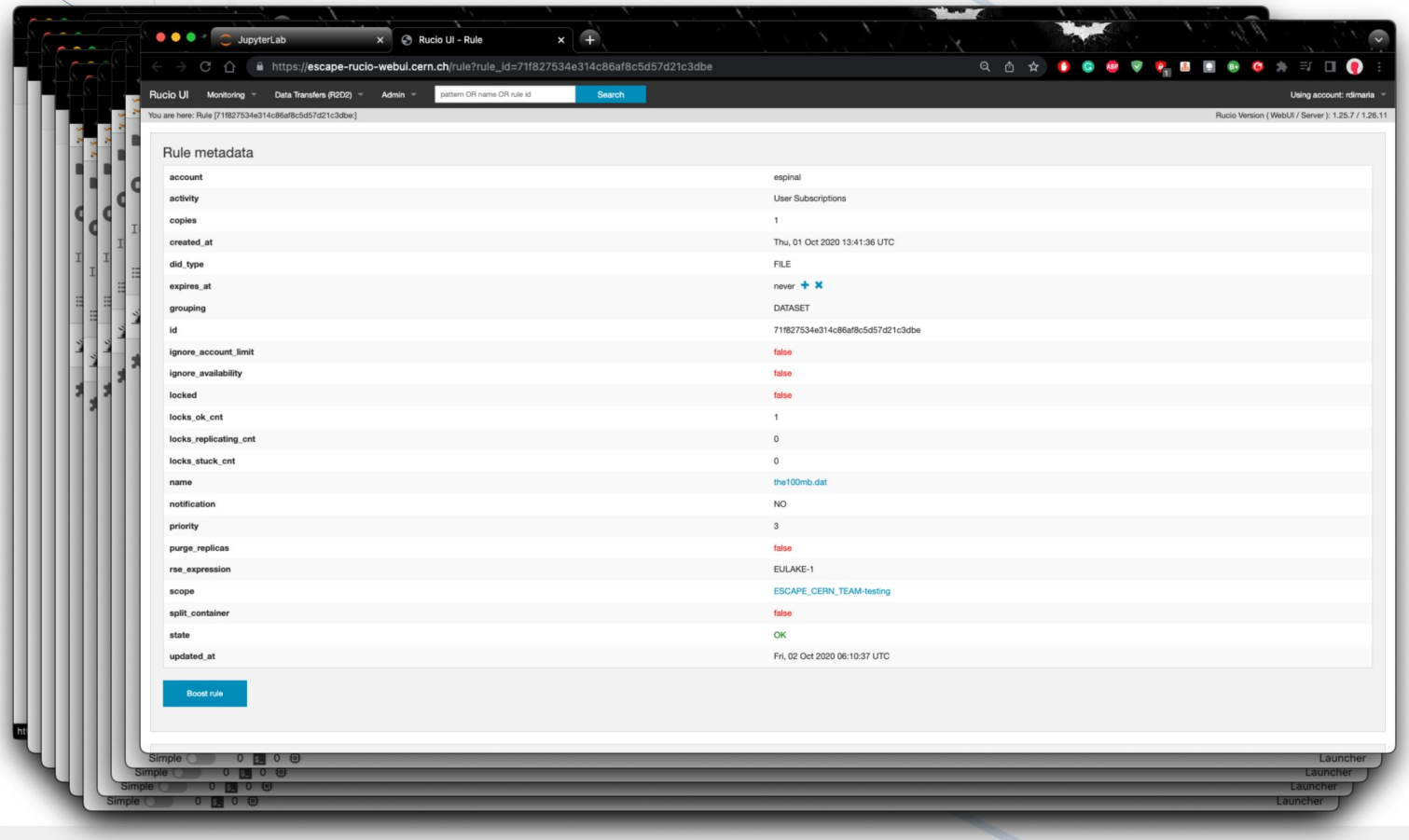
Funded by the European Union's Horizon 2020 - Grant N° 824064

Data Browsing



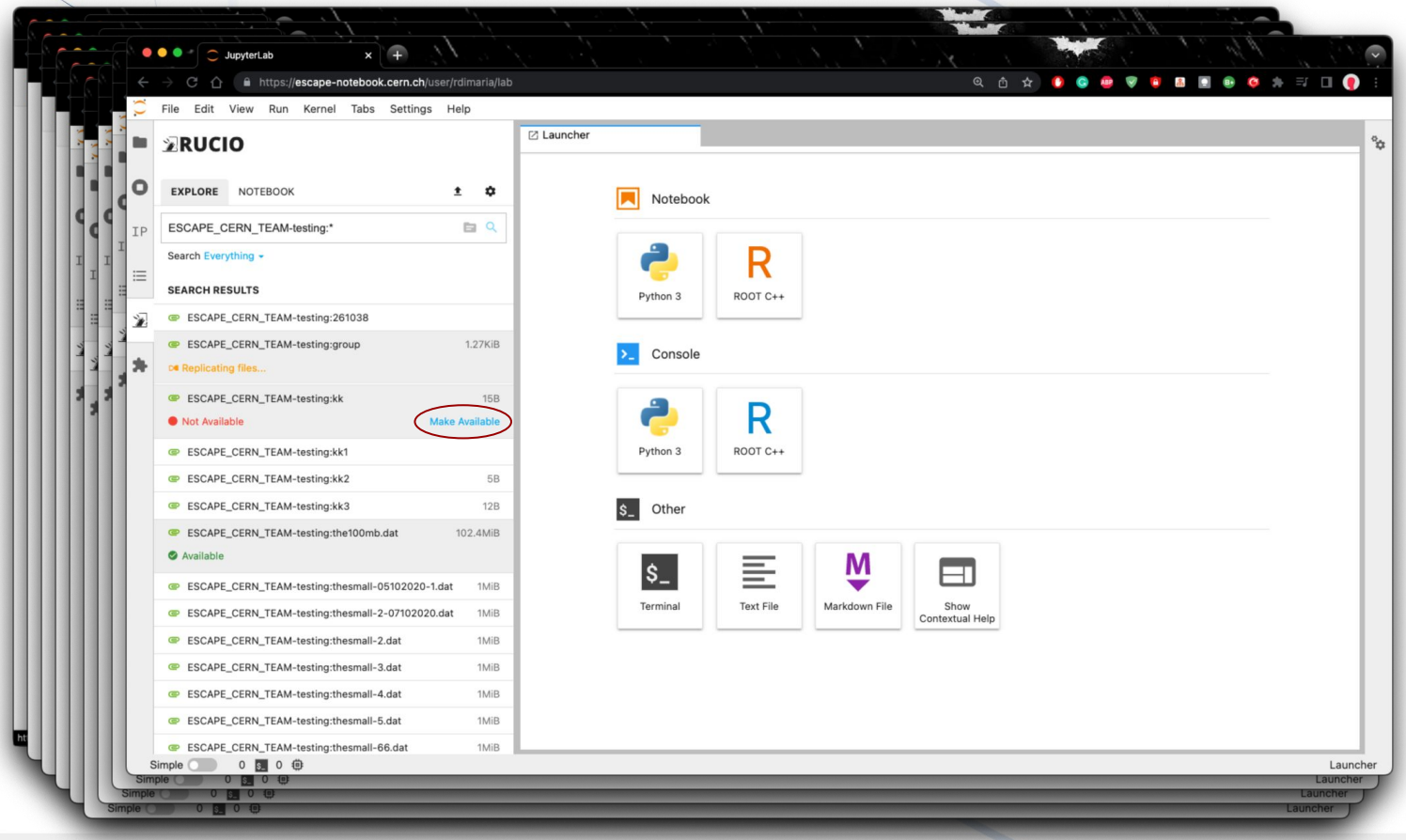
Funded by the European Union's Horizon 2020 - Grant N° 824064

Redirection to ESCAPE Rucio WebUI



Funded by the European Union's Horizon 2020 - Grant N° 824064

Make Data Available “Locally”

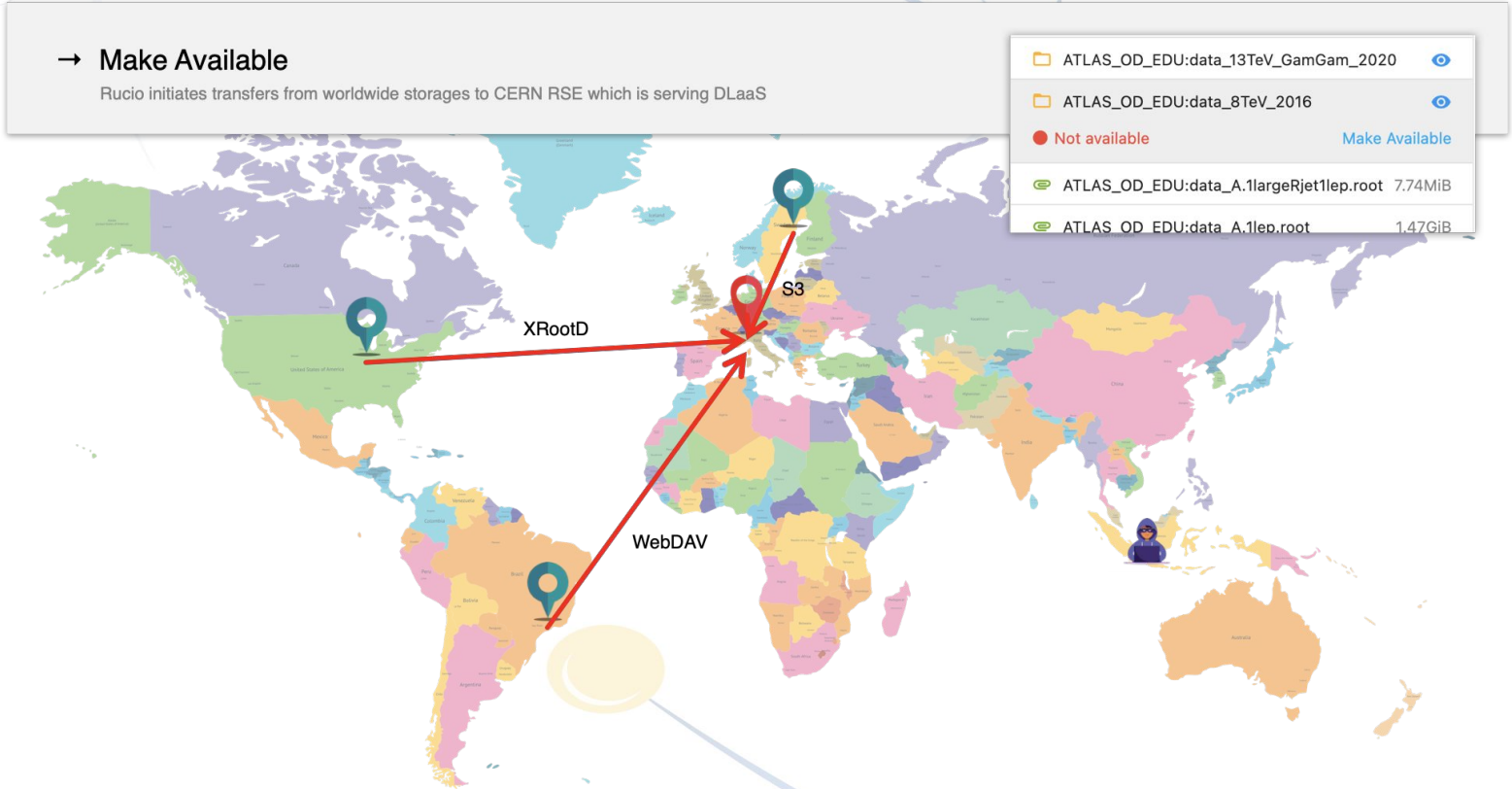


Funded by the European Union’s Horizon 2020 - Grant N° 824064

Make Data Available “Locally”

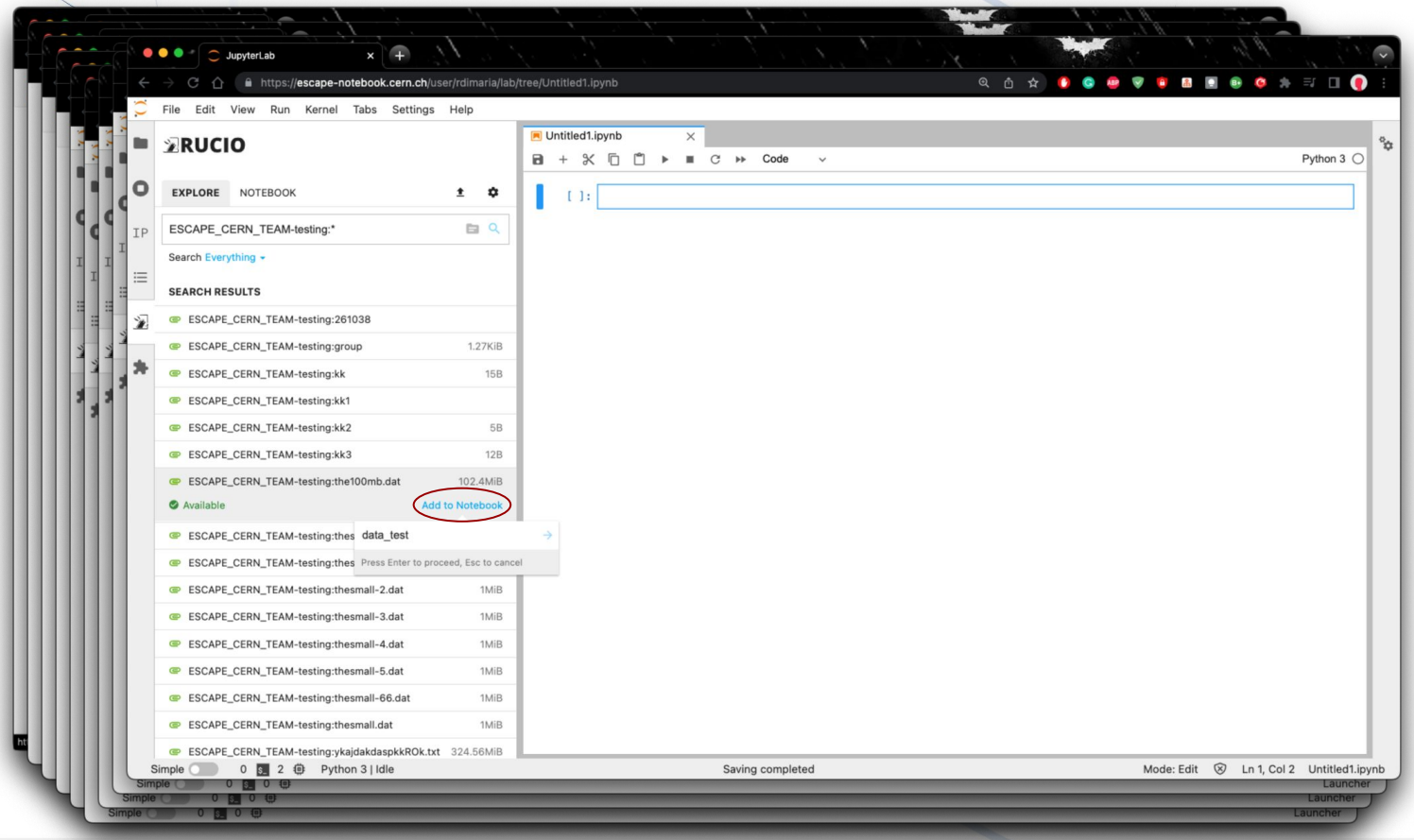
→ **Make Available**

Rucio initiates transfers from worldwide storages to CERN RSE which is serving DLaaS



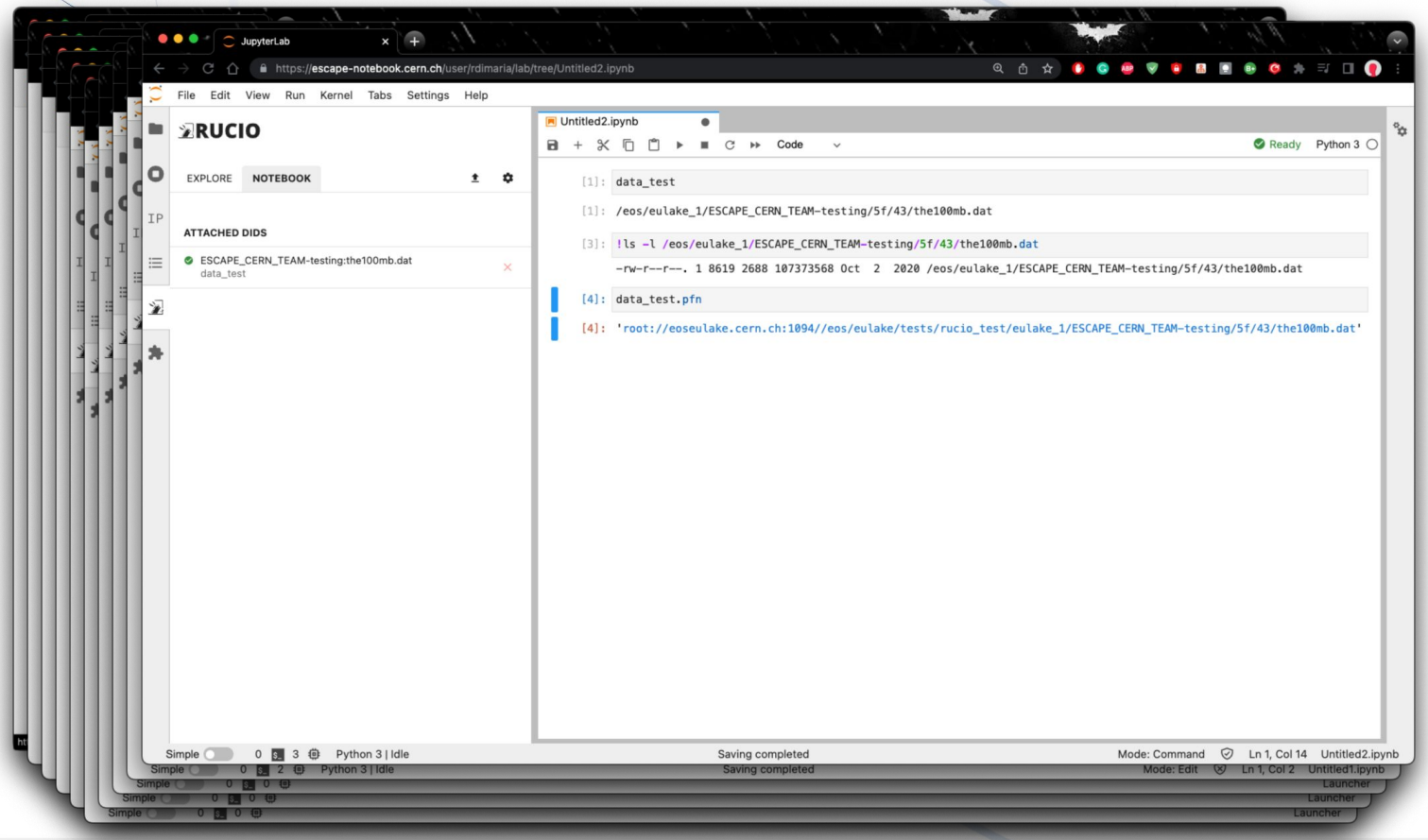
Funded by the European Union’s Horizon 2020 - Grant N° 824064

Add Data to Notebook



Funded by the European Union's Horizon 2020 - Grant N° 824064

Use Data for Analysis

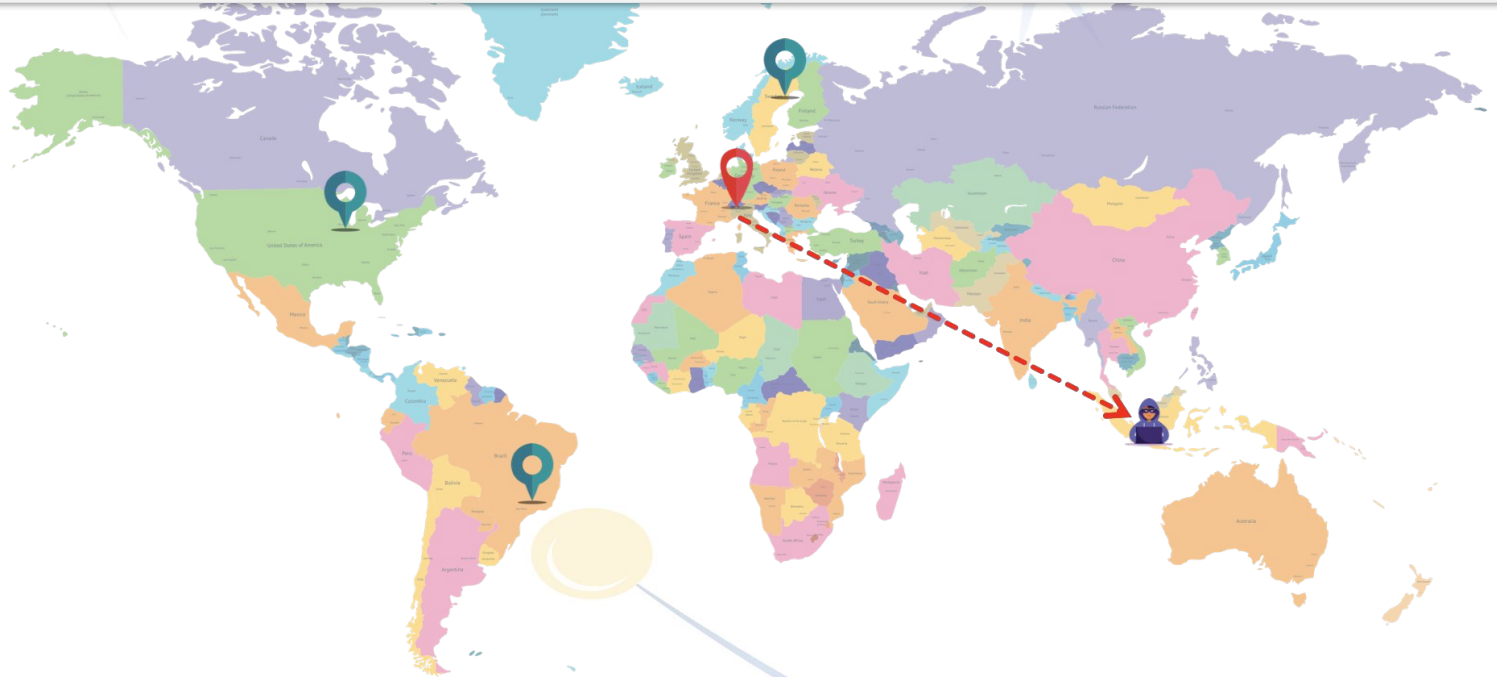


Funded by the European Union's Horizon 2020 - Grant N° 824064

DataLake-as-a-Service for Open Science

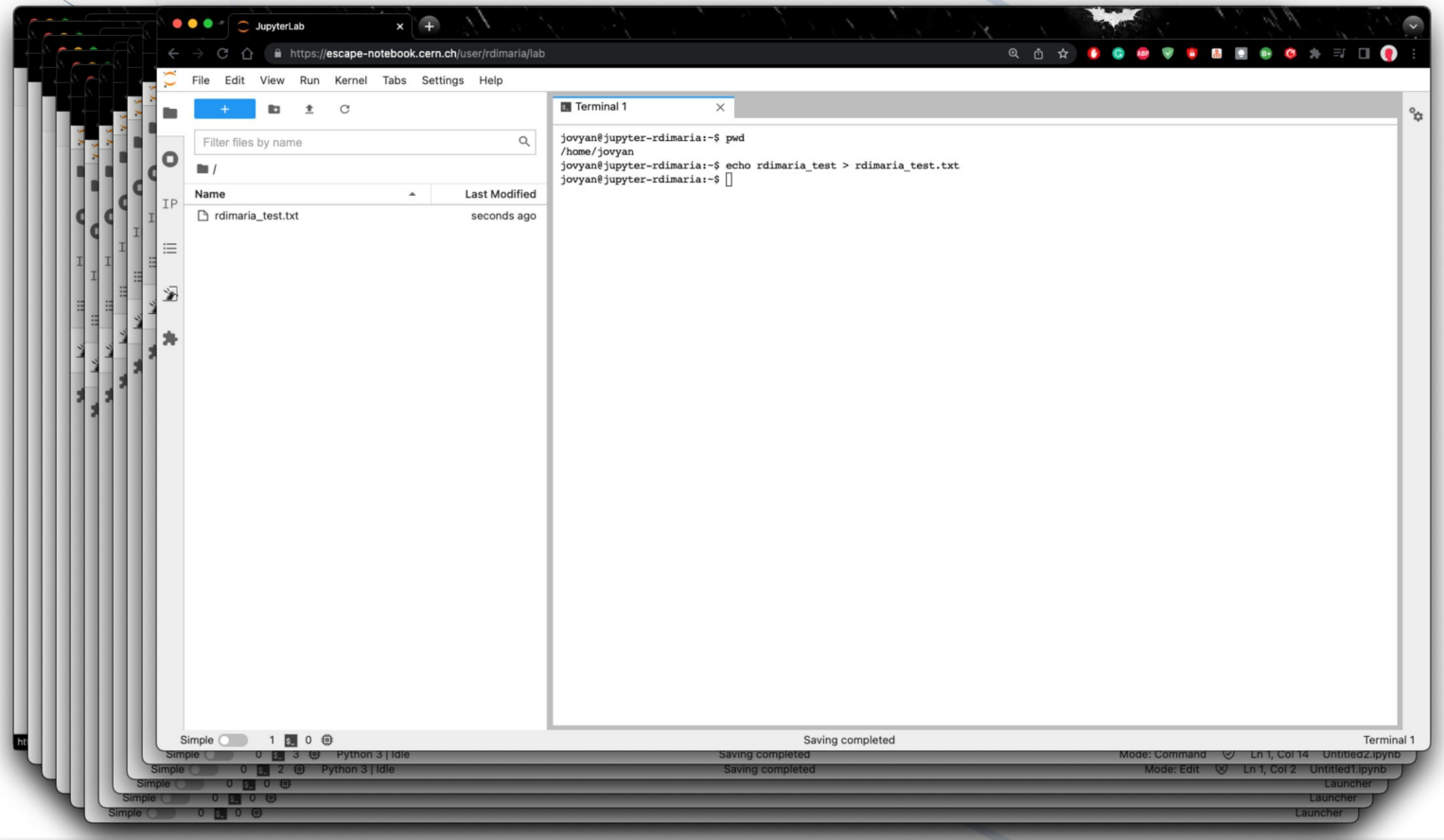
→ **Scientists' Analysis Workflows Read Data**

The code runs on the Notebook server at CERN, and the output is shown to the user



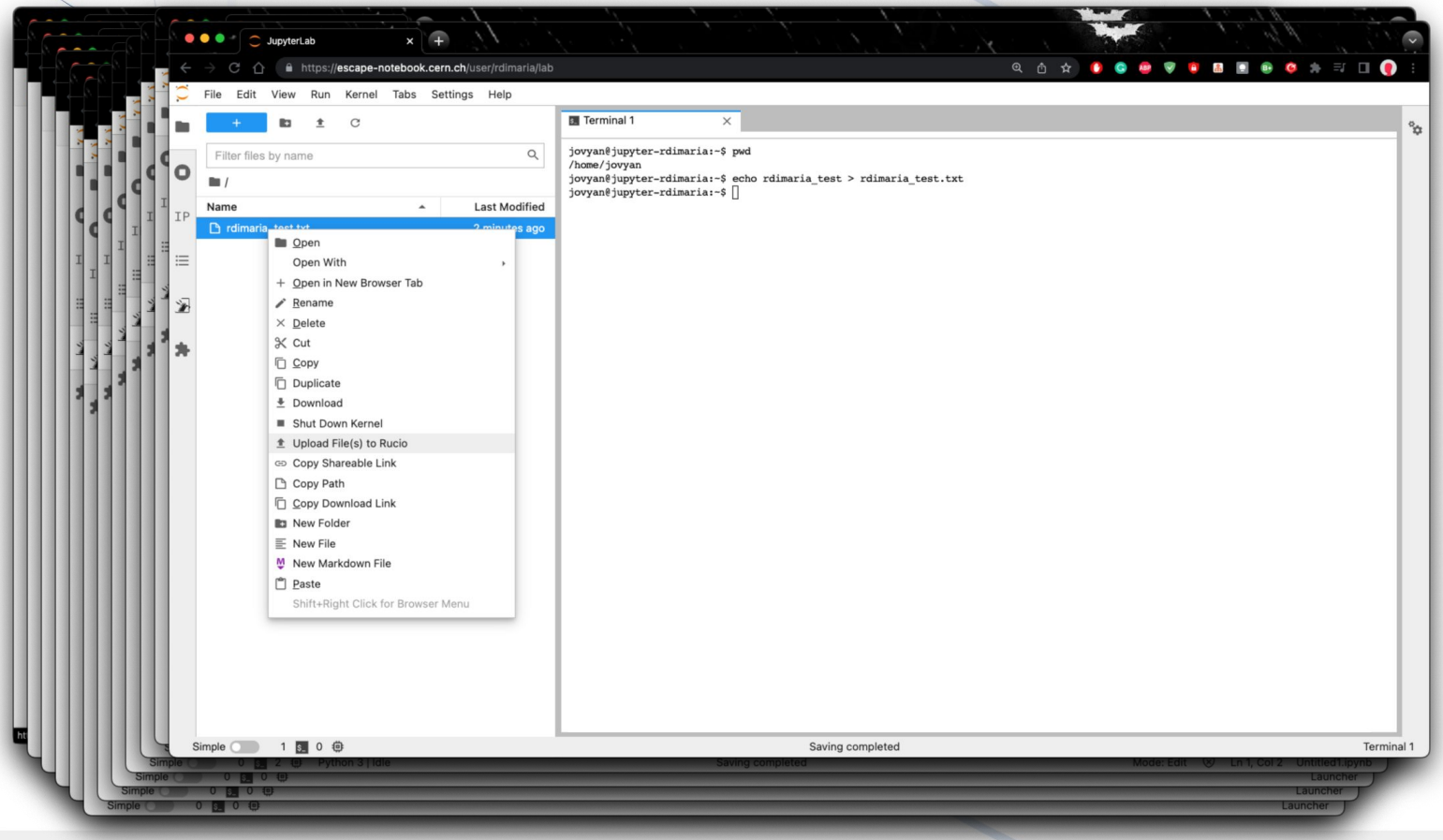
Funded by the European Union's Horizon 2020 - Grant N° 824064

Output of Analysis (Small-Sized Files)



Funded by the European Union's Horizon 2020 - Grant N° 824064

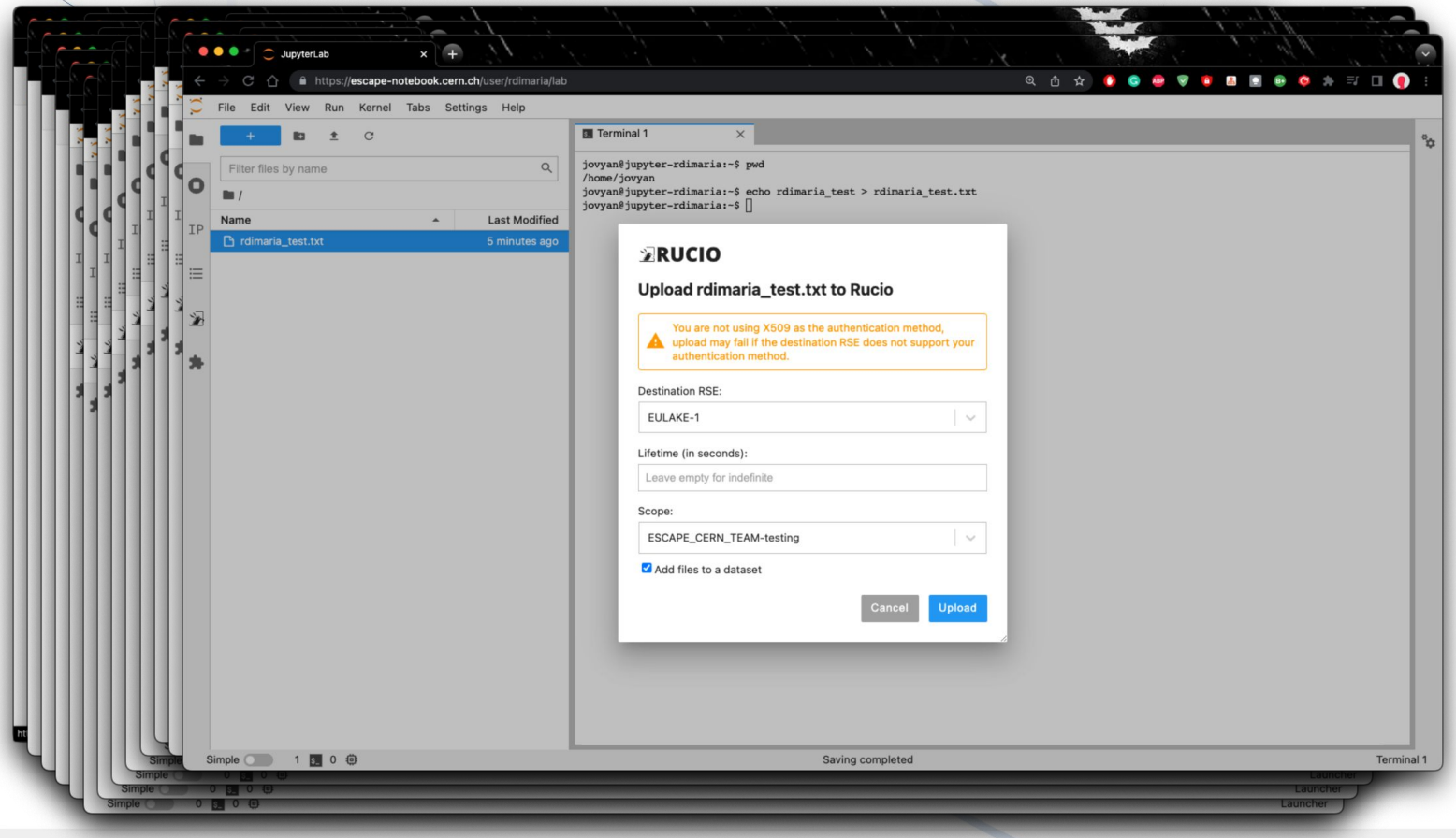
Data Preservation Use Case



Riccardo.Di.Maria@cern.ch

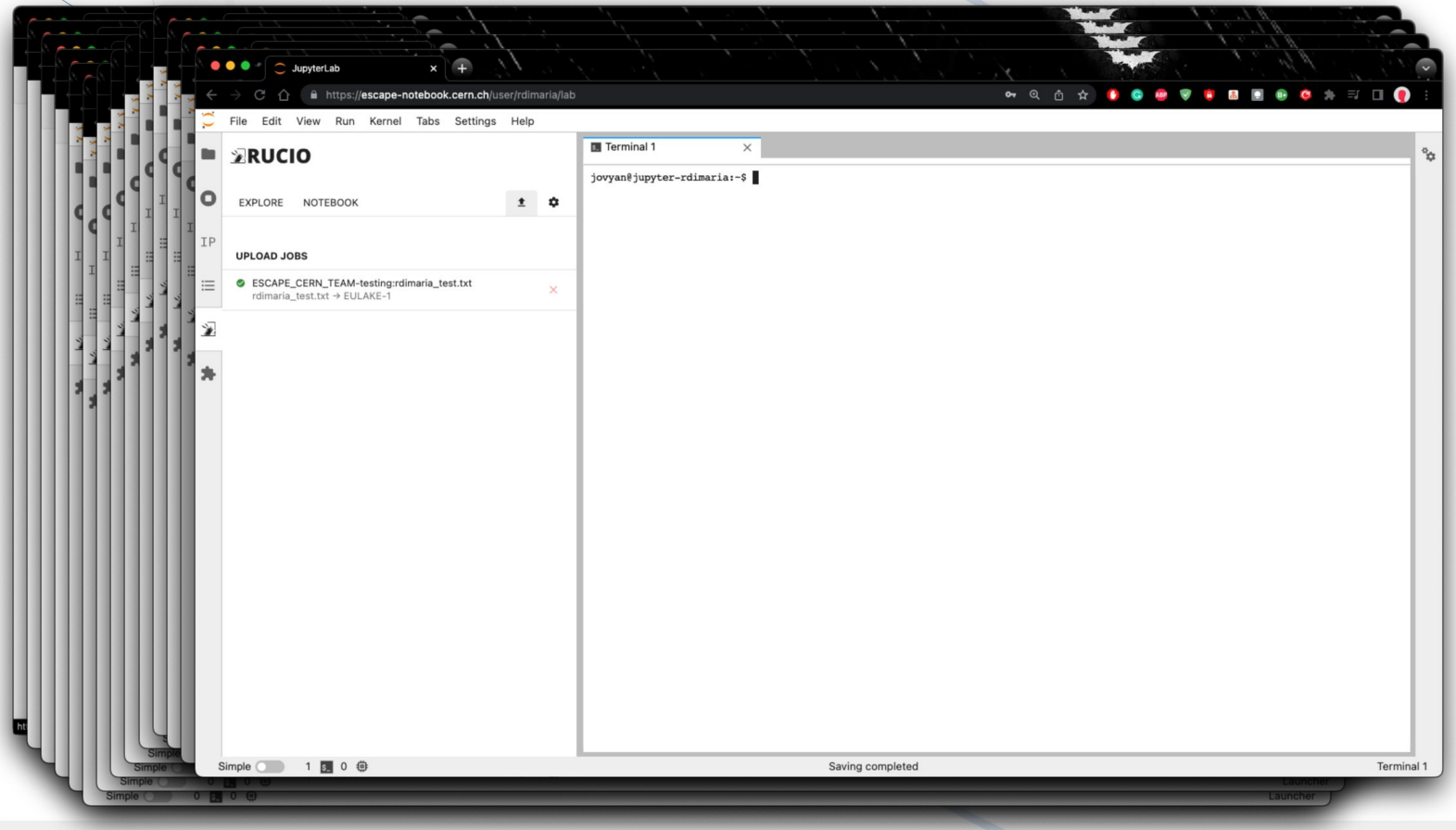
Funded by the European Union's Horizon 2020 - Grant N° 824064

Data Preservation Use Case



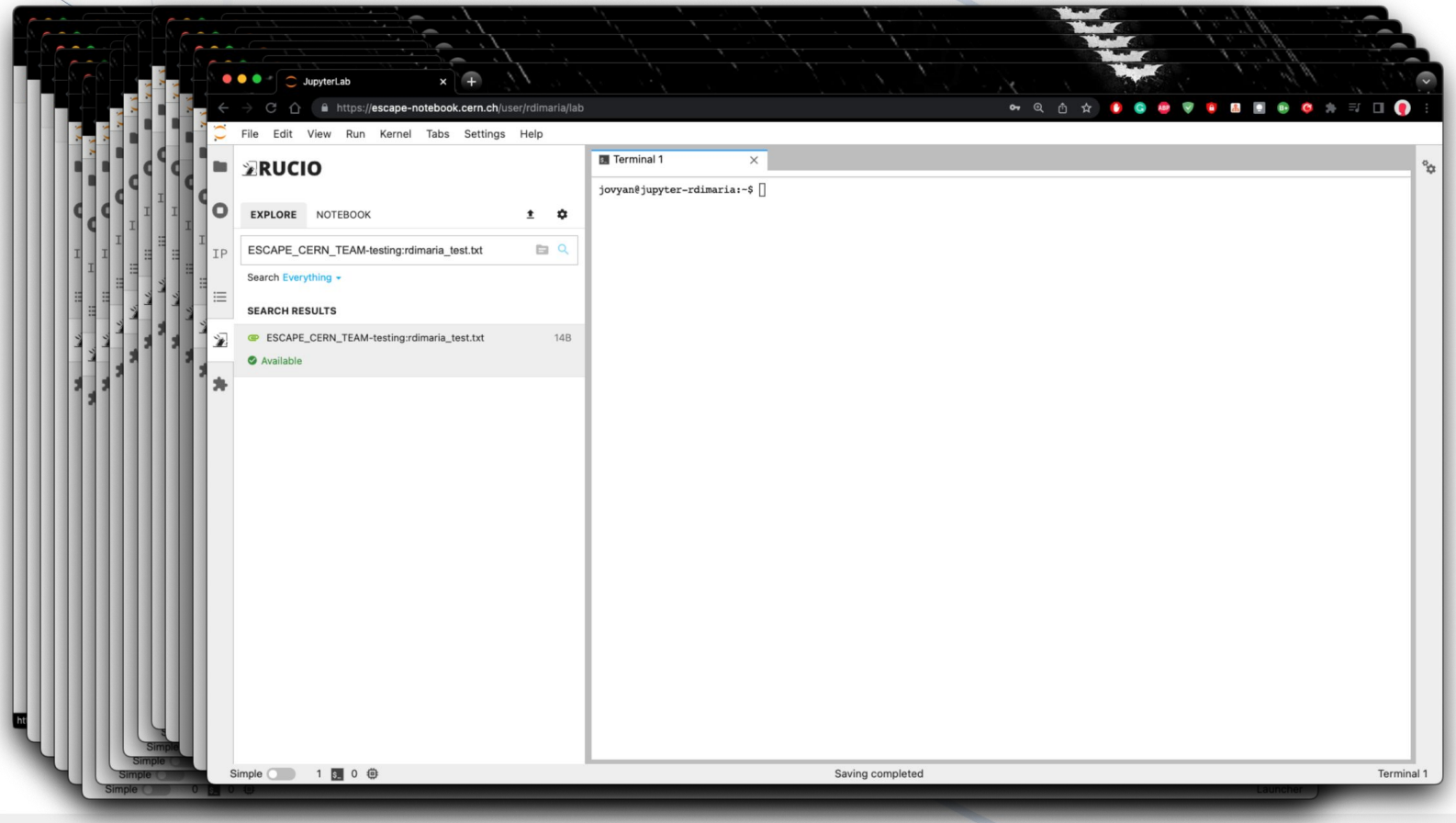
Funded by the European Union's Horizon 2020 - Grant N° 824064

Data Preservation Use Case



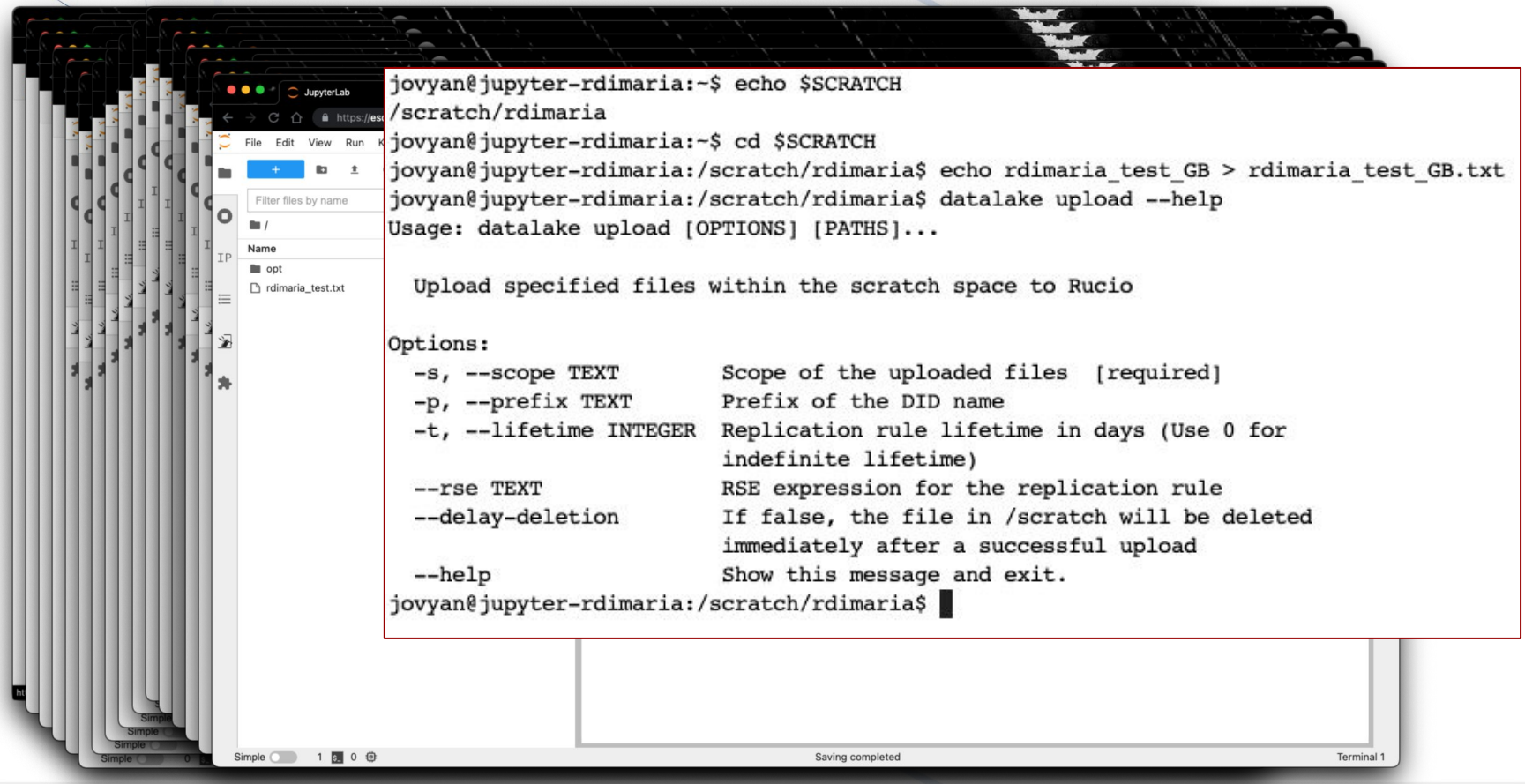
Funded by the European Union's Horizon 2020 - Grant N° 824064

Data Preservation Use Case



Funded by the European Union's Horizon 2020 - Grant N° 824064

Output of Workflow or Analysis (Large-Sized Files)



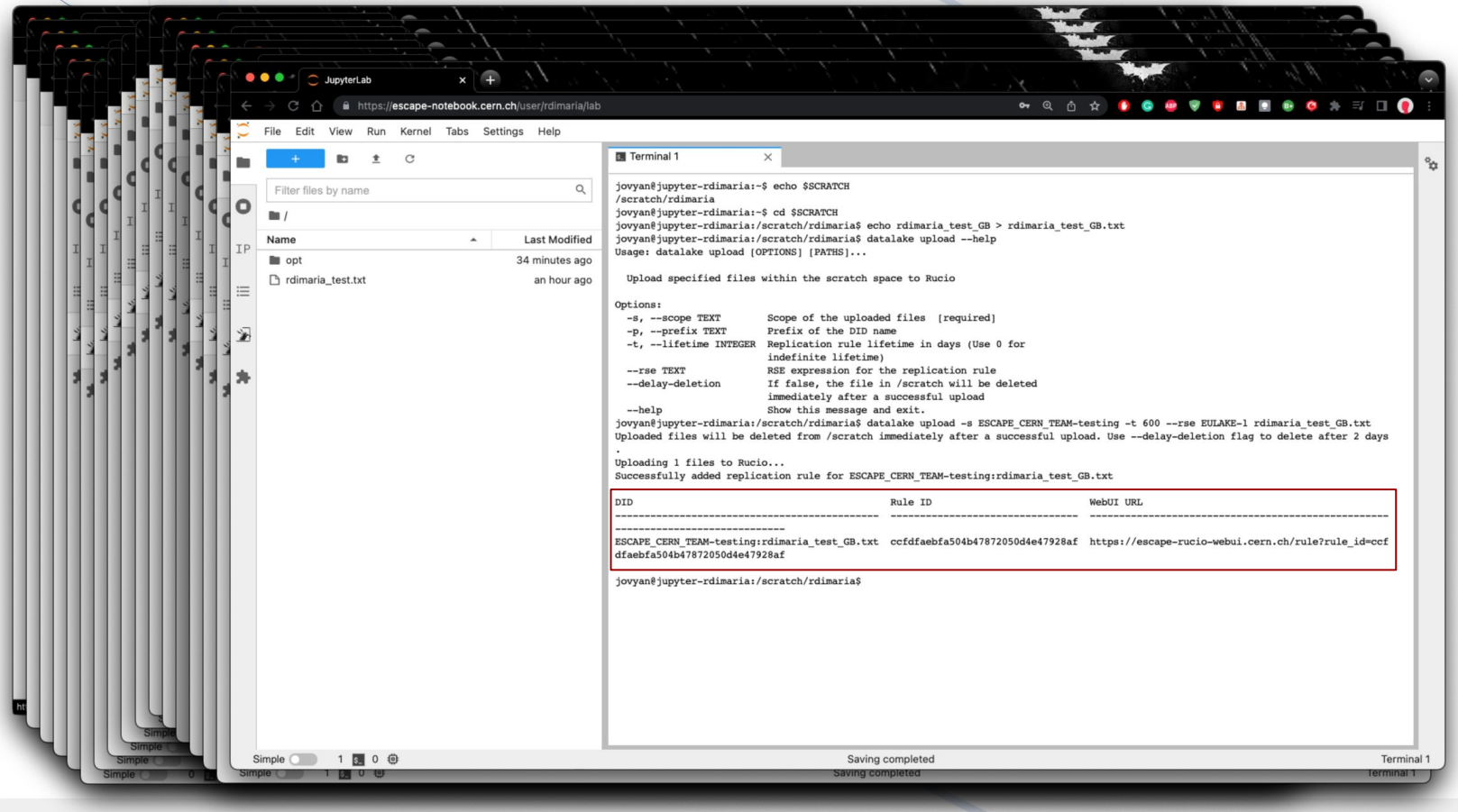

```
jovyan@jupyter-rdimaria:~$ echo $SCRATCH
/scratch/rdimaria
jovyan@jupyter-rdimaria:~$ cd $SCRATCH
jovyan@jupyter-rdimaria:/scratch/rdimaria$ echo rdimaria_test_GB > rdimaria_test_GB.txt
jovyan@jupyter-rdimaria:/scratch/rdimaria$ datalake upload --help
Usage: datalake upload [OPTIONS] [PATHS]...

  Upload specified files within the scratch space to Rucio

Options:
  -s, --scope TEXT        Scope of the uploaded files  [required]
  -p, --prefix TEXT       Prefix of the DID name
  -t, --lifetime INTEGER  Replication rule lifetime in days (Use 0 for
                          indefinite lifetime)
  --rse TEXT              RSE expression for the replication rule
  --delay-deletion        If false, the file in /scratch will be deleted
                          immediately after a successful upload
  --help                  Show this message and exit.
jovyan@jupyter-rdimaria:/scratch/rdimaria$
```

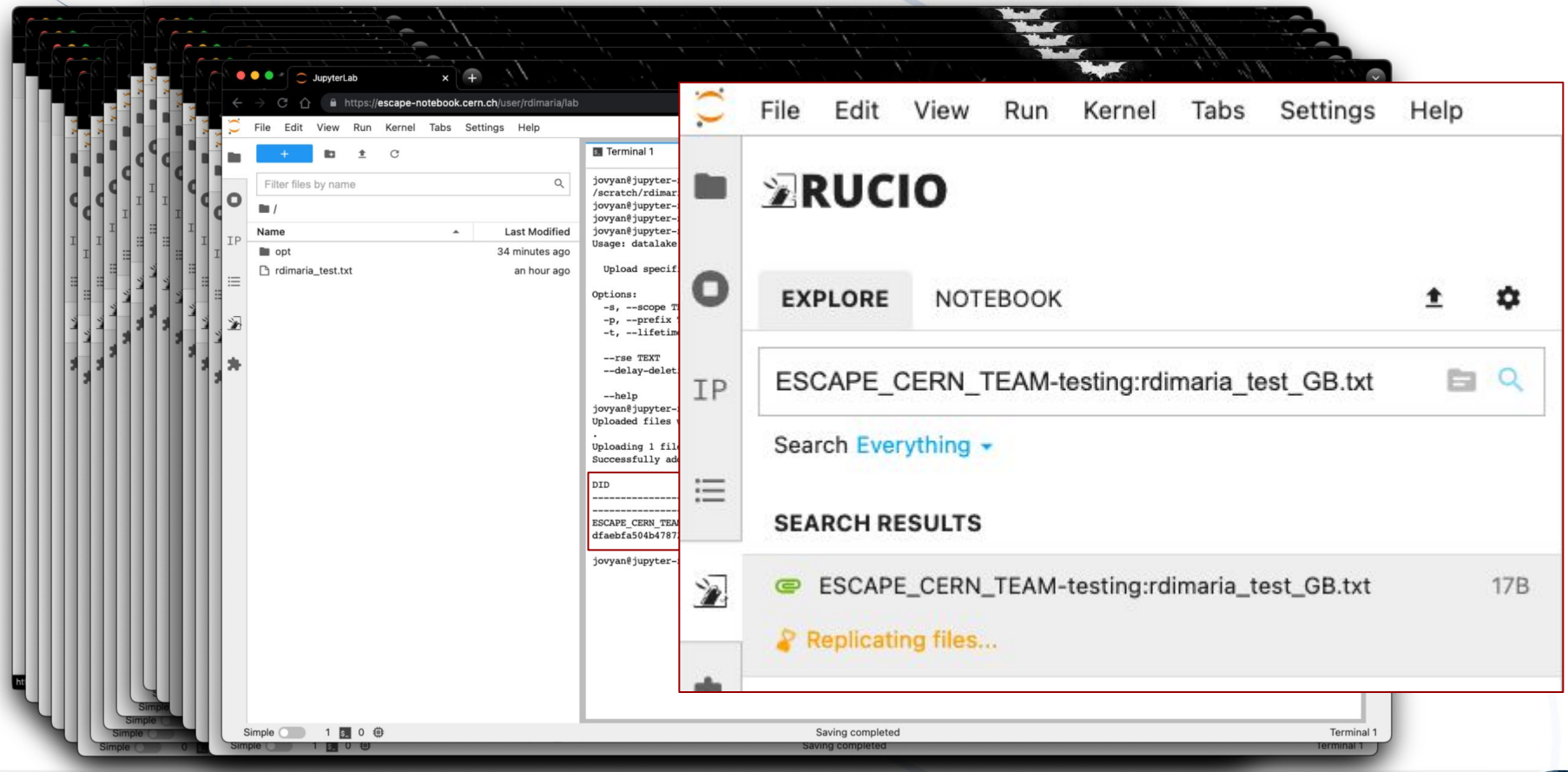

Funded by the European Union's Horizon 2020 - Grant N° 824064

Data Preparation and Processing Use Case



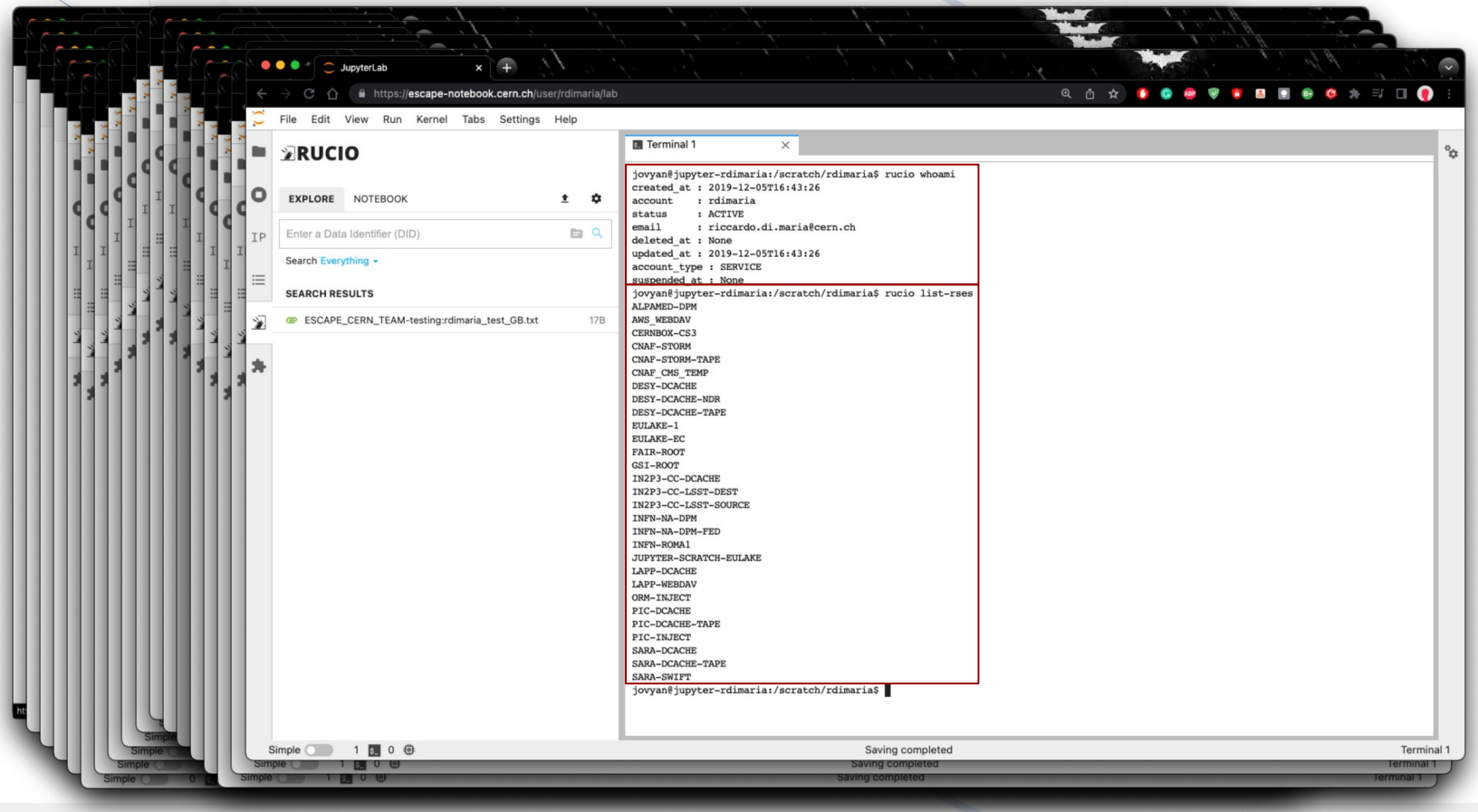
Funded by the European Union's Horizon 2020 - Grant N° 824064

Data Preparation and Processing Use Case



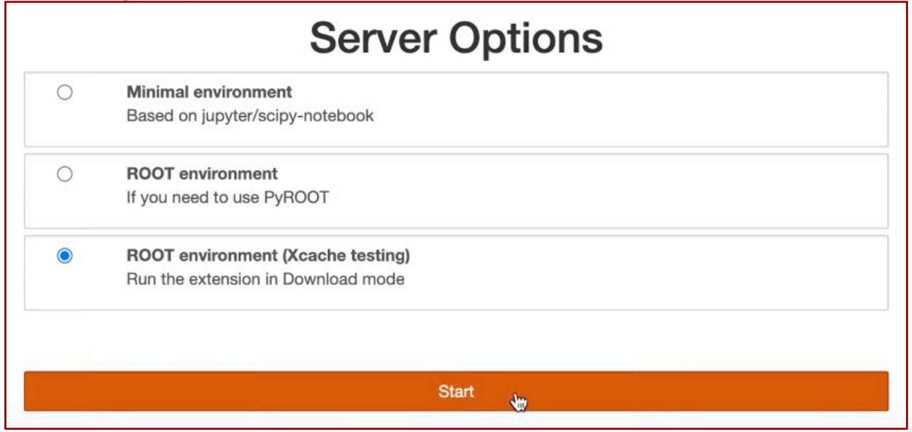
Funded by the European Union's Horizon 2020 - Grant N° 824064

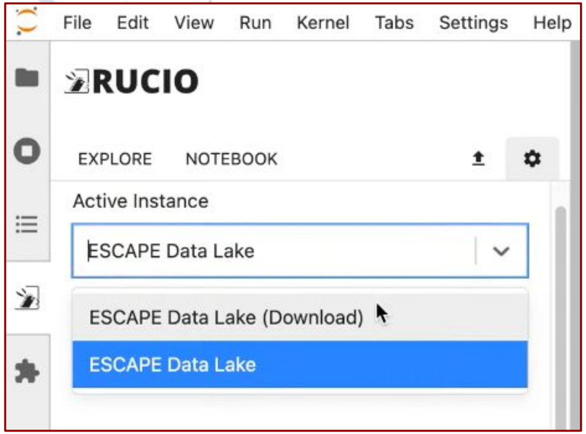
Rucio CLI Available

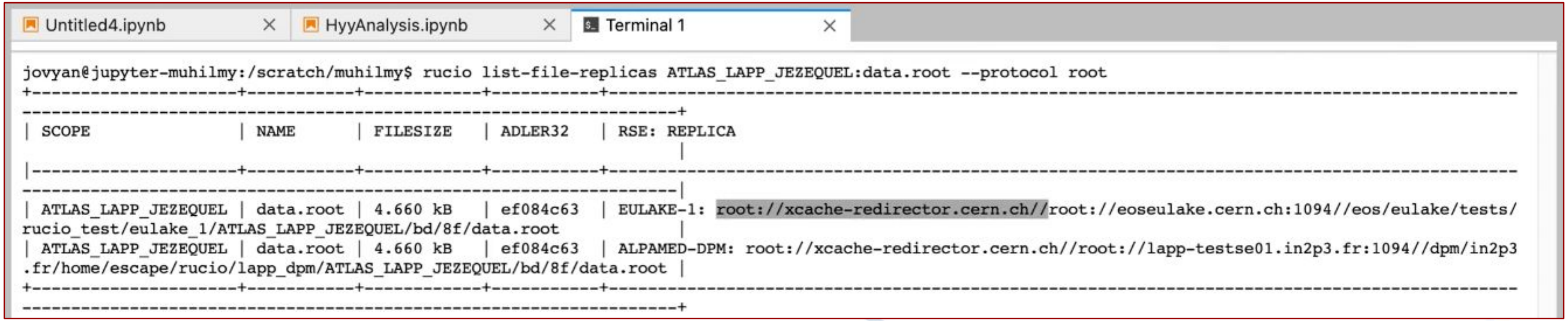


Funded by the European Union's Horizon 2020 - Grant N° 824064

XCache Integration







Funded by the European Union's Horizon 2020 - Grant N° 824064

DLaaS Use Cases

- Data discovery and access
- Submitting jobs to external services (remote computing)
  - conveniently browse data in Rucio through the extension
  - access file PFN directly from the Notebook
- Data preparation and processing
  - prepare/process data and upload back to the Data Lake
- Data preservation
  - produce data and upload to the Data Lake

Funded by the European Union's Horizon 2020 - Grant N° 824064

Future Developments and Prospectives

- Additional kernel compatibility (currently, only Python supported)
- Token-support for direct download and upload → OIDC integration ongoing at storage level
- Integration with content delivery and caching layer (successfully tested only at small scale)
  - XCache can be integrated to allow faster file download → transparent for the end-user
- Multi-VO or off-site (CERN) deployment, and distribution model for sciences
- **DLaaS** interesting for both **aficionados** and **newcomers** of **Rucio**
  - community-driven “development and operation”
    - needs and requirements of different experiments and sciences
  - addressing long term sustainability beyond ESCAPE mandate
    - ongoing proposal to establish a **Special Interest Group**

Funded by the European Union’s Horizon 2020 - Grant N° 824064

Conclusion and Next Steps

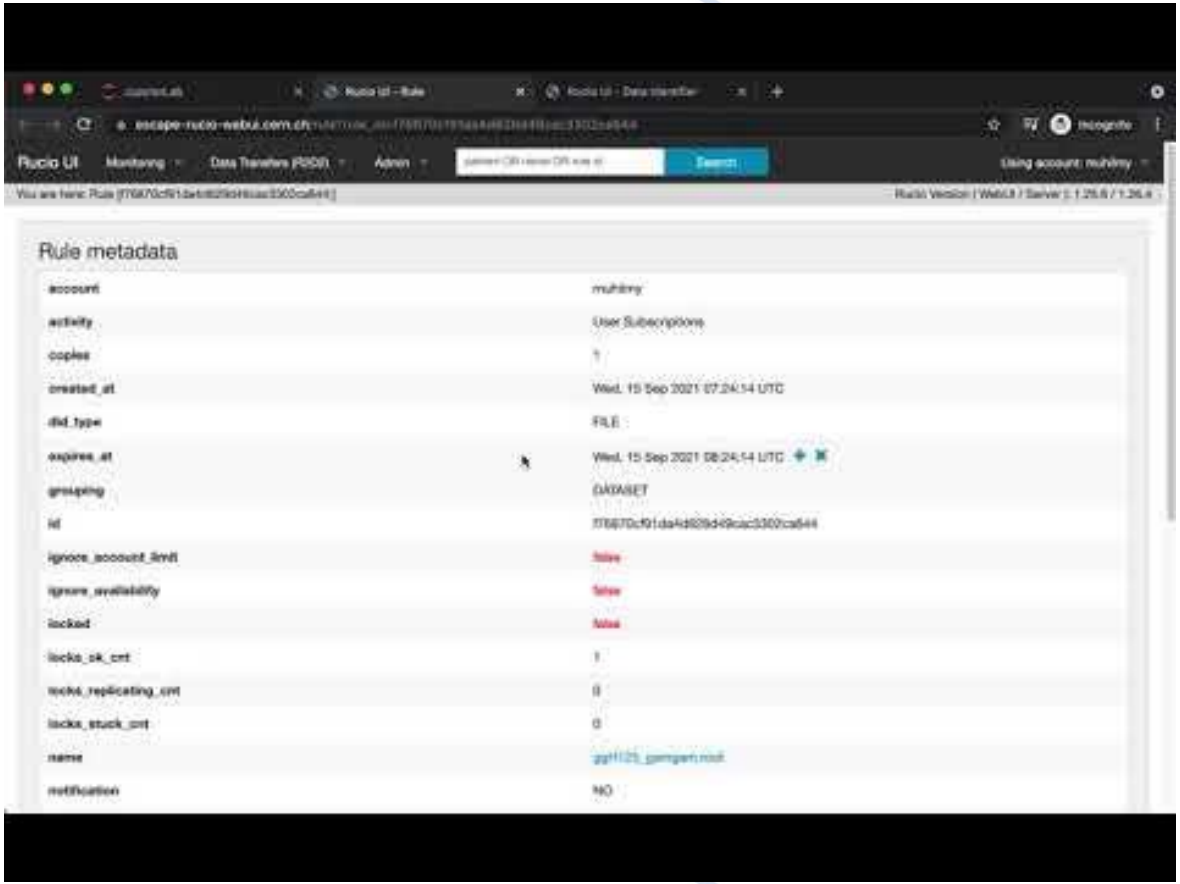
- **DLaaS** hides the complexities of the Data Lake from the end users (so that they are happy and productive!)
  - interesting for both **aficionados** and **newcomers** of **Rucio**
  - community-driven development, hence driven by the needs of different experiments and sciences
- ESCAPE managed to pilot and prototype a Data Lake infrastructure fulfilling functional data management needs of flagship ESFRIs from several scientific disciplines
  - sensible technologies choice from WLCG environment and LHC experiments
  - successful assessments of the Data Lake in 2020 and 2021
    - pivotal to test model and concepts for several communities:
      Astro-particle Physics, Electromagnetic and Gravitational-Wave Astronomy, Particle Physics, and Nuclear Physics **pursuing together** FAIR and open-access data principles
    - exploring non-HEP-standard scenarios and collaboration with other communities, e.g. PaNOSC, ExPaNDS, CS3MESH4EOSC
- ESCAPE end in 2022 → addressing long term sustainability

Funded by the European Union's Horizon 2020 - Grant N° 824064

Backup Slides

Funded by the European Union's Horizon 2020 - Grant N° 824064

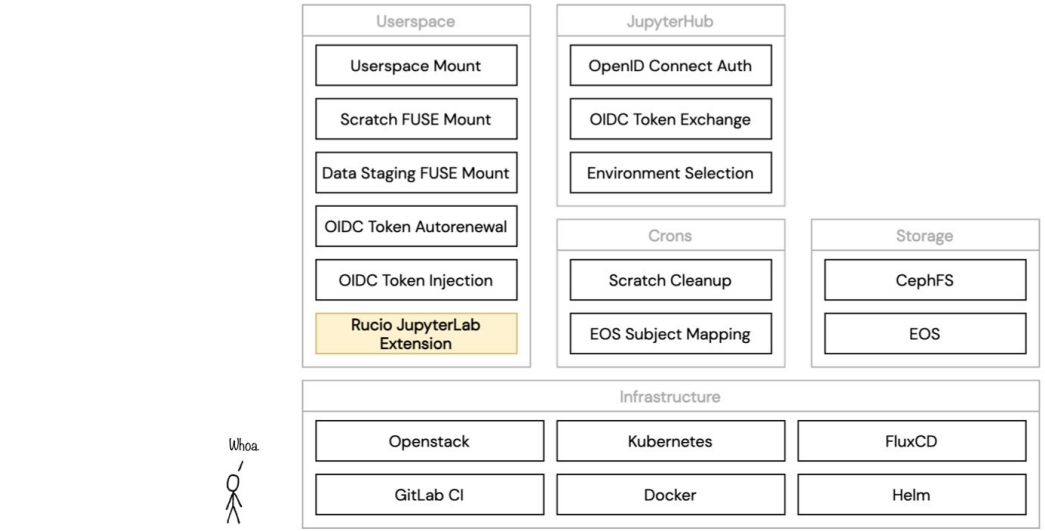
DataLake-as-a-Service for Open Science



Funded by the European Union's Horizon 2020 - Grant N° 824064

Backup Slides

The DataLake as a Service



CERN openlab

Muhammad Aditya Hilmy

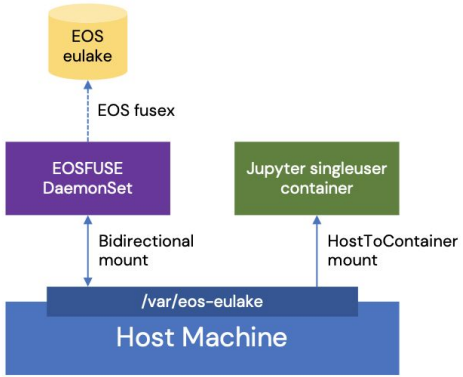
6

Funded by the European Union's Horizon 2020 - Grant N° 824064

Backup Slides

FUSE mount to EOS eulake

- There are two FUSE mounts to the same EOS instance:
  - /eos/eulake_1 → /eos/eulake/tests/rucio_test/eulake_1
  - /scratch → /eos/eulake/tests/jupyter-scratch
- FUSE mount is implemented using k8s DaemonSet, mounting to a folder in the host, with Bidirectional mount propagation
- Singleuser containers bind to the mount folder, with HostToContainer mount propagation
- Uses OAuth2 authentication
  - ESCAPE IAM user is mapped to EOS user using crons



Funded by the European Union's Horizon 2020 - Grant N° 824064

Backup Slides

OAuth2 in EOS FUSE mount

- In the singleuser container:
  - JWT is stored in a file in the following format:
    - `oauth2:<jwt>:<token-introspection-endpoint>`
    - Example: `oauth2:eyJ...:iam-escape.cloud.cnaf.infn.it/userinfo`
    - Note: token introspection endpoint doesn't have the "https://" part
    - The token file must have at most 0600 permission
  - An environment variable needs to be set:
    - `OAUTH2_TOKEN=FILE:/path/to/token/file`
- In the EOSFUSE DaemonSet container:
  - EOS FUSEx daemon (eosxd) needs to be configured for SSS authentication
  - SSS keytab needs to be present

Docs: https://eos-docs.web.cern.ch/using/oauth2.html

Funded by the European Union's Horizon 2020 - Grant N° 824064

Backup Slides

Singleuser container setup

Some things need to happen:

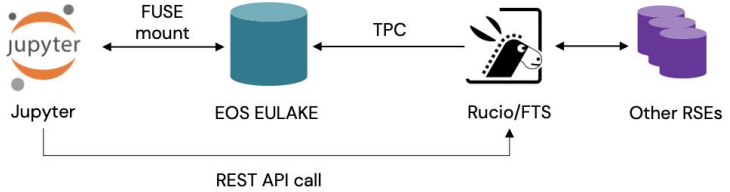
- OAuth token exchange (eos-eulake and rucio)
  - Uses a modified version of SWAN's KeyCloakAuthenticator
- Enable token autorenewal
  - Uses swanoauthrenew
- Write token files to /tmp
- Set OAUTH2_TOKEN env for EOS authentication
- Write rucio.cfg file

Funded by the European Union's Horizon 2020 - Grant N° 824064

Backup Slides

Making files available

- Replica mode: uses Third Party Copy (TPC)
- EULAKE-1 is a Rucio Storage Element and is FUSE-mounted to /eos/eulake_1
- When "Make Available" is clicked:
  - The extension creates a replication rule to move requested files into EULAKE-1
  - Lifetime is set to 7 days (configurable by service admins)
  - Rucio will move the files to EULAKE-1
  - Once the replication status is OK, the extension translates the Physical File Name into local path
    - `root://eoseulake.cern.ch:1094//eos/eulake/tests/rucio_test/eulake_1/file → /eos/eulake_1/file`
  - File is accessible as if it were local



Funded by the European Union's Horizon 2020 - Grant N° 824064

Backup Slides

Uploading files in scratch (..technically EOS eulake)

When “datalake upload” is run:

- The script translates local path to full Physical File Name:
  - `/scratch/muhilmy/file → root://eoseulake.cern.ch:1094//eos/eulake/tests/jupyter-scratch/muhilmy/file`
- The file in scratch is added to the Rucio replica catalogue
- A replication rule is created to move the files from scratch space to a destination storage
- Rucio will move the files to the destination storage
- When the replication status is OK, Rucio will delete the file in scratch
- A cron job will run every 24h to delete files (and folders) older than 2 days old that might not be in the Rucio catalogue

Funded by the European Union’s Horizon 2020 - Grant N° 824064