The Virtual Research Environment

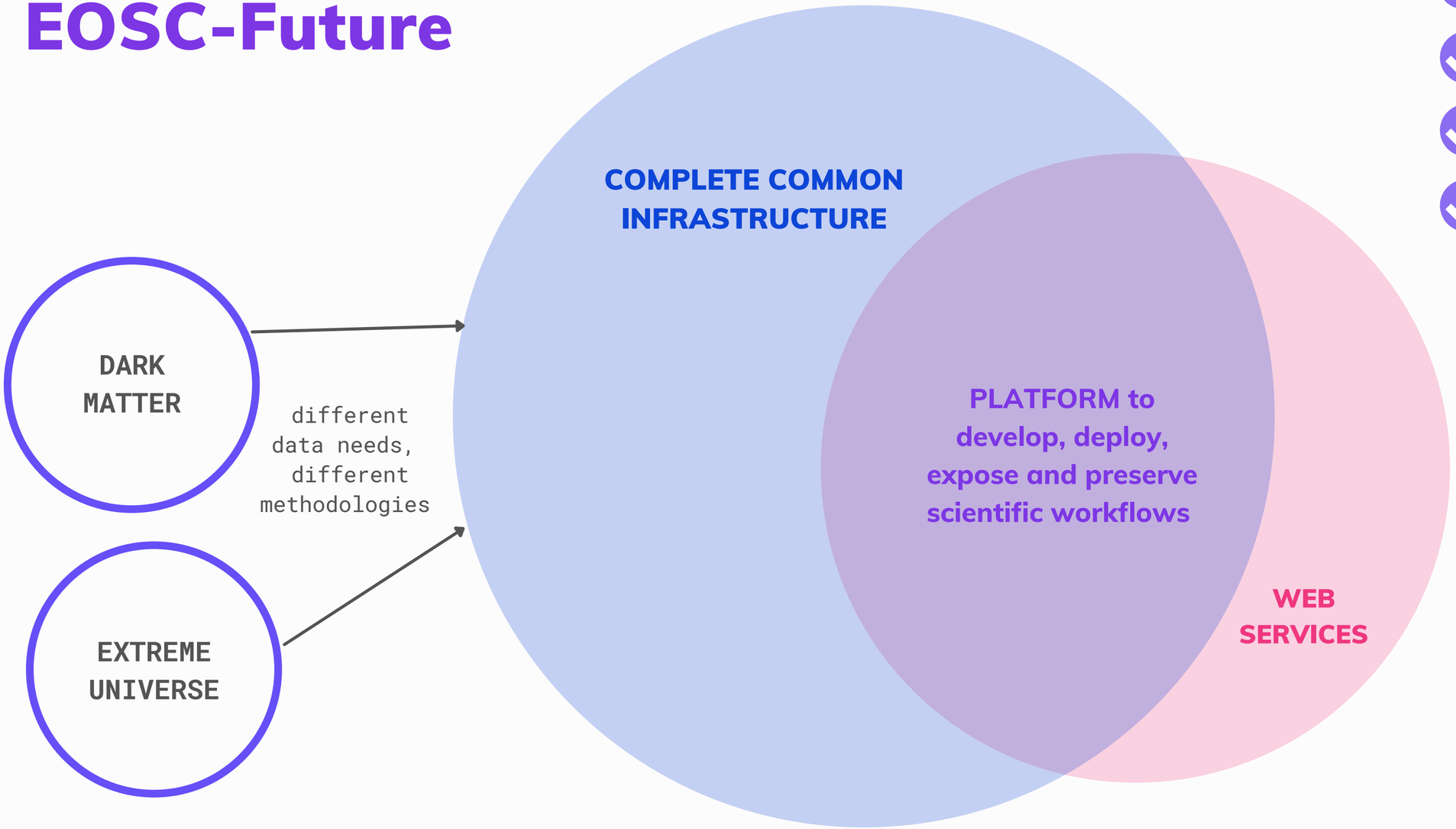
Developing an online collaborative platform to serve astro-particle physics communities

Elena Gazzarrini
Alba Vendrell Moya
Enrique Garcia

EOSC-Future progress meeting, 21-22 July 2022

EOSC-Future

- OPEN SOURCE SCIENCE
- COLLABORATION
- FAIR DATA PRINCIPLES
- ANALYSIS PRESERVATION



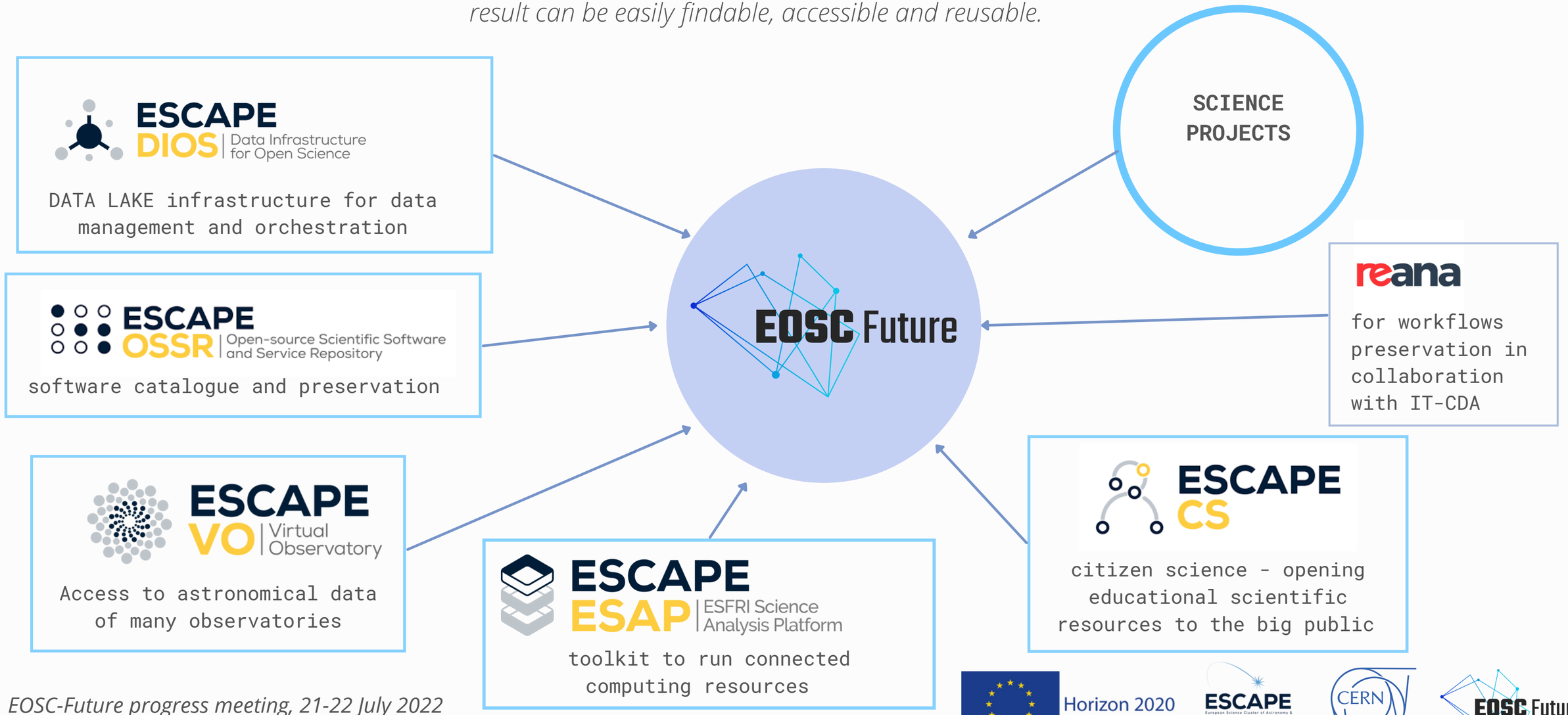
OBJECTIVE

The project will serve as a science driver for other communities

Horizon 2020 ESCAPE CERN EOSC Future

The VRE

*A collaborative **online platform** where interested researchers members will be able develop their **Science Projects**. It will host and implement the services provided by the **ESCAPE WPs 2, 3, 4, 5 and 6** to provide an environment where the digital content of a scientific result can be easily findable, accessible and reusable.*



Horizon 2020
ESCAPE
CERN
EOSC Future

VRE concept

Generation and simulation of events

Experimental DATA

Data processing

Including reconstruction and calibration path

Analysis

Including background subtraction, estimation and statistical analysis

Interpretation of results

COMBINATION of results from different experiments

COMPARISON of results with other searches

from ESCAPE experiments

KM3NeT, ATLAS EXPERIMENT, DARKSIDE, ASTRON, EGO, cta

FTS, kubernetes, SCIENTIFIC DATA MANAGEMENT RUCIO

ESCAPE's Data Lake running on a K8s cluster:
storage orchestration
+
data management

jupyter

WebUI:
DataLake-as-a-Service
(credits to Riccardo Di Maria, Muhammad Aditya Hilmy)

+

reana

a platform to reproduce workflows running on various computing backends

Horizon 2020, ESCAPE, CERN, EOSC Future

VRE principles

WebUI interface

To facilitate onboarding. Good documentation is key.

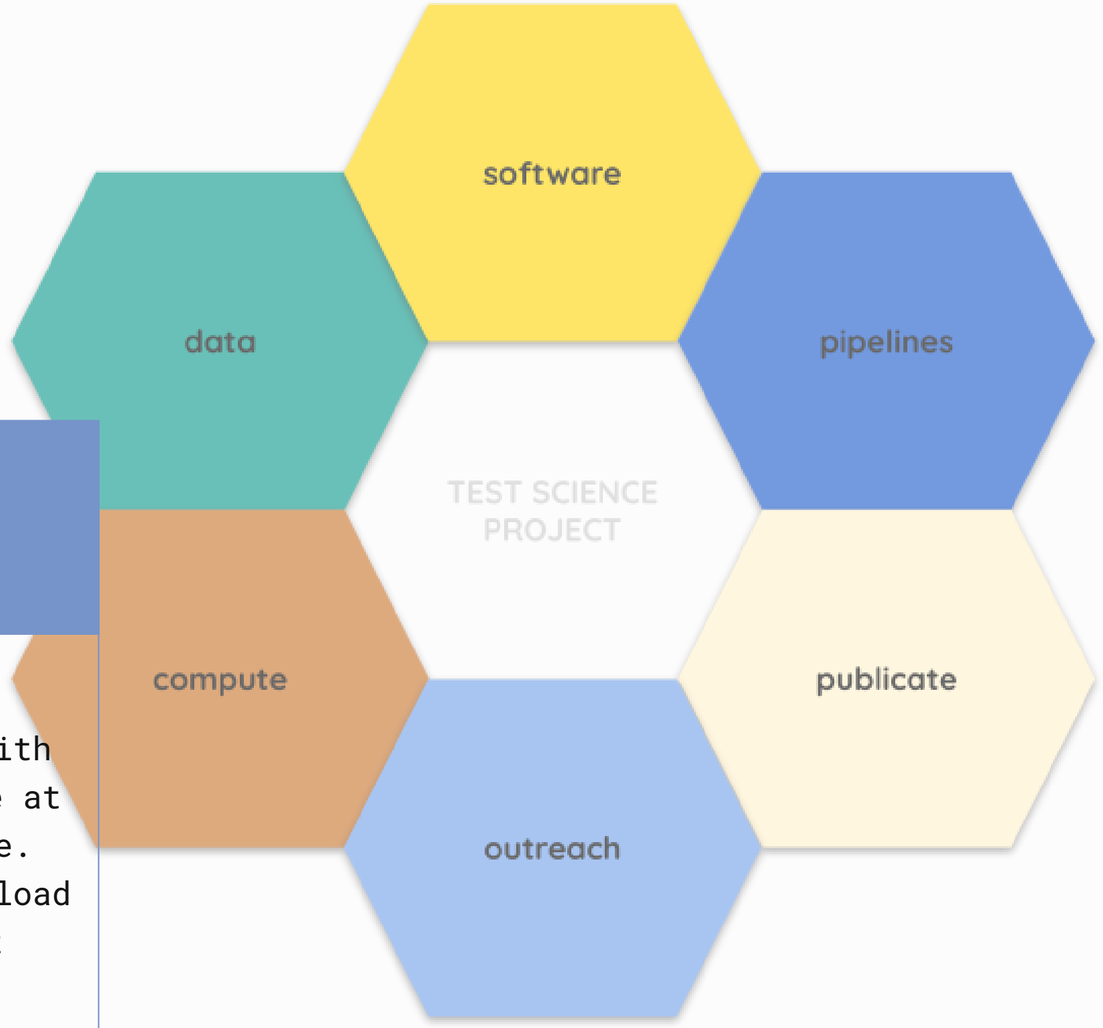
VRE principles

WebUI interface	Scalability
To facilitate onboarding. Good documentation is key.	Possibility of running jobs at various scales. Collaboration to make it easy to re-deploy the service on external clusters.

Horizon 2020
ESCAPE
CERN
EOSC Future

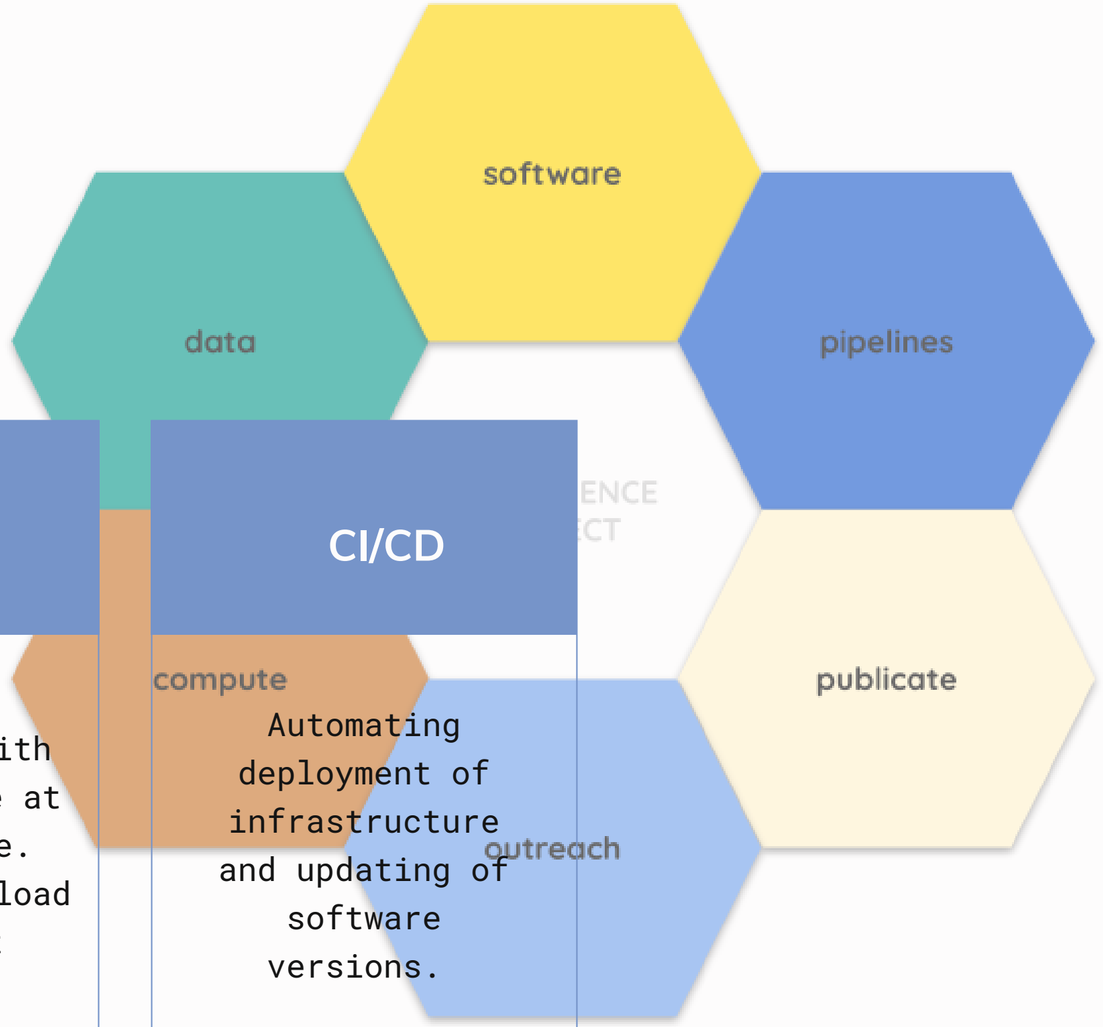
VRE principles

WebUI interface	Scalability	Portability
To facilitate onboarding. Good documentation is key.	Possibility of running jobs at various scales. Collaboration to make it easy to re-deploy the service on external clusters.	Many users interacting with infrastructure at the same time. Members can upload the content directly.



Horizon 2020
ESCAPE
CERN
EOSC Future

VRE principles



WebUI interface	Scalability	Portability	CI/CD
To facilitate onboarding. Good documentation is key.	Possibility of running jobs at various scales. Collaboration to make it easy to re-deploy the service on external clusters.	Many users interacting with infrastructure at the same time. Members can upload the content directly.	Automating deployment of infrastructure and updating of software versions.

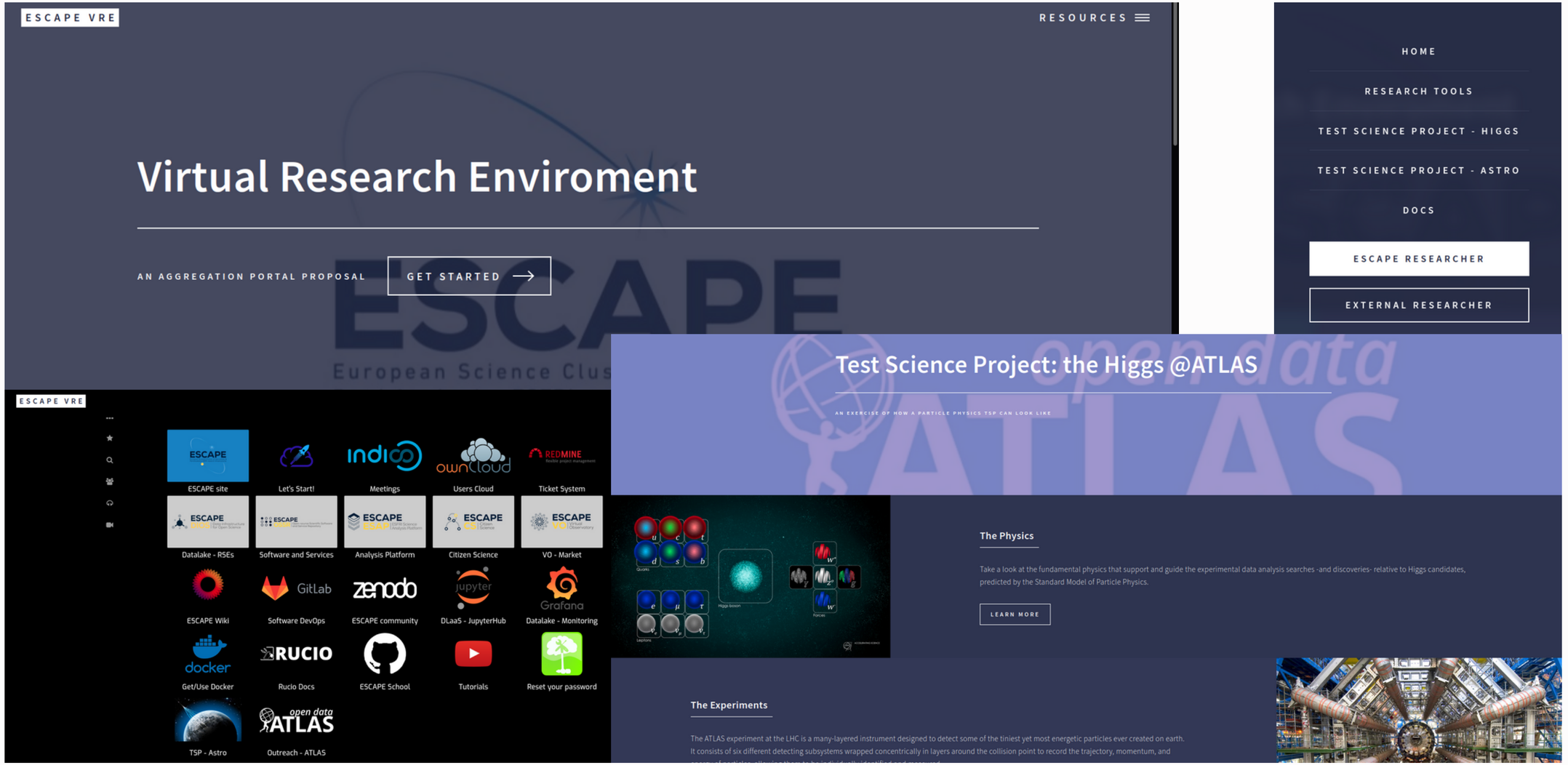
Horizon 2020
ESCAPE
CERN
EOSC Future

VRE principles

WebUI interface	Scalability	Portability	CI/CD	Flexibility
To facilitate onboarding. Good documentation is key.	Possibility of running jobs at various scales. Collaboration to make it easy to re-deploy the service on external clusters.	Many users interacting with infrastructure at the same time. Members can upload the content directly.	Automating deployment of infrastructure and updating of software versions.	Supports different formats of data, containerization techniques, job submission protocols

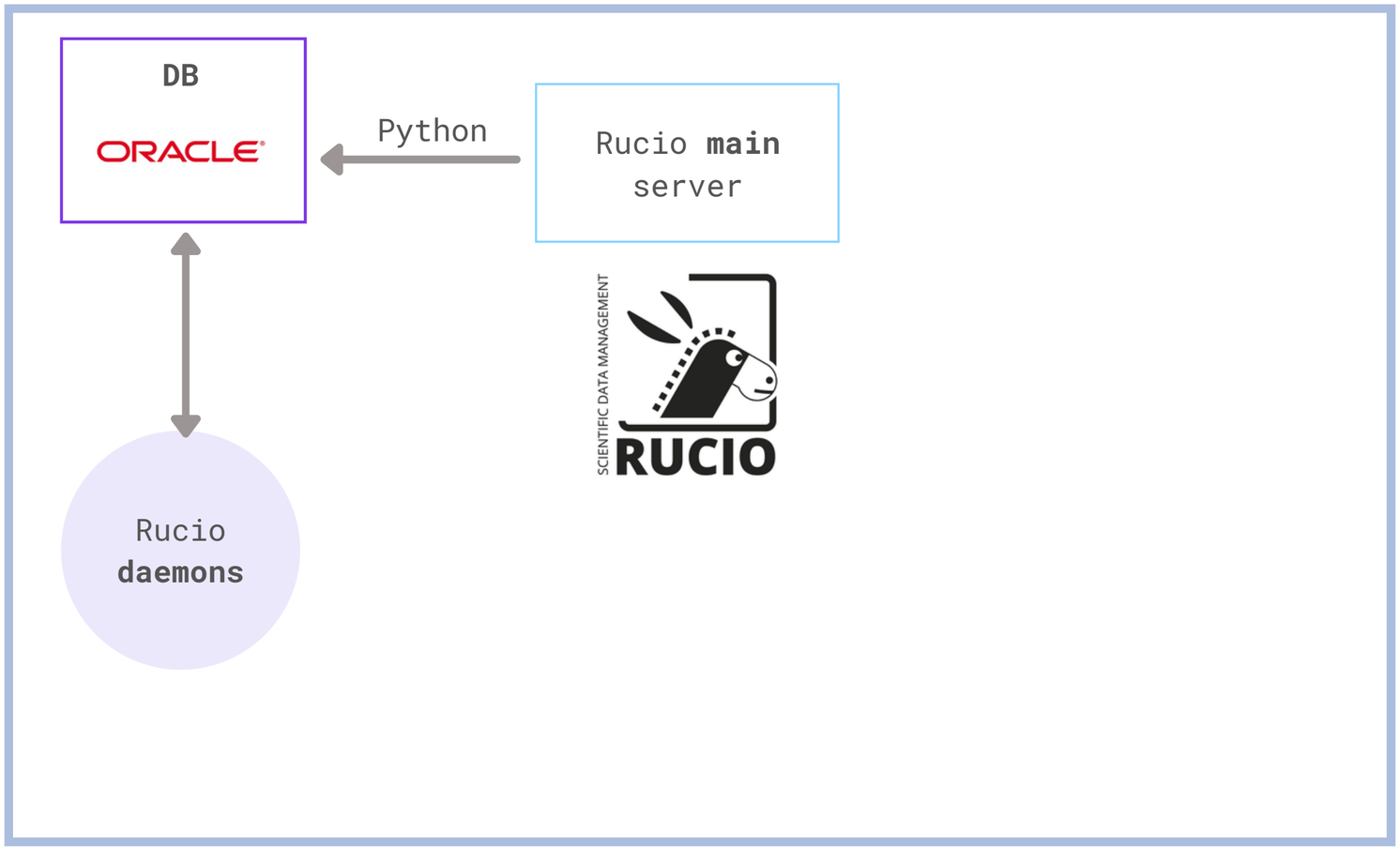
Horizon 2020
ESCAPE
CERN
EOSC Future

VRE landing page



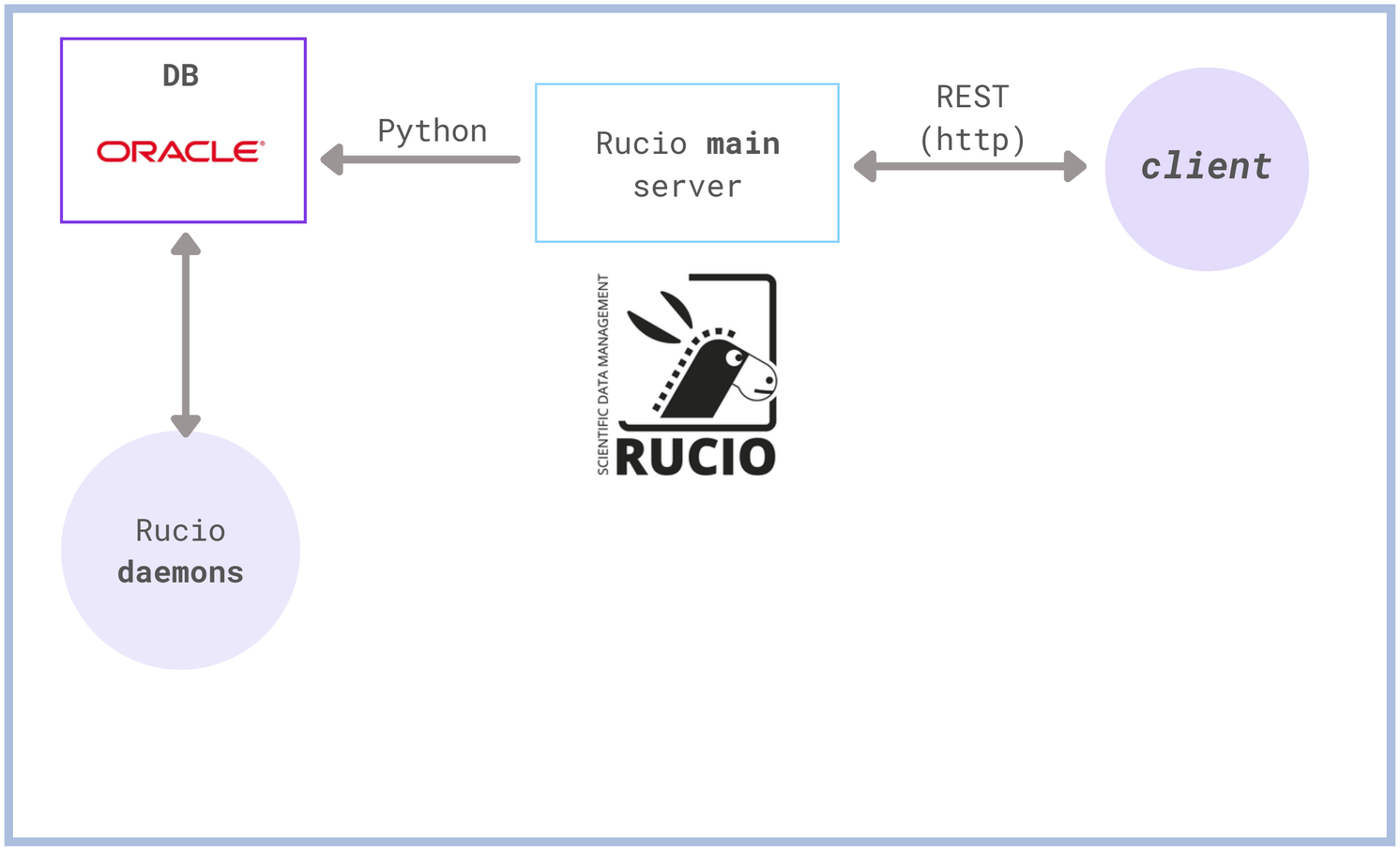
credits to Arturo Sanchez Pineda

Infrastructure overview



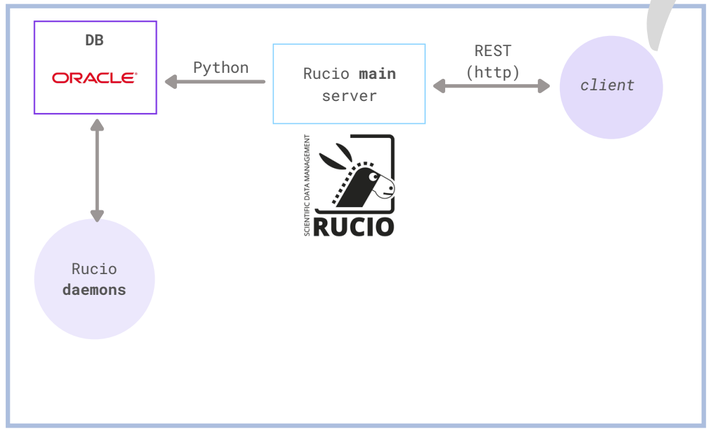
Horizon 2020
ESCAPE
CERN
EOSC Future

Infrastructure overview



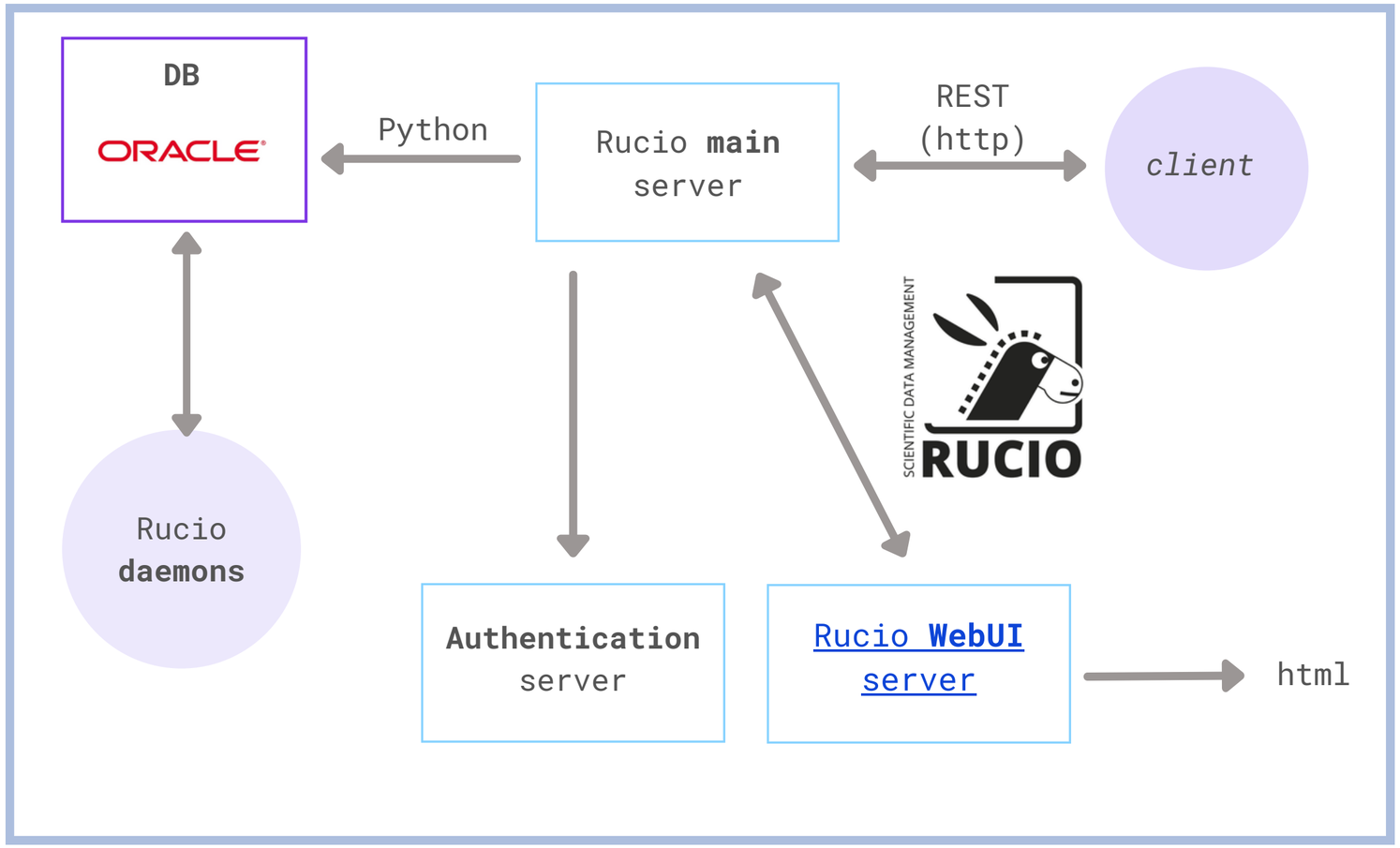
Horizon 2020
ESCAPE
CERN
EOSC Future

Infrastructure overview



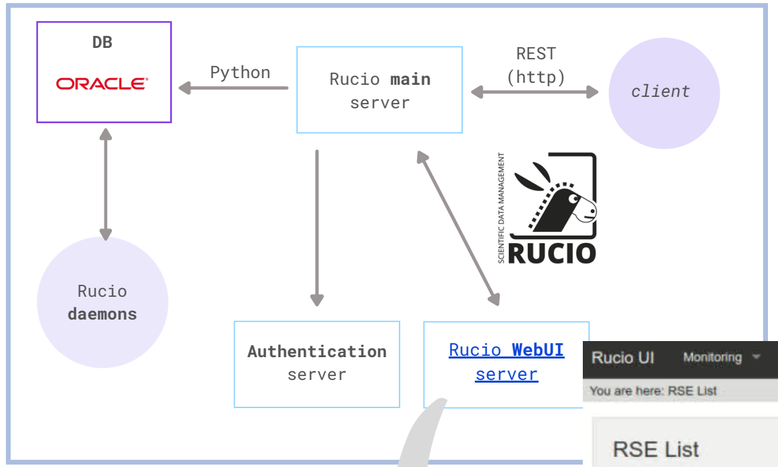

```
(rucio-cli) root@egazzarr-rucio:~#
```

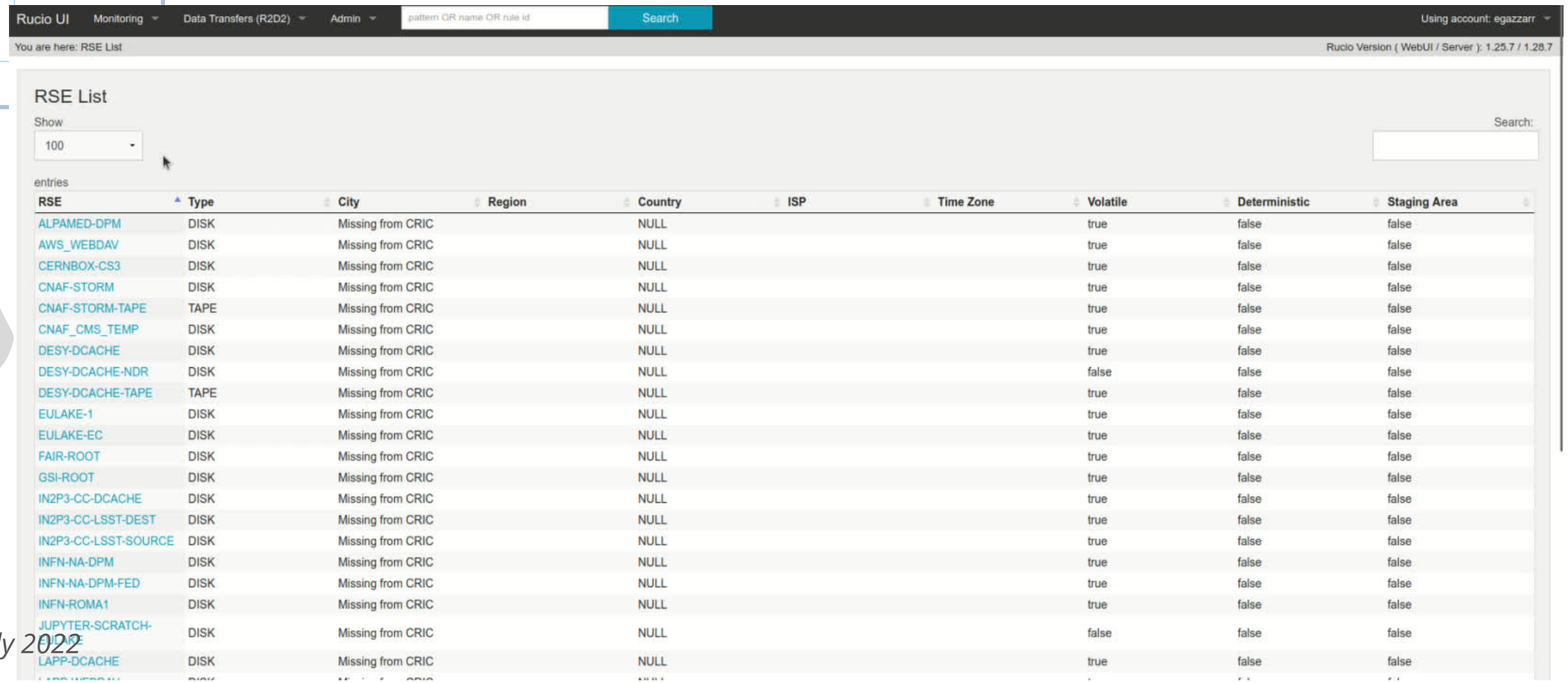
Infrastructure overview



Horizon 2020
ESCAPE
CERN
EOSC Future

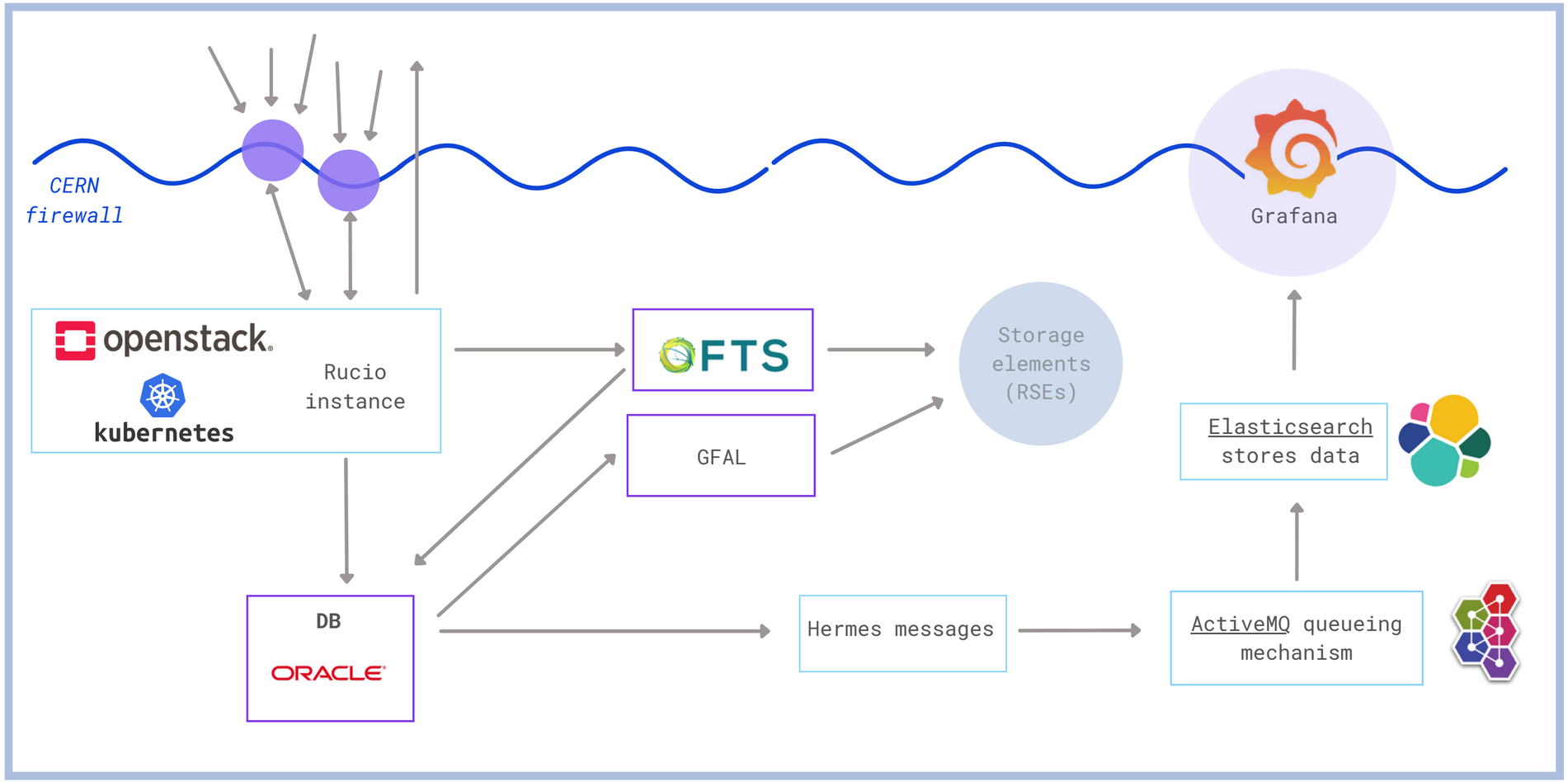
Infrastructure overview





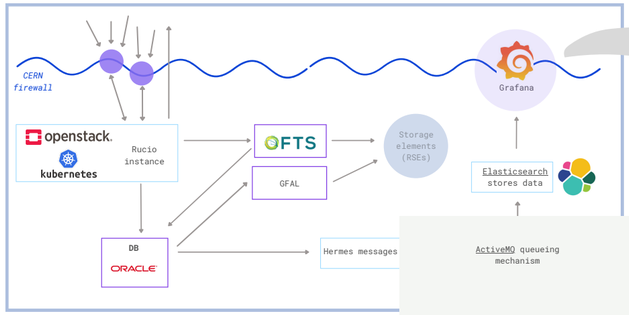
RSE	Type	City	Region	Country	ISP	Time Zone	Volatile	Deterministic	Staging Area
ALPAMED-DPM	DISK	Missing from CRIC		NULL			true	false	false
AWS_WEBDAV	DISK	Missing from CRIC		NULL			true	false	false
CERNBOX-CS3	DISK	Missing from CRIC		NULL			true	false	false
CNAF-STORM	DISK	Missing from CRIC		NULL			true	false	false
CNAF-STORM-TAPE	TAPE	Missing from CRIC		NULL			true	false	false
CNAF_CMS_TEMP	DISK	Missing from CRIC		NULL			true	false	false
DESY-DCACHE	DISK	Missing from CRIC		NULL			true	false	false
DESY-DCACHE-NDR	DISK	Missing from CRIC		NULL			false	false	false
DESY-DCACHE-TAPE	TAPE	Missing from CRIC		NULL			true	false	false
EULAKE-1	DISK	Missing from CRIC		NULL			true	false	false
EULAKE-EC	DISK	Missing from CRIC		NULL			true	false	false
FAIR-ROOT	DISK	Missing from CRIC		NULL			true	false	false
GSI-ROOT	DISK	Missing from CRIC		NULL			true	false	false
IN2P3-CC-DCACHE	DISK	Missing from CRIC		NULL			true	false	false
IN2P3-CC-LSST-DEST	DISK	Missing from CRIC		NULL			true	false	false
IN2P3-CC-LSST-SOURCE	DISK	Missing from CRIC		NULL			true	false	false
INFN-NA-DPM	DISK	Missing from CRIC		NULL			true	false	false
INFN-NA-DPM-FED	DISK	Missing from CRIC		NULL			true	false	false
INFN-ROMA1	DISK	Missing from CRIC		NULL			true	false	false
JUPYTER-SCRATCH-	DISK	Missing from CRIC		NULL			false	false	false
LAPP-DCACHE	DISK	Missing from CRIC		NULL			true	false	false

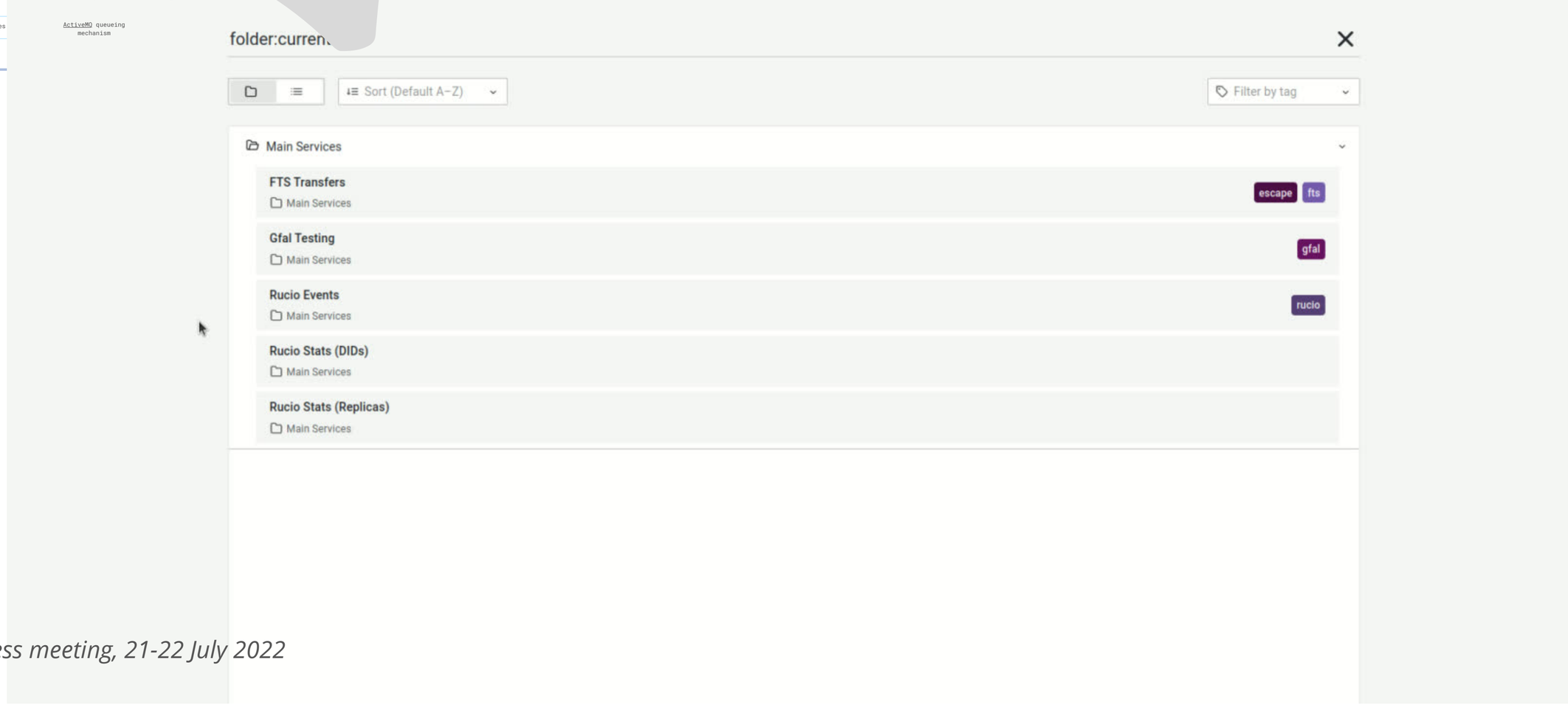
Networking + monitoring overview



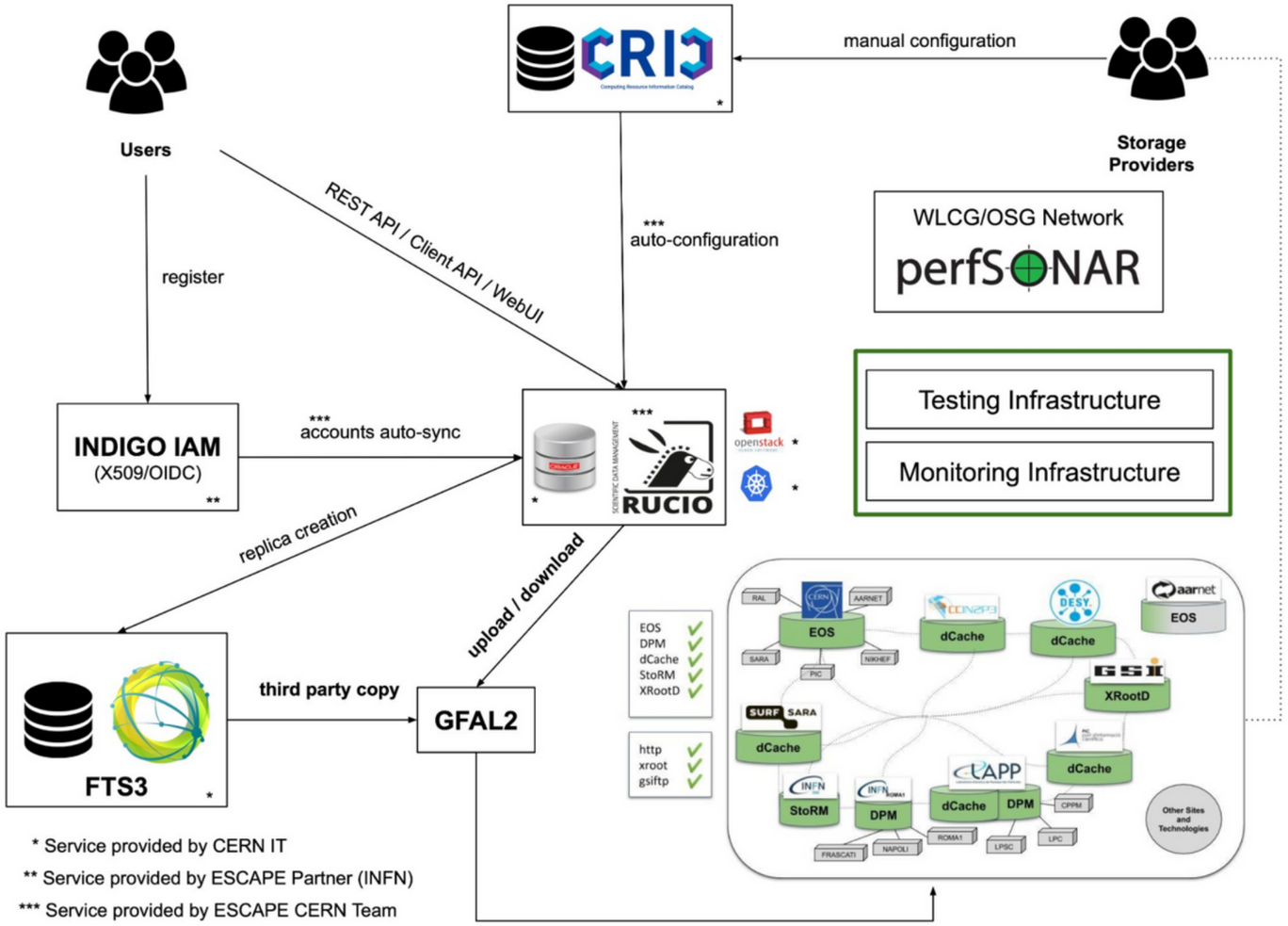
Horizon 2020
ESCAPE
CERN
EOSC Future

Networking + monitoring overview





The Data Lake architecture

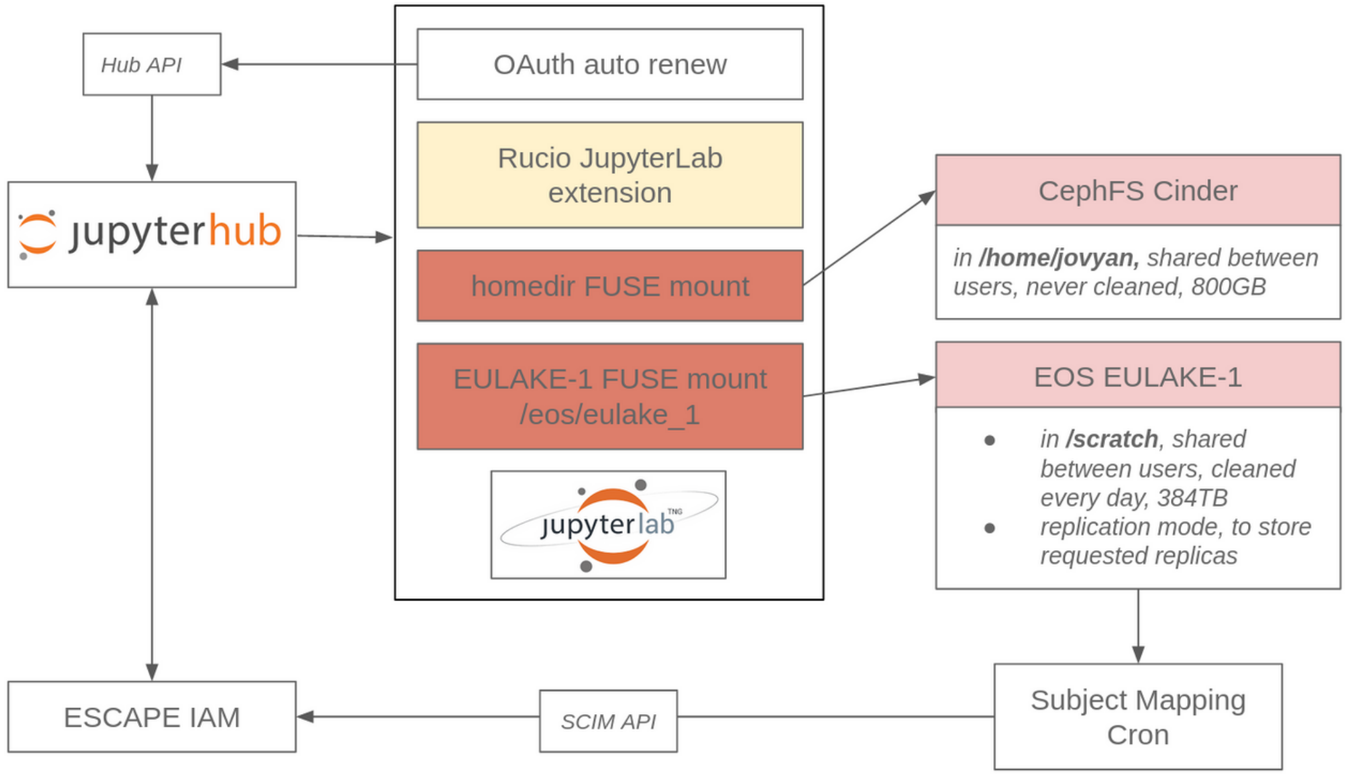


A federated data infrastructure

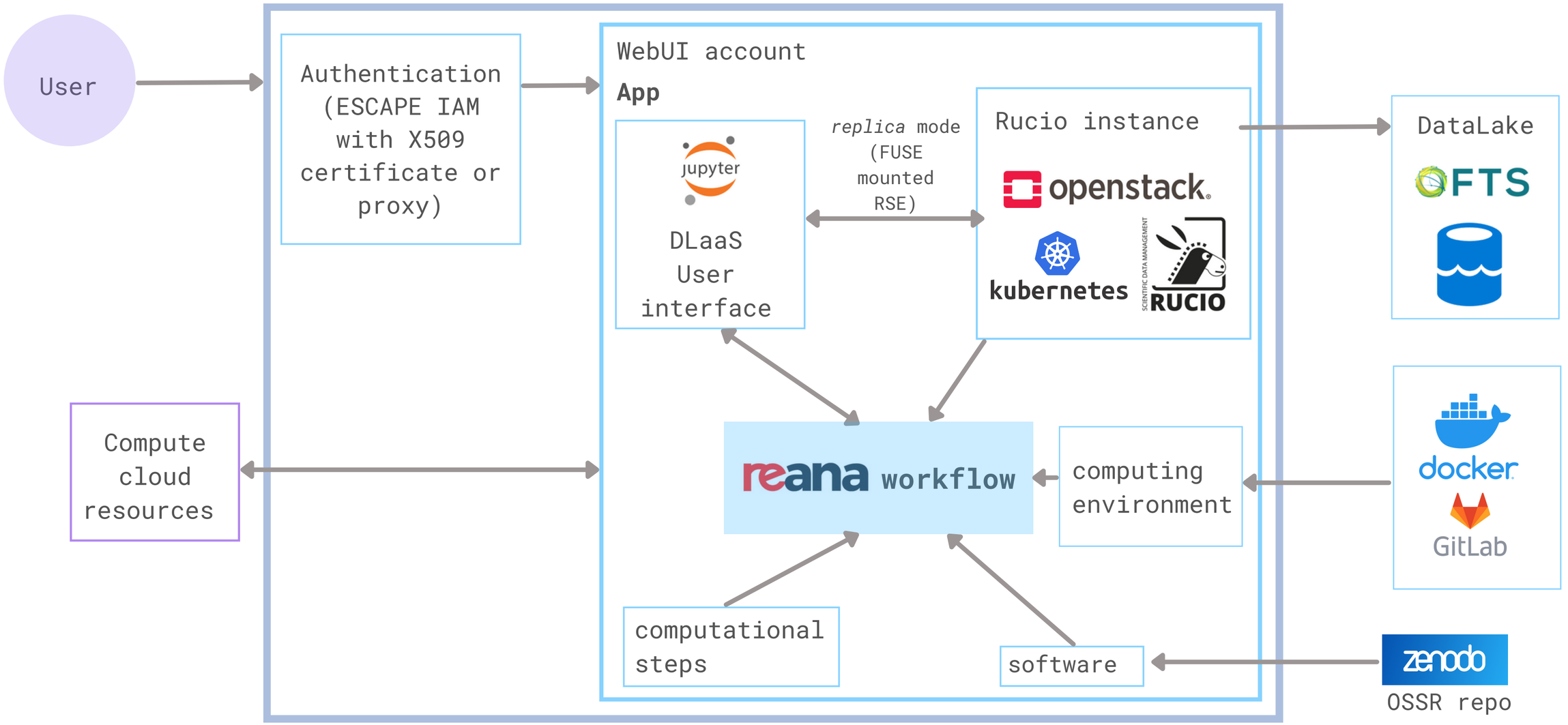
- **storage** services (RSEs) across Europe
- **data orchestration** with Rucio, policy-driven data management
- **file transfer** and with FTS3, multiple protocols (davs, xrootd, gsiftp)
- grid file access library with GFAL2
- Identity and Access Management (IAM) deployed at INFN/CNAF - *new Rucio version supporting token-based auth*
- CRIC Global Information Catalogue

Horizon 2020
ESCAPE
CERN
EOSC Future

A step further - Data Lake as a Service (DLaaS)

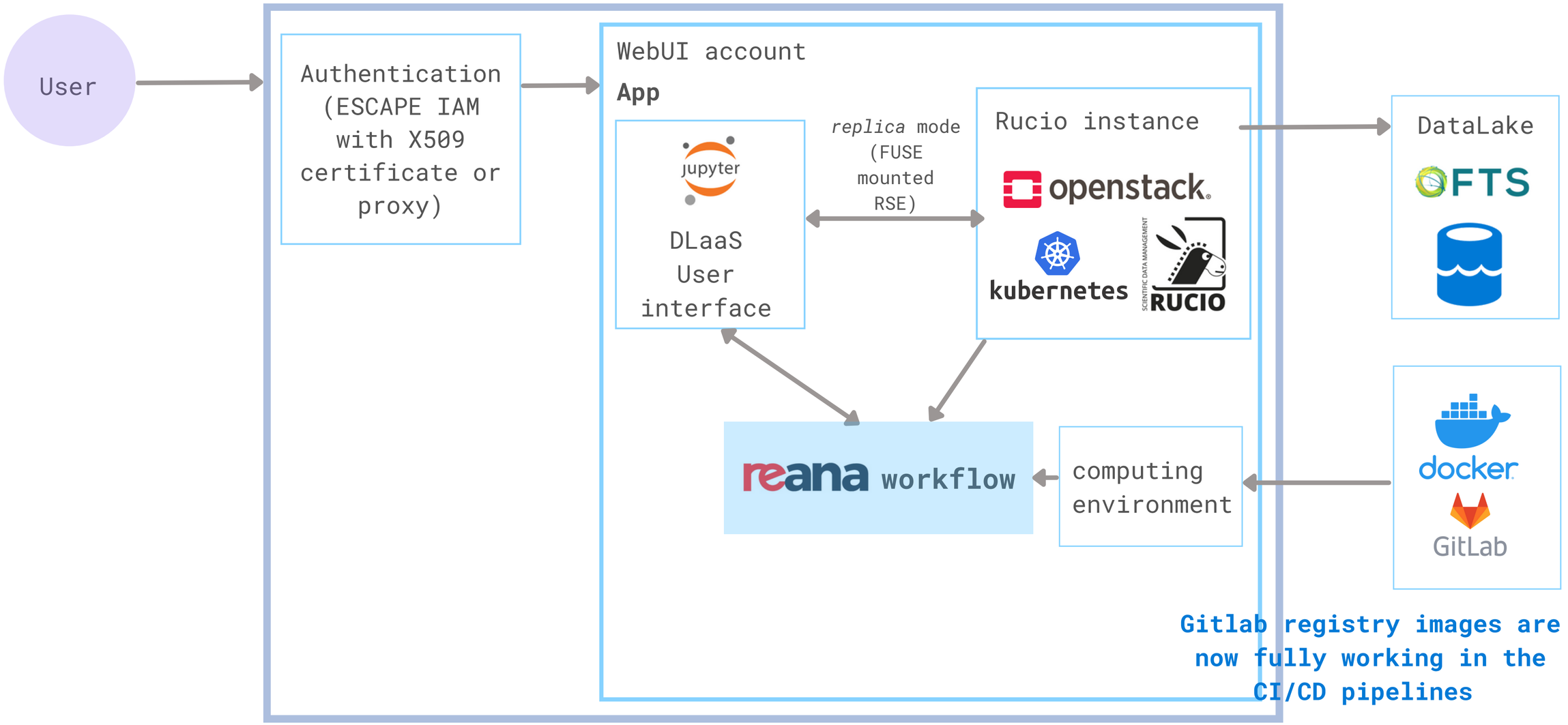


The VRE server



Horizon 2020
ESCAPE
CERN
EOSC Future

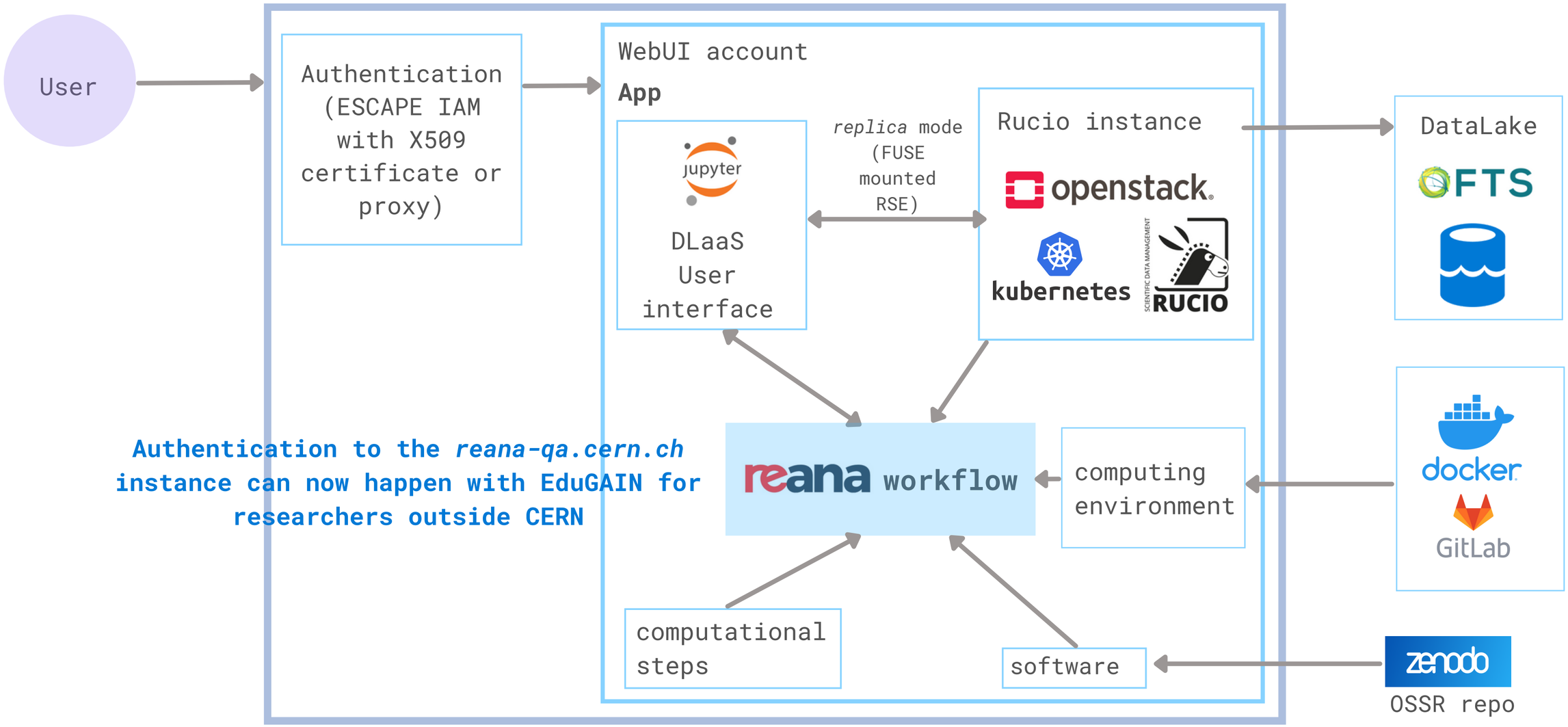
The VRE server



Gitlab registry images are now fully working in the CI/CD pipelines

Horizon 2020
ESCAPE
CERN
EOSC Future

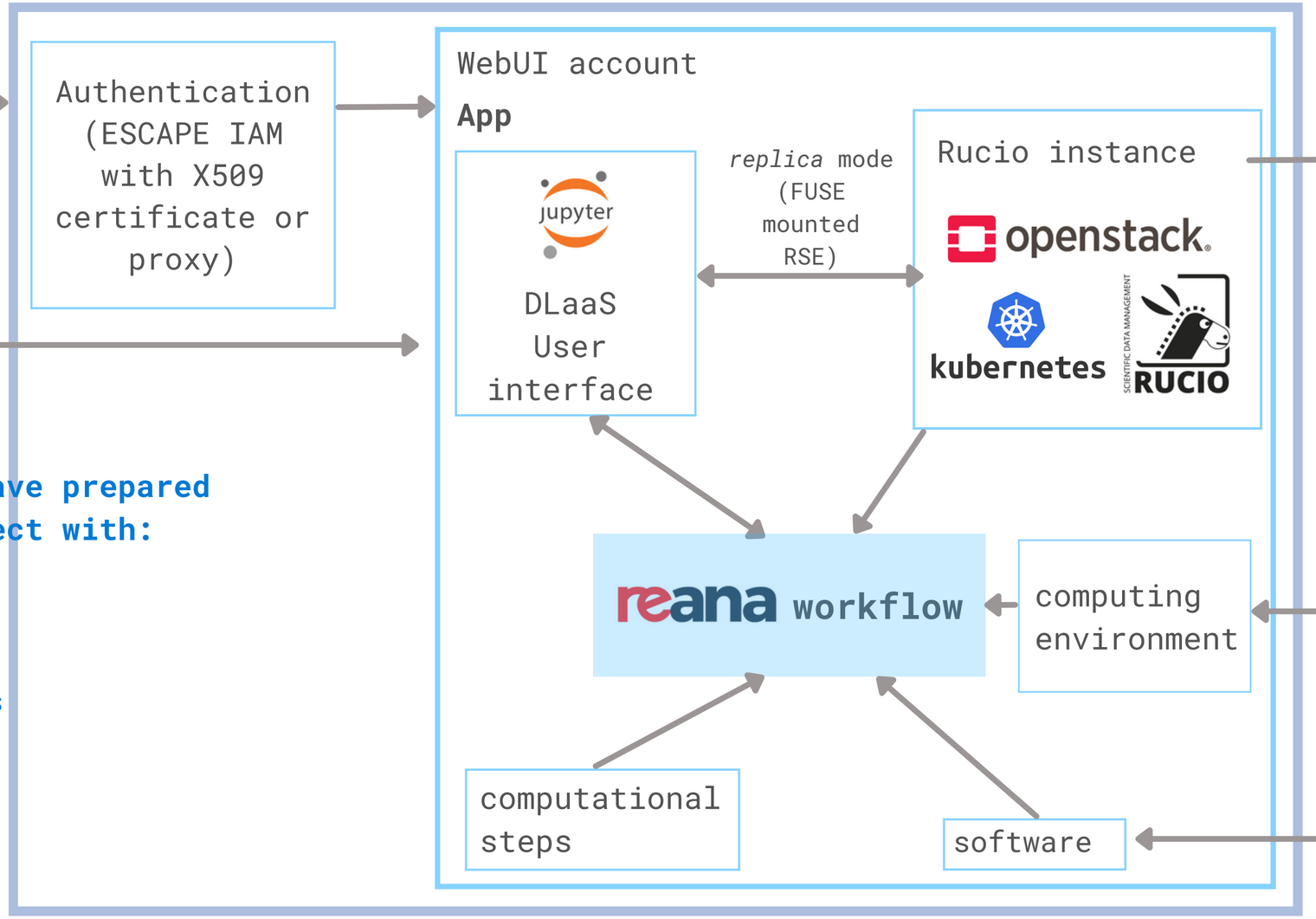
The VRE server



Authentication to the *reana-qa.cern.ch* instance can now happen with EduGAIN for researchers outside CERN

Horizon 2020
ESCAPE
CERN
EOSC Future

The VRE server



INFN-CLOUD-CNAF have prepared an OpenStack project with:

- **500vCPU**
- **1000GB**
- **10TB storage**
- **10 Floating IPs**

for 1 week

Horizon 2020
ESCAPE
CERN
EOSC Future

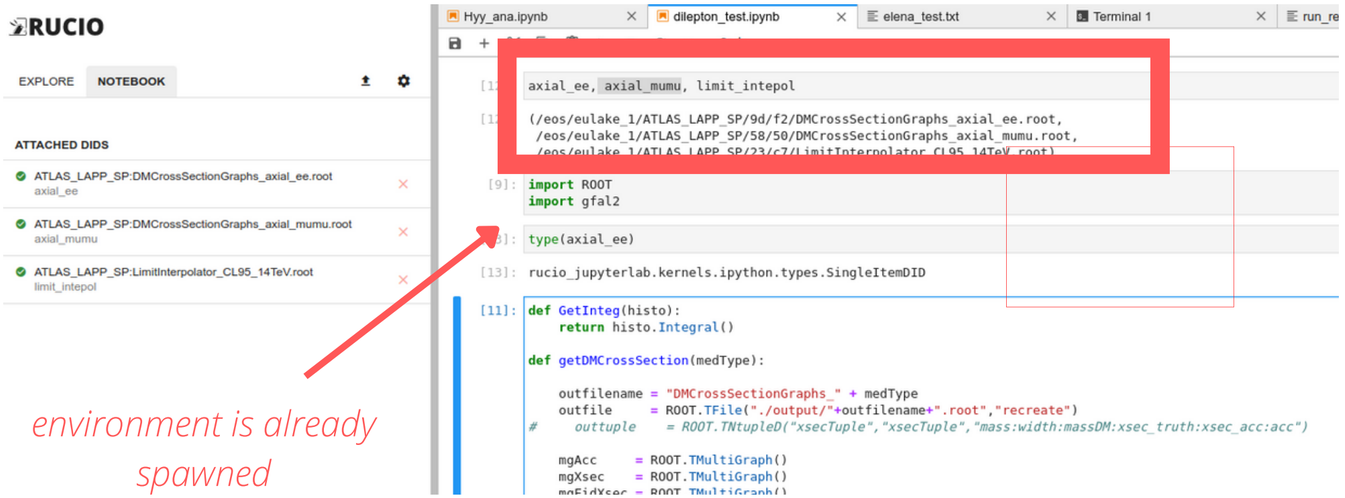
AN EXAMPLE

Starting to connect the dots

Horizon 2020
ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures
CERN
EOSC Future

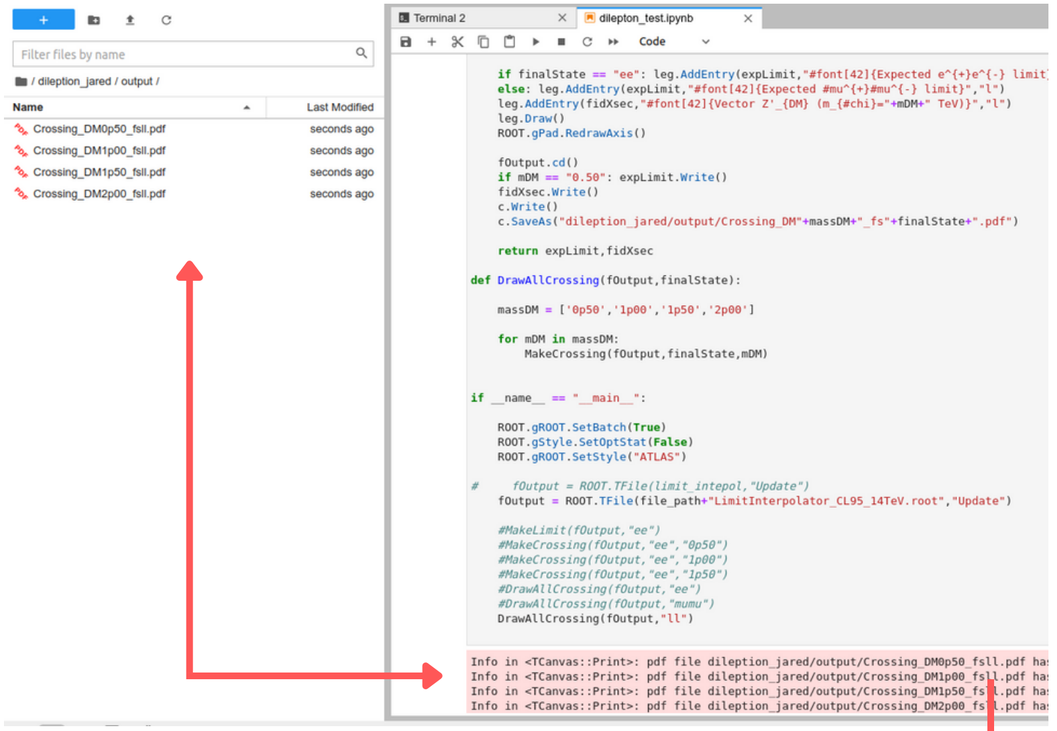
ATLAS Dark Matter Reinterpretation - Dilepton Resonance

1. Import files as variables into notebook

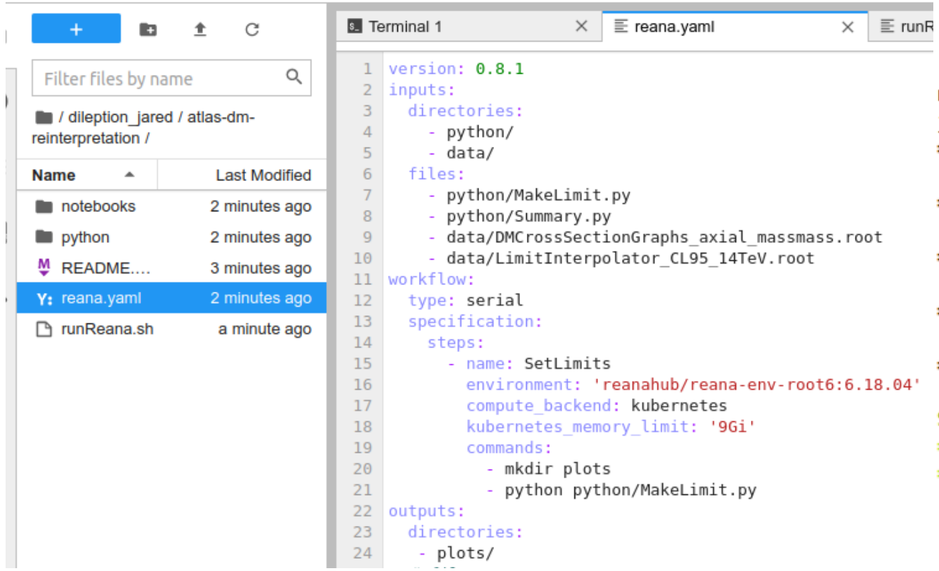


environment is already spawned

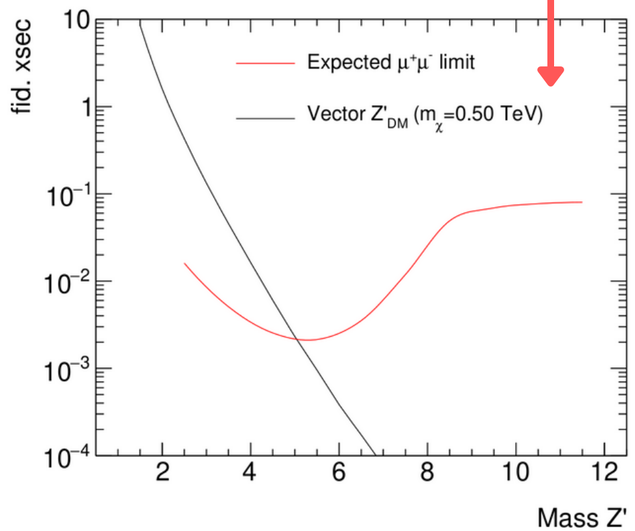
2a. Output generation



2b. REANA automatises workflow execution



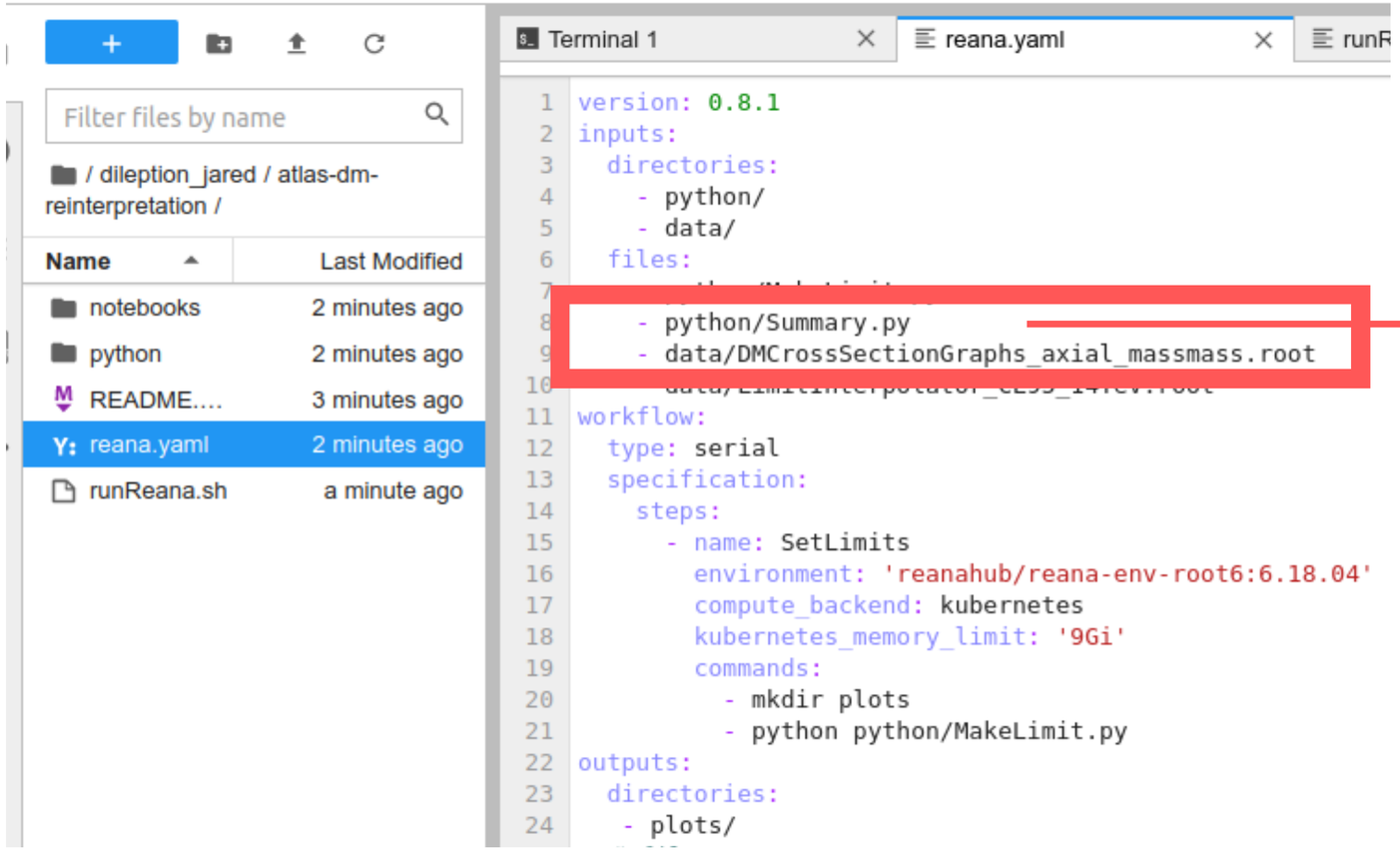

```
notebooks  python  README.md  reana.yaml  runReana.sh
jovyan@jupyter-egazzarr:~/dileption_jared/atlas-dm-reinterpretatio
==> Verifying REANA specification file... /home/jovyan/dileption_j
  -> SUCCESS: Valid REANA specification file.
==> Verifying REANA specification parameters...
  -> SUCCESS: REANA specification parameters appear valid.
==> Verifying workflow parameters and commands...
  -> SUCCESS: Workflow parameters and commands appear valid.
==> Verifying dangerous workflow operations...
  -> SUCCESS: Workflow operations appear valid.
==> Verifying compute backends in REANA specification file...
  -> SUCCESS: Workflow compute backends appear to be valid.
SettingLimits.1
==> SUCCESS: File /python/MakeLimit.py was successfully uploaded.
==> SUCCESS: File /python/Summary.py was successfully uploaded.
```



Summer project

Welcome Agisilaos Kounelis !

AIM: be able to use the Data Lake data from the Reana cluster, without the restriction of having data locally.



substitute with
scope:data

(from Rucio DID logic)

Final remarks and next steps

1. To do:
   a. connect with IVOA (WP4) and ESAP (WP5)
   b. summer project for data fetcher/wrapper
   c. summer project for Zenodo plugin for environment import (similar to Rucio one)

1. As we want to run on EOSC-provided resources, are we in need of re-deploying the Data Lake cluster and another REANA cluster?

1. Federated authentication for all users and for both DL and Reana, ongoing discussion on accessing reana CERN resources

1. How to deal with scalability?

References

- Rucio
- Rucio Escape WebUI
- K8s cluster CI/CD Helm + Flux configurations
- Cluster testing
- Grafana monitoring
- VRE webpage (in progress)
- VRE documentation
- VRE onboarding
- VRE scientific analyses + docker images for notebooks

Thank you! Questions?