

Data discovery, analysis and reproducibility in Virtual Research Environments

Enrique García¹, Giovanni Guerrieri¹, Rubén Pérez¹, Michael Zengel¹, Georgy Skorobogatov², Andrés Tanasijczuk³, Hugo Gonzalez¹ and Xavier Espinal¹

¹ CERN, ² ICCUB Barcelona, ³ UCLouvain

CHEP 2024 | Track 9 | 24 October



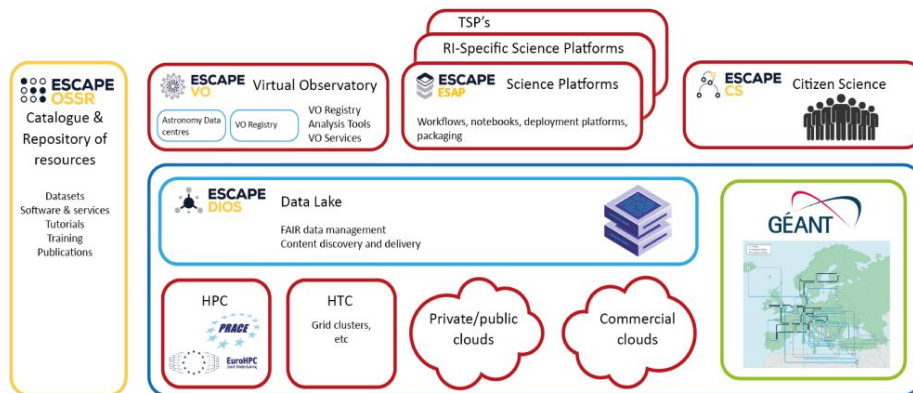
Outline

- Context
- The CERN Virtual Research Environment
- Demo video
- Summary & next steps

Context - ESCAPE EU Project



- Starting in 2019, the ESCAPE collaboration brought together various physics communities (ESFRIs, RIs, ERIC...)
 - Tackle and solve common computing issues
 - Data Management
 - Software and Service repository
 - Science platforms
 - Citizen Science
 - IVOA
 - From an Open Science perspective
- Project ended in 2023 → ESCAPE Open Collaboration
 - In addition to main project outcomes
 - Various initiatives and communities were naturally established



Consolidate ESCAPE outcomes



- It is fundamental to keep the momentum and communication to foster collaboration on common projects.
- Cascading grants via OSCARS project.
 - Understand the outcome of the first grant's call.
 - Put communities together to build a common basic roadmap.
 - Identify and aggregate specific community needs.
 - Both within a cluster (ESCAPE) and a between clusters (PaNOSC, SSHOC, ENVRI-FAIR, EOSC-Life).

Disclaimer

What this presentation **IS** going to show

- One of many analysis facility "concepts"
- Results of a bottom-up approach from (ESCAPE) user needs
- Built fully Open Source

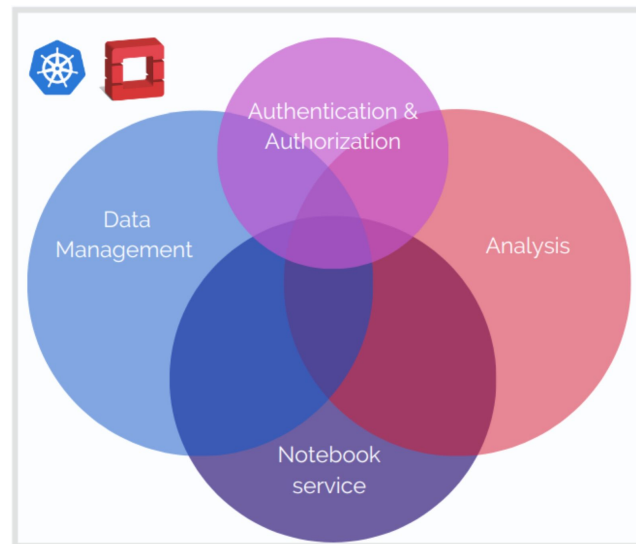
What this work **IS NOT** trying to do

- Solve everybody's needs (collaborations, experiments etc...)
 - Might not be compatible with certain computing architectures (bottom-up vs top-down).
- Provide an unique / monolithic solution

CERN Virtual Research Environment



- ESCAPE outcome
 - The place where most of the data-related R&D activities happened.
 - PoC platform provisioning different solutions for users:
 - Built on top of a common scientific tool: Jupyter
 - Software environment & repositories
 - Connection with resources: EOS, CVMFS
 - Ease interacting with underlying services



- To date CERN VRE acts as a **middleware interface** allowing scientific users to access different services and computing resources.

CERN Virtual Research Environment



- K8s cluster using OpenStack @ CERN
- Services installed via helm
 - ESCAPE Rucio Instance - ESCAPE Data Lake
 - Reana VRE instance
 - JupyterHub
- AAI: ESCAPE Indigo IAM
- Pre-loaded software environments via Docker images
- CVMFS
- EOS and CEPHFS storages



IAM for **escape**



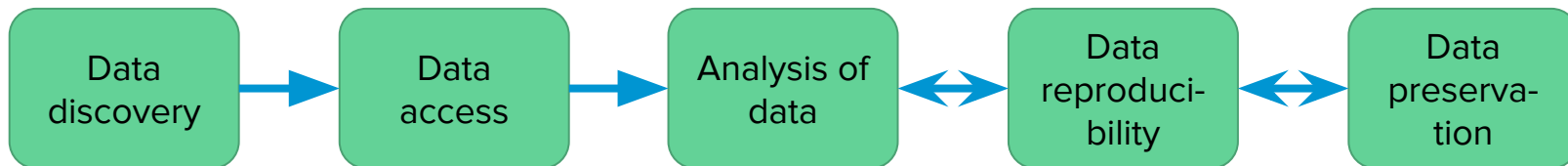
docker



Scientific analysis stages



Scientific analysis stages



Services accessible via the CERN VRE



The screenshots illustrate the reana workflow management and execution environment. The top-left window shows the 'CONNECT TO REANA' page with the following details:

- Server Name: `https://reana-vre.cern.ch`
- Access Token: `.....`

The top-right window shows a workflow named 'test_ET.1' in a 'Notebook' environment, which is 'finished in 2 minutes'.

The bottom-left window shows a list of workflows:

Workflow Name	Status
test_ET #1	finished
test_demo_GA #1	finished
test1 #1	failed
test #2	failed
test #1	failed

The bottom-right window shows a terminal window with a file download list:

File	Size	Last modified
reana.yaml	133 Bytes	17/10/2024, 14:43:44
psds_csds_1002646016-1002648064.npz	2.13 MiB	17/10/2024, 14:44:57
parameters_1002648064-1002650112.npz	1.08 KiB	17/10/2024, 14:44:24
point_estimate_sigma_1002641920-100...	766.42 KiB	17/10/2024, 14:43:38
psds_csds_1002648064-1002650112.npz	2.13 MiB	17/10/2024, 14:44:57
parameters_1002648016-1002648064_f...	1.08 KiB	17/10/2024, 14:44:24
parameters.ini	951 Bytes	17/10/2024, 14:43:45
point_estimate_sigma_1002643968-100...	766.49 KiB	17/10/2024, 14:45:08

Scientific analysis stages



+



Services accessible via the CERN VRE



The image displays three overlapping screenshots of the Zenodo web interface, accessed via a terminal window. The terminal window title is `jovyan@b5ea29c50901: ~`.

Left Screenshot: Search Results

Showing Results from "ESCAPE OSSR"

Title	Resource Type	Date Published
cds-astro/aladin-lite: 3.5.1-beta	Software	2024-09-18
GammaLearn	Software	2024-08-21
cds-astro/mocpy: v0.16.2	Software	2024-07-26
MOC Lib Rust, MOCCLI, MOCWasm and MOCSet	Software	2023-12-11
R3BRootGroup/R3BRoot: R3BRoot Release June 2024	Software	2024-06-20
G-Tomo	Software	2024-06-03
timewise-sup: The Timewise Subtraction	Software	2024-05-05

Middle Screenshot: Upload Page

Upload

Sandbox

Select Files to Upload *

\$HOME

- dataset 14/10/2024 4096 B
- reana.yaml 14/10/2024 0
- codemeta.json 14/10/2024 0
- README.md 14/10/2024 0
- work 09/09/2024 4096 B

4 files Expand

Digital Object Identifier

Right Screenshot: Confirmation Page

Confirmation

Title: Software upload test

Resource Type: software

DOI: (automatic)

Description: None given.

Creators:

- Enrique G (CERN)

Files:

- /home/jovyan/README.md
- /home/jovyan/codemeta.json
- /home/jovyan/reana.yaml
- /home/jovyan/dataset

Sandbox: No

Edit Confirm

Scientific analysis stages can be performed in the VRE



+



Demo (video x1.5)

[permanent link](#)



Sign in with OAuth 2.0

Summary

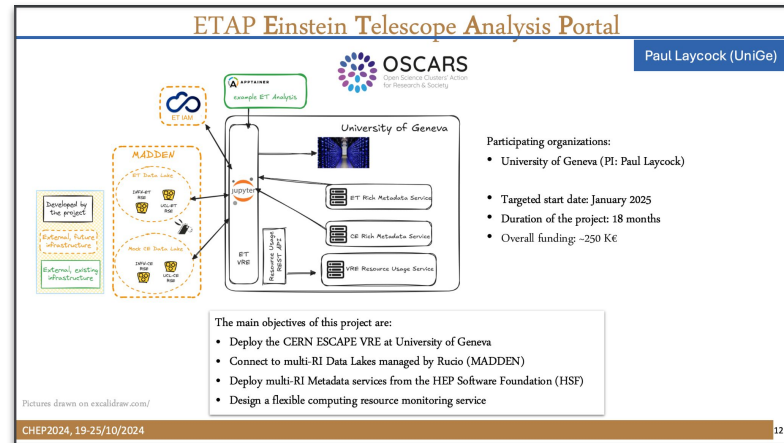


- CERN VRE presents an Analysis Facility concept
 - Allows performing most of the stages of a complete scientific analysis
 - All from the localhost of a Jupyter session - compatible with a remote / cloud-based architecture
 - Accessing and discovering data via the **Rucio** Plugin
 - (re)Triggering a workflow on a **Reana** instance
 - Publishing results on **Zenodo**
- Fully modular
 - Can help establishing the basic tools for certain communities

Future of the CERN VRE

- Demonstrator will continue
 - As a PoC and test place for other communities
 - As a place for R&D activities
 - Potential Open Science platform
 - [ET Workshop](#) Feb '24
- Maintain the community active and provide support / expertise
 - Lia L. Track 1 [presentation](#) on behalf ET community

The screenshot shows a web browser window with the URL `agenda.infn.it/event/38405/contributions/218806/`. The page is for an event on 20-23 févr. 2024 at BV Grand Hotel Assisi. The main content is for the 'ET Introduction Mock Data Challenge' on 20 févr. 2024, 17:20, lasting 1h 20m in the Auditorium. The speaker is Elena Cuomo. The description states: 'MDC hands on: The lecture will highlight the procedure for accessing the Einstein Telescope Mock Data Challenge (MDC). It will cover fundamental signal processing techniques applied to noise data and provide insights into the detection of astrophysical signals. Software requirement: code and data in a web browser, but we need to request the AAI credentials for all the students to ESCAPE VRE team.' There are links for 'Documents de présentation' including 'ET-MDC-introduction handson.pdf' and 'ET-MDC-Tutorials'.





Thanks for your attention

enrique.garcia.garcia@cern.ch; escape-cern-ops@cern.ch

VRE links:

- Access to the VRE <https://jhub-vre.cern.ch/>
- VRE documentation <https://vre-hub.github.io/>
- GitHub organisation <https://github.com/vre-hub>

Back up slides

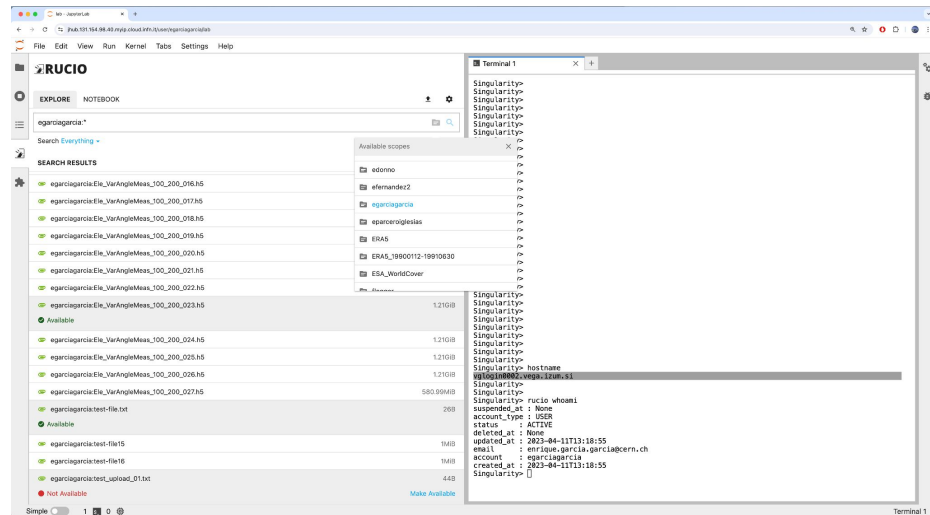
Rucio JupyterLab Extension

- Official Rucio component
- Compatible both
 - jupyterlab v3
 - jupyterlab v4
- Integrated into interTwin - EU project infrastructure
- Code:

<https://github.com/rucio/jupyterlab-extension>

- Documentation:

<https://vre-hub.github.io/docs/extensions/rucio-jupyterlab/>



Reana Extension

- 2024 Summer Student project

pypi v1.0.0

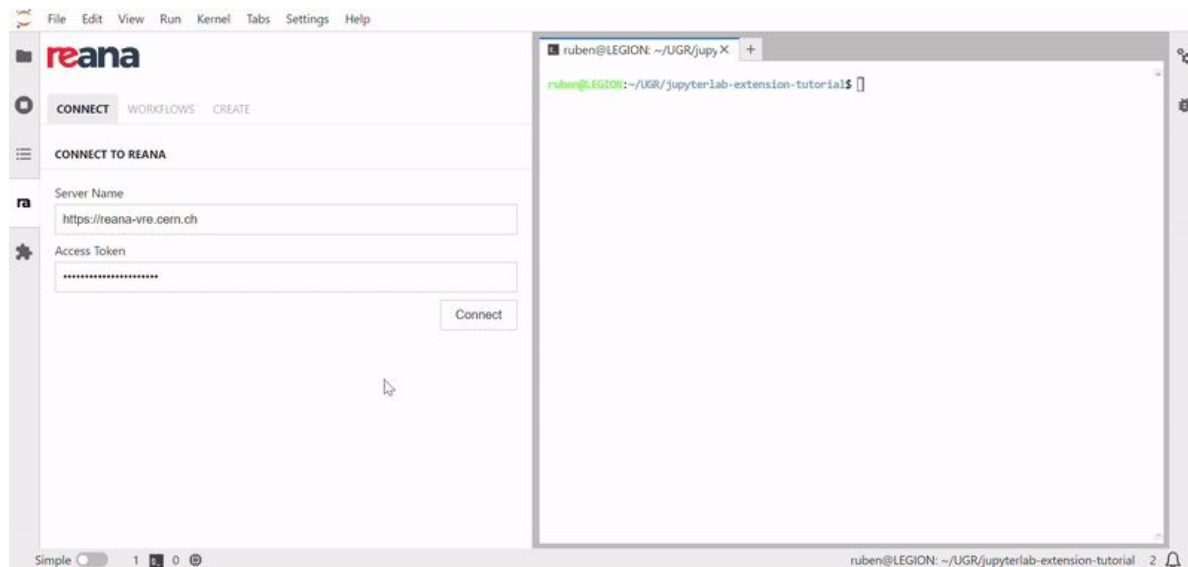
DOI 10.5281/zenodo.1336643

- Code:

<https://github.com/vre-hub/reana-jupyterlab-extension>

- Documentation:

<https://vre-hub.github.io/docs/extensions/reana-jupyterlab/>



Zenodo JupyterLab Extension

- 2024 Summer Student project
 - Few functionalities still under dev
 - V1.0.0 before the end of the year

- Code:

<https://github.com/vre-hub/zenodo-jupyterlab-extension>

- Documentation:

<https://vre-hub.github.io/docs/extensions/zenodo-jupyterlab/>

