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



- 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



/RE@

• 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



Scientific analysis stages







Services accessible via the CERN VRE

↓).	→ C	ab Q 🛧 🔿 🖒 ettings Help	New Chrome availa	able :						
	₽RUCIO	Terminal 2 × +		°0						
0	EXPLORE NOTEBOOK 1	jovyanejupyter-garcia:~\$ jovyanejupyter-garcia:~\$ jovyanejupyter-garcia:~\$		ð						
≔	ATLAS_LAPP_SP:*	jovyan@jupyter_garcia:~\$ jovyan@jupyter_garcia:~\$ jovyan@jupyter_garcia:~\$								
2	SEARCH RESULTS	jovyan@jupyter-garcia:~\$ jovyan@jupyter-garcia:~\$ jovyan@jupyter-garcia:~\$								
*	 ATLAS_LAPP_SP:DMsummary.dilept • ATLAS_LAPP_SP:DMCrossSec 25.79KiB 	jovyan@jupyter-garcia:~\$ rucio list-scopes test ET_OSB_MDC1 CMS_CFERN_AGC								
	 ATLAS_LAPP_SP:DMCrossSec 25.79KiB ATLAS_LAPP_SP:DMCrossSec 25.79KiB 	ATLAS_LAPP_SP KM3NET_ECAP_SP EG0_INFN_GW								
	 ATLAS_LAPP_SP:final_histos.r 4.64MiB ATLAS_LAPP_SP:LimitInterpol 26.38KiB 	LAPP-MUSIRY RIS_TESTS jovyan@jupyter-garcia:~\$ rucio list-dids ATLAS_LAPP_SP:filte	er 'type=all' ++							
	ATLAS_LAPP_SP:merged_D 167.13MiB	SCOPE:NAME	[DID TYPE]							
	ATLAS_LAPP_SP:merged_D 167.13MiB	<pre>ATLAS_LAPP_SP:DMCrossSectionGraphs_axial_ee.root ATLAS_LAPP_SP:DMCrossSectionGraphs_axial_mumu-z.root ATLAS_LAPP_SP:DMCrossSectionGraphs_axial_mumu_root</pre>	DIDType.FILE							
	ATLAS_LAPP_SP:testFC.root 4.64MiB ATLAS_LAPP_SP:testFC.root 4.64MiB	ATLAS_LAPF_SP:Netrossocctionophs_dxldt_mund.toct ATLAS_LAPF_SP:final_histos.root ATLAS_LAPF_SP:final_histos.root ATLAS_LAPF_SP:minitInterpolator_CL95_14TeV.root ATLAS_LAPF_SP:merged_DM_axial_ee_gDM_1p0.root ATLAS_LAPF_SP:merged_DM_axial_mumu_gDM_1p0.root ATLAS_LAPF_SP:metstFc.root	DIDType.DATASET DIDType.FILE DIDType.FILE DIDType.FILE DIDType.FILE DIDType.FILE DIDType.FILE DIDType.FILE							
	Simple 2 2 0 (iii)	jovyan@jupyter-garcia:~\$ []	Term	ainal 2						
L	2 Numbrie 2 Na 0 1@5		leini							

Data discovery, analysis and reproducibility in VREs - CHEP 2024 - 24 Oct - Enrique Garcia

Scientific analysis stages







Data discovery, analysis and reproducibility in VREs - CHEP 2024 - 24 Oct - Enrique Garcia

Services accessible via the CERN VRE



💭 File	Edit View Run Kernel	Git -	Tabs Settings Help		_		
• re	ana			🖸 Launcher +		File Edit View Run Kernel Git Tabs Settings Help Cauncher + %	
Осом	NNECT WORKFLOWS CREAT	E		test_ET.1	0	CONNECT WORKFLOWS CREATE test_ET.1	
Scon	INECT TO REANA	С	File Edit View Run Kernel Git	Tabs Settings Help	•>	← Notebook	
i≣ Ser	ver Name		reana		≔	© test_ET #1 finished in 2 minutes Finished: 17/10/2024, 14:46:00 step 13/13	
htt	tps://reana-vre.cern.ch	0	CONNECT WORKFLOWS CREATE		ze		옥 ☆ む 💩 🗄
Acc	cess Token	٠	YOUR WORKFLOWS		ra	Ster / download_data :]	°6
ra	•••••	=	Search	٩ (٦	æ	download_ddata sm.ucl.ac.be/MDC1/ddta/E2/E-E2_STRAIN_DATA+ O connect woeknows create tatal 2	ĕ
'n		ze	Status all - Sort by newest first -		*	job: download_data bas download_data 	7 14:43 ET_tests 17 16:10 Work 17 16:17 test_ET.1
*		ra	test_ET #1	\odot finished		Com download_data :cism.ucl.ac.be (et-origin.cism.ucl.ac.be) 38 HTT	i 1 3.npz 6.noz
		*	test_demo_GA #1	⊙ finished		run_pygwb psds_tcst_parts/set/set/set/set/set/set/set/set/set/s	
			test1 #1	× failed		200K 09 36 10 reax speed 133 Bytes 170(0224, 54-43.84.1 256K 09 52 36 0 parax speed 133 Bytes 170(0224, 54-43.84.1 386K 09 52 36 0 parax speed 133 Bytes 170(0224, 54-43.84.1 360K 09 52 36 0 parameters, 1002645054-1002644064-1002 170(0224, 54-44.24.1 460K 09 46 10 parameters, 1002645064-1002650112, L. 108 KB 170(0224, 54-44.24.1)	
	test		test #2	× failed		458K 09 52 point_extinue_signe_1002641020-100_ 796.12 KiII 177002024, 44453 C 598K 09 42 point_extinue_signe_1002640084-1002860112,ret 2:13 Mi8 177002024, 44453 C 558K 09 42 point_extinue_signe_1002640084-1002640084-1002860112,ret 2:13 Mi8 177002024, 44453 C	
			test #1	× failed	Ļ	imple 2 1 (a) (b) parameters ini 991 Bytes 1770 02024, 44.3344 D parameters init_sign_1002643906+100	
Simple	💶 2 🛐 1 @; 🚸		К	< 123 > >I		Python F	
		s	imple 🔵 2 🛐 1 @ 🚸			Simple 2 g 1 @ *	X Terminal 3 16 🚊

Scientific analysis stages





Services accessible via the CERN VRE







Login Search Upload	
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





Data discovery, analysis and reproducibility in VREs - CHEP 2024 - 24 Oct - Enrique Garcia

Demo (video x1.5)

 \leftrightarrow \rightarrow C 25 jhub-vre.cern.ch/hub/login

💭 jupyterhub

<u>permanent link</u>

@☆ ♪! 🕘 :



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

← → C 😨 agenda.infn.i	/event/38405/contributions/218806/	🖈 🖸 🚭 🗄
20–23 févr. 2024 BV Grand Hotel Assisi Fuseau horaire Europe/Rome		Entrer le texte à rechercher Q
Pueseu horaire EuropeRome Accueil Ordre du jour Accommodation & Registration Liste des contributions Liste des participants	ET Introduction Mock Data Challenge 20 few: 2024 17:20 3 20 few: 2024 17:20 3 20 few: 2024 17:20 3 20 few: 2024 17:20 4 20 few: 2024 17:20 5 2024 17:20 5 20 few: 2024 17	In Tolescope Mack Data Challenge (MDC). It revoke resigns into the detection of the AAI credentials for all the students to
	EFADO-introduction handson pdf # EFADO-Tutorials	





Thanks for your attention

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

VRE links:

- Access to the VRE <u>https://jhub-vre.</u>
- VRE documentation
- GitHub organisation

https://jhub-vre.cern.ch/ https://vre-hub.github.io/ 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/

C 😂 Bub.131.154.98.40 mylp.cloud.infn.it/usen/eganciagarcia/ab			a 🖈 🚺 🗅
ile Edit View Run Kernel Tabs Settings Help			
RUCIO		Terminal 1 × +	
EXPLORE NOTEBOOK	± 0	Singularity> Singularity> Singularity> Singularity>	
egarciagarcia:*	🖻 Q	Singularity> Singularity>	
Search Everything +	Available scopes	× P	
SEARCH RESULTS	🖿 edonno	0	
egarciagarcia:Ele_VarAngleMeas_100_200_016.h5	efernandez2	10 10 10 10 10 10 10 10 10 10 10 10 10 1	
egarciagarcia:Ele_VarAngleMeas_100_200_017.h5	🖿 egarciagarcia	0 0	
egarciagarcia:Ele_VarAngleMeas_100_200_018.h5	🖿 eparceroiglesias	2	
garciagarcia:Ele_VarAngleMeas_100_200_019.h5	E ERAS		
egarciagarcia:Ele_VarAngleMeas_100_200_020.h5	Ea ERA5_19900112-19910630		
egarciagarcia:Ele_VarAngleMeas_100_200_021.h5	ESA_WorldCover	· · · · · · · · · · · · · · · · · · ·	
egarciagarcia:Ele_VarAngleMeas_100_200_022.h5	Pro Annon	P Singularity	
egarciagarcia:Ele_VarAngleMeas_100_200_923.h5 Available	1.21G/B	Singularity Singularity Singularity	
egarciagarcia:Ele_VarAngleMeas_100_200_024.h5	1.216/8	Singularity>	
egarciagarcia:Ele_VarAngleMeas_100_200_025.h5	1.216/8	Singularity-	
egarclagarcla:Ele_VarAngleMeas_100_200_026.h5	1.216/8	Singularity> hostname vglogin8002.vega.izum.si	
egarciagarcia:Ele_VarAngleMeas_100_200_027.h5	580.99MiB	Singularity> Singularity>	
egarciagarcia:test-file.txt	268	suspended_at : None	
Available		status : ACTIVE	
egarciagarcia:test-file15	1MiB	updated_at : 2023-04-11T13:18:55	
egarciagarcia:test-file16	11418	account : egarciagarcia created at : 2023-04-11713:18:55	
egarciagarcia:test_upload_01.txt	448	Singularity> [
Not Available	Make Available		

Reana Extension

 2024 Summer Student project







• 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
 - \circ V1.0.0 before the end of the year

File Edit View Run Kernel Git Tabs Settings Help jovyan@b5ea29c50901: ~ × + zenodo (base) jovyan@b5ea29c50901:~\$ [] 0 € ≔ Search... ze Records Communities Search Showing Results from "ESCAPE OSSR" × Title **Resource Type Date Published** cds-astro/aladin-lite-Software 2024-09-18 3.5.1-beta Software 2024-08-21 GammaLearn cds-astro/mocpy: v0.16.2 Software 2024-07-26 MOC Lib Rust, MOCCLi, Software 2023-12-11 MOCWasm and MOCSet R3BRootGroup/R3BRoot: 2024-06-20 R3BRoot Release June Software 2024 G-Tomo Software 2024-06-03 timewise-sup: The Timewise Subtraction 2024-05-05 Software Simple 🔵 1 👩 🛞 🊸

• Code:

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

• Documentation:

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