DLaaS working group

Organising future workforce activities

Enrique Garcia, Alba Vendrell Moya, Elena Gazzarrini and Giovanni Guerrieri
What is the DLaaS

A Jupyter environment that connects the ESCAPE Data Lake with a computing facility.

Created by WP2 to:

- Provide **easier interaction** with the RUCIO instance
- Connect DL with **compute** resources + **storage** resources
- Start **collecting** ESCAPE services under a JupyterHub (server) environment
Current status
EOSC-Future science project

COMPLETE COMMON INFRASTRUCTURE
- RUCIO INSTANCE
- SOFTWARE REPOSITORY
- COMPUTING FACILITY

PLATFORM to develop, deploy, expose and preserve scientific workflows

WEB SERVICES

OBJECTIVE
Providing an environment to access ALL digital content related to a scientific result

OPEN SOURCE SCIENCE
COLLABORATION
FAIR PRINCIPLES
ANALYSIS PRESERVATION

DARK MATTER
different data needs, different methodologies

EXTREME UNIVERSE
Virtual Research Environment (VRE)

A platform to collect all ESCAPE services and facilitate EOSC-Future project development
Looking into the future

We envision three different scenarios:

■ **User** - researcher
  ● I would use the DLaaS to run my analysis workflows

■ **Developer** - help on technical side to maintain and expand the DLaaS capabilities
  ● Maintain the K8s cluster component
  ● Connect to large computing resources (through Reana)
  ● Scale the infrastructure to allow more users
  ● Authenticate to DLaaS and Reana with the same token

■ **Provider**

■ **Learner** - anyone
  ● I am interested in the technologies and would like to reproduce the DLaaS in an environment which is not CERN specific

■ **Other?**
Round table

1. Why do you want to join?
2. What do you expect from these monthly meetings?
3. What is your expertise? How could you contribute?
Back-up slides
DLaaS into specifics

Hub API

OAuth auto renew

Rucio JupyterLab extension

homedir FUSE mount

EULAKE-1 FUSE mount /eos/eulake_1

CephFS Cinder

in /home/jovyan, shared between users, never cleaned, 800GB

EOS EULAKE-1

- in /scratch, shared between users, cleaned every day, 384TB
- replication mode, to store requested replicas

ESCAPE IAM

Subject Mapping Cron

API
Looking into the future

1. FLEXIBLE REPRODUCIBILITY
   a. Ability to change variables within REANA configuration files
   b. Modular analyses: different environment and software distribution for each step