

Aidan Campbell (ESRF)  
Jamie Hall (ILL)


# EOSC integration and common portal for remote data analysis services



# Quick overview

*“The objective of this work package is to make data analysis services available through cloud hosted services and on the EOSC.*

- *It must be possible to choose, control and execute analysis services remotely*
- *User interface for local and remote execution is identical or at least similar*
- *In the context of FAIR principles, a complementary objective is to support traceability, persistent identification, and reproducibility of the data analysis process from raw data to publication data”*



# EOSC integration and common portal for remote data analysis services

## Task 4.3

*“We will provide a service that will allow a user to remotely analyse their data from any facility via a common portal. Federated authentication and cross facility data transfer will be required for the common portal to reach its full potential.*

*These requirements will be addressed in WP6:*


- *Task 6.3 (Data availability for the services)*
- *Task 6.4 (Authentication Authorisation Infrastructure)*

*In this task we will implement a facility connector for the remote desktop services and adapt these services for the EOSC AAI and data transfer mechanisms.”*



# Common portal

- *“We will extend the single site remote desktop portal for the selection of the compute and data providers”*
- *“Each participating facility will implement a connector that allows the common portal access to their compute infrastructure and manage the transfer of data”*
- *“A user will be able to select any facility and start remotely analysing their data via a single interface”*
- *“This common platform, based on existing solutions, should also provide the possibility of directly archiving and sharing results after a user has completed their data analysis using the services provided”*



# EOSC Authentication and Authorisation Infrastructure

- *Implement the technical solutions chosen in WP6 to authenticate the users in the different layers of the portal architecture*
  - *Portal access, data access and compute resources*
- *Authentication should allow the different providers to grant the proper authorisation and to implement service usage accounting*
- *We will review the security of the system to identify risks and necessary security measures*



# Data sharing

- Modify the compute services to benefit from the work in WP6 concerning the movement of data
- Integrating the data transfer solution, we will allow users to transparently work on data, irrespective of its location, and perform data analysis



# What do we currently have?

- VISA
  - Virtual Infrastructure for Scientific Analysis
- CALIPSOplus project
  - Convenient Access to Light Sources Open to Innovation, Science and to the World.
  - <http://www.calipsoplus.eu/>



# VISA

- Give users access to compute resources to analyse their data
- Users can analyse their data using a remote desktop inside their web browser
- Compute resources can be shared i.e. work together collaboratively
- Currently supports OIDC for the AAI layer
- Written in Angular for the frontend and Java for the backend
- Due to be made public to ILL scientific visitors in the coming weeks





# CALIPSOplus - Key Features

Common portal between synchrotrons for the purpose of :

- Common interface UI/UX
- Access to experimental data from our datahubs.
- Jupyter Notebooks for analysing and visualising data
- Containers and virtual machines with pre-installed software



# Jupyterhub

*JupyterHub is the best way to serve Jupyter notebook for multiple users. It can be used in a classes of students, a corporate data science group or scientific research group. It is a multi-user Hub that spawns, manages, and proxies multiple instances of the single-user Jupyter notebook server.*

Three prototypes have been developed which can be analysed and implemented at each synchrotron according to their needs.

SudoSpawner

OAR/SLURM spawner

Kubernetes spawner



# Jupyterhub

## SudoSpawner

- Single computer, GPU access, unable to install additional software at user level

## OAR/SLURM spawner

- Scalable service, GPU access, unable to install additional software at user level

## Kubernetes spawner

- Scalable service, GPU access, ABLE to install additional software at user level



# Common analysis services goals

- Analyse data remotely
  - Transparently (irrespective of where the data and compute resources are hosted)
- Collaborate with your experimental team
- Record your analysis process
  - Long term archive
- Publish your analysis results / process / environment
  - Make these findable (FAIR)
- Share resources between compute environments
  - Scripts, data etc.
- Learning resources
  - Tutorials, videos, screencasts etc. (WP8)
- Exploring / finding analysis procedures / examples



# Common API

Common API between each participating institute that allows:

- Provisioning of compute resources (irrelevant of the compute infrastructure)
- Searching data via the API provided by WP3
- Access to data (services provided by WP6)
- Publishing / sharing analysis results
- KPIs/Metrics/Accounting (links to WP7)
- Build on microservices to help collaborative development

API acts as a *facility connector* for the common portal



# Common portal

Centralised portal that allows:

- Connects to each institutes API
- Common interface UI/UX (taking into account the PaNOSC branding guidelines)
- Allows access to Remote Desktops or Jupyter Notebooks
- Authentication / authorisation handle by services provided in WP6
- Developed using a well supported component based framework e.g. React or Angular
- All code to be made open source (Github)
- Share components



# Proposed Roadmap

1. Define a common API between the institute partners and build on experience from the CALIPSO project
2. Go over current use cases
  - a. <https://github.com/panosc-eu/panosc/blob/master/Work%20Packages/WP4%20Data%20analysis%20services/resources/visa-use-cases.md>
3. Define the services that will plug together the common portal:
  - a. Accounting / Quotas / KPIs
  - b. Remote desktop access
  - c. Notebook Access
  - d. Authentication service (plugs in to WP6 AAI)
4. Agree on common tools and languages
5. Start to agree on tasks that can be worked on collaboratively



# Open questions

1. How can we better facilitate working together more collaboratively?  
Are the tools we are using working to our advantage?
2. How can we distribute resources? Development, system administration, infrastructure etc.
3. Are there any projects that we can contribute back to?
4. How do we try and get the other cross-projects involved? ExPANDS, LENS, LEAPS , ESCAPE etc.
5. Any other ideas?



Thank you.  
Any questions?

