

Deliverable

Deliverable Number	D24.7
Deliverable Title	Organisation of a workshop to present the results of the DaaS demonstrator and obtain feedback from users on how this approach fits the current needs.
Lead Beneficiary	PSI / ESRF
Authors	A.Götz (ESRF), V.Piffer (PSI), D. Webster (PSI), D.Palmisano (ELETTRA), G.Kourousias (ELETTRA), F.Schlutzen (DESY), D. Salvat (ALBA)
Type	Report
Dissemination Level	Public
Due date of delivery	Month 36 - postponed to Month 42

Abstract

JRA2 activity within the CALIPSOplus project was presented as the first experience of the PaN facilities with Data Analysis as a Service at the ExPaNDS – PaNOSC technical workshop on October 8th – 9th, 2020. The workshop counted the presence of 80 participants.

This report is intended to briefly describe the content delivered and the feedback obtained.

Introduction

Light sources are generators of big volumes of complex scientific data and their users need assistance in analysing the scientific data.

The mission of CALIPSOplus Joint Research Activity 2 (JRA2) was to develop a prototype remote Data Analysis As a Service (DAAS) portal and platform for users to:

- access their experimental data
- use pre-packaged Data Analysis software available at each institute
- access onsite computer resources to assist with data reduction and processing

JRA2 and its activities have proved to be a small step in the right direction (also given the new challenges presented by the COVID-19 pandemic). It fostered collaboration and exchange on data analysis as a service and the demonstrator helped some users with generic and some specific data analysis tasks.

The long-term goal was to prepare for bigger projects in the future to solve the data analysis issue.

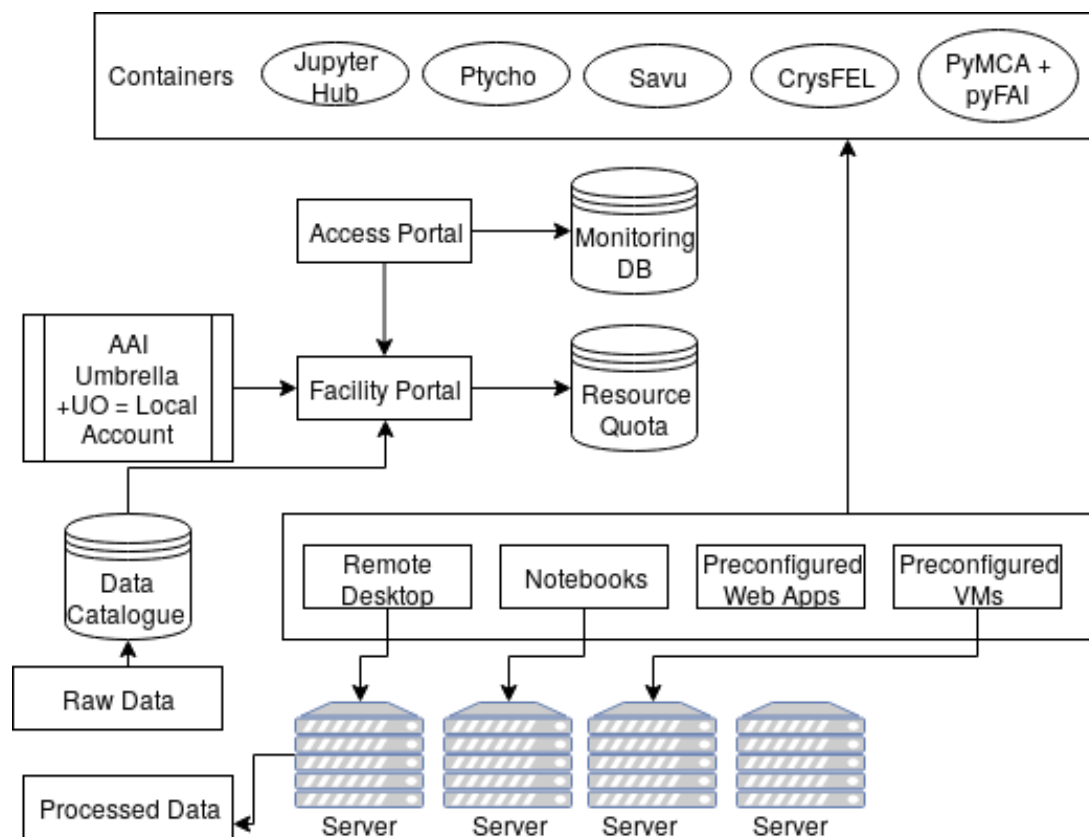


Fig. 1: CALIPSOplus DAAS Portal Architecture

The details of the architecture and goals can be found in the Blueprint (cf. deliverable D24.2¹) and in the paper² presented at the ICALEPCS 2019 conference in New York.

Final Workshop

As the ExPaNDS / PaNOSC forerunner project on Data Analysis as a Service paradigm, the JRA2 of CALIPSOplus joined the first technical workshop of the two projects with a dedicated two-hour session on October 8th, 2020. The workshop timetable and presentations was organised jointly by ESRF, CELLS, ILL and PSI.

ExPaNDS/PaNOSC and CALIPSOplus Technical Coordination Workshop : "I - The Portal Architecture test experience"

8-9 October 2020
Europe/Madrid timezone

Overview

Timetable

Contribution List

Registration

Participant List

Timetable

Thu 08/10
Fri 09/10
All days

Print
PDF
Full screen
Detailed view
Filter

CALIPSOplus session
Introduction
Technical Talks
×

13:00	Introduction to the project (Andy Goetz)	13:00 - 13:15
	Aidan Campbell (ESRF)	13:15 - 13:30
	Integration of Umbrella - Jean-François Perrin (ESRF)	13:30 - 13:40
	Alex Camps (ALBA)	13:40 - 13:50
	Dario Palmisano (ELETTRA)	13:50 - 14:00
14:00	Dan Webster (PSI)	14:00 - 14:10
	Johannes Reppin (DESY)	14:10 - 14:20
	Open discussion - Lessons learned (Chair: Mirjam van Daalen)	14:20 - 15:00

After a brief introduction by Andy Götz (ESRF) on the project and its status, each partner presented an update on the activities within their facility.

Detailed information on the workshop, including the agenda of the whole event and respective contributions can be found at the following link: <https://indico.cells.es/event/249/>

¹ <http://www.calipsoplus.eu/wp-content/uploads/2019/03/D24.2.pdf>

² <https://doi.org/10.18429/JACoW-ICALEPCS2019-WEPHA057>

JRA2 - Introduction

Talk by Andy Götz (ESRF) :

<https://indico.cells.es/event/249/contributions/1052/>

Andy Götz, co-leader of JRA2, gave an overview of the project. He explained who the partners were and the different tasks and who was in charge of them. He explained how the JRA2 had demonstrated the need and usefulness in providing a platform for discussing computing needs and strategy at photon sources. Even if JRA2 received modest funding it was a step in the right direction and fostered collaboration and exchange on data analysis as a service. This proved to be extremely useful and clairvoyant when the COVID-19 pandemic unexpectedly arrived in 2019 and put a stop to travel (see section on COVID-19 below). The development and deployment of a demonstrator helped sites explore the topics of data analysis as a service and to improve their local infrastructures. JRA2 was always meant to be a prototype and not be a production service. JRA2 paved the way to larger projects which have since been funded as part of the EOSC - PaNOSC³ and ExPaNDS⁴.

Prototype of Data Analysis as a Service Platform

Talk by Aidan Campbell (ESRF) and Dan Webster (PSI) :

<https://indico.cells.es/event/249/contributions/1053/>

Aidan Campbell (ESRF) and Dan Webster (PSI) presented the DAAS portal prototype architecture and technology choices. The main aim was to provide a platform for Jupyter notebooks, containers and virtual machines for data analysis as a service. They presented the status of the portal developments and deployment at the ESRF and PSI.

ESRF - Status of Portal

The ESRF used the JRA2 work package to develop a prototype and then eventually production ready service for Jupyter notebooks. The notebook service was integrated into the slurm batch scheduler at the ESRF based on the work done at DESY. The service has been made available to all ESRF scientists and users as part of the remote services for data analysis offered by the ESRF following the COVID-19 pandemic. The service can be found at <https://jupyter-slurm.esrf.fr> and provides the following interface:

³ <https://panosc.eu>

⁴ <https://expands.eu>

Server Options

Simple

Advanced

Architecture

Intel (x86_64)

Intel Xeon
Partition: nice

IBM Power (ppc64le)

IBM POWER9
Partition: p9gpu

CPUs

Minimum

1 core(s)

Medium

14 core(s)

Maximum

Entire node

Launch JupyterLab ☐

Job duration

List of available resources:

Current Status		
Partition	# nodes	# avail
nice	55	17
p9gpu	7	0

Start

► More info...

The service provides access to the Intel CPU and Power 9 CPU+GPU clusters. Over 200 individual scientists have used the service at the ESRF and at a recent training there were 100 participants. The training material is available here: <https://gitlab.esrf.fr/silx/training/jupyter-slurm> and the code for the jupyter-slurm integration on the ESRF gitlab: <https://gitlab.esrf.fr/silx/jupyterhub-slurm>.

ESRF deployed prototype services for the containers and virtual machines on a prototype OpenStack platform. The next step will be to move to a production Kubernetes platform. This work will be done as part of PaNOSC.

Integration of Umbrellaid

Talk by *Jean-François Perrin (ESRF)* :

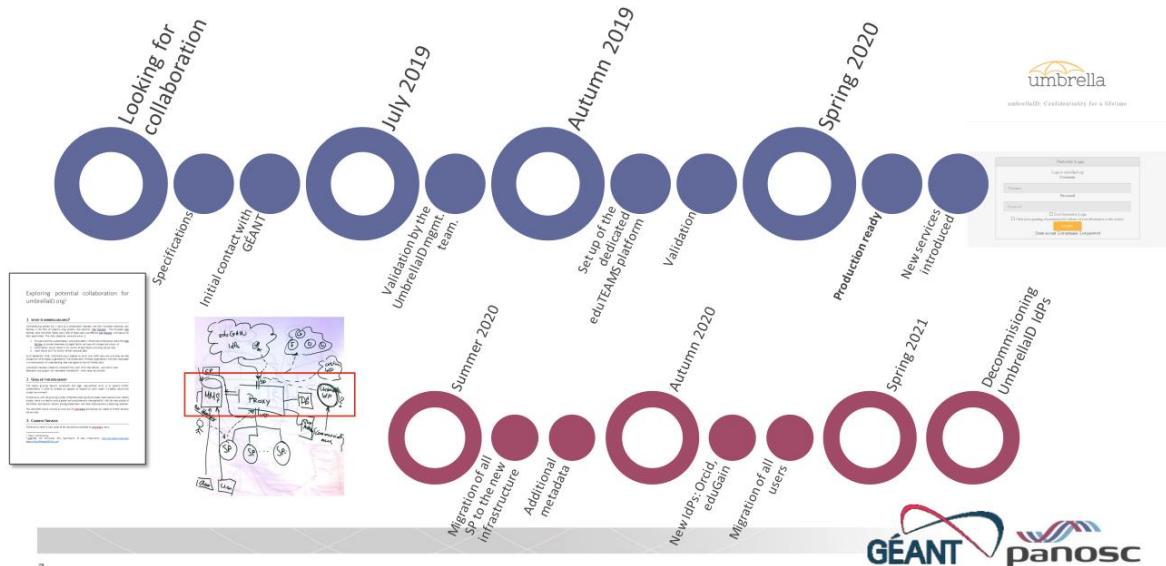
<https://indico.cells.es/event/249/contributions/1054/>

Jean Francois Perrin presented the developments on the Umbrellaid identity provider which was extended as part of JRA2 and has been adopted and made sustainable by PaNOSC with the help of GÉANT⁵. Thanks to these efforts Umbrellaid is now AARC and EOSC AAI compliant and ready. In plain words this means Umbrellaid can be used as an EOSC identity while still being a community identity provider. The next steps include stopping the local Identity Providers (IdP) and switching to the GÉANT IdP. The list of attributes will be extended to support more attributes necessary to make Umbrellaid

⁵ <https://www.geant.org/>

GDPR compliant and support more services. The timeline for the UmbrellaID changes is presented below.

UmbrellaID timeline of changes



ALBA - Status of Portal

Talk by *Alex Camps (ALBA)* :

<https://indico.cells.es/event/249/contributions/1055/>

Alex Camps presented the work by ALBA on the prototype. As one of the main developers of the portal prototype, they contributed heavily to the code development on github⁶. They deployed the portal locally but have not put it in production due to lack of manpower to deploy a production-ready portal and metadata catalogue (ICAT).

DESY - Status of Portal

Talk by *Johannes Reppin (DESY)*:

<https://indico.cells.es/event/249/contributions/1058/>

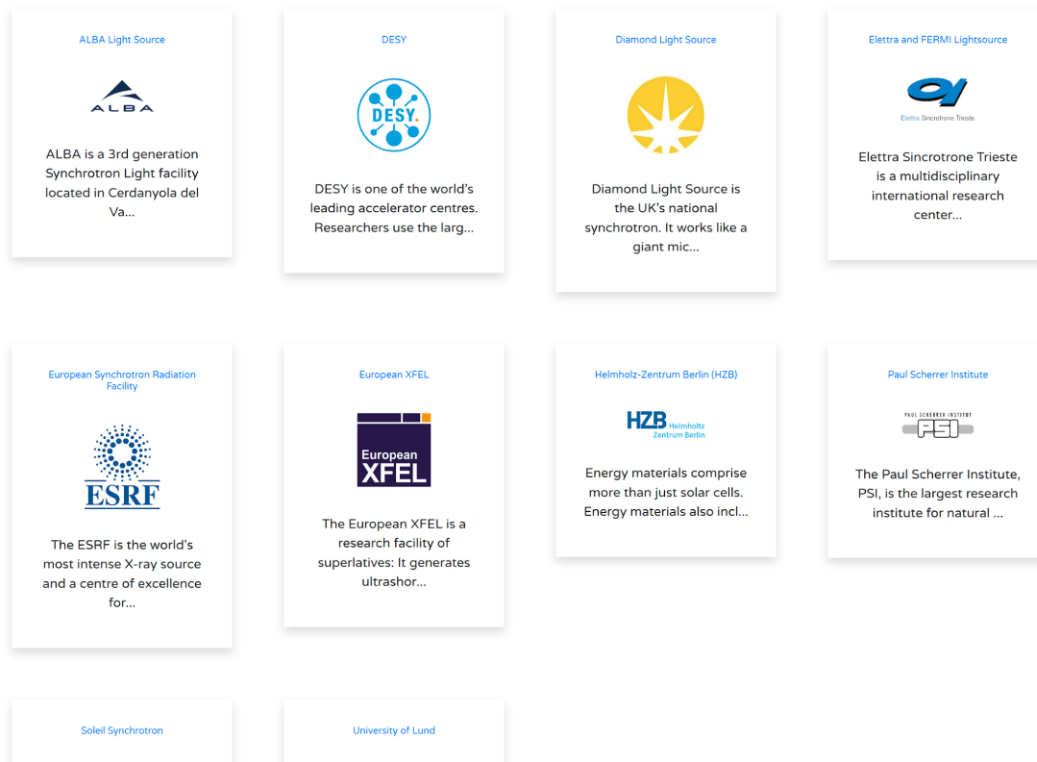
Johannes Reppin presented the work by DESY deploying the portal. Due to the fact that DESY has an Openstack and Kubernetes cluster in production it could test the full deployment of the prototype with all the services. DESY already had a jupyter-slurm service in production since 2018 with over 250 different users per month and up to 120 concurrently running sessions. The Helm chart used at DESY to deploy the portal prototype is available on their gitlab: <https://eosc-pan-git.desy.de/calipsoplus/calipsoplus-portal-helm>. The portal was integrated with the DESY Keycloak

⁶ <https://github.com/Calipsoplus>

Single-Sign-On (SSO) AAI. DESY is working on establishing a metadata catalog (Scicat), and evaluating users experience and acceptance through a pilot instance serving an AI project.



Partners involved



CalipsoPlus © 2019

Fig. The CALIPSOplus common portal landing page installed at DESY - <https://calipsoplus.desy.de/>

PSI - Status of Portal

Talk by *Dan Webster*

The CALIPSOplus portal was deployed on OpenShift 4.x, running on vSphere – the cluster comprised three masters, and two worker nodes. The deployment itself was orchestrated via GitLab CI/CD, adding a DevOps/GitOps element to the pipeline. (Fig.2)



Sessions from proposal 20180101

test 1_D
this is a description...
BL11

	Code	Start	End
Launch! ▾	S0011	10/10/2018 10:10	10/11/2018 18:10
base_image	S0012	10/10/2018 11:10	10/12/2018 18:10
base_jupyter	S0013	10/11/2018 12:10	10/13/2018 18:10
base_image_ubuntu	S0014	10/12/2018 13:10	10/14/2018 18:10
link_google			
ptychoshelves			

Fig.2: example of CALIPSOplus portal at PSI

The use case was cSAXS / Ptychography, based on the in-house MATLAB package “ptychoshelves”, including a custom C-based solver. This was wrapped in a Linux container compatible with the remote access methods offered by the portal (RDP/VNC), and this container was then integrated into the portal catalogue, for rapid instantiation. (Fig.3)

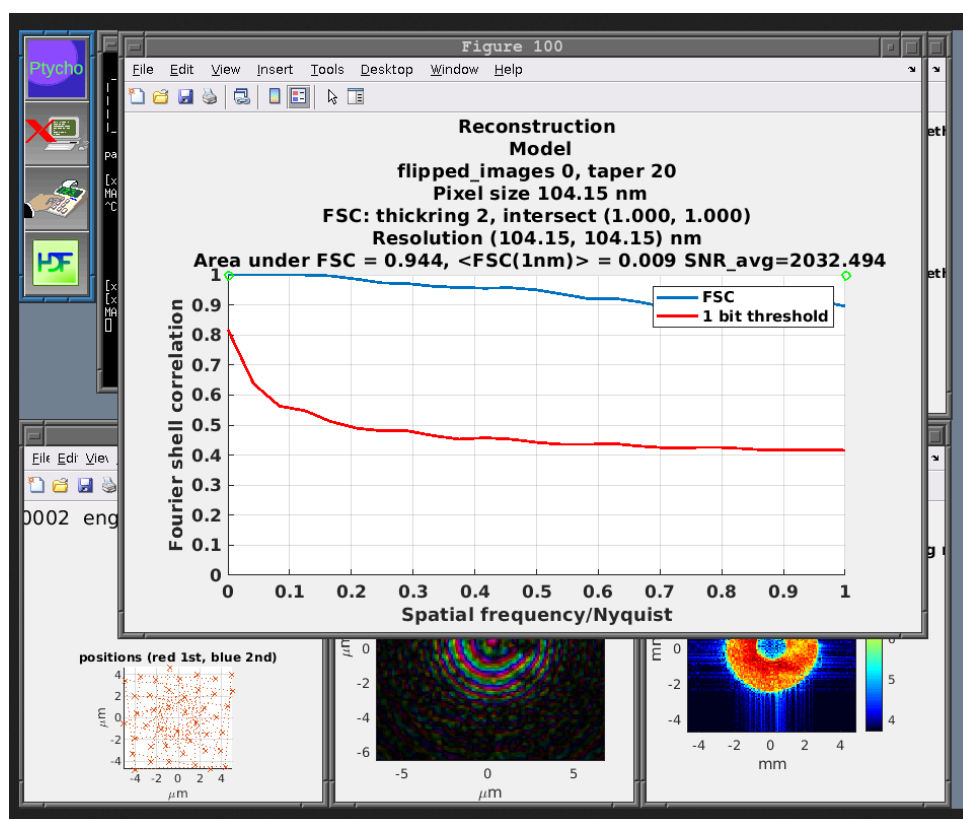


Fig.3: Container running ptychoshelves

ELETTRA - Status of Portal

Talk by Dario Parmisano (ELETTRA)

The overall feedback from the users of the synchrotron Elettra and its free electron laser FERMI was positive. Having said that, selected users were informed in advance that the local deployment was in beta and not in production level. The deployed hardware was of limited performance, the available applications were hand-picked (mostly Jupyter and PyMCA) and the authentication and data access were not fully implemented. Users could benefit from interacting directly through a web browser. This is of particular relevance during the COVID-19 pandemic and the increased requirements for remote working.

An issue raised was the one of limited desktop surface in case of remote desktop (VNC) to beamline computers with multi-monitor setups. A much appreciated feature and perceived advantage for the DaaS approach was using remote hardware resources, common software deployments and less reliance on VPN and multiple gateways. This feedback is encouraging for other ongoing projects in which Elettra is participating (ExPaNDS, for example) and that build on the DaaS experience and blueprint of JRA2.

User Feedback

Feedback from users was positive and demonstrated the need for such services as documented in deliverable D24.5⁷. The JRA2 has come at just the right time to help photon sources face up to the massive move towards remote experiments following the outbreak of the COVID-19 pandemic.

COVID-19 Pandemic

The pandemic has changed the way photon sources and many other aspects of our lives operate. With the wide travel ban, it is currently not possible for users to travel to do experiments. All institutes are obliged to provide a remote experiment service. Due to the remote aspect of the users, they need remote services for accessing and analysing their data. The JRA2 Data Analysis as a Service portal is a perfect fit for this. The experience gained by developing the portal and in some cases the remote services, which were put into production have proven to be extremely useful and timely to address the new modes of operating due to COVID-19.

Conclusion

As a general response, CALIPSOplus JRA2 has been very successful in bringing together sites to collaborate and share a prototype portal for providing Data Analysis as a Service. The results are very promising and enabled the issues to be identified and addressed. The project has produced a blueprint

⁷ http://www.calipsoplus.eu/wp-content/uploads/2019/05/190506_Deliverable_D24.5_b.pdf

design as well as use cases and recipes for containerising data analysis packages. The documents and source code produced will be useful for the future projects.

The Jupyter service for remote data analysis is being generalised at most sites and other services, such as metadata catalogues, containers and virtual machines for data analysis are either in production or under test. Umbrellald has been made sustainable and will enable sites to connect via Umbrellald to EOSC services with eduTEAMS.

The JRA2 identified some difficulties in providing DAAS at photon sources. The main difficulty encountered in providing DAAS services in production is the lack of a clear data analysis policy at all the sites, as was shown by the two surveys that were prepared and shared with all partner facilities in 2019 and 2020. The photon sources have not yet clearly defined the level of service they want to provide to users for data analysis. The results is that most services are provided without a well defined level of service users can expect and most data analysis support activities are carried out by beamline scientists in their spare time without dedicated staff for these tasks. A related difficulty and bottleneck is the lack of manpower dedicated to developing and operating the infrastructure required for data analysis as a service in production. During the JRA2 collaboration and at the workshop it was clear that lack of manpower applies to most areas of computing at photon sources.

The workshop highlighted the potential of and strong need for further collaborative developments as part of the PaNOSC and ExPaNDS projects and beyond (LEAPS and EOSC). The next step for the CALIPSOplus partners is to participate in the activities of these two projects with the goal of offering remote data analysis services during and after the COVID-19 pandemic.