Ref. Ares(2022)9011213 - 30/12/2022

ExPaNDS
European Open Science Cloud Photon and Neutron Data Services

D4.5 Deployment of EOSC Analysis Services for the EOSC within the EOSC Hub

Document Control Information

Settings	Value
Document Identifier:	D4.5
Project Title:	ExPaNDS
Work Package:	WP4
Document Author(s):	S. Schoen (DESY), M. Ounsy (SOLEIL), P. Millar (DESY), K. Pozsa (PSI), Z. Matej (MAX IV Laboratory), P. Fuhrmann (DESY), C. Reynolds (Diamond), N. Soler (Alba), O. Knodel (HZDR), U. Konrad (HZDR), F. Lang (UKRI), A. Manzi (EGI), M. Schuh (DESY), A. Barty (DESY)
Document Reviewer(s):	S. Servan (DESY), D. Spruce (MAX IV), S. Lambert (UKRI)
Responsible Partner:	DESY
Doc. Issue:	0.1
Dissemination level:	Public
Date:	30/12/2022

Abstract

We present the final deliverable in ExPaNDS related to the deployment of analysis services into EOSC, the establishment of portable, facility independent data analysis services for five real-world use cases, training material for operation of services, and harmonisation with related past and future initiatives such as PaNOSC and NFDI. Specifically, this document documents deliverable D4.5 of the Horizon 2020 ExPaNDS project.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Table of contents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Document log

Issue	Date	History of changes
1	12/12/2022	Initial version
2	15/12/2022	Review version
3	28/12/2022	Reviewed and finalised by Coordinator (PF)

Acronym/Abbreviation	Definition
API	Application Programming Interface
CI/CD	Continuous Integration/Continuous Delivery
CrystFEL	suite of programs for processing diffraction data acquired "serially" in a "snapshot" manner
DESY	Deutsches Elektronen-Synchrotron
DLS	Diamond Light Source
ELBE	Center for High-Power Radiation Sources at HZDR
EOSC	European Open Science Cloud
ExPaNDS	European Open Science Cloud Photon and Neutron Data Service
HZDR	Helmholtz-Zentrum Dresden-Rossendorf
IDAaaS	Intelligent Data Analysis as a Service
ISIS	Neutron and Muon Source, UK
LEAPS	League of European Accelerator-based Photon Sources
LENS	League of advanced European Neutron Sources
NFDI	Nationale Forschungsdateninfrastruktur
NRI	National Research Infrastructure
PaN	Photon and Neutron
PaNOSC	*Photon* and Neutron Open Science Cloud
PSI	Paul Scherrer Institut
SANS	Small Angle Neutron Scattering
SLA	Service Level Agreement
TELBE	Terahertz facility @ ELBE
VISA	Virtual Infrastructure for Scientific Analysis
WP	Work Package

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Executive Summary

We describe data analysis services deployed into the EOSC as a part of the ExPaNDS project. We selected **five real-world use cases** from the scientific community which are representative of typical analysis workflows performed by the user community and deployed them as portable workflows into the EOSC environment.

This document builds on the foundations and directions set earlier in the project, including the selection of representative data sets and workflows (D4.2[1]), development of a continuous testing framework (D4.3[2]), and deployment of prototype services (D4.4[3]).

The five focus use cases we describe in this document are:

1. 4D Full field tomography
2. Terahertz Spectroscopy
3. Electron scanning diffraction imaging
4. Neutron Small Angle Scattering
5. Serial crystallography

These use cases are made portable across research infrastructures by either creating portable Jupyter notebooks or bundling software into containers for execution at any facility. The methods for sharing are designed to be representative examples that can be re-used for sharing other workflows outside the scope of this document. Descriptions of how this is achieved are briefly described here without going into too much technical detail. Further details have been provided to ExPaNDS WP5 for inclusion in the ExPaNDS training material for use for additional use cases by other researchers. In return, for four out of the five use cases we provide a sketch of the workflow from our WP5 training platform.

These use cases are deployed into the EOSC by deploying existing NRI data analysis infrastructure into the EOSC so that the data analysis services can be located through the EOSC marketplace portal. The services ExPaNDS has deployed into the EOSC marketplace are:

1. PSI - PSI Remote Desktop Service (NX)[4]
2. Diamond - Diamond Remote Desktop (NX)[5]
3. UKRI - ISIS Data Analysis as a Service (NX)[6]
4. MAX IV Laboratory[7] -

[1] https://doi.org/10.5281/zenodo.4558708
[2] https://doi.org/10.5281/zenodo.5718671
[3] https://doi.org/10.5281/zenodo.6305000
[4] https://marketplace.eosc-portal.eu/services/remote-desktop-service
[5] https://marketplace.eosc-portal.eu/services/diamond-remote-desktop
[6] https://marketplace.eosc-portal.eu/services/isis-data-analysis-as-a-service
[7] As the legal entity of MAX IV is Lund University, registering MAX IV as a EOSC Service provider requires a slightly more complex procedure, which our partners are still working on with the help of EGI, see 5.4.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

5. ALBA - Registered as a service provider (services not yet registered)
6. SOLEIL - Registered as a service provider (services not yet registered)
7. HZDR - PaN-training catalogue service[8]
8. DESY - VISA EOSC service[9]

Access to these services is generally available after completion of a user service provision agreement and/or access agreement, the details of which vary depending on the facility.

Finally, we anchor these activities into current and future activities, not only with regard to EOSC-related activities such as EOSC-Future and EOSC-Synergy, but also initiatives in the LEAPS/LENS community of European PaN facilities and in national initiatives like the German NFDI[10].

1. Introduction

This document is the final deliverable in work package 4 of the Horizon 2020 ExPaNDS project. It builds on the foundations and directions set earlier in the project, including the selection of representative data sets and workflows (D4.2[11]), development of a continuous testing framework (D4.3[12]), and deployment of prototype services (D4.4[13]).

As stated in the proposal: The purpose of WP4 is to provide Photon and Neutron users with the ability to find and run analysis workflows against the EOSC aligned data services. This will consolidate the reusability of the workflows, which is key to fully developing the FAIR principles for Photon and Neutron data. Integrating these existing data analysis services with EOSC services such as browser driven remote desktops and Jupyter analysis services helps to ensure that they are Accessible and Interoperable. Finally, standardising against these EOSC analysis services greatly increases the reusability and reproducibility of the underlying algorithms used to understand the FAIR data. For reproducibility, and as a basis for further continuous development, WP4 selected reference Photon and Neutron data sets.

Deliverable D4.4 described fully five challenging data analysis pipelines implementation as remote data analysis services taking into account the needs of representative scientific communities while supporting the diversity of the institutions' existing computing infrastructures. Each of these five use cases has been developed by a different facility and, in order to achieve sustainable solutions, all project partners based their implementations on concepts and environments which are typical for state-of-the-art open-source projects. GitHub, GitLab and CI/CD or on already available core services in the EOSC Marketplace.

The coordination of efforts by ExPaNDS helped its partner-facilities to provide PaN scientists in Europe with standardised data analysis pipelines available through the EOSC portal and

[8] https://marketplace.eosc-portal.eu/services/pan-training-catalogue
[9] https://visa.desy.de
[10] https://nfdi.de
[11] https://doi.org/10.5281/zenodo.4558708
[12] https://doi.org/10.5281/zenodo.5718671
[13] https://doi.org/10.5281/zenodo.6305000

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

to conceive an overall modular architecture (see Deliverable D1.6[14]) around generic APIs. They also opened the road to enlarge the audience of facilities using the VISA portal.

Chapter 2 provides an explicit description of the user relevance of these activities. In chapter 3 we describe our strategy for integration of existing analysis services into EOSC. Chapter 4 details the deployment of selected reference analysis services as specific use cases, with Chapter 5 providing reference to linked training material coordinated with WP5. Additional details of analysis services deployed into EOSC are provided in chapter 6. In Chapter 7 we describe how ExPaNDS is harmonised with other related initiatives. Although this is a technical deliverable, we briefly discuss the necessity and status of common service agreements and we are elaborating on interoperability and access policies (Sections 8 and 9 respectively)

2. Importance for the PaN end user

Before the advent of the FAIR principles, data processing and analysis was carried out locally by users: a research team would transfer the raw data acquired at a given facility to their home institution, together with a minimal set of metadata often simply captured using pen and paper. Reprocessing and analysis would then be performed at the home institution, typically on tabletop PCs or using the facility local HPC if present. The management of the raw data and any derived data associated with it was left to the group that had performed the measurement and was often not considered as a priority once the corresponding scientific publication had been released. In the last decade however, some techniques like tomography and macromolecular crystallography (MX) have witnessed a massive increase in the volume of data produced during a typical shift, rendering data transfer and analysis very tedious. Data analysis as a service (DAaaS) platforms in PaN facilities and their EOSC-centralised access emerged as a good solution to avoid such transfer but also convey additional advantages if well designed:

- **Reproducibility**: richly annotated data processing workflows allow (re)users to easily understand the crucial aspects and parameters affecting the underlying scientific question. This could be particularly important in the case of publication review.

- **Provenance**: containers and Jupyter notebook can allow (re)users to track which software and versions were used, together as the architecture of the data-processing/analysis pipeline.

- **Interoperability**: scientific projects nowadays tend to go towards multimodality, integrating the output of different types of experiments performed on different instruments. There is no reason for users to stick to one particular facility so that datasets related to the same scientific project are often located in different PaN facilities. Centralising the access to the DAaaS platforms of these facilities via the EOSC-marketplace is therefore releasing the burden on users of having to remember different access points and procedures. Hopefully, the concept of portable workflow will allow an easier integration of datasets collected at different places in the future.

[14] https://doi.org/10.5281/zenodo.6958045

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

3. Strategy for the deployment of national RI analysis services within the EOSC

The experimental data from PaN research experiments are reaching sizes nowadays that are becoming ever harder to handle. The volumes of data produced makes it impractical for data to be exported and handed to the user on an external hard-drive or other physical media. Moreover, the transfer of data to the user's home institution through normal internet connections, not to mention the subsequent storage and archiving of data, becomes uneconomic and unsustainable. For all these reasons, it is necessary to move to a situation where analysis as a service is available where the experimental data of the user was originally recorded and stored.
Within the ExPaNDS project, multiple analysis pipelines have been made available by the national research partners within the EOSC. The pipelines give the user the means to perform analysis on their experimental data. Many techniques are similar between facilities, thus the analysis workflow should be portable across facilities, independent of the details of underlying compute infrastructure.

A key point of effort in WP4 has been making workflows portable between facilities. To achieve this we used a two-pronged approach:

1. Workflows executable in Jupyter notebooks are by design largely self-contained scripts. These can be deployed as code to be imported into and executed within the Jupyter service. The key here is to make Jupyter services accessible via EOSC with authenticated access to storage and compute infrastructure on the back end. This can be achieved either through a web interface, via virtual machines, or using remote desktop services.
2. Non-Jupyter workflows are bundled into containers to make them executable on web-based infrastructure. Here, the challenge is the requirement for authenticated access to large data volumes stored outside the container, and the necessity to write data back to the file system. The process by which this is achieved was described in previous deliverables. Here, the key to deployment of workflows in EOSC is (a) have access to an appropriately built container environment, and (b) access to either a virtual machine or remote desktop service via EOSC.

The use cases presented here cover both cases, for real-world software in use by researchers at the facilities. Typically we connect a remote desktop or virtual machine service to the EOSC marketplace that provides authenticated access to compute and storage infrastructure on the backend. Within these environments, the RI owned workflow can be applied on the data. However the environments are not restricted to a specific workflow. To allow interoperability of external pipelines within the specific remote desktop environment, the workflows themselves are developed in a generic way, based on either jupyter notebooks or docker containers. Figure 1 briefly depicts a high level view of the relationship between the local analysis services and the EOSC marketplace.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

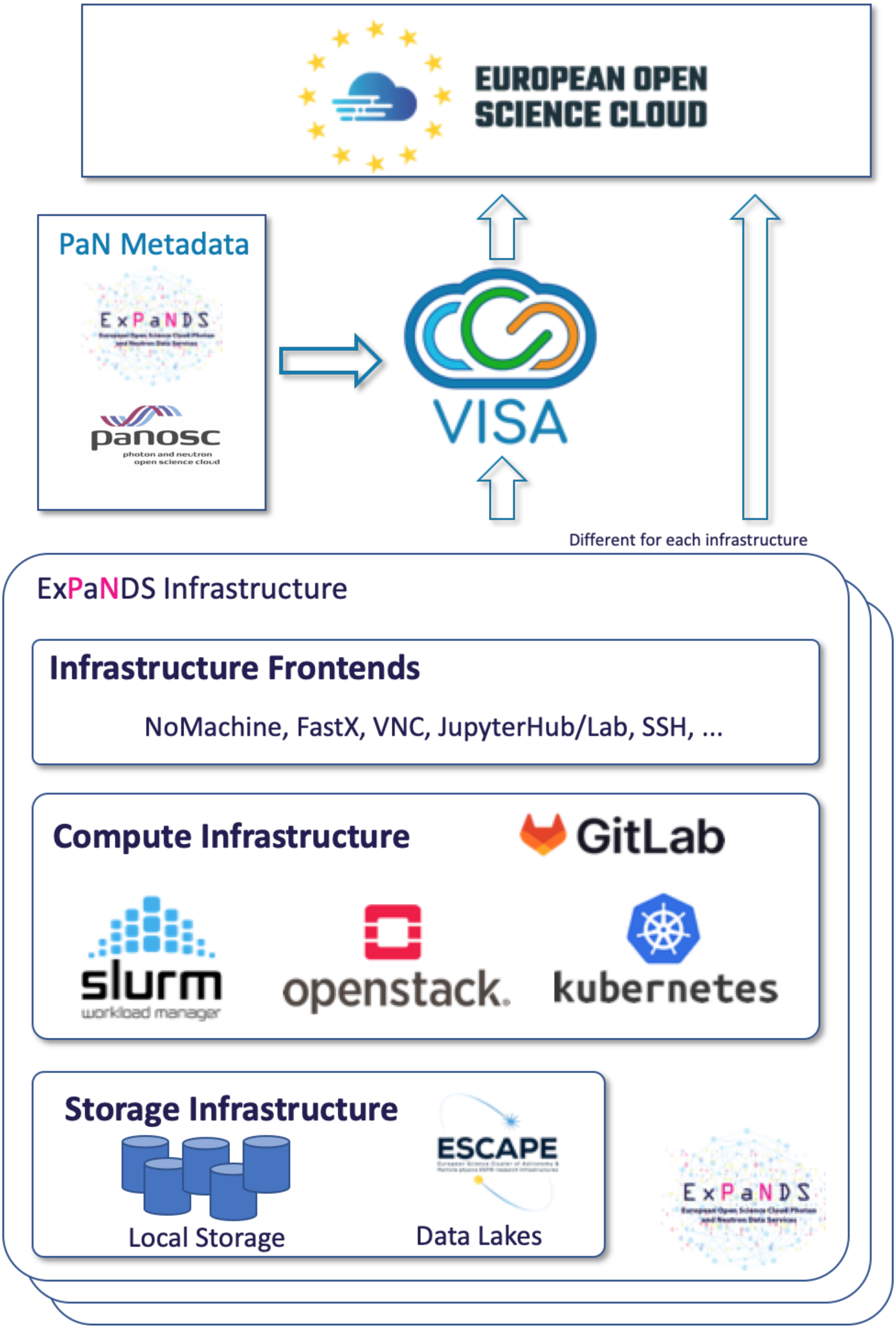


Figure 1: Bundling of software in containers for execution on horizontal infrastructures and associated technical challenges.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

4. Real world reference use cases deployed in EOSC

In this section we describe how the reference use cases are rendered by our PaN training platform and how they can be accessed through the EOSC marketplace.

Technique	Partner	PaN Training	EOSC Portal access
4D Full field tomography	PSI	✅[15]	✅ PSI, DESY, Diamond (See 4.1)
Terahertz Spectroscopy	HZDR	✅[16]	✅ Via PaN Training Portal[17]
Electron scanning diffraction imaging	DIAMOND	❌	✅ Diamond Remote Desktop Service[18]
Small-angle neutron scattering	UKRI	✅[19]	✅ ISIS Data Analysis as a Service, DAaaS[20]
Serial crystallography	DESY	✅[21]	✅ DESY VISA Portal Service[22]

4.1. 4D Full field tomography (PSI)

The analysis itself is contained in the Jupyter notebooks[23]. These notebooks have been demonstrated to run on the following services registered in the EOSC marketplace:

- PSI Remote Desktop Service - EOSC Marketplace (eosc-portal.eu)
- PAN Notebook - EOSC Marketplace (eosc-portal.eu)
- DLS Remote Desktop Service - EOSC Marketplace (eosc-portal.eu)

Other EOSC services that support the execution of Jupyter notebooks and have the appropriate environment setup will support the execution of this analysis pipeline. The GitHub repository contains information about the environment necessary for the execution of the code contained in the notebooks.

[15] https://pan-training.eu/workflows/backup-fork-of-full-field-tomography-at-psi-wip
[16] https://pan-training.eu/workflows/telbe
[17] https://marketplace.eosc-portal.eu/services/pan-training-catalogue?q=PaN-Training+Catalogue
[18] https://marketplace.eosc-portal.eu/services/diamond-remote-desktop?q=Diamond+Remote+Desktop
[19] https://pan-training.eu/workflows/small-angle-neutron-scattering-sans
[20] https://marketplace.eosc-portal.eu/services/isis-data-analysis-as-a-service
[21] https://pan-training.eu/workflows/containerized-serial-crystallography-visa-workflow
[22] https://marketplace.eosc-portal.eu/services/desy-visa-portal
[23] https://github.com/paulscherrerinstitute/tomorec

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Accessing the workflow

Guidance on how to execute the pipeline using [PSI Remote Desktop Service - EOSC Marketplace][24]:

- Follow the instructions in the "Manual" to connect to the remote desktop
  NOTE: you need a PSI account to be able to access the PSI infrastructure
- Log into the PSI JupyterHub service
- Follow the detailed instructions in the github repo (the below is a summary):
  - Install the python kernel from the conda environment files (use the non-gpu version)
  - Bring your own data (or use the [PSI reference dataset][25]
  - Run the analysis

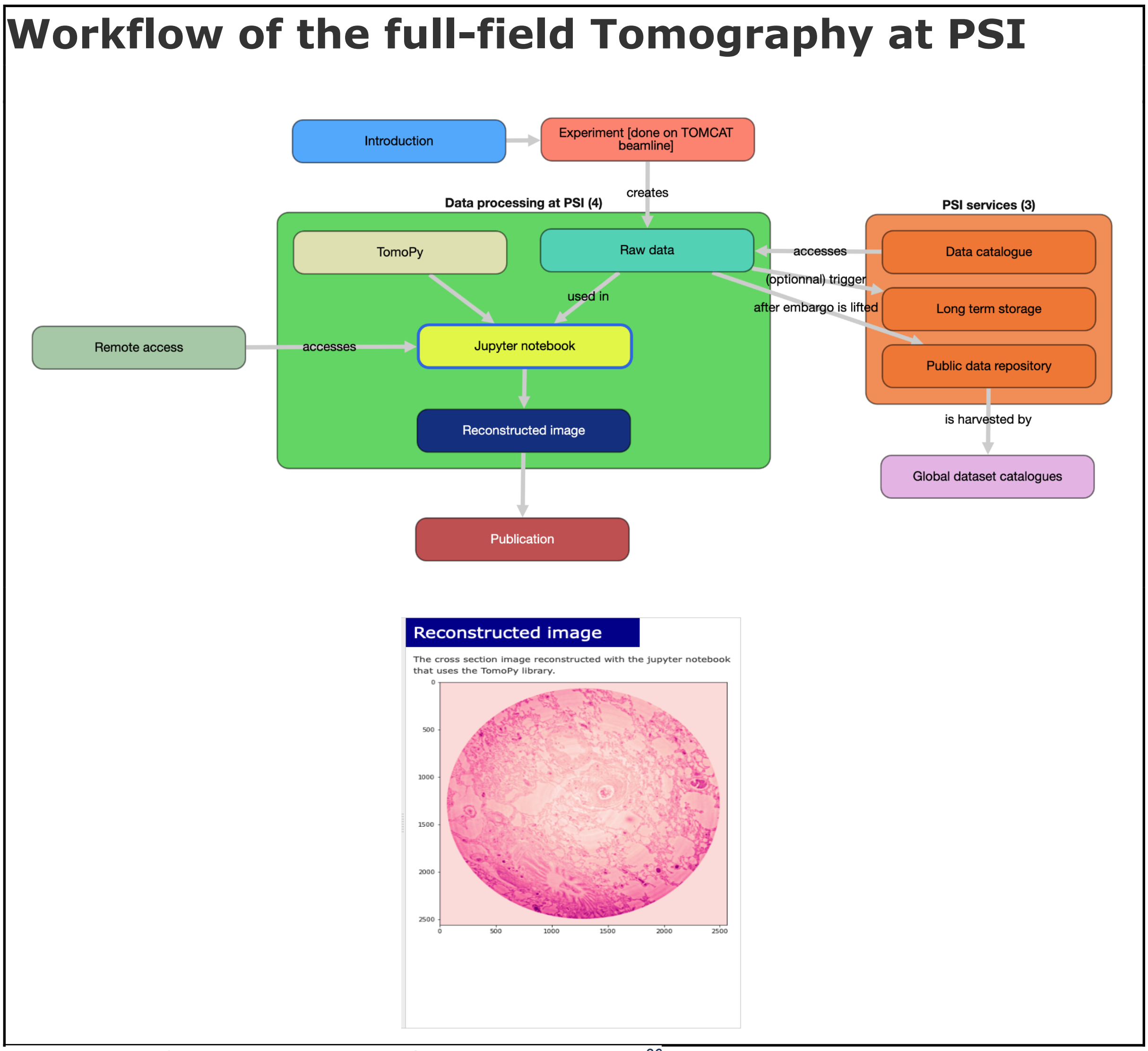


Figure 2: Full-field Tomography at PSI at the PaN training[26]

[24] https://marketplace.eosc-portal.eu/services/remote-desktop-service?q=PSI+Remote+Desktop+Service
[25] https://doi.psi.ch/detail/10.16907/d699e1f7-e822-4396-8c64-34ed405f07b7
[26] https://pan-training.eu/workflows/backup-fork-of-full-field-tomography-at-psi-wip

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

4.2. Terahertz Spectroscopy (HZDR)

The example use case from HZDR is the Terahertz (THz) Spectroscopy workflow. The data used in the workflow is generated at the radiation source ELBE (Electron Linac for beams with high Brilliance and low Emittance) at the Helmholtz Centre Dresden Rossendorf (HZDR), which can produce several kinds of secondary radiations. The THz radiation is available at the TELBE beamline.

The TELBE spectroscopy workflow is based on a Jupyter notebook, which provides the possibility to handle the raw data from the TELBE experiment used in a recent Nature publication. All data and software products are available online and the analysis workflow can also be accessed through EOSC services with any academic/social account. The HZDR computing infrastructure is at the moment only available for users from institutions within the Helmholtz AAI authorization service.

The main components and resources required for the TELBE spectroscopy workflow are:

- The analysis software (Jupyter notebook) is available on GitHub[27]
- The dataset is published in the HZDR's data repository RODARE[28]
- The analysis can be performed at the HZDR datacenter using a Jupyter notebook service[29]
- All these steps are introduced in the form of a schematic training workflow on our PaN-training catalogue[30] (registered as EOSC service)

In addition to HZDR's data repository the metadata-catalogue SciCat[31] is under development and will provide the experiment-specific metadata publicly findable in addition to the pure data publication.

Accessing the workflow

To enable the Terahertz Spectroscopy workflow of the TELBE beamline, components of the infrastructure at HZDR have been registered as EOSC services. Our smaller analysis workflows are collected in PaN training and thereby findable using the search for training materials in the EOSC Portal[32]. Through this we have the training catalogue as entry point from the EOSC marketplace to find, access, comprehend our workflows:

- Enter the EOSC marketplace and search for "PaN training".
- Access the "PaN Training Catalogue" and search "TELBE".
- Go to "Training Workflow on Terahertz Spectroscopy at TELBE" and further to the training workflow TELBE Terahertz Spectroscopy workflow.

[27] https://github.com/hzdr/TELBE-raw-data-evaluation
[28] https://doi.org/10.14278/rodare.276
[29] https://hifis-gpu.hzdr.de/service/
[30] https://pan-training.eu/workflows/telbe
[31] http://scicat.hzdr.de
[32] Not all of our materials are currently available in the new catalogue search portal.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

- The schematic workflow starts with an introduction followed by references to the analysis software, dataset of RAW data, and the final scientific publication.
- In the training workflow the step "Perform analysis service" provides a direct entry point in two jupyter notebook instances (the HZDR Jupyter notebook service[33] and EGI Binder[34]). The spectroscopy workflow notebook "sorting_binning.ipynb" is automatically loaded into the notebooks.
- Select for instance EGI Binder and the spectroscopy workflow will be deployed on EGI resources. The dataset is downloaded and the analysis should look similar to the following:

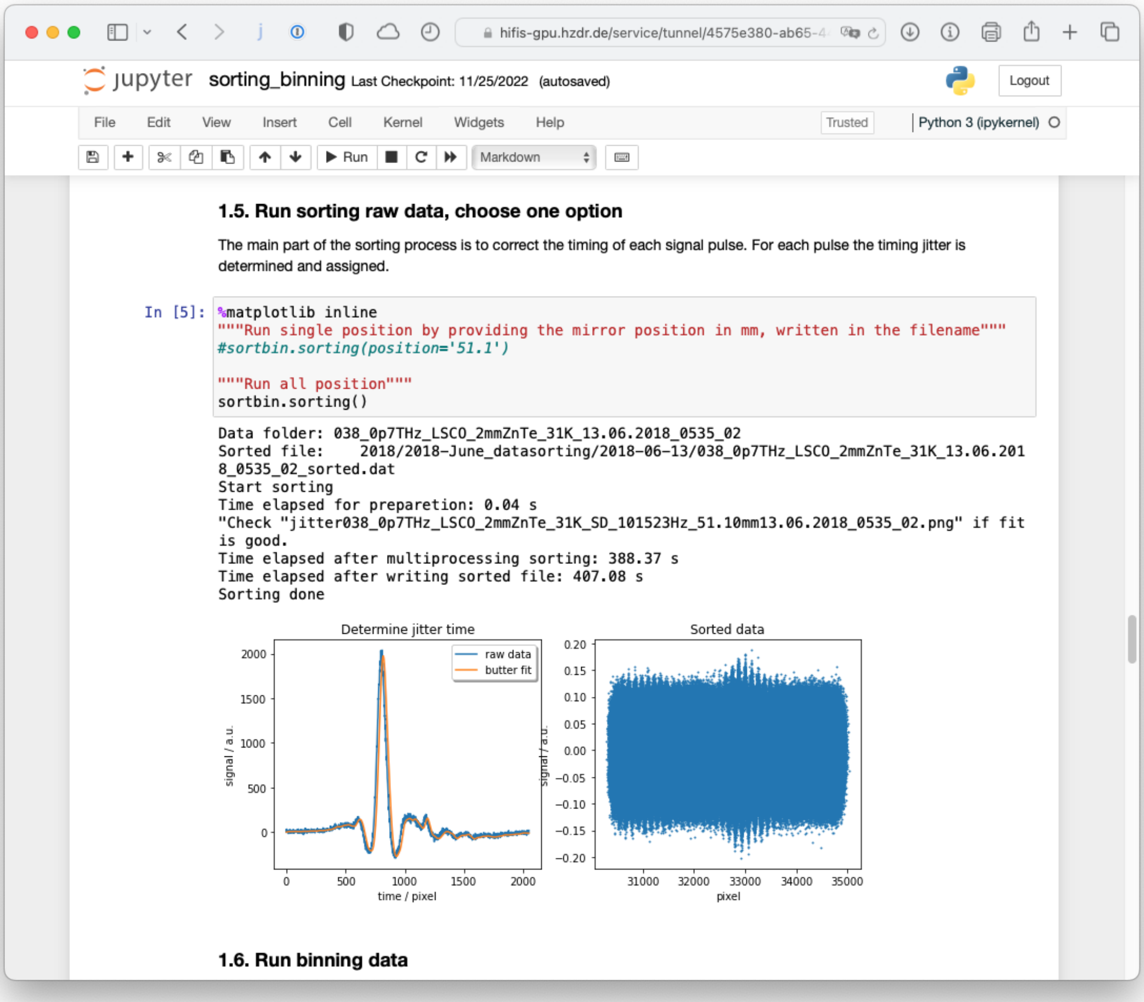


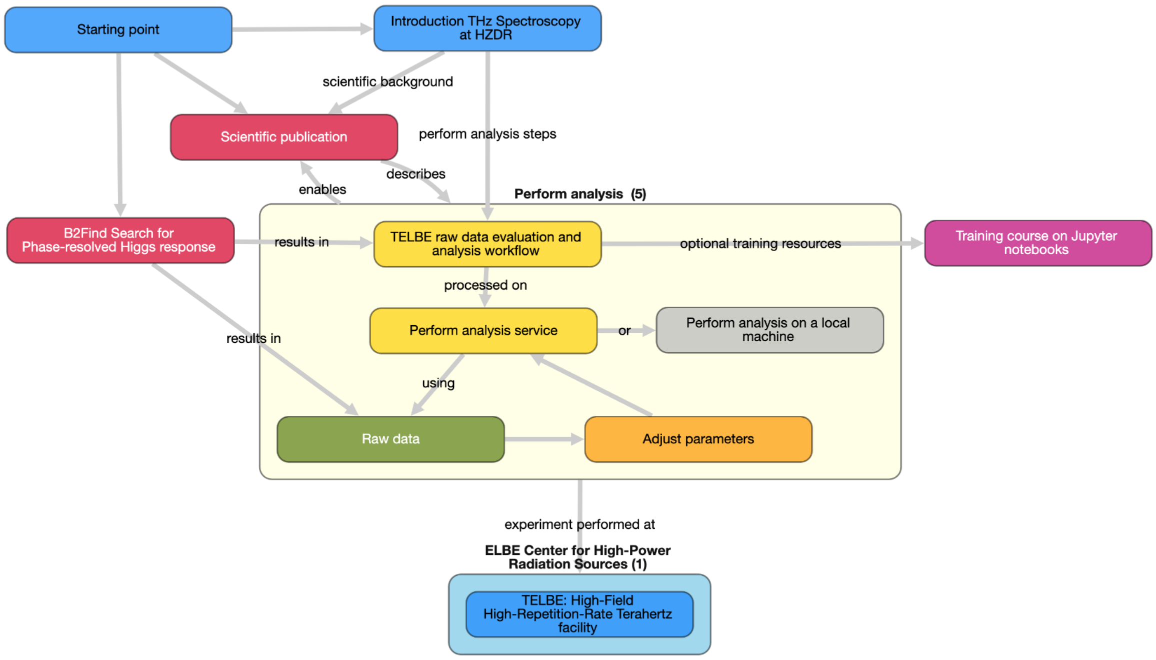
Figure 3: Analysis Step of the TELBE spectroscopy notebook on the HZDR jupyter notebook service[35]

- The function of the notebook cells and the important parameters are described in the notebook itself and all steps and parameters can be changed and evaluated.

[33] https://hifis-gpu.hzdr.de
[34] https://binder.notebooks.egi.eu
[35] *https://hifis-gpu.hzdr.de/service/*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Workflow of the TELBE Terahertz Spectroscopy



The radiation source ELBE (**E**lectron **L**inac for beams with high **B**rilliance and low **E**mittance) at the Helmholtz Centre Dresden Rossendorf (HZDR) can produce several kinds of secondary radiations. THz radiation is one of them and can be used with a typical pulse frequency of 100 kHz as a stimulation source for elementary low-energy degrees of freedom in matter. It was used in the \"Phase-resolved Higgs response in superconducting cuprates\" publication (DOI: 10.1038/s41467-020-15613-1). The raw data for this publication can be accessed on RODARE (DOI: 10.14278/rodare.277) and will be used to reproduce the figures from the publication.

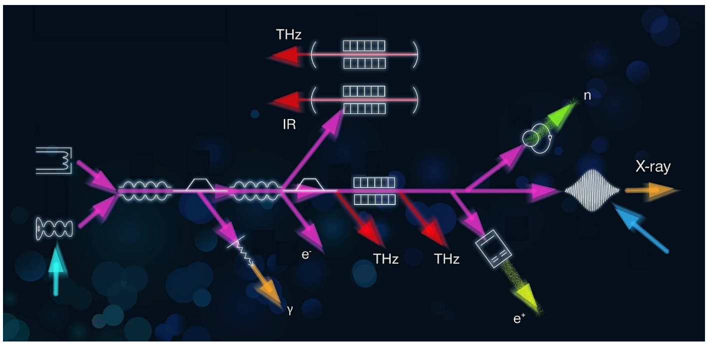


Figure 4: TELBE Terahertz Spectroscopy workflow[36]

[36] https://pan-training.eu/workflows/telbe

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

4.3. Electron scanning diffraction imaging (Diamond)

The reference data set and Jupyter notebook public access. They should be downloaded using a web browser *after* logging in to the Diamond Remote Desktop Service[37].

Accessing the workflow

To execute the Jupyter notebook, one can use the Diamond Remote Desktop Service registered on EOSC Marketplace. This can be located using the search, or following this URL[38].

The instructions describe how to access Diamond systems using the NX remote desktop tool. There is a reference video[39] in the service marketing section that shows users how to do this. Users will need a valid Federal ID (FedID) from Diamond to access this resource. FedIDs are created when visits are booked, so you must be on an ongoing or historical visit.

Once logged into the remote desktop tool, open a browser and load the DIAMOND Jupyterhub[40]. Select a notebook flavour (small should be sufficient) and click to spawn the notebook. Once the notebook has been spawned, open the notebook provided by the reference data set download. You can browse for the notebook in the default file browser that starts when your notebook is spawned.

Once the notebook is open, select the kernel “Python 3.7 - EPSIC [DLS Conda]”. Edit the file input path defined in the variable sID to point to the location you downloaded the reference data set, e.g. sID='/home/fedid/MOF_data.hspy'.

Run the cells in turn. You should see plots similar to figure 5.

[37] https://dlsltd.sharepoint.com/:u:/s/GRA0046-PRV/ER6elCBerHxDlCfYClBeUFsBakNbDsREMRobNW9VvrUXaQ?e=HYEm24

[38] https://marketplace.eosc-portal.eu/services/diamond-remote-desktop?q=Diamond+Remote+Desktop

[39] https://www.youtube.com/watch?v=Y6FHWuy0Oug

[40] https://jupyterhub.diamond.ac.uk

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857641.

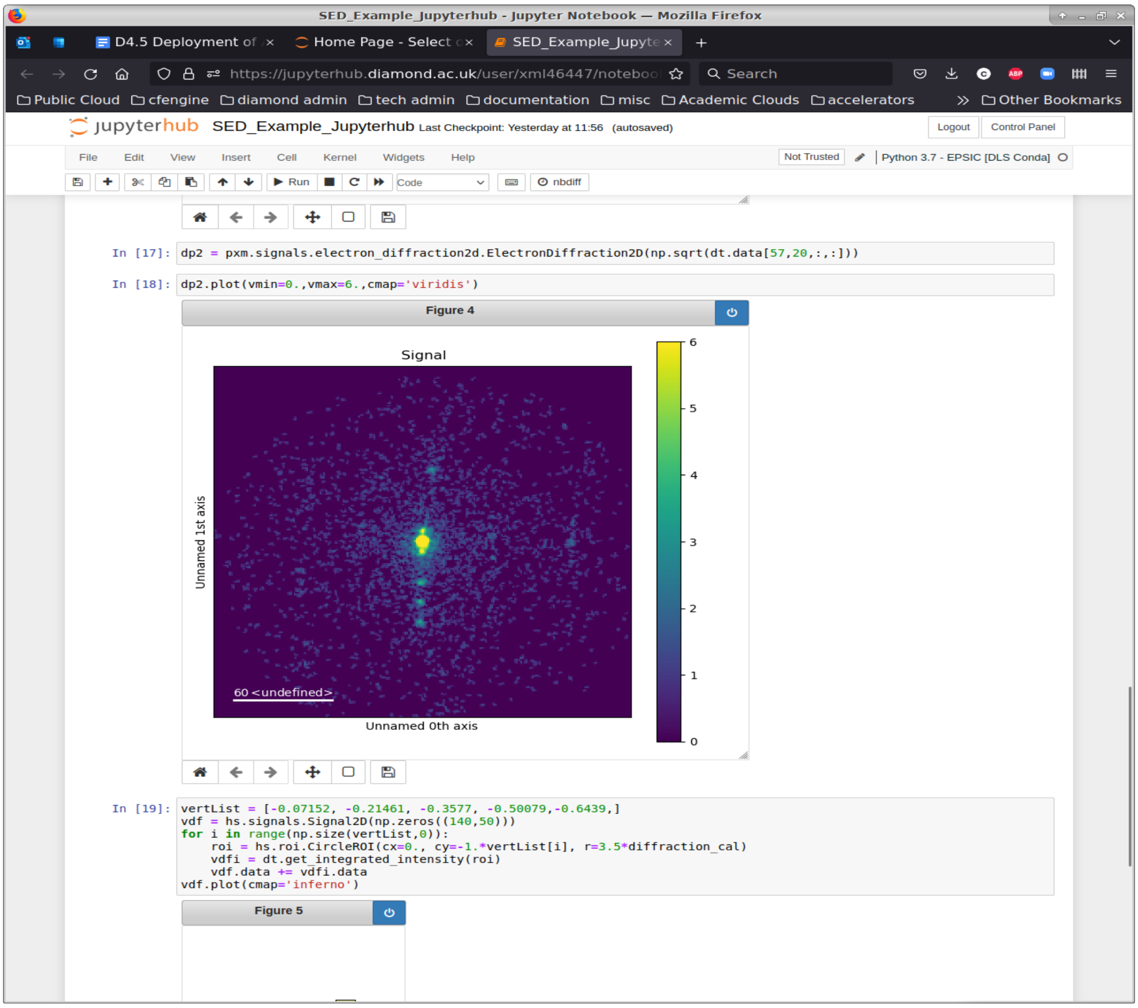


Figure 5: The electron scanning diffraction imaging notebook running on DLS Jupyterhub

4.4. Small-angle neutron scattering (UKRI)

The Small-angle neutron scattering (SANS) analysis pipeline is fully public, together with the associated raw and metadata, and can be run out-of-the-box on our local remote data analysis service registered in the EOSC marketplace. It has also been added as an example workflow to the PaN training catalogue.

- Original publication occurred as part of D4.2[41]
- The raw and metadata, together with an annotated Jupyter notebook to run the analysis pipeline and reference output, and documentation about the underlying experiment and associated scientific publication, are published at GitHub[42].
- The analysis pipeline/workflow and its scientific context is documented in detail and published as an example workflow in the PaN training catalogue[43].

[41] https://doi.org/10.5281/zenodo.4558708
[42] https://github.com/DAaaS-reference-data/SANS
[43] https://pan-training.eu/workflows/small-angle-neutron-scattering-sans

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

- The analysis pipeline was prepared in such a way that it can be run on our local remote analysis service without any changes needed, and at other facilities with minimal dependencies and changes required.
- The local remote analysis service (called ISIS Data Analysis as a Service) has been registered in the EOSC marketplace[44].

Accessing the workflow

How to use the ISIS Data Analysis as a Service (IDAaaS) to run the SANS workflow using the reference data set:

- Go to the EOSC marketplace and search for "ISIS" or "ISIS Data Analysis as a Service" (or similar) and click on the "webpage" link of the service. This will forward you to the service.
- Click on the "Login" button, which will forward you to the STFC User Office. Either create an STFC User Office account on this website or login with your existing account's credentials. Once logged in, you should be brought back to the IDAaaS webpage.
- Create a remote "workspace" by following

  `Workspaces > New Workspace >ALF >`
  `Excitations Powder > Create Workspace`

  A number of other workspace types would also work, but for the reference data set this one will do.
- Open the workspace by clicking "Tab" or "Window" when hovering over it with the mouse.
- Inside the workspace open a terminal: click on the start menu

  `Applications > System > Terminal.`

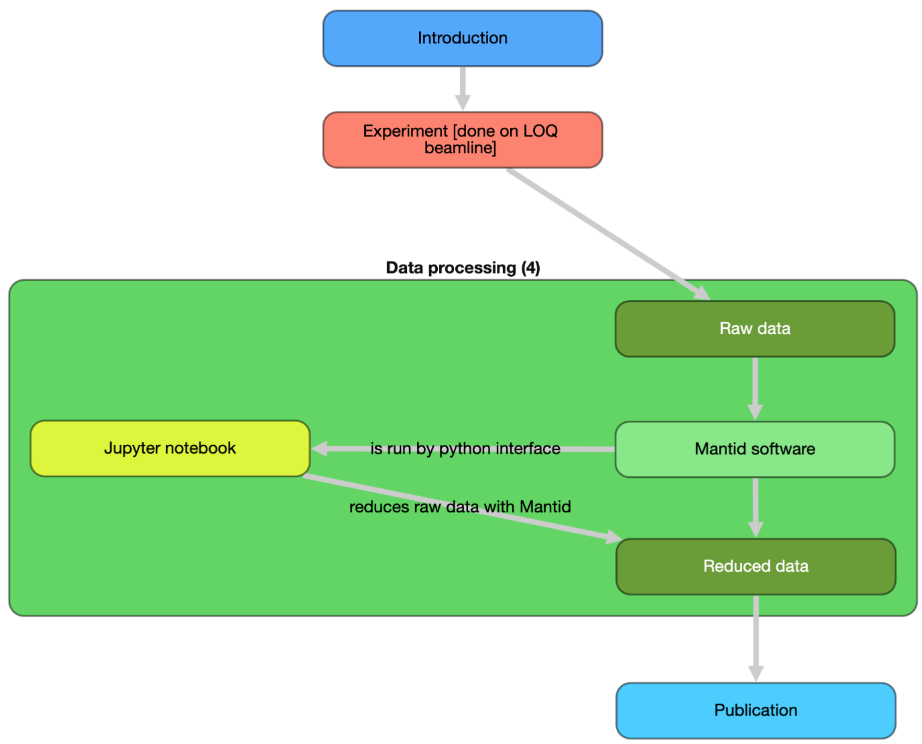
  Download the reference data set by running the following inside the terminal:

  `git clonehttps://github.com/DAaaS-reference-data/SANS.git`
- Open Jupyter by clicking on the start menu: Applications>Software>Jupyter notebook. Navigate to the "ExPaNDS_reduction.ipynb" inside the downloaded directory of the reference data set and open this notebook.
- Check that the Python kernel has "mantid" in its name. If not, change the kernel by clicking Kernel>Change kernel>mantid-*, where any one of the mantid kernels can be chosen.
- Run the cells in the notebook, by selecting each cell and either clicking "Run" or pressing the shift+return keys simultaneously. Running the whole notebook should take about one minute.
- The annotated workflow is also available within the PaN training catalogue[45].

[44] https://marketplace.eosc-portal.eu/services/isis-data-analysis-as-a-service
[45] https://pan-training.eu/workflows/small-angle-neutron-scattering-sans

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Workflow of Small Angle Neutron Scattering



This use case employs the data reduction software Mantid ,the most used data reduction software at ISIS and adopted by a number of other facilities. Furthermore, it demonstrates a typical workflow to turn raw data into reduced data, which can then be used by subsequent analysis.The main purpose of this use case is to highlight a bread-and-butter analysis workflow for analysing neutron data using the Mantid software. At a high level this workflow is similar for analysing data from a whole range of other neutron experimental techniques and is the most common analysis workflow at the ISIS Neutron and Source.

Figure 6: Small-Angle Neutron Scattering Workflow[46]

46 https://pan-training.eu/workflows/small-angle-neutron-scattering-sans

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

4.5. Serial crystallography (DESY)

The serial crystallography analysis pipeline can be started within any **VISA instance** that has either CrystFEL pre-installed or **git** and **Apptainer** available. VISA is registered as a service in the EOSC marketplace and can either be started from the corresponding DESY service[47] or directly from its website[48].

Currently, only testing data is available through DESY's dCache. With the advent of the metadata-catalogue SciCat, which is currently being developed at DESY, each user will have access to the experiments and data according to his GAMMA[49] account in the future. Next to experimental data, open data will also be available within VISA.

- A schematic of the workflow is published in the PaN Training Portal[50]:
- The VISA service is available in the EOSC marketplace[51]
- A CrystFEL docker image is available at the DESY Gitlab[52]
- And optionally command wrappers for the image are also available at the DESY Gitlab[53]

Accessing the workflow

Access and Authorization

VISA on visa.desy.de is openly accessible by anyone with at least one of the following credentials: DESY credentials, EGI Check-in, Helmholtz AAI or GitHub. When logging in the first time, a stable user identifier (uid) for unix-like operating systems is created and linked to the account, in case no uid was preexistent. The initial quota of VISA instances a user can create is zero and can be upgraded to a quota of two by email request.

Running the workflow

To run the serial crystallography workflow, the following steps can be used. CrystFEL will be started within a VISA instance with Apptainer pre-installed. An helper script will create and mount a CrystFEL Apptainer image from the gitlab docker registry and put easy-to-use wrapper scripts in place:

- Enter the EOSC marketplace and search for DESY VISA service and follow the link.
- Click the "Sign in" button and enter your account credential.

[47] https://marketplace.eosc-portal.eu/services?providers=116
[48] https://desy.visa.de
[49] Data management of DESY users (https://gamma-portal.desy.de)
[50] https://pan-training.eu/workflows/containerized-serial-crystallography-visa-workflow
[51] hhttps://marketplace.eosc-portal.eu/services/desy-visa-portal
[52] gitlab.desy.de:5555/thomas.white/crystfel/crystfel
[53] https://gitlab.desy.de/silvan.schoen/containerizedapplication

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

- Push the “Create a new instance” button.
- Push “Search for experiments” and choose one of your experiments. This will automatically add linked experimental data to the session (work in progress). Alternatively activate “Instance not associated with any specific experiments”.
- Choose “VISA_Apptainer” image and select desired hardware.
- Optionally choose screen- and keyboard-layout.
- Accept terms and conditions and create the instance.
- The instance will now be built. After the build process finished, connect to the instance.

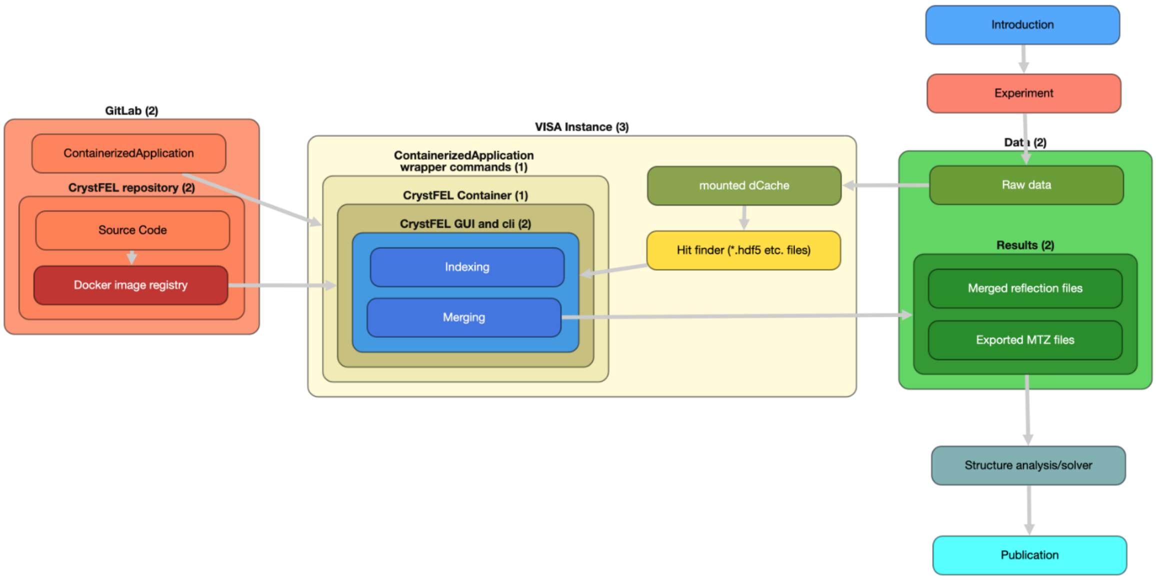
Within the VISA instance:

- Clone containerized application:
  ```
  git clone
  https://gitlab.desy.de/silvan.schoen/containerizedapplication
  ./target
  ```
- Add data directory to wrapper:
  ```
  cd ./target
  nano config.sh
  ```
  Add “/home/data/data/crystfel-tutorial” to bind_directories
- Install wrapper scripts
  ```
  source ./install_script.sh
  cd ..
  ```
- CrystFEL can now be used as described by the workflow PaN Training Portal[54].

[54] https://pan-training.eu/workflows/containerized-serial-crystallography-visa-workflow

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857641.

Workflow of Serial Crystallography



Experiments generate up to 150 TB per day of data saved at the measurement facility. Such large datasets are impractical for users to take home. Subsequent analysis needs to be performed remotely making it attractive for deployment as a cloud-like use case. Involving EOSC in the analysis and re-use of this data is an appropriate use case.

Serial crystallography is a beam-line technique for collecting information on the structure of a protein without growing large protein crystals. Instead, a large number of small protein crystals are held in a pulsed X-ray beam. In a second step, the series of produced images are used to reconstruct a precise 3-D image of the protein structure. Serial crystallography is the preferred technique for obtaining diffraction data of proteins at room temperature, where radiation damage from the X-ray beam starts rapidly. The standard software for analysing serial crystallography is “CrystFEL”.

The proposed workflow was rendered in a standard fashion, which would allow it to be easily adopted by arbitrary systems or also other containerized applications. The only requirements are an Apptainer installation on the system and a Docker or Singularity/Apptainer image of the application, as well as an adjustment of the configuration file for the wrapper script.

Figure 7: Serial Crystallography workflow [55]

[55] https://pan-training.eu/workflows/containerized-serial-crystallography-visa-workflow

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857641.

5. Portfolio of training material

One of the goals of WP5 is to train research scientists to better use available computational research infrastructures to address critical research questions and to understand the computational workflows for specific experiments. In order to achieve this, our training catalogue PaN-training.eu provides, besides the registration of materials, the possibility to show interrelationships between different steps, resources, services or materials in the form of a training workflow. PaN-training is available as an EOSC service and the service tags are regularly updated based on the workflow keywords to make these resources findable within the EOSC. In the EOSC catalogue search portal resources from PaN-training are also findable under the category “Trainings”.
For our analysis services within ExPaNDS we created example training workflows to combine all resources necessary to understand and also to perform the analysis steps using EOSC services provided by our RIs.
For instance our workflows start typically with an introduction where further materials such as videos, websites with descriptions or background information are references. Building on this, we show where the data is available and on which EOSC resource the analysis can be performed in an interactive way.

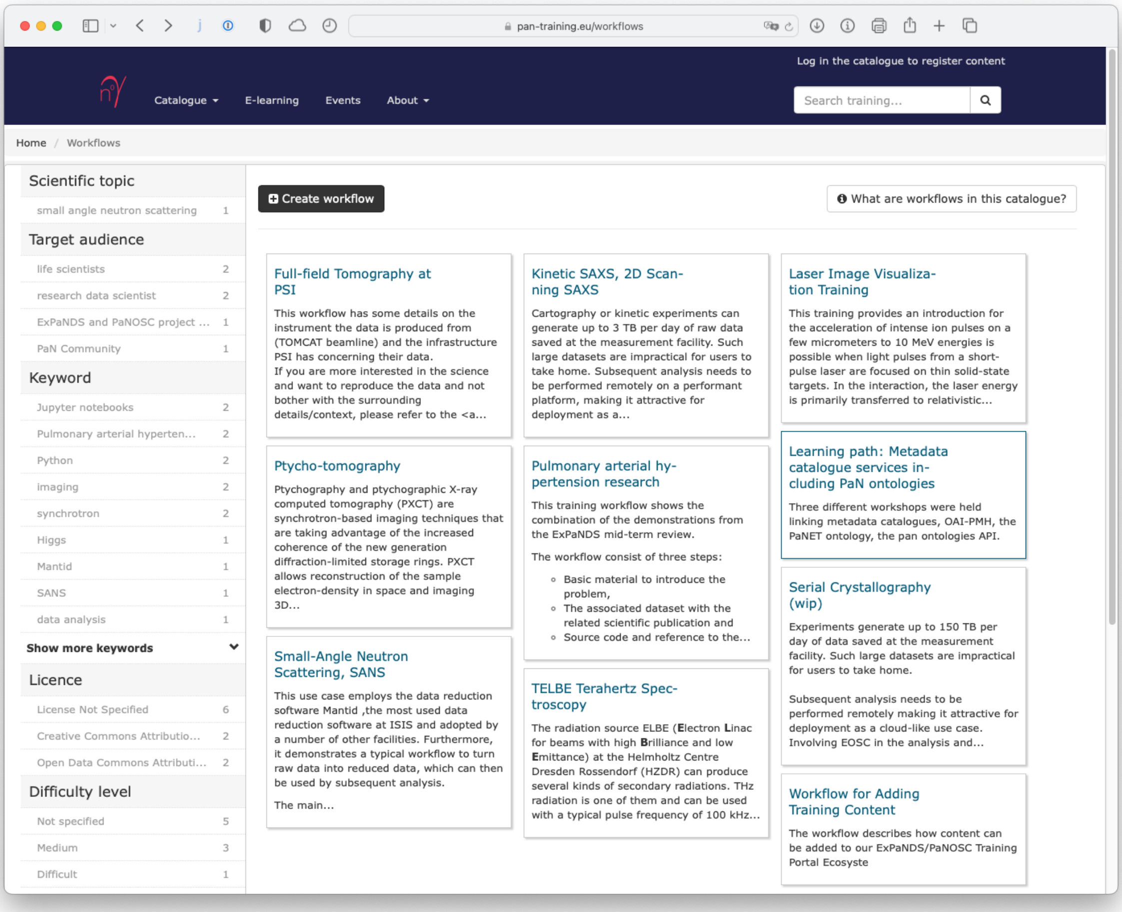


Figure 8: Overview of current training workflows in PaN training[56]

[56] https://pan-training.eu/workflows

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857641.

6. General analysis services deployed into EOSC by ExPaNDS partners

In this section we describe the generic data analysis platforms deployed into EOSC by the ExPaNDS partners. These generic data analysis services can be used to execute many possible data analysis workflows, including the example workflows described in section 4.

Partner	EOSC Provider	EOSC Services
PSI	✅	✅ PSI Remote Desktop Service
DIAMOND Light Source	✅	✅ Diamond Remote Desktop Service
UKRI	✅	✅ SIS Data Analysis as a Service, DAaaS[57]
Max IV Laboratory (Lund Uni)	❌(*)	❌ MAX IV data analysis services (*)
ALBA	✅	❌ NoMachine service (*)
SOLEIL	✅	❌ VISA and data catalogue services (*)
HZDR	✅	✅ PaN-training.eu
DESY	✅	✅ Jupyter and VISA service (see below)

(*) Publishing in the EOSC marketplace planned before the end of project funding. (See section 6)

6.1. PSI

The PSI Remote Desktop Service - EOSC Marketplace[58] is a generic remote desktop service that provides access to HPC resources as well as beamline consoles. Researchers can access a JupyterHub instance or submit jobs via Slurm. This means that the infrastructure is capable of supporting the various use cases identified as part of ExPaNDS.

However, these services are not open access and potential users are required to have a PSI account and the necessary authorization to use the individual services available within PSI.

6.2. Diamond

The Diamond Remote Desktop Service is registered on EOSC Marketplace. This can be located using the search, or the following URL[59]:

Instructions on the landing page describe how to access Diamond systems using the NX remote desktop tool. There is a reference video in the service marketing section that shows users how to do this. Users will need a valid Diamond FedID to access this resource. FedIDs are created when visits are booked, so you must be on an ongoing or historical visit.

57 https://marketplace.eosc-portal.eu/services/isis-data-analysis-as-a-service
58 https://marketplace.eosc-portal.eu/services/remote-desktop-service?q=PSI+Remote+Desktop+Service
59 https://marketplace.eosc-portal.eu/services/diamond-remote-desktop?q=Diamond+Remote+Desktop

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

6.3. UKRI

The ISIS Data Analysis as a Service (IDAaaS) is available on EOSC. This can be accessed by going to the EOSC marketplace, searching for “ISIS” or “ISIS Data Analysis as a Service” (or similar), click on the “webpage” link of the service. This will forward you to the service. Click on the “Login” button, which will forward you to the STFC User Office. Either create an STFC User Office account on this website or login with your existing account’s credentials. Once logged in, you should be brought back to the IDAaaS webpage.

6.4. MAX IV Laboratory

EOSC provider inclusion criteria requires the provider registration process is finished by providers’ legal entities first. Lund University is the legal entity for MAX IV Laboratory. Lund University is a member of the EOSC Association but it is currently not registered as the EOSC Provider. Thus MAX IV data analysis services are not accessible directly through the EOSC marketplace. They can be reached via the common European Photon and Neutron Open Data Search Portal[60].

6.5. ALBA

Alba is registered as a service provider on the EOSC marketplace and will provide authenticated access to remote desktop via NoMachine in the near future. The next steps consist in registering external access to our Jupyterhub and VISA services, once they are in production and open to external users.

6.6. SOLEIL

SOLEIL is registered as a service provider on the EOSC marketplace and is in the process of preparation to onboarding VISA and data catalogue services once they are in production and available to external users.

6.7. HZDR

The HZDR is registered as a service provider on the EOSC and with PaN-training.eu the first productive service is available. Further HZDR services which may be registered as EOSC services in the future are:

- Installation of an OpenStack infrastructure in order to provide the virtual machines for the data analysis workflow. VISA on top of the infrastructure is currently under development.
- Based on that infrastructure container instances are generated that run the Jupyter Notebooks.
- User authentication is provided by federated AAI (Helmholtz AAI and future UmbrellaID).
- The FAIR data repository RODARE was made available to EOSC B2FIND with its own community as EOSC service.

[60] European Photon and Neutron Open Data Search Portal

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857641.

- The federated HZDR GitLab instance is used for common development and testing of all components of the analysis workflows. The latest version of the workflow is also published open accessible on GitHub.
- A SciCAT data catalogue prototype is being deployed and evaluated to manage metadata for Terahertz Spectroscopy data. This approach is discussed within the DAPHNE consortium of the German NFDI (National Research Data Infrastructure).

6.8. DESY

The VISA EOSC service at DESY enables users to start VISA instances for their respective experiment. Enter the EOSC marketplace and search for DESY VISA service and follow the link. Click the “Login” button and enter your account credentials to gain access.

The default VISA environment has Jupyter Lab pre-installed. This makes it possible to start Jupyter-based workflows out of the box within the VISA instance. All instances are shareable, which makes collaborative workflows within the project during or after an experiment easily possible. If the analysis relies on more complex workflows, including applications or graphical user interfaces, a VISA environment with Apptainer pre-installed is choosable by default. The user has to provide a containerized version of his application, which can be downloaded and run from within the VISA instance. If containerization proves to be difficult for specific workflows, the user can also ask for a custom VISA image to be created, which will then have all requirements and applications pre-installed. This image will be uploaded to the EGI Fed-Cloud application database[61] for the specific Virtual Organisation (VO) and thus be made available to the user via the EOSC service.

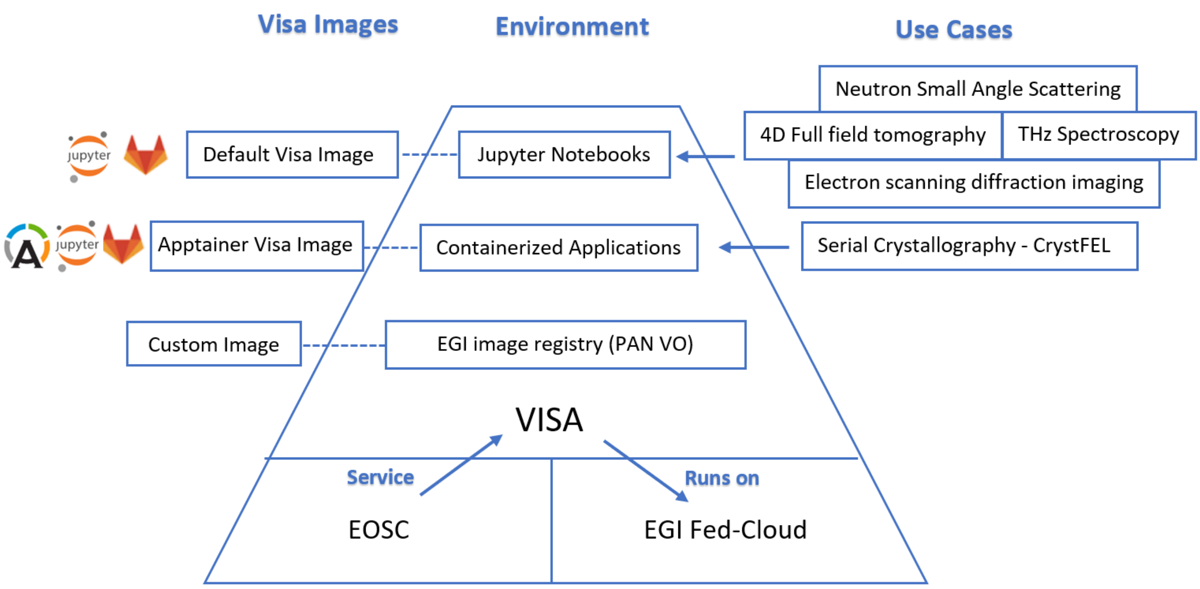


Figure 9: A VISA instance can be initiated through the EOSC service. Depending on the application, either a generic or a more specific image will be used for the instance.

[61] https://appdb.egi.eu/

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857641.

7. Harmonisation with other relevant initiatives

7.1. PaNOSC

The adoption of VISA in ExPaNDS is one of the major alignments with the PaNOSC project where it has been chosen as the reference Remote Data Analysis platform and was described there in deliverable 4.3[62]. VISA has been initially developed at the Institut Laue-Langevin (ILL) in Grenoble, France aiming to provide a solution to offer *Data Analysis as a Service* in the cloud. PaNOSC and ExPaNDS enlarged the audience of facilities using the VISA platform. The institutions operating these installations depend on the availability and the reliability of the product in the operation of their installations over 20 to 30 years. Therefore there exists a strong need to guarantee the long-term sustainability of the platform. That is the reason why we are now working on establishing a Memorandum of Understanding (MoU) declaring the commitment of the partners to the long-term development and maintenance of the VISA platform.

7.2. EOSC Future

EOSC Future[63] is an EU funded project for establishing a platform for open science and FAIR Data. In an early demonstration they showed a use case based on serial crystallography. Here, data from one facility (DESY) was provided to be used at another facility (ESRF). The subsequent analysis was done by CrystFEL. The containerized form of the software was used, which was initially developed within ExPaNDS. In a further step within EOSC Future, a training tool for CrystFEL is planned, which provides the necessary data and uses Python wrapper functions within a Jupyter notebook to simplify a typical workflow for new users and ease up the learning curve. The notebook will also be applicable within the Jupyter environment of VISA instances.

7.3. ESCAPE

The ESCAPE[64] project targets storage and computing solutions for Exabyte-scale datasets managed at Particle Physics and Astronomy facilities in the context of the European Open Science Cloud. It implements a Data Lake model, connecting modular storage systems at the partner institutes with a reliable network and providing access with a common AAI service. It deploys Rucio[65] to orchestrate data management and locality.

In the context of the ESCAPE Data Lake, DESY operates a dCache[66] storage system and registers it as a Rucio Storage Element (RSE). Rucio allows users to declare rules that define the availability of replicas of a given dataset at a given site. The virtual machine instances managed by the VISA portal mount the same filesystem to which Rucio copies the data as part of the ESCAPE Data Lake and give users the option to work with their datasets in customised interactive analysis environments. Both, the remote desktop and the

[62] DOI 10.5281/zenodo.7333306
[63] https://eoscfuture.eu/
[64] https://projectescape.eu/
[65] https://rucio.cern.ch/
[66] https://dcache.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

JupyterLab interface exposed by VISA, give transparent access to the data, abstracting away the Data Lake infrastructure for users who don't need to be aware of the federated storage architecture and just want to focus on data analysis.

Scientists who manage replicas and data locality find an interface to work with Rucio directly from the JupyterLab, which has been developed in the ESCAPE project and is installed in VISA instances using the JupyterLab Extension Framework. This provides users with a graphical user interface to interact with the ESCAPE Data Lake from VISA instances.

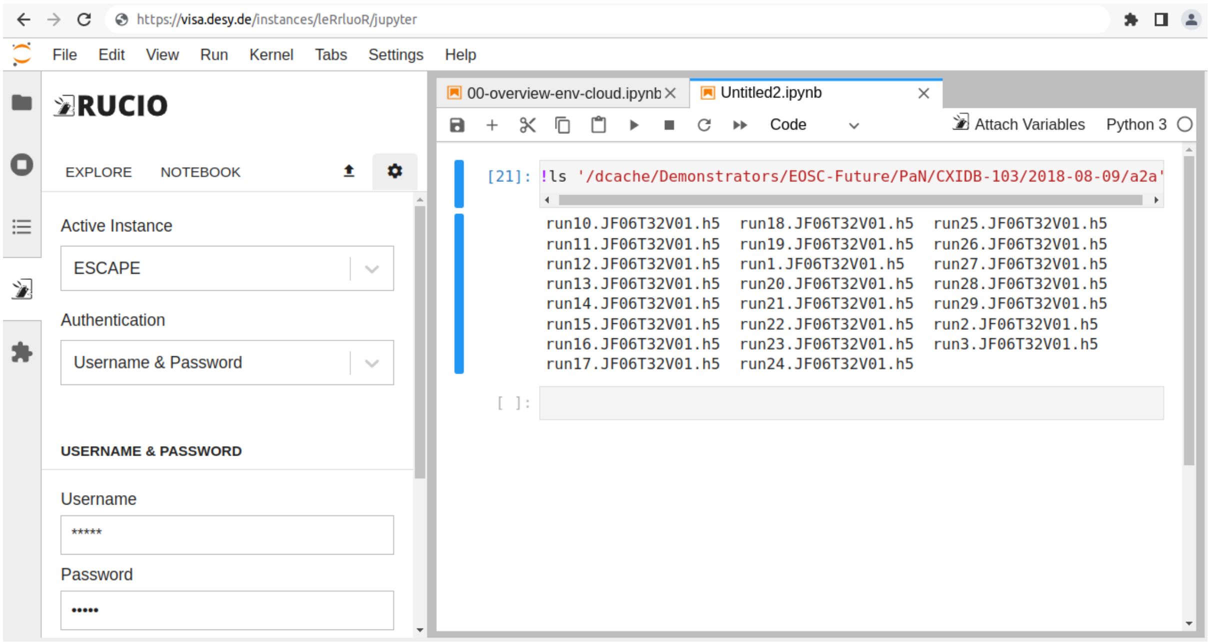


Figure 10: Accessing ESCAPE Data Lake storage within VISA through the RUCIO JupyterLab extension.

7.4. EGI-ACE

EGI-ACE[67] is a 30-month project coordinated by the EGI Foundation with a mission to empower researchers from all disciplines to collaborate in data- and compute-intensive research through free-at-point-of-use services by implementing the EOSC compute platform. The EGI-ACE services are offered through the mechanism of Virtual Access (VA) which is a new financial instrument to reimburse the access provisioning costs to access providers. This instrument is provided by the EC to increase the sharing of research infrastructures and services that otherwise would not be available to international user groups. VA access is open and free access to services through communication networks to resources needed for research, without selecting the researchers to whom access is provided.

DESY, the ExPaNDS coordinator, is participating in EGI-ACE providing cloud resources (being part of the EGI Federated Cloud).

The participation of the coordinating partner in the EGI-ACE project has been beneficial to better understand the VA mechanism as a new way to fund resource access within the EOSC which is going to be used also for the procurement calls published by the EC in December 2022.

[67] https://www.egi.eu/project/egi-ace/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

The funding using the VA mechanism could be used to provide access to resources available in PaN facilities to international resource groups that otherwise would not be available to access, given the limitation on the compute resources. It can be used in two ways: PaN facilities can offer specialised resources to partner facilities or facilities can use the EC procurements to utilise pre-procured resources provided by industry. This may also help harmonising the access to the compute resources, solving the current heterogeneous landscape.

7.5. EOSC-Synergy

ExPaNDS and EOSC-Synergy organised joinded meetings during the early autumn of 2021. The general aim was to increase recognition of EOSC-Synergy project activities and tools developed by EOSC-Synergy for quality assurance of research projects software and EOSC services on one side and ExPaNDS scientific application cases and ExPaNDS testing and validation framework on the other side. The collaboration resulted in an ExPaNDS contribution to a workshop organised by EOSC-Synergy within the EGI Conference 2021[68].

Improved understanding of the guidelines, tools and services provided by EOSC-Synergy was very valuable for the ExPaNDS project, in particular for identifying gaps and shaping the "ExPaNDS Testing and Validation Framework"[69]. It was evident that guidelines, tools and services provided by EOSC-Synergy were highly complementary to the already implemented parts of the ExPaNDS framework. More precisely whereas the ExPaNDS project was solely focusing on the validation of scientific software environments and scientific data analysis workflows related to the ExPaNDS application cases, EOSC-Synergy approach is covering EOSC service testing and validation in a much wider scope, including beside others service quality assurance on the IT infrastructure level.

Deployment and coverage of service quality testing can vary between the ExPaNDS sites, however in general the ExPaNDS project was taking advantage of EOSC-Synergy achievements on three levels. First, software specifically developed for the ExPaNDS project is following the EOSC-Synergy guidelines for the software quality assurance. Next the software environment and scientific workflows testing and validation are profiting from the software level. On the top level, tools and services developed by EOSC-Synergy for the general testing of EOSC services can be deployed on the IT infrastructure of ExPaNDS partners.

7.6. ELIXIR

The PaN-training catalogue is a derivative of ELIXIR's TeSS (Training eSupport System). To reach a higher level of sustainability, further joint developments together with ELIXIR are planned in the area of workflow representations. In particular, the training workflow feature is of utmost importance for the sustainable description of an experiment's context, as shown in section 3. A next step could be to collaborate with ELIXIR in order to implement an execution of the analysis steps in the training workflows with VISA.

[68] EOSC-Synergy workshop at EGI Conference 2021
[69] ExPaNDS Testing and Validation Framework

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Here, better training with additional information can be made possible by integrating VISA deeper into the training catalogues. In this context in- and export of common workflow languages into training workflows and the other way round are possible future joint developments. With such modification the transfer of a productively used scientific workflow to a training workflow is a very low barrier for our scientists and the training catalogue workflows contribute even more to the reproducibility of the analysis workflows described in this document.

An automated publication of our workflows in ELIXIR's workflowhub.eu with a publish button in pan-training is also an opportunity to promote the use of workflows in the PaN community and to make our workflows and methods more findable. In the long term, the PaN community can benefit from stronger cooperation with ELIXIR.

7.7. LEAPS/LENS

The LEAPS[70] and LENS[71] initiatives (respectively the League of European Accelerator Photon Sources and League of European Neutron Sources) represent both National and European photon and neutron large scale facilities. Both have working groups related to computing and information technology generally. The aim of these working groups is to synchronise and harmonise the IT efforts across institutions and nations. Many of the lead participants in ExPaNDS and PaNOSC are also active in the respective LEAPS and LENS computing working groups, to the point that membership largely overlaps. During the time of PaNOSC and ExPaNDS, joint meetings of those projects largely performed the role of LEAPS/LENS IT working group meetings. With those EU projects coming to an end, perpetuating the successful elements of those projects under the umbrella of LEAPS/LENS has naturally been on the agenda of recent meetings. In particular, the continuation of PaNOSC/ExPaNDS efforts and the furtherance of open data from facilities has achieved prominence in the LEAPS/LENS IT strategy roadmap. This is a very promising mark of commitment for sustaining the efforts of these two EU projects into the future.

7.8. NFDI (DAPHNE4NFDI)

DAPHNE4NFDI[72] is one of 19 consortia receiving funding as part of the German National Research Data Infrastructure program (NFDI[73]) and combining the German large scale photon and neutron research infrastructures with the user community. The goal of DAPHNE4NFDI is to create a comprehensive infrastructure to collect, process and publish research Data from Photon and Neutron experiments according to the FAIR principles. This perfectly aligns DAPHNE4NFDI with the objectives of ExPaNDS and PaNOSC on the European and National levels.

DAPHNE4NFDI is coordinating with both ExPaNDS and PaNOSC to identify synergies and cases where the output of both projects can be further sustained. In particular, the extension of open data catalogues to ingest metadata from a wider range of instruments combined with

70 https://leaps-initiative.eu/
71 https://lens-initiative.org/
72 https://www.daphne4nfdi.de/english/
73 http://nfdi.de

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

analysis infrastructure for deployment on horizontal infrastructures will be further developed and deployed. To this end we are actively building a network between the task areas of DAPHNE4NFDI and the work packages of ExPaNDS. Furthermore, use cases and workflows, ExPaNDS partner facilities have experience with, provide invaluable opportunities for DAPHNE4NFDI facilities to develop and deploy their own services.

8. Common Service Level Agreements between PaN RI's.

Besides the technical implementation and integration of the PaN services into the EOSC marketplace, it would significantly simplify their usage by other facilities and/or individual scientific users, if a common agreement could be reached on the terms of usage and the service level provided by the facilities. Although the framework of the marketplace enforces a minimum set of "Principles of Participation" and offers multiple levels of compliance to choose from, the PaN community unfortunately has not yet reached a stage of a common policy.

At the time of this writing, the operation and maintenance of the available services are handled by the different facility providers with a variety of support levels depending on local user support constraints (human resources, hardware/software IT skills, etc.) or even constraints based on the mandate of the facility by their stakeholders.

However, facilities are taking advantage of being involved in ExPaNDS and PaNOSC and launched discussions to evaluate possibilities to agree on common policies for services that are registered in the EOSC marketplace. The EGI service level agreements[74] could serve as a role model. As the first step, negotiations started, agreeing on a MoU between PaN sites and facilities providing services as a foundation of more legally binding agreements. Making services, developed and deployed in the context of ExPaNDS/PaNOSC funding, is essential for some of our partner facilities as they already build their computing infrastructure on those outcomes. It is therefore mandatory to sustain those services through official collaborations.

9. Interoperability and Access Policies

The "EOSC Interoperability Framework" (EOSC IF) report, published by the EOSC Executive Board Working Groups FAIR and Architecture[75], discussed interoperability in relation to the fact that "research data usually need to be integrated with other data; in addition, the data need to interoperate with applications or workflows for analysis, storage, and processing".

The efforts made in the ExPaNDS project in relation to the FAIR principles are aligned as much as possible with the general principles of the EOSC-IF on what are generally considered to be four levels of interoperability: technical, semantic, organisational and legal.

[74] https://documents.egi.eu/public/ShowDocument?docid=2371&version=41

[75] https://op.europa.eu/en/publication-detail/-/publication/d787ea54-6a87-11eb-aeb5-01aa75ed71a1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

The technical operability is underpinned by the adoption of:

- the Nexus standard data format[76] for data exchange
- a standard federated “PaN Search API” for data findability collaboratively developed with PaNOSC
- a unified way for data analysis software containerization using Docker or Singularity/Apptainer
- a standard remote data analysis platform (VISA)

Semantic interoperability is an ongoing process already launched by the adoption of a common metadata framework and a shared ontology (PaNet).

The organisational interoperability is driven by the permanent collaborative work done in the frame of the LEAPS and LENS consortia.

The legal interoperability is taken care of by the efforts of data policy harmonisation between the facility partners and by the adoption of similar Creative Commons licences (CC BY).

[76] http://www.nexusformat.org/

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857641.

10. Conclusion

This deliverable documents the integration of existing PaN NRI data analysis services into EOSC, as a means of providing PaN scientists in Europe with standardised access to data analysis services through the EOSC ecosystem.

Due to the existence of a heterogeneous collection of existing analysis services at the NRI partners, we adopted a modular architecture in which each facility connected their existing web portals and data analysis facilities directly into EOSC. Prototype analysis services have been set up and tested in different configurations within this infrastructure exploiting either Jupyter notebooks or remote desktop services running either Jupyter notebook code, or custom software deployed in containers so that the end-user can use the workflows in the same way in any participating facility. One of the tangible outcomes of ExPaNDS was that the participating partners learned how to make their software interoperable not only between each other's facilities, but also on third party infrastructure. This outcome will be absorbed into standard working practices and used in future software deployments.

These outcomes have been passed on to WP5 training so as to enrich the community shared knowledge base on this topic. Nevertheless, continued investment into training is important in order to propagate the outcomes of ExPaNDS into the community. In particular, lessons learned for the user community, such as how to deploy software in a portable manner, are sustainable and can be further developed to benefit the wider user community based on experience gained in this project. Further work is also required to increase the FAIR-ness of data, for example by improved metadata harvesting from instrumentation and user logbooks, publication of data PIDs, and opening of raw data on publication.

One very concrete outcome of ExPaNDS is to have increased the profile of open data and interoperable analysis pipelines within the facilities. As, due to limited resources, the top of facility priorities is naturally serving immediate needs of facility users. By having a dedicated project on the topic of FAIR data handling, data from our PaN facilities are now in a position to open up to the rest of the outside world in a coordinated, interoperable fashion. Without ExPaNDS, this may not have happened any time soon.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.