PaNOSC
Photon and Neutron Open Science Cloud
H2020-INFRAEOSC-04-2018
Grant Agreement Number: 823852

Deliverable: D6.5 - Report on EOSC Integration.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Project Deliverable Information Sheet

Project Reference No.	823852
Project acronym:	PaNOSC
Project full name:	Photon and Neutron Open Science Cloud
H2020 Call:	INFRAEOSC-04-2018
Project Coordinator	Andy Götz (andy.gotz@esrf.fr)
Coordinating Organization:	ESRF
Project Website:	www.panosc.eu
Deliverable No:	D6.5
Deliverable Type:	Report
Dissemination Level	Public
Contractual Delivery Date:	30 November 2022
Actual Delivery Date:	4 January 2023
EC project Officer:	Flavius Alexandru Pana

Document Control Sheet

Document	Title: Report on EOSC Integration
	Version: FINAL
	Available at: https://doi.org/10.5281/zenodo.7504250
	Files: 1
Date	4 January 2023
Authorship	Written by: Jean-François Perrin
	Contributors: Jayesh Wagh, François Sturma, Aidan Campbel, Cédric Vinet, Bertrand Lanson, Andrea Manzi, Giuseppe La Rocca
	Reviewed by: Jordi Bodera, Andy Götz
	Approved: Andy Götz

List of participants

Participant No.	Participant organisation name	Country
1	European Synchrotron Radiation Facility (ESRF)	France
2	Institut Laue-Langevin (ILL)	France
3	European XFEL (XFEL.EU)	Germany
4	The European Spallation Source (ESS)	Sweden
5	Extreme Light Infrastructure Delivery Consortium (ELI-DC)	Belgium
6	Central European Research Infrastructure Consortium	Italy

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

	(CERIC-ERIC)	
7	EGI Foundation (EGI.eu)	The Netherlands

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Table of Content

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Report on EOSC Integration

Introduction

Highlight collaboration with other Projects and partners

The Photon and Neutron community (PaN) gathers scientific analytical facilities such as Synchrotrons, Lasers, Free Electron Lasers and Neutron sources and their user communities.

WP6 has played a pivotal role in integrating services and other technical developments from PaNOSC work packages (WP3, WP4, WP5) in the EOSC ecosystem. The integration activities involved collection of requirements, setting up multiple pilots and technical collaborations with PaN RIs, EOSC cluster projects, e-infrastructures and other EOSC related projects. Some highlights of collaborations with other projects and partners are listed below.

- GÉANT has supported PaN community immensely to implement UmbrellaID AAI. This collaboration is still continuing in the framework of the EOSC Future[1] and AEGIS projects.
- While exploring suitable data transfer solutions, WP6 actively reached out to EGI to implement OneData pilot.
- VISA, the Virtual Infrastructure for Scientific Analysis platform was developed by ILL. This background development was further enhanced under WP6 with active involvement of PaNOSC and ExPaNDS partners.
- WP6 investigated software provisioning solutions with the other science clusters, and adopted the approach used in the ESCAPE cluster[2].
- WP6 made use of the OCRE project[3] frame contract for commercial cloud procurement activities.

This document complements previous deliverables by presenting an overview of the integration of UmbrellaID AAI, Data transfer services, Data analysis services and Commercial cloud procurement activities in the EOSC.

[1] https://eoscfuture.eu/

[2] https://projectescape.eu/

[3] https://www.ocre-project.eu/funding-opportunities/ocre-funding-and-procurement-call

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

AAI

UmbrellaID (https://umbrellaid.org/what.html) has been serving the AAI needs of the PaN community. The direct advantage of the umbrellaID environment for the user is that he/she has only one account for all partner facilities and once logged in he/she can access the services at these partner facilities without a need for a new identification (Single Sign On).

An *UmbrellaID workshop: Keycloak*[4] was organised in May 2022 to support the PaN community IT personnel for the implementation of UmbrellaID AAI for their research infrastructure users.

Over the course of the PaNOSC project, engaging with GÉANT and adopting the eduTEAM platform, UmbrellaID has evolved significantly. It is now ready for the EOSC AAI architecture model. As described in *D6.3 Integration of the PaN AAI into the EOSC*[5], all new services developed by the PaN community, even beyond PaNOSC partners, can be integrated into EOSC AAI with minimum efforts. By continuing this cooperation with GÉANT and other communities in the scope of EOSC-Future and AEGIS[6], we should be able to establish an EOSC AAI where interoperability between the communities will not be a challenge anymore.

Data transfer

Three use cases were considered to identify a suitable data transfer solution for PaN RI in deliverable *D6.1 Data Hub.*[7]

	Use cases	Data transfer solution
1	A RI wants to archive its experimental data in a remote data centre.	Rclone
2	A facility user wants to transfer a large dataset from an RI's archive to a remote computing centre or her or his home pc.	Globus
3	A user wants to access a data analysis service, data has to be available "transparently".	OneData

[4] https://indico.psi.ch/event/12701/

[5] https://zenodo.org/record/5913472#.Y4htYITMK3A

[6] https://wiki.geant.org/display/AARC/AEGIS

[7] https://zenodo.org/record/5912966#.Y5c_uoTMK3A

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Use Case #1 - A RI wants to archive its experimental data in a remote data centre.

For the 1st use case, three solutions (FTS, Rucio[8] and Rclone) were evaluated through a pilot setup between ILL and UKRI-STFC. Rclone appeared to be the most suitable solution compared to the FTS and Rucio which require substantial infrastructure modifications. Rclone, an open source solution, is the most versatile and suitable solution for cold archiving use cases.

Use Case #2 - An RI user wants to transfer a large dataset from the RI's archive to a remote computing centre or her or his home pc.

For this 2nd use case, there are very few solutions available on the market that could fit this strong and increasing community need as datasets tend to become bigger every year. For example, datasets of 10TB or more have become routine for X-Ray tomography experiments. Two solutions: IBM Aspera[9] and Globus[10] were evaluated through a pilot setup at ESRF. Both solutions are suitable, however Globus has been adopted by the ESRF and the different PaN RIs. IBM Aspera also passed the evaluation successfully and has been identified as a potential candidate for adoption but the cost has been a concern for general deployment in the community of RIs. However, IBM Aspera is used in push mode (I.e. the data are pushed using the Aspera client from the RI repository to the user's remote storage) for users requiring the IBM Aspera solution.

Globus is a solution that relies on the open GridFTP protocol which also allows to achieve great transfer rate performance. Despite the fact that the users need to install a client on their computer it remains simple to use, triggering a transfer can be done in a few seconds by any user without specific training needs. For the administrators this is also a relatively easy solution to implement. The solution also provides various connectors that can support different cloud providers (azure, google, AWS) and various storage protocols (Posix based, S3, ...). Regarding access authorisation, the solution could rely on the ACL of the POSIX filesystem of the repository which is extremely convenient and easy to handle for the system administrators. The Globus web interface also allows transfers between two end points that the users have access to. The authentication to the web interface can be managed by the identity and access management of the RI or the community one, which also simplifies the

[8] https://rucio.cern.ch/

[9] https://www.ibm.com/products/aspera

[10] https://www.globus.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

work for the administrators and the users. Globus is also extremely resilient, if connectivity is lost between two endpoints the protocol will halt the transfer and wait for both endpoints to be available again to resume the transfer from where it was left without having to start from the beginning, this is really important especially because of the very large volume of data (sometimes hundreds of Terabytes/per transfer) that are transferred and the time necessary to complete it. A Globus subscription is recommended to gain support privilege and additional features but not mandatory.

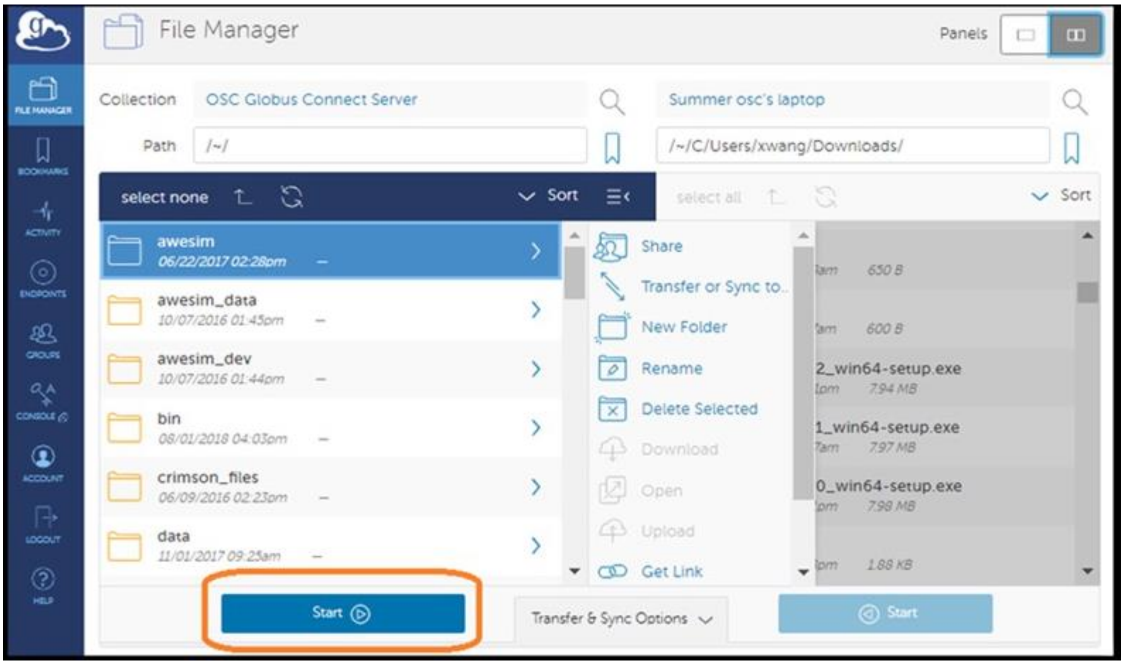

Figure 1 Globus user facing interface

IBM Aspera is a solution that relies on the proprietary protocol ascp, developed by IBM. The throughput and the performance of the transfer depends on the commercial plan chosen. Every time a new user wants to download data using Aspera, his account has to be created on the Aspera cloud and given access to a certain space containing the data, the ACL of the filesystem do not apply. Automation of the processes required to create users and give them access which requires more adaption and complexity for the RI repository administrators. The users need also to create an IBM account, which represent an extra step for users.

Use Case #3 - A user wants to access a data analysis service, data has to be available “transparently”.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 823852

For the 3rd use case, two solutions (FTS and OneDATA[11]) were evaluated through a pilot setup between ESRF and EGI. This use case is currently not actively requested by the PaN user community but will be necessary to support EOSC implementation especially when the data repository and the data analysis service are operated by different organisations or communities (i.e. EOSC service composability).

To complement deliverable D6.1 as we have pursued the effort with the OneData solution. New features have been added by the developers like the ability to choose which user ID (UID) will be used for accessing the storage of the RI repositories, but also the integration of community AAI authorisation mechanisms. Despite these efforts and the very positive appraisal of the user environment the pilot was not fully conclusive. Concerns have been raised that the central catalogue approach will have difficulties to scale the 100s of billions of files produced by the RIs of the community, as the indexation of all the files in a single experiment was taking days to be indexed.

The OneData development team added a missing feature that the ESRF required, the ability to choose which UID to use to access the storage in the documentation.

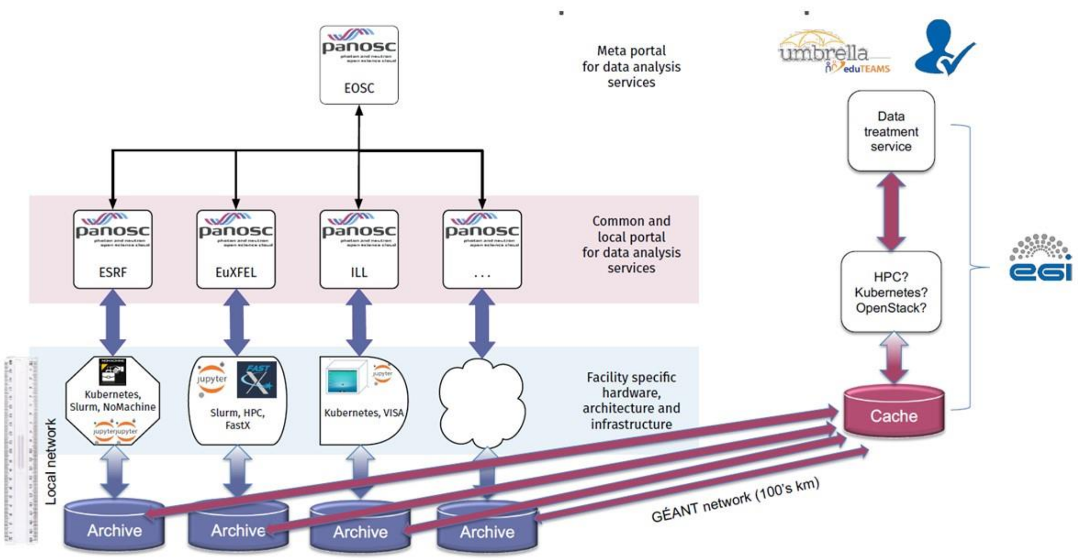

Figure 2: A facility user wants to transfer a large dataset from an RI's archive to a remote compute

This use case remains largely unanswered as it needs a global approach at the level of EOSC. In 2022, there is still a gap between the mechanisms to access and authorise access to data access in the communities and the solutions envisaged at the scale of EOSC. We hope to be

11 https://wiki.egi.eu/wiki/EGI_Opendata_platform

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

able to close this gap in EOSC-Future or in future EOSC related projects as this is a key element for service composability. The CERN developed Data lake is a possible solution which could satisfy this use case. It should be tested between PaN RIs in the future.

As an outcome of the three pilots, it is recommended that the users are trained on data transfers using community solutions such as Globus. A more global data transfer solution that can work across RIs and communities is still needed. In the future, machine actionable data transfer solutions will also be investigated.

Data analysis services

Jupyter Notebooks

One of the main objectives of the project was to implement production level (TRL9) data services based on Jupyter notebooks which can be used by the community to execute, share and replicate research at each RI. In addition to deploying a Jupyter service at each PaNOSC partner PaNOSC also had access to the EGI Notebooks[12] service, a service-like environment based on the Jupyter technology[13], offering a browser-based tool for interactive data analysis, and the EGI Binder[14] service to allow users to replay complex calculations, simulations, and visualisations scenarios by importing Notebooks and their runtime environment and sharing them with a single link. The EGI Notebooks is a multi-user service that can scale on demand, being powered by the compute services of EGI.

The EGI Notebooks service provides the Jupyter interface for notebooks, with the following added features:

- Easy access: Login with any eduGAIN or social accounts (e.g. Google, Facebook).
- Software distribution with the CernVM-File System (CVMFS)[15].
- Transparent access to the EGI DataHub spaces as regular files, and other 3rd. party services like the EUDAT B2DROP[16].
- Deployed on the cloud providers of the EGI Federation.

When the project started EGI provided a dedicated custom set-up of the EGI Notebooks and Binder services in order to let members of the PaN community to use UmbrellaID accounts to log-in. This dedicated set-up was replaced when the UmbrellaID[17] community AAI for the

[12] https://www.egi.eu/service/notebooks/
[13] https://jupyter.org/
[14] https://binder.notebooks.egi.eu/
[15] https://cernvm.cern.ch/fs/
[16] https://www.eudat.eu/catalogue/b2drop
[17] https://umbrellaid.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Photon and Neutron (PaN) community was integrated into the EGI AAI Check-in[18] service (the e-Infrastructure proxy).

The EGI Notebooks and Binder services operated by EGI was useful but limited by the lack of access to data which were too large to move. In the majority of cases the PaN users relied on the locally deployed Jupyter services provided by WP4 which had access to users data. Jupyter helped PaN researchers to move on the reproducibility spectrum and make modelling/analytics code, as well as referenced datasets, shareable and re-executable with anyone.

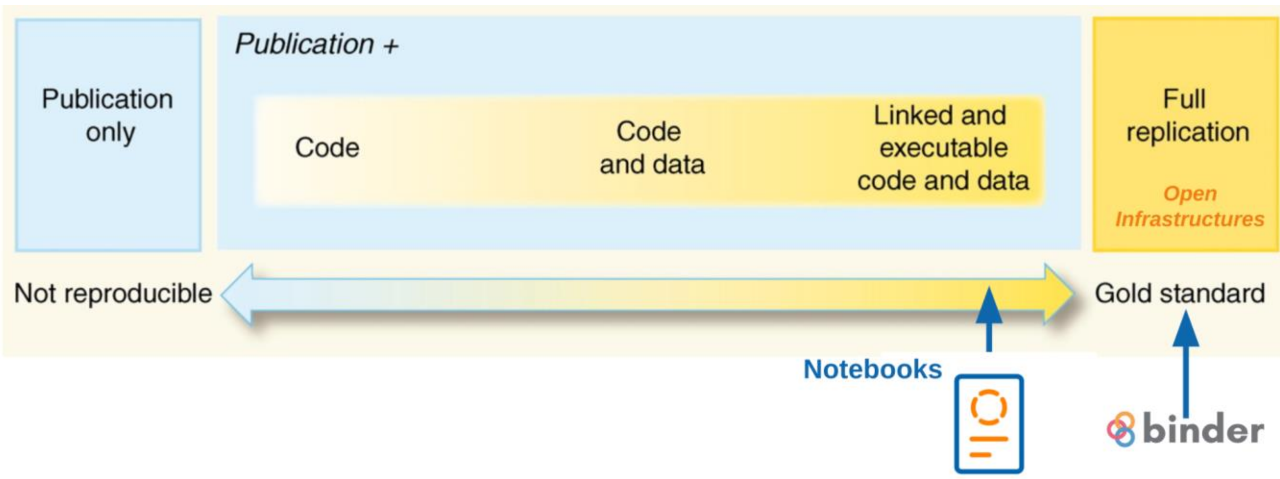

Figure 3 Support the reproducible spectrum with the Jupyter Notebooks and Binder services.

VISA

While the development of VISA in terms of functionality and code took place in WP4, WP6 was in charge of supporting the deployment of the underlying infrastructure for the different partners. The initial infrastructures, based on the OpenStack cloud computing platform, were mainly defined to support the creation of the virtual machines (VM) with CPUs and RAM capacities. With the advancement of the deployments in the RIs, it became evident that access to GPGPU and to HPC cluster resources were also necessary to fit the needs of users. WP6 has addressed these needs and integrated these types of resources in VISA capabilities when needed, typically at the ESRF A40 GPGPUs are now available for selected beamline users and a small HPC cluster has been setup to allow running software that necessitate parallel computing infrastructure.

GPU

[18] https://www.egi.eu/service/check-in/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Besides providing CPU-only instances, we can also provide GPU accelerated instances by leveraging the *Cyborg* component of Openstack. The advantage of using Cyborg compared to creating images with pre-defined PCIe devices attached is that we are more flexible in the way the resources are allocated. Cyborg treats all the PCIe devices from our hosts as a resource pool, and allocates from this pool whenever an instance requires it.

This flexible approach allows us to add various types of PCIe devices in the resource pool, without having to change much from the overlaying configuration. The following modifications need to be done on the bare-metal fleet to attach these devices on the VISA instances:

- IOMMU needs to be enabled in the kernel (without it, it is not possible to attached PCIe devices to virtual machines)

```
AMD: amd_iommu=on iommu=pt
Intel: intel_iommu=on iommu=pt
```

- The devices need to be made unavailable to the host system in order for them to pass correctly to instances. For this, we're using the vfio-pci modules:

/etc/modules

```
vfio-pci
```

/etc/modprobe.d/vfio.conf

```
options vfio-pci ids=0000:21:00.0,0000:81:00.0
softdep nouveau pre: vfio-pci
```

/etc/initramfs-tools/modules

```
vfio-pci
```

/etc/initramfs-tools/scripts/init-top/vfio.sh

```
#!/bin/sh
PREREQ=""
DEVS="0000:21:00.0 0000:81:00.0"

for DEV in $DEVS; do
        echo -n "vfio-pci" > /sys/bus/pci/devices/$DEV/driver_override
        echo -n "$DEV" > /sys/bus/pci/drivers/vfio-pci/bind
done
exit 0
```

- Once this is done, cyborg is then able to list the different devices on each host:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

```
$ openstack resource provider list
+--------------------------------------+------------------------+
| uuid                                 | name                   |
+--------------------------------------+------------------------+
| 030c0758-e76b-4858-9f9c-7bc06df3e617 | psykokwak              |
| e86bc441-9d2d-37e0-89c3-31541fada919 | psykokwak_0000:81:00.0 |
| 43c6935b-3702-37df-b72f-500b74921446 | psykokwak_0000:21:00.0 |
+--------------------------------------+------------------------+
```

- The last thing to do is create a profile (a type of PCIe device to attach), and add it to any openstack flavor we want:

```
$ openstack flavor set --property 'accel:device_profile=gpu_a40' esrf.gpu.a40.xlarge
```

HPC

The VISA platform is suitable for most scientific workflows and software data analysis packages. However, some users need parallel code execution of HPC computing clusters to process their data. To accommodate these needs, we added a compute cluster managed by the SLURM workload orchestrator to VISA. The users have access to a small amount of CPU and GPU nodes through SLURM in addition to their VISA instance. This cluster relies heavily on CVMFS for the software provisioning, as explained in the next chapter.

Software provisioning

With the introduction of additional Data Analysis Services open to the community and in particular VISA infrastructures, which comes besides the existing compute infrastructures (HPC, HTC, online analysis dedicated machines) already in place in the RIs, software deployment on the different compute nodes using classical methods started to represent a heavy workload for the IT teams in charge of the compute resources maintenance in the RIs. This workload has been identified as an obstacle for VISA take off.

Additionally, we wanted to benefit from the participation of EGI in the WP6 work package to study alternative models of providing resources for our community and the possibility to provide such Data Analysis Service by non-RI operators, in other word to host Data Analysis Infrastructure outside the RI premises for the benefits of all scientific communities.

The PaN RIs, with their scientific instruments, address diverse scientific fields (from biology to innovative materials, nuclear physic to cultural heritage and so forth) because of this diversity we have to deploy very different processing software (more than 150 are currently referenced in our software catalogue[19]) for users. This software, when installed on compute nodes usually require additional libraries and there are often incompatibilities between the

[19] https://software.pan-data.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

different software and their libraries.

We identified two different needs to address: the simplification of software deployment and the availability of software to compute nodes even if the software repositories and the computers are remote. We started early on in the project to look for technical solutions in order to solve these challenges. We benefited from the experience of EGI in this domain, but also from the numerous exchanges with the other clusters and in particular with GSI, one of the ESCAPE partners.

Two technical solutions have been identified: CVMFS and light container technology that are explained hereafter.

CVMFS as software distribution channel

The CernVM-File System (CernVM-FS)[20] provides a scalable, reliable and low-maintenance software distribution service. CernVM-FS is implemented as a POSIX read-only file system in user space (a FUSE module) on Linux based machines. Files and directories are hosted on standard web servers and mounted in the universal namespace /cvmfs. The system transfers data and meta-data on demand and verifies data integrity by cryptographic hashes.

During the PaNOSC project, ESRF has installed a central CVMFS component (Stratum-0 and Stratum-1 servers) dedicated to the PaN community, which allows publishing and distributing software. The publication process has been transparently integrated with the ESRF local Gitlab Continuous Integration (CI) in order to automate the containers creation and deployment on the CVMFS infrastructure to eventually facilitate the software updates triggered by the software developers.

Container technology for packaging software

Container technologies have become the standard way to provide services in the IT industry. In the PaN community use cases we require solutions to package software that could remove dependencies between the software and the operating system where these software are executed. Quite often scientific software depends on computing libraries that are installed on the system, these dependencies are often quite strong and prevent the installation of different releases of the same software. Using “light” containers (i.e. simplify technology that does not involve the network stack or other aspects necessary to maintain services) we

[20] https://cernvm.cern.ch/fs/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

can encapsulate the software and all its dependencies in a single transportable file that could be executed by any system where the application to run the container is installed.

Different container technologies for software exist like AppImage, Snap or Flatpak but we selected Apptainer (formerly Singularity) as it is an open source technology specifically developed for HPC. Apptainer[21] supports natively high-performance interconnects, such as InfiniBand as well as graphic accelerators (GPUs) and MPI by utilising a hybrid MPI container approach where MPI exists both inside and outside the container.

Using containers, we gain software portability and reproducibility. Apptainer/Singularity images can be prepared by one partner and executed by the whole community, the same software image will provide the same results as long as it is executed on the same processor architecture (typically x86_64) independently of the operating system or installed libraries. The technology also allows different versions of the software to run on the same system, as there is no more dependency on the system. This feature also allows us to preserve and run different releases of software on modern systems and to verify results or reproduce analysis that were performed years ago without having to install an ageing and unsupported operating system. This is an important asset toward FAIR and especially reproducible data analysis workflows.

All WP6 partners were convinced by this approach and have started to migrate their suite of software to Apptainer/Singularity containers. By the end of PaNOSC, the ILL had migrated all the scientific software used on their VISA instance to Apptainer containers. The ESRF migrated half of the most used software and developed an automatic pipeline embedded into the Continuous Integration (CI) of the ESRF local GitLab repository to foster this migration to containers. Other partners have also started adopting this approach.

The only real difficulties toward a comprehensive vision where all data analysis software would be containerised concerns large software suites that are used by some communities. These suites group different software and are prepared to ensure a consistent suite of software for a specific scientific domain community. A typical example is the PXSOFT[22] maintained by the ESRF for the structural biology community or the well known CCP4[23] for the Macromolecular X-Ray Crystallography, these software suites could have a very large size (dozens of GB) and host many executables. The size of these suites makes it impractical to be containerised; they need to be reorganised to benefit from the container's improvements; this work is beyond the scope of PaNOSC WP6 tasks.

Containerisation of software brings real benefits for system maintenance, scientific analysis

[21] http://apptainer.org/

[22] https://www.esrf.fr/UsersAndScience/Experiments/MX/Software/PXSOFT

[23] https://www.ccp4.ac.uk/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

reproducibility and for decoupling from infrastructures i.e. software could run everywhere without the intricacy of specific system preparation. This represents a key element toward the realisation of EOSC and the compatibility of services.

CVMFS and containers technology, all together

With CVMFS and Containerised, software packaged by one RI could be consumed locally on the RI infrastructure (workstations, HPC cluster, VISA, ...) or remotely by any independent compute infrastructure. They provide transport and isolation mechanisms which allows service composability for the software layer.

We have successfully tried to implement the concept of consuming software prepared by one RI in another RI infrastructure. ESRF has exposed one of their CVMFS repository (hpc.esrf.fr) hosting software in the form of singularity images to DESY and Soleil, 2 partners of the ExPaNDS project, the result of the technical evaluation was very positive. We are now discussing a trust framework that should address: safety of code execution, release planning and licensing questions in order to envisage to set up a federated software repository for the community with software ready to be used. Discussions will continue beyond the end of PaNOSC and will hopefully convince other communities with similar efforts to join forces.

Current state of EOSC core integration

At the end of PaNOSC, EOSC Core is still made of 3 services namely AAI, service monitoring and helpdesk.

As explained in the first section, the PaNOSC AAI is fully part of the EOSC AAI federation. This is a success story of the collaboration between PaNOSC and GÉANT.

Service monitoring and helpdesk core services were opened for the scientific communities in September 2022. Even if these services appear appealing, they came too late to be adopted the PaNOSC project. Nevertheless, we plan to take benefits of the offer for some community services, for example UmbrellaID the community AAI would clearly take advantage of a central professional helpdesk platform.

Cloud procurement adoption

An evaluation of commercial cloud procurement was carried out during the last year of the project. This work clearly highlighted some procurement difficulties between the procurement rules in place in the RIs and the commercial model proposed by the cloud providers, to adapt to the current financial rules of the ESRF we had to downsize the scope of the evaluation.

Despite advertising the service to the PaNOSC partners only one use case was identified. Hereafter you will find the report of a pilot made with a scientist using commercial cloud to test the Alphafold software for Structural Biology X-Ray experiments on different hardware configurations.

> *In order to create multiple virtual machines with the same software installed and the same configuration, we created a single virtual machine at the beginning and installed all the software needed for AlphaFold. Once we tested it and confirmed that it was working as expected, we could create an Amazon Machine Image of the virtual machine which could then create more virtual machines with everything preconfigured.*
>
> *As we would test thousands of protein sequences, we needed multiple virtual machines with GPUs to finish the predictions faster and begin the followup predictions. This required us to share information between virtual machines. Amazon provides a Network File System (NFS) through AWS EFS which allows us to share a filesystem between all the virtual machines and grow it only to the size that we needed automatically while we added more files.*
>
> *We had initially tried to test Alphafold on virtual machines with multiple Nvidia A100 (40G) GPUs however, Alphafold did not support multiple GPUs at the same time. As it would have been a waste to only use one out of the eight GPUs per virtual machine, we created multiple virtual machines with a single Nvidia A100 GPU at a much lower cost.*
>
> *One of the biggest difficulties was getting virtual machines with the GPU configuration that we needed. As thousands of people and organisations were asking for similar virtual machines with GPUs, we found that getting a reservation for some of the higher end GPUs was very difficult and borderline impossible in some cases. This required us to try to create virtual machines in various data centres across Europe, Asia and the US. This also posed another problem whereby we have to pay to transfer data between multiple data centres. As this would have been very expensive, we moved all of our data to Virginia where Amazon has a bigger data centre and could provide a higher chance of us getting the virtual machines with the GPUs we needed.*

While the public cloud environments offer an extensive set of technical solutions and resources for scientists, It also stresses the need of IT specialists to support scientists in their work in order to get the full benefits of such a vast environment.

Conclusion

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

The PaNOSC project was highly beneficial for the PaN RIs and user community. WP6 was a facilitator for the integration of the community service into EOSC but also for the introduction of new services for the RIs. As an example, the community AAI has been fully revamped to meet the latest AAI standards and was ready on time for the creation of the EOSC AAI.

Jupyter notebook analysis services and in particular VISA has been deployed by all the RI partners and adapted to the specific use case of the RIs. The packaging of software for these services inside the PaN infrastructures has been largely simplified thanks to the collaboration with the ESCAPE partners.

The standard data transfer use cases have found appropriate solutions to help users and RIs facing the challenge of transferring large data volumes reliable to user facilities. Unfortunately, the pilot for 3rd party data transfer use case, did not enable a satisfactory solution to be identified. More work with all EOSC communities together on a common specification and solution is required.

Finally, WP6 demonstrated that commercial cloud services require strong IT support in order to be adopted successfully by scientists.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852