PaNOSC

Photon and Neutron Open Science Cloud

H2020-INFRAEOSC-04-2018

Grant Agreement Number: 823852

Deliverable: D1.9 Mid-year summary 4

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:	D1.9
Deliverable Type:	Report
Dissemination Level	Public
Contractual Delivery Date:	31/05/2022
Actual Delivery Date:	31/05/2022
EC project Officer:	Flavius Alexandru Pana

Document Control Sheet

Document	Title: Mid-year summary 4
	Version: 1
	Available at: https://github.com/panosc-eu/panosc
	Files: 1
Authorship	Written by: Jordi Bodera
	Contributors: Juncheng E, Andy Götz, Rudolf Dimper, Tobias Richter, Fabio Dall'Antonia, Carsten Fortmann-Grote, Jean-François Perrin, Ornela De Giacomo, Thomas Holm Rod and Nicoletta Carboni
	Reviewed by: Jayesh Wagh
	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	ELI European Research Infrastructure Consortium (ELI-ERIC)	Belgium
6	Central European Research Infrastructure Consortium (CERIC-ERIC)	Italy
7	EGI Foundation (EGI.eu)	The Netherlands

Table of Content

Introduction

This document summarises the progress achieved in the project since the last management report until the month 42 of project execution (May 2022). As such, this document and the one that will follow on month 48 will complement the previous reports (at months 6, 12, 18, 24, 30 and 36) to continue providing a regular update on the project, its management and comparing its current status with what was forecasted.

Executive Summary

Overall, the project is advancing according to plan. A face-to-face meeting (https://indico.esss.lu.se/event/2862/) attended by partners, ExPaNDS' project members and cluster representatives took place in November 2021. The partners are engaged in the project, participate in regular meetings and recruitment has been successful. Our main document repository proves continuous activity: https://github.com/panosc-eu/panosc

A project review took place early in 2022. The feedback obtained has been very positive, stating that "the project has fully achieved its objectives and milestones for the period" and "has delivered exceptional results". Since then, the project team has been focusing on work to deliver the remaining deliverables, achieve the last milestones and ensure that the project's outcomes are sustainable.

Summary of Executive Board meetings

The sixth EB meeting took place on 1st of April 2022 via zoom. The meeting was chaired by O. de Giacomo (CERIC-ERIC). The agenda of the meeting was:

1. Approval of the agenda
2. Approval of the minutes from the last meeting
3. Report from the coordinator
4. Internal financial report
5. Work remaining until the end of the project
6. Post-PaNOSC collaborations
7. Any other business

After approving the agenda, the meeting proceeded with the report from the coordinator (the approval of the minutes from the last Executive Board meeting took place via email after the meeting).

A. Götz presented the feedback from the project review and highlighted the good progress the project is making. The recommendations from the project review were also covered. He explained the impact and outcomes of the project and his satisfaction with the commitment from work package (WP) leaders.

J. Bodera presented a draft Grant Agreement amendment that the project team has been working on (mostly

to formalise the fact that ELI-ERIC no longer has a legal link with ELI-NP) and clarified that the proposed amendment does not change the project's timescales or budget. The Project Board endorsed the proposed Grant Amendment.

The current financial status of the project was briefly discussed, naturally leading into post-PaNOSC collaborations and future funding opportunities. A. Götz reported that calls for proposals concerning EOSC cluster projects are expected to be launched soon and some of these could help build on the outcomes of PaNOSC.

The Project Board agreed to organise a final meeting to discuss the project's sustainability and thanked R. Dimper for his contribution to the project ahead of his retirement.

Summary of Project Management Committee meetings

The joint annual conference and meeting of the PaNOSC and ExPaNDS (European Open Science Cloud Photon and Neutron Data Service) projects did not take place in the last quarter of 2021 as planned due to the pandemic. An online event was instead organised (the European Photon and Neutron EOSC Symposium[1]) and a face-to-face meeting took place on 29th and 30th of November 2021 with key stakeholders of PaNOSC and ExPaNDS.

Meeting in Copenhagen 29th and 30th November 2021

The one-day meeting was a good opportunity for all WP leaders to share the status of their respective WPs, major pending actions and main concerns. In addition, some time was reserved at the end to go through the upcoming Periodic Report and sustainability.

WP1 - Management: J. Bodera reported the current level of spending and compared it with forecasts, reminded everyone of the upcoming Periodic Report, the Grant Agreement amendment planned for 2022 and of that the project should show via real use cases, Key Performance Indicators (KPIs) and the amount of open data available to users the value that is delivering.

WP2 - Data Policy and Stewardship: A. Götz reported on the present status of WP2 in M36. According to the planning of the project proposal WP2 has officially completed. Nonetheless the work is ongoing due to the long delays required to get data policies updated and for new facilities to publish open data due to the embargo period of 3 years. At the time of the Copenhagen meeting only ESS had fully implemented a DMP template generator, while the ESRF has implemented DMP templates, the other facilities are planning to follow suit. A summary of ongoing activities was presented which will ensure the full implementation of the objectives of WP2 by the end of the project. Activities which will

[1] https://indico.psi.ch/event/11435/

continue after PanOSC were identified e.g. training users in how to publish data and the advantages thereof, registering data repositories with publishers and learned societies. The collaboration with ExPaNDS were noted especially the common efforts on data policies and DMPs.

WP3 - Data Catalogue Services: T. Richter highlighted that the work package has to submit three deliverables in the period until next summer. The expected content of the three reports was outlined and activities have been identified that still need to be completed by partners so all deadlines can be met. Tobias also gave an overview of tasks that would benefit our user community without being required by a deliverable. Some of these connect to other work packages, like sustainability and analysis services.

WP4 - Data Analysis Services: F. Dall'Antonia (not able to be present in person due to CoViD-10 travel restrictions) reported on the status of the work package, highlighting VISA deployment and search portal development (with WP3) as main tasks for the remaining duration of the project, with the deliverable 4.3: "remote desktop and Jupyter service deployed at EOSC" being the deliverable due end of May. Partners gave optimistic feedback on the feasibility to meet the deliverable requirements and associated milestone. Concerning the realisation of a federated VISA instance, hosted at EGI for EOSC, the conclusion was that this project has low chances to go beyond its pilot phase (proof-of-concept status) and sustainability of a production version would require extra funding, for instance through the EOSC Future project.

WP5 - Virtual Neutron and x-raY Laboratory (VINYL): Due to the soared number of COVID-19 infections in Germany and Denmark, the European XFEL COVID-19 task force rejected J. E's travel to Copenhagen. Instead of meeting other group leaders in person, J. E did the presentation online. He reported the status of WP5 while highlighting the ongoing efforts, which includes the task plan for the WP5 sprint in December, to meet the deadline of D5.3 and MS5.3. The development tasks of libpyvinyl, SimEx-Lite, McStasScript and instrument database were assigned to the group members. In his summary, he concluded that WP5 was in line with the planning and that the upcoming sprint will ensure that deliverables and milestones are delivered on time.

WP6 - EOSC Integration: JF. Perrin presented the technical work done and planned in 2022 to improve the EOSC integration of the IT infrastructures within the community. He presented the evolution of the community AAI with the extension of metadata sets, the authorisation mechanism and the engagement with EOSC-Future to deliver the EOSC federation AAI. Data transfer use cases and the associated solutions identified in the course of the project were also discussed alongside software provisioning technology that could also improve the interoperability between the RIs. These technologies were also presented in the scope of the VISA platform deployments that are taking place in the different RIs, they not only ease the integration of VISA in RI's infrastructure but could potentially transform it as a more generic tool (i.e. not tied to the RI were data were produced) by providing solution to bring data and software to the hosting infrastructure.

WP7 - Sustainability: The WP leader, O. De Giacomo started with an update on the latest activity of

the WP but focusing especially on the next deliverables and how to achieve a good maturity in the discussions, since the deliverables are a reflection of the rest of the work being done on the project and the perspective for a future deployment and maintenance of the common solutions. The topic of sustainability goes across all the activities of the project and emerged naturally in all WPs, when they asked how the outputs would be maintained in the future. By the time this meeting was held in Copenhagen, feedback had already been collected from some of the LEAPS members and directors of pan-European RIs through the Association of European-Level Research Infrastructures Facilities (ERF-AISBL). The general feedback from this meeting is that there is no clear common strategy nor full understanding on what providing FAIR data and data services fully entitles, especially in the EOSC. Most facilities are ready and committed to provide services for their own users, but the model to provide costly resources beyond the own user base, free at the point of use, is still far from being clear. The meeting was extremely useful to collect feedback of all the other WPs, on their progress and findings, to review the business models that were under discussion as an internal task of the WP. In particular, it was useful to discuss together with ExPaNDS, since the strategy for the future foresees the use of common tools by all PaN facilities, not just PaNOSC partners.

WP8 - Staff and User Training: The WP8 leader, T.H. Rod, gave an update on WP8 beginning with outlining the work done in WP8. This included additional work that was not anticipated at the time of planning but which nonetheless are important and can be considered to be in the scope of the EOSC (e.g. creation of a training portal with ExPaNDS and integration of UmbrellaID). The major concern raised by the WP leader was that of getting the photon sources in PaNOSC and ExPaNDS to adapt the e-learning platform whereas a major highlight was the collaboration with ExPaNDS on a joint training portal for photon and neutron sources. This platform both contains a training catalogue provided by ExPaNDS and an e-learning platform provided by PaNOSC. Moreover, activities for the next 12 months were outlined. The major future activity was the organisation of a summer school and development of content in the e-learning platform for that summer school.

WP9 - Outreach/Communication and Dissemination/Impact: N. Carboni reported on the progress of the activities carried out in WP9, as well as on the data based on the WP's main KPIs. The communication and dissemination actions foreseen in the WP roadmap up to the end of the project were also presented, in particular towards wider user engagement, as well as to stimulate further collaboration with other EOSC-related projects. Inputs from the various WP leaders have been collected to fine-tune the communication/dissemination actions based on the WPs' specific needs. It was also agreed that WP9 would support WP6 in updating the content for promoting UmbrellaID across the PaN user community.

Meeting 08/12/2021

The first PMC meeting after the Copenhagen meeting took place on the 8th of December and it started with a debrief and general conversation about the meeting in Copenhagen. This was followed by an update on the milestones and deliverables due at the end of November and that were late. A tentative schedule of meetings for January 2022 was also agreed, which enabled a close follow-up of the work done for the Periodic Report.

The summary notes of this meeting[2] (as for all the others) are available in GitHub[3].

Meeting 05/01/2022

The first meeting of 2022 was used to go in detail through all the milestones and deliverables that were due in November 2021. At this point, some of them have been submitted/achieved, while a few were still outstanding.
This meeting was also used to share and organise the work for the Project Review and its Periodic Report (due 29th of January) and a Project Brief (to be submitted by 10th of January).

Meeting 12/01/2022

At this point, only three deliverables and two milestones from November were outstanding and WP leaders committed to completing the work before the end of January. The submission of the project brief was reported as well.

Meeting 19/12/2022

Only a deliverable and milestone were still delayed by the time the meeting took place and the project members in charge of them committed to complete the work within the next 10 days. The attendees confirmed that they were not aware of any conflict of interest with the proposed reviewer for the Project Review and discussed progress and pending actions regarding the Periodic Report.

Meeting 26/01/2022

Only one milestone was still delayed by this date due to a small technical issue. The Periodic Report (due 29/01/2022) was by this time nearly completed and most of the time of this meeting was used to go through its content and identify the last small contributions that were pending.

Meeting 09/02/2022

During this meeting, it was reported that the last milestone still due from November 2021 had been achieved and that the Periodic Report was successfully submitted. First discussions about a possible face-to-face/hybrid meeting over the summer took place and news from the EOSC and EOSC Future were shared. WP leaders also reported on the progress achieved on their respective WPs.

Meeting 23/02/2022

This meeting focused on a general review of WP8. The WP leader T. Holm Rod and A. Götz reported on the progress achieved, the main pending tasks and challenges. A review was made of the tasks and deliverables. The list of outcomes which need to be maintained were identified. These include the two platforms pan-

[2] https://github.com/panosc-eu/panosc/blob/master/Work%20Packages/WP1%20Management/Meetings/Project%20Management%20Committee/2021-12-08-PMC/2012-12-08-PMC_summary.md

[3] https://github.com/panosc-eu/panosc/tree/master/Work%20Packages/WP1%20Management/Meetings/Project%20Management%20Committee

learning.org and pan-training.org (by ExPaNDS) which were in the process of being merged into a joint training platform. The main challenge was and will be the production of quality training material and updating it regularly. The material for photons was identified as still lacking in quantity even if it is slowly increasing. Photon facilities were strongly encouraged to upload existing material and to adapt the training platform for their facilities. A lack of compute resources was identified which should be addressed during PaNOSC by the cloud resources to be procured in WP6.

Meeting 02/03/2022

The meeting started with J. Bodera reporting the feedback obtained from the EC on the Periodic Report submitted in January. The report needed to be resubmitted within 15 days explaining some minor deviations for ESS and EGI. After these initial remarks, the meeting focused on sustainability and the best approach to assess the value proposition and ensure that services do not disappear after the project (and funding) ends.

Meeting 09/03/2022

During this meeting, the possibility of organising a joint PaNOSC- ExPaNDS meeting in Prague was discussed. WP leaders also provided with an update on progress and challenges.

Meeting 30/03/2022

During this meeting, J. Bodera reported that the resubmitted Periodic Report has been accepted and explained the recommendations that will need to be addressed. T. Richter reported that D3.3 was submitted on time. The draft Grant Agreement amendment that was due to be presented at the upcoming Executive Board meeting in April was discussed.

Meeting 13/04/2022

J. Bodera reported on the Executive Board meeting that took place on the 1st of April and it was agreed to formally hold a vote regarding the draft Grant Agreement amendment in the next PMC meeting. A survey from the EOSC Association was discussed and news from the EOSC Future project and ExPaNDS were reported.

Meeting 20/04/2022

This meeting focused on the EOSC Association survey, deciding what would be the Key Exploitable Results (KER) from PaNOSC to be included and the best way to proceed.

Meeting 27/04/2022

The draft Grant Agreement amendment was validated via a formal vote and J. Bodera was mandated to engage with the Project Officer to obtain feedback and submit the amendment. The state of the upcoming milestones and deliverables was discussed and reports were received regarding the planned face-to-face meeting, the EOSC Association survey, ExPaNDS and the EOSC.

Meetings in May 2022

Due to the proximity of these meetings to the date on which this deliverable is due, unfortunately they cannot be reported in detail. It is expected that the meetings will focus on the deliverables and milestones due at the

end of May, the organisation of the face-to-face meeting in June and progress achieved with the Grant Agreement amendment. The minutes of all meetings are available in GitHub (https://github.com/panosc-eu/panosc/tree/master/Work%20Packages/WP1%20Management/Meetings/Project%20Management%20Committee).

Progress of work packages

Work Package 1: Management

Just before the meeting in Copenhagen the last WP1 deliverable (D1.8) was submitted.

As the project entered the last year of execution, WP1 focused on the following actions:

- Coordination between partners and organisation of regular PMC meetings
- Organisation of the Executive Board meeting
- Coordination with ExPaNDS
- Representing PaNOSC at different meetings/conferences
- Communication with the PO
- Liaise with other EOSC initiatives (EOSC Association, EOSC Future and other cluster projects)
- Project monitoring
    - Following progress towards submitting deliverables and milestones
    - Internal financial report
    - Answering ad hoc queries and supporting WP leaders
- Work with WP leaders and partner representatives for the Periodic Report
- Document the recommendations and assign a responsible team member for each one
- Follow-up the implementation of recommendations
- Contribute to the organisation of face-to-face meetings
- Support mailing lists and GitHub management
- Lead the work for the Grant Agreement amendment

The regular meetings and work relationships built since the start of the project enable the WP1 - Management to function well, being proactive and ensuring coordination among project members and with ExPaNDS.

The last six months of project execution will be key to ensure not only that all deliverables are submitted, milestones achieved and a final Periodic Report submitted, but supporting at the same time the work to show and measure the difference PaNOSC is making to the Photon and Neutron (PaN) community.

Work Package 2: Data Policy and Stewardship

The planning for tasks related to Work Package 2 officially stopped in month 36, i.e. November 2021. However, the work continues as some objectives require longer approval processes e.g. updating existing data policies, and the needs of data stewardship require effort constantly.

Concerning the data policies, the main change has been that ELI-ERIC officially approved their FAIR data policy based on the PaNOSC framework and published it on zenodo: Weeks, Allen, Szabó,Gábor, Hvězda, Roman, Gliksohn, Florian, & Ivănoaica, Teodor. (2022). ELI ERIC Data Policy (As adopted by the Extreme Light Infrastructure ERIC General Assembly on 17 March 2022.). Zenodo. https://doi.org/10.5281/zenodo.6515903. The facilities which already had an open data policy before PaNOSC (ESRF, ESS, ILL) have an ongoing process to update their data policies. In the case of the ESRF, the most significant change over the past year has been the increased publishing of processed data. The most significant addition has been Human Organ Atlas: https://human-organ-atlas.esrf.eu which has attracted a lot of interest with multiple researchers downloading the open data and analysing it. The Human Organ Atlas is an excellent example of FAIR data being re-used. It will serve as an inspiration for other scientific domains of the advantages of making data open and FAIR.

The second main topic which has progressed significantly over the last 12 months has been the Data Management Plans. The DMP DS Wizard (https://ds-wizard.org/) developed by Elixir (a member of EOSC-Life) has been adopted by WP2 as the standard tool. ESS has done significant work in preparing the template and providing a solution to deploy DS Wizard. All WP members helped review the template. The ESS and ESRF have deployed and tested the DS Wizard with real users. The ESRF users were positive about the added value provided by the tool and as a consequence a DMP service is being deployed for all ESRF users. The new service will be provided for users who have to prepare a DMP for funders when data are produced at the ESRF. All ESRF proposals will automatically receive a DMP prefilled with information concerning the ESRF data policy and the data produced during their experiment. The DMP will be used primarily to inform users, they will be encouraged to complete the DMP but not be obliged to fill in the DMP. The DMP will be used as a tool to help users. The other members of PaNOSC are planning on similar approaches for implementing DMPs.

European XFEL has established a working group that will coordinate and oversee the update of the Scientific Data Policy, including the introduction of the Data Management Plans. The group falls under the responsibility of the IT and data management (ITDM) and in close collaboration with the data analysis group (DA) and consists of one project lead and one technical lead, including data experts as well as instrument scientists. The group will analyse and report on the current technical implementation, including ad-hoc provisions which fall outside of the published Scientific Data Policy and steps to mitigate any shortage in capacity; propose updated data policy and iterate on it with identified stakeholders; define an adoption process plan for the data policy, including all necessary consultations; prepare the adoption of Scientific Data Policy by Management Board and potentially the Council Ensure that technical steps necessary to implement the Scientific Data Policy are realised, considering the need to optimise data storage efficiency.

Issues which still need addressing in WP2 are to ensure data DOIs are produced at all facilities and that open data are available at all sites. Those sites which have adopted an open data policy recently are encouraged to identify users who have published their results to open their data before the end of the embargo period. Another issue which still needs addressing is training and encouraging users to cite data DOIs. This is a practice which is still not being widely adopted in the PaN community. Training users to do this is an ongoing process.

The ESRF has gone through the process of certification of their data repository (https://data.esrf.fr) by CoreTrustSeal (https://www.coretrustseal.org/) and is waiting for the final approval. Other facilities are encouraged to do the same as part of their ongoing process of implementing trustworthy repositories.

Work Package 3: Data Catalogue Services

In this period, the work package has two deliverables due with the final one following two months after. For that reason, there was a lot of focus on converging the efforts, especially tasks needed to successfully complete the deliverables. The first one, D3.3 Catalogue Service (federation of search APIs), was submitted at the end of March. This deliverable marks an important achievement. It defines the functionality baseline of the PaNOSC Domain Specific Search as one of the core pillars of our open/FAIR data infrastructure. Between the partners there is now a better, shared understanding of data curation tasks, dataset taxonomies and the types of queries on the data repositories that might be relevant. Following the deployment of the Federated Search Service and the associated Web Frontend we have seen significant momentum to add more features and search functionality. This corroborates the usefulness of the iterative development approach that has been taken. Before more features will be considered, though, we need to ensure there are a significant number of compliant implementations of the common API in its most up to date version. And the quality of the search results depends not only on the correct implementation of the search API and the scoring. The metadata corpus held in the local data catalogues needs to follow the best practices and agreed mappings for queries involving experimental technique, parameters, roles, etc.

That means in terms of software development tasks, we expect to be able to resolve a number of smaller issues in the common code base as they arise during the course of the PaNOSC project and for the coming future. Especially operational stability and fault tolerance are aspects that have not had much attention in the proof of concept operations so far. But the main tasks should be to ensure the local catalogues have a fully compliant connection to the federated search and the local curation of metadata is in a good shape at least for new datasets entering the repositories.

The second deliverable D3.5 NeXus Metadata Schema reports on the positive impact of the work package on the NeXus adoption of the partner sites and the effect our contributions had to the community at large. The “Show and Tell” sessions held to survey the status quo at the beginning of the project were very useful specifically to the “newer” facilities, such as ESS, ELI and XFEL. For example the ILL could offer a lot of experience in recording different techniques in somewhat similar ways. XFEL, ESS and ELI have made or prepared individual contributions to NeXus definitions. So the work package had an “upward” effect on the NeXus community. However at least the same amount of work was dedicated to tools and local integrations that are shared as open source projects. The software can be re-used to better write NeXus files with a higher degree of compliance (nexus constructor for example) and commonality or to better visualise and process files using common code (h5web and scipp for example). This is a benefit to the photon and neutron community as a whole.

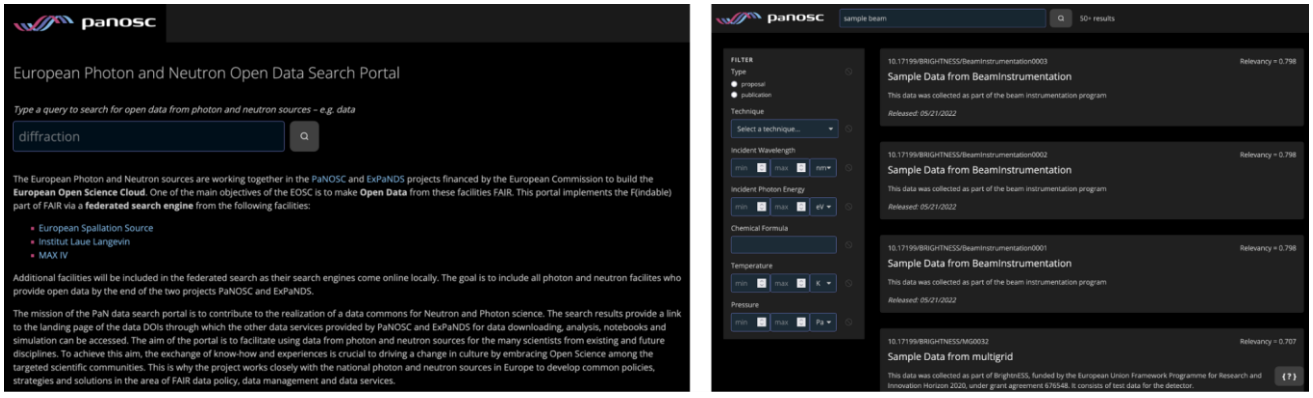
The remaining deliverable in the WP is focussed on metadata quality and the local implementation of the data catalogues at partner sites (OAI-PMH harvesting, integration of data sources, search endpoints and compliance with API definition, etc.) Work on that will start immediate after this reporting period.

Work Package 4: Data Analysis Services

VISA development by ILL colleagues (https://github.com/ILLGrenoble) has continued with version 2.2.1 as latest product. Since the last report, additional features for the VISA administration user interface were added, a more generic user authentication management was implemented and improvements for the Jupyter service were made.

The deployment at PaNOSC sites has seen progress: ESRF have an agreement with two of their beamlines as first users of a production instance of VISA. The microservice deployment has been accomplished, and required software for the needs of the scientists has been installed into the virtual machine image by making use of a Singularity container approach. ESS has VISA ready as service using virtual machine instances based on the demo image. As such, it is already used by the ESS instrument data scientists and software developers. European XFEL have a deployed VISA instance (visa.xfel.eu) interfacing with the DESY OpenStack cloud infrastructure, where microservice (re-)deployment is managed by a Gitlab/Kubernetes continuous integration workflow. The installation is currently providing three possible virtual machine images in demonstration state, one providing crystallographic software. Deployment of VISA at CERIC-ERIC with an interface to ProxMox infrastructure for the virtual machine instances is close to completion.

As the result of an in-person developers meeting at ESRF in April, the federated search portal has been given a new front-page and web interface design, and is accessible under the URL https://data.panosc.eu. Integration of the web frontend with the federated search API and the metadata catalogue API endpoints has progressed since then.



Federated search portal home page (left) and result page with filter options (right).

Other developments and use cases: The h5web application has been fully integrated to the ESRF data portal as a generic service to browse the Nexus structure of HDF5 files and visualise the data. PaNOSC facilities have meanwhile deployed the h5web tool at their own sites, either for graphical HDF5 display from a JupyterHub-based file browser or even with integration to their data catalogue. Posing a different solution for a similar use case, CERIC-ERIC have published h5nuvola (https://gitlab.elettra.eu/panosc/h5nuvola), a framework for the purpose of HDF5 file browsing and visualisation in a web front-end with general-purpose graphical plug-ins, and/or automatic redirection to dedicated applications for context-specific data analysis tasks (e. g. X-ray fluorescence). h5nuvola is integrated to the user office system VUO at the Elettra synchrotron.

Work Package 5: Virtual Neutron and x-raY Laboratory (VINYL)

The activities in PaNOSC WP5 have focused on developing further the python library “libpyvinyl” which will serve as a foundation for scattering simulation packages, and in this way harmonize their user experience. ESS took part in design, development and documentation of libpyvinyl as well as in the McStasScript package which is a python API for the McStas neutron simulation tool that was started in PaNOSC WP5, and development of this API has continued. This year it was moved onto the libpyvinyl foundation and many new features were added. These new features range from recognizing common McStas errors to full tools that help build complex sample environments. McStasScript also got a fully rewritten documentation presented on a github hosted webpage which have had a large influence on adoption of the code. Several talks were held, mostly in virtual meetings, also with the aim of increasing the userbase of the software. ESS also contributed to the design of the instrument database (see below) which will be the most common entry point for users that need to just run a simulation, and thus potentially has a much larger userbase.

The **instrument database** has been revisited with the newer release of the libpyvinyl library. Instruments at facilities are now described as a python script following the libpyvinyl API and using McStasScript and SimEx as backends and collected in a github repository. A detailed guide for contributors is provided to allow instrument experts to populate further the repository with detailed descriptions of existing instruments.The committment of experts from the different facilities ensures the mid and long term sustainability of the repository and its usage. Users running the simulations are facilitated by the common libpyvinyl API that harmonizes and hides the complexities of the single backends and by a dedicated API for accessing the instrument database and the instrument descriptions.Further flexibility is offered to the facilities that are able to clone and maintain their own repository for specific use cases, but still using the same API to access the simulations.

At CERIC, work from Task T5.3 (Photon and Neutron beamline simulation data services) continued in the scope of which CERIC has already developed and demonstrated Oasys as a web service. Further integration with the rest of the WP5 services was pursued. This specifically calls for a better linkage between the graphical user interface (GUI) of Oasys and a scripting interface of the notebooks (Simex). An easy-to-deploy solution has been found with the support of CERIC's WP6 where a desktop can be run within the Jupyter lab instance. This allows for interoperability of notebooks and traditional GUI programs, both sharing the user- and file-space (i. e. datasets, libraries, authentication). Furthermore, the whole solution has been provided as a Docker container. Even though the initial prototype has included only Oasys software (as per project call), we have correctly predicted that the solution is generic, as McStas has afterwards also been successfully packaged using the same tech stack. The container is already running as a part of WP8 pan-learning portal. This has exposed our previous work on the beamline database to a wider audience and is paving the way towards integration of Simex and Oasys, of which a small demonstrator has already been developed. Considering the above, CERIC is tackling the final task, Task T5.5 (Integrated simulation data workflows), by connecting and integrating all the so far developed pieces both within the Oasys ecosystem as well as fullfilling the Panosc objectives of simulation codes interoperability and fully web-based access.

At European XFEL, the design and implementation of the BaseData and BaseFormat classes for libpyvinyl were carried out. The BaseData class is a representation of the data flow between two calculators. A specialized Data class can be created for a kind of data with similar attributes based on the abstract BaseData class. The

abstract class provides useful helper functions and a template for the Data interface so that no specific read and write functions are needed to develop the specialized Data class. Only the required keys of data need to be defined. A Specialized Data class connects to multiple specialized format classes, which are responsible for communications between the python objects in the Data class and the data stored in the specific format on the hard disk. EuXFEL also contributed an example instrument to libpyvinyl as its integration test and a demonstration of its usage. The simple example shows how one can create a complex simulation instrument easily with libpyvinyl. The relevant parts of the documentation of libpyvinyl were also contributed by European XFEL. Adopction of the latest libpyvinyl in SimEx-Lite, the lightweight python API for the platform to simulate the experiments at advanced photon sources, is ongoing. The users can then simulate the single-particle imaging experiment easily with the python package.

Work Package 6: EOSC integration

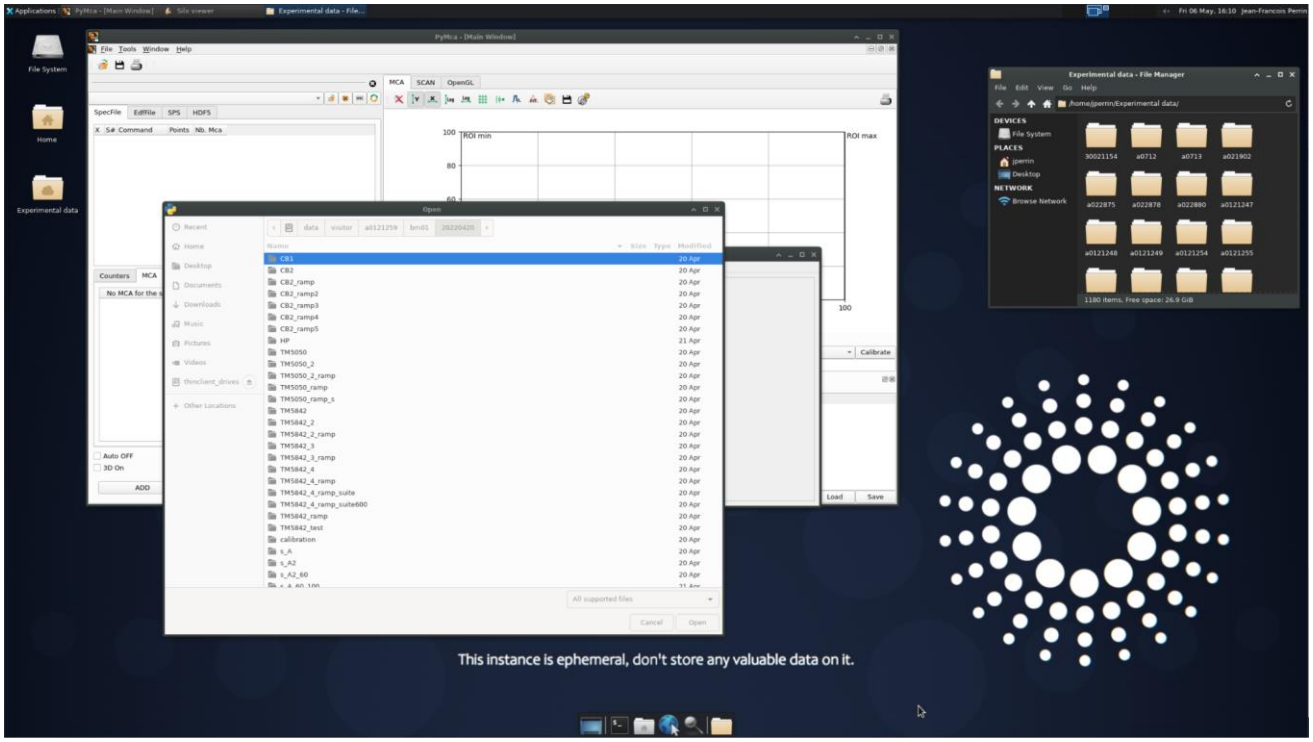
After the adoption of the data transfer solutions by the user community, the last year of the WP6 efforts are dedicated to the onboarding of researchers to PaNOSC services, more precisely UmbrellaID and VISA.

Efforts on UmbrellaID are concentrated on the comprehensive outreach to the PaN community RIs as well as actively participating in the development of the EOSC AAI federation as driven by the EOSC Future project.

An UmbrellaID online workshop https://indico.psi.ch/event/12701/ was organised in collaboration with the ExPANDS project in May 2022. This workshop was dedicated to the PaN community IT staff members who want to enable community AAI for their users. This workshop included presentations on UmbrellaID, EOSC AAI, and a hands-on session on connecting services to UmbrellaID. This was a very successful training attended by more than 40 infrastructure experts and software developers from the PaN community.

The cloud procurement using the OCRE framework has been completed and ESRF is now expecting to have amazon cloud access very soon. ESS, ELI and ESRF have provided use cases for this access.

VISA service has been deployed at the ESRF and has received very positive feedback for training purposes and data analyses from the ESRF scientists; it should be open to the large community of users in June 2022. For ESRF it was necessary to adapt VISA to include access to GPUs and to HPC clusters in order to meet users requirements. These developments are now available for others involved in deploying VISA. WP6 has also been following up with other PaN RIs for deploying VISA service.



Screenshot of VISA instance at ESRF.

While deploying VISA we also quickly understood the need to facilitate software provisioning, We survey the the technologies used by the other communities and retained for further exploration the use of light containers like Apptainer (formerly Singularity) and CVMFS (CERN VM FS) for making software available to RIsl HPC and remote clouds.

WP6 activities were also presented at the Science Mesh Workshop "Global Platform for Scientific Collaboration" organised by the H2020 CS3MESH4EOSC project in January 2022.

Work Package 7: Sustainability

This work package kept tracking closely the development of the other activities of the project. The deliverable on metrics and costs of the EOSC was finalised, and then reviewed internally due to the feedback received by the project officer. On the basis of this deliverable, a shorter publication is being prepared, to improve the presentation of the information. Although the suggestion was to prepare a blueprint, this may not be possible. However, the publication focuses on the main outcomes and obstacles, and describes the methodology used, to make it as similar as possible to a blueprint. We expect that this publication, once ready, will increase the visibility of the information contained in the deliverable.
In addition, the business models presented or hypothesised in the past were reviewed in function of the developments and the context. These business models are being tested with the canvas approach and will be defined for consultation with the directors of facilities and other relevant funders. This feedback will be used to define the proposal for the sustainability of the PaN EOSC (last deliverable).
In parallel to these activities, the partners have contributed by keeping a straight liaison activity with the other four cluster projects, and actively contributing to some task forces of the EOSC Association. In particular, the partners were actively involved in the task force “sustaining EOSC” where one stream studies the sustainability of the federation of data. To this stream, PaNOSC partners have contributed with the point of view of the Research Infrastructures that run physical infrastructures (not only data), in defining the costs involved and possible models for the sustainability. This exchange has also been very informative for WP7, since the task

force collected use cases of communities that managed to be sustainable, therefore good practices examples.

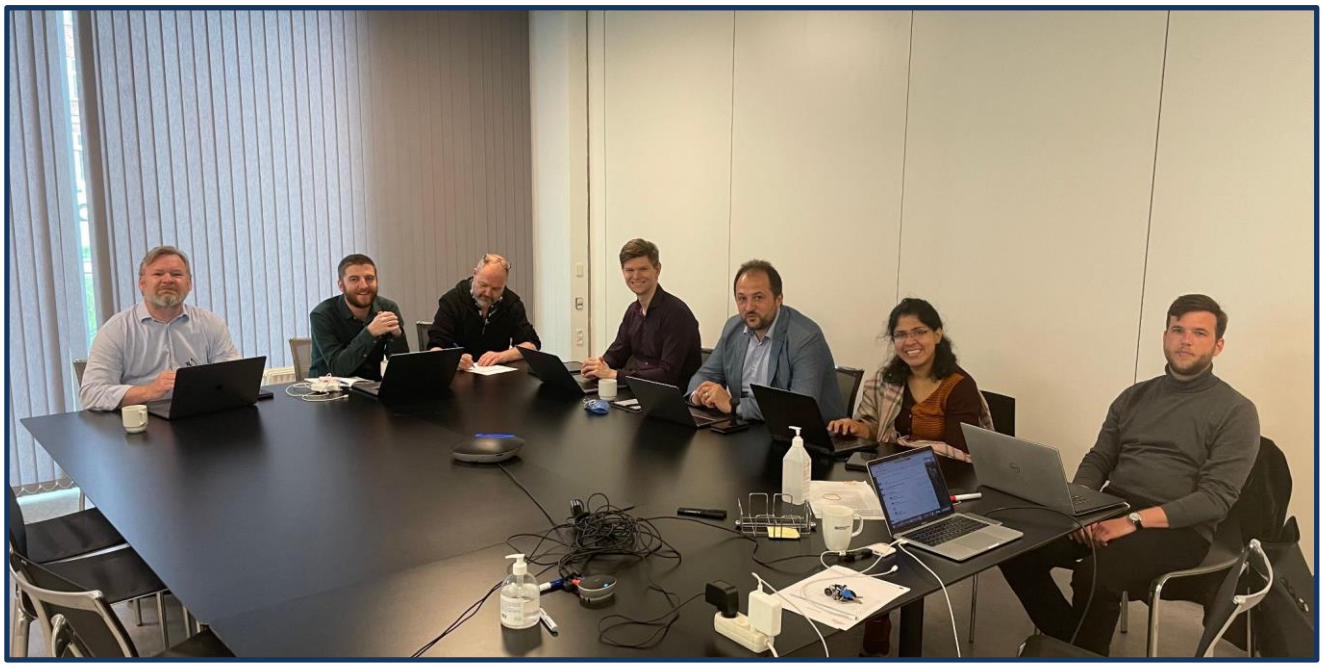
Work Package 8: Staff and User Training

During the last six month we have seen an increased upload of material for photon sources, particularly in relation to MS8.5: e-learning virtual facilities, which have resulted in a number of Jupyter notebooks executable from the e-learning platform. These in turn form the seeds for a subsequent summer school (Task 8.8). A Hercules 2022 course: Introduction to OASYS is now developed as a moodle course. Furthermore, engagement beyond the PaNOSC/ExPANDS community has been shown with utilisation by staff at Vrije Universiteit Amsterdam to deliver a course including executable Jupyter Notebooks.

A major achievement is the launching of the training portal https://pan-training.eu in collaboration with ExPaNDS. The training portal contains both a training catalogue delivered by ExPaNDS and an e-learning platform, pan-learning.org, provided by PaNOSC. Moreover, the entire platform, incl. pan-learning.org, has been redesigned. As a result pan-learning.org is now leaner and easier to maintain; The WordPress layer, inherited from https://e-neutrons.org has for instance been removed. As a result of the merge with the training portal, https://pan-learning.org now redirects to https://e-learning.pan-training.eu. A major and related achievement is that UmbrellaID can now be used for authentication and authorization of users of training portal.

Deliverable 8.1 was completed end of 2021. The corresponding report gives an overview of the current status of implementing FAIR at the PaNOSC partner facilities including associated training and concludes with lessons learned.

A workshop (https://indico.esss.lu.se/event/3015/) took place in Copenhagen 11th-12th May, 2022 with the sole purpose of organising a PaNOSC summer school (Task 8.8). The two major resource holders in WP8, ESS and ELI as well as CERIC-EIRC, participated face-to-face, whereas ILL and ESRF joined online.



WP8 workshop in Copenhagen. From left to right: T.H. Rod, A. McCluskey, P.K. Willendrup, M. Bertelsen (ESS), T. Ivănoaica, M.U. Kahaly (ELI), A. Hafner (CERIC-ERIC). Photo taken by P. Aulin (ESS) who also participated.

As a result of the workshop, the summer school is now scheduled to take place September 12th-16th in Szeged, Hungary at ELI-Alps. A programme with speakers was drafted and subsequently approved by PMC at a meeting May 18th, 2022. The summer school will make use of blended learning using the developed https://pan-training.eu system for hands-on exercises. Deliverable 8.3 (Teaching material for users) has been delayed to match summer school given that they are intrinsically linked – as additional teaching material is being developed for the school.

Finally, the leadership of WP8 has been handed over to Andrew McCluskey, an Instrument Data Scientist at ESS, who also chaired the aforementioned workshop.

Work Package 9: Outreach/Communication and Dissemination/ Impact

In the frame of WP9, the regular activities carried out in the previous period have been further implemented between November 2021 and May 2022.

These implied continuous interaction with all WP leaders to promote the main outcomes and events in the various WPs through the project's online channels and mailing lists.
Jointly with ExPaNDS, a proposal to take part in the ICRI2022 conference with a satellite event has been submitted. If accepted, our 3rd PaN EOSC Symposium would be added to the ICRI2022 programme, thus increasing the possibility of also attracting ICRI's participants to our final event.
Another proposal was submitted to the science dissemination event ESOF 2022, taking place in Leiden in July 2022, with the title: *Open Data for healthier societies: a virtuous cycle?*, which was selected to be included as an online session under the theme Healthy Societies as part of the ESOF2022 Programme.
Moreover, WP9 has been collaborating with PaNOSC WP1 and with ExPaNDS, to organise a face-to-face meeting in the last semester of the project, and involving WP leaders of the two projects to discuss the next steps to be taken before the projects end.

For a wider outreach to the user community of users of photon and neutron facilities, user meetings and workshops have been targeted also in this semester, and the scientific data management system developed within PaNOSC, and specifically at ELI-ERIC, was presented at the ELI Alps user workshop in November 2021.
To further promote the activities, disseminate the results of the project among the PaN user community, and discuss sustainability, contacts with the LENS and LEAPS initiatives have also been continuous.
Several users learnt about OASYS, one of the services developed within PaNOSC for experiments' simulation and remote access, in a dedicated session at the Hercules school 2022.

In line with the plan to increase the visibility and enhance the adoption of FAIR data practices across the PaN community, some of the services developed in the project have been presented through video interviews and demos – published on the PaNOSC YouTube channel and the website, and distributed via PaNOSC and the partners' social media channels, newsletters and mailing lists:

- Video interview with Andrea Lorenzon on H5nuvola for HDF5 data visualisation and analysis (+DEMO): https://youtu.be/BRYKtunF98k
- Video interview with Dr Claire Walsh on the Human Organ Atlas: https://youtu.be/VqGvjOpmsBo

In addition, in WP9, interviews with PaN users on their views on Open Data and Open Science were released:

- Video Interview with Dr Stella d'Ambrumenil on Open Data: https://youtu.be/eMf9R7nE52o
- Video interview with Alessandra Gianoncelli on FAIR data: https://youtu.be/5KAzaFp1jPk
- Video interview with Mousumi Upadhyay Kahaly on FAIR data: https://youtu.be/eHFYl9HEInU

Four out of the five above-listed videos have also been included in the Women in Science section of the PaNOSC website, to highlight the contribution of women to scientific research and to the adoption of FAIR data practices.
On the basis of the metrics collected on PanOSC's YouTube channel, as of 3 May 2022, all PaNOSC published videos reached over 3988 views, with a 28,6% increase over the previous reporting period.

Also, eight new factual examples (reaching a total of 30) of the use of the services being developed in the project were published to showcase current practices in data stewardship, data transfer, (remote) data analysis, and data and experiments' simulation. One of the newly published use cases relates to VISA, the Virtual Infrastructure Scientific Analysis. Initially developed as part of PaNOSC, VISA has been seen as a flagship product for performing remote data analysis with its benefits being widely adopted. However, the platform is not limited to Photon and Neutron facilities, and can also be used in other research domains.
VISA was also presented at the 14th International Conference on Synchrotron Radiation Instrumentation[4] held in March 2022, which targeted the worldwide synchrotron radiation and free-electron laser community.

Collaboration with other EOSC projects has also been continued. PaNOSC contributed to the Science Mesh workshop in January 2022, targeting researchers, educators, data curators and analysts, as well as system developers, providers and administrators, and policy makers, to highlight the needs of ESFRI clusters and their related infrastructures in terms of data synch and share. In the same month, PaNOSC also took part in the 3rd ESFRI-EOSC Workshop on RIs and EOSC, which brought together ESFRI, ESFRI RIs and EOSC stakeholders, to showcase and better comprehend the EOSC concept, including the Open Science and "FAIR" policy agenda, and the vision for the future, along with the EOSC value proposition for its users, and ensure an optimal federation of clusters and RIs with EOSC.
Finally, in April 2022, PaNOSC was invited to join the closing session of the SSHOC conference, where Rudolf Dimper highlighted the need to join forces across clusters to reach interdisciplinarity, to exchange knowledge and expertise, and speaking up as a single voice to stakeholders.
Over the entire reporting period, PaNOSC contributed actively to the Science Cluster coordination meetings. The main recurrent subjects concerned sustainability of the project outcomes, structuring of the interactions with the Commission and the EOSC Association, and the active participation in the EOSC Future project.

Finally, KPIs in WP9 have been continuously monitored, with the goal of further fine-tuning communication and dissemination actions within the project.

[4] https://www.sri2021.eu/

Comparison between actual and forecasted project status

The following list shows all deliverables and milestones with their current status:

Milestone or Deliverable Id	Name	Due Date	Status
D1.1	Project initiation documentation	31-Jan-2019	Submitted
D1.2	Data Management Plan	31-May-2019	Submitted
D1.3	Mid-year summary 1	31-May-2019	Submitted
D1.4	Report of annual workshop 1	30-Nov-2019	Submitted
D1.5	Mid-year summary 2	31-May-2020	Submitted
D1.6	Report of annual workshop 2	30-Nov-2020	Submitted
D1.7	Mid-year summary 3	31-May-2021	Submitted
D1.8	Report of annual workshop 3	30-Nov-2021	Submitted
D1.9	Mid-year summary 4	31-May-2022	Submitted
D1.10	Report of annual workshop 4	30-Nov-2022	
D2.1	PaNOSC data policy	31-May-2020	Submitted
D2.2	DMP Template	30-Nov-2021	Submitted
D2.3	Guidelines on implementing Data Policy	30-Nov-2020	Submitted
D2.4	Integration of the policy	30-Nov-2021	Submitted
D3.1	API definition	31-May-2020	Submitted
D3.2	Demonstrator implementation	31-Mar-2021	Submitted
D3.3	Catalog service	31-Mar-2022	Submitted
D3.4	Implementation Report	31-Jul-2022	
D3.5	NeXus Metadata Schema	31-May-2022	Submitted
D4.1	Report data analysis capture	30-Nov-2019	Submitted
D4.2	Prototype remote desktop and Jupyter service	31-May-2020	Submitted
D4.3	Remote desktop and Jupyter analysis service deployed at EOSC	31-May-2022	Pending
D4.4	Publicly accessible Demonstrator	30-Nov-2022	
D5.1	Prototype simulation data formats	30-Nov-2019	Submitted
D5.2	Documented simulation APIs	30-Nov-2020	Submitted
D5.3	Documented simulation tasks executable	31-May-2022	Submitted
D5.4	Software tested and released including interactive simulation and analysis workflow	30-Nov-2022	
D6.1	Data-hub	31-May-2020	Submitted
D6.2	Compute cloud	30-Nov-2021	Submitted
D6.3	AAI	30-Nov-2021	Submitted
D6.4	Software catalogue	30-Nov-2020	Submitted
D6.5	Report on EOSC integration	30-Nov-2022	
D6.6	Integration of RIs in EOSC	30-Nov-2022	
D7.1	Photon and Neutron EOSC Stakeholder Feedback	31-May-2020	Submitted
D7.2	Photon and Neutron EOSC metrics and costs model	30-Nov-2021	Submitted
D7.3	Photon and Neutron EOSC Business model reference document	31-May-2022	Pending
D7.4	Photon and Neutron EOSC Sustainability plan	30-Nov-2022	
D8.1	Report on lessons learned and future prospects for adopting best practises data stewardship at the PaNOSC facilities	31-Jul-2021	Submitted
D8.2	Report on lessons learned for adopting the e-learning platform at the PaNOSC facilities, task 8.4	31-May-2022	Pending

D8.3	Teaching material for users of PaNOSC services, FAIR principles, and the PaNOSC facilities accessible in the e-learning platform at pan-learning.org, task 8.5-7	31-May-2022	LATE
D8.4	Closing report including report from summer school, task 8.8	30-Nov-2022	
D9.1	PaNOSC's Communication and Dissemination Plan	30-Jun-2019	Submitted
D9.2	PaNOSC's Website	31-May-2019	Submitted
D9.3	PaNOSC's repository for internal communications	28-Feb-2019	Submitted
D9.4	Dissemination and Outreach activities	30-Nov-2022	
D10.1	POPD - Requirement No. 1	31-Dec-2018	Submitted
MS1.1	Project Initiation Stage completed	31-Jan-19	Achieved
MS1.2	First Annual Report	30-Nov-19	Achieved
MS1.3	Second Annual Report	30-Nov-20	Achieved
MS1.4	Third Annual Report	30-Nov-21	Achieved
MS1.5	Final Annual Report	30-Nov-22	
MS2.1	First version of PaNOSC DP Framework	30-Nov-19	Achieved
MS2.2	Adoption of PaNOSC DP framework	30-Nov-20	Achieved
MS2.3	Implementation of PaNOSC DP framework	30-Nov-21	Achieved
MS3.1	Survey of Catalogue APIS and Roadmap to EOSC Integration	30-Nov-19	Achieved
MS3.2	Anthology Feedback to API Tasks	30-Nov-19	Achieved
MS3.3	Catalogue Integration Best Practices Meeting	31-May-21	Achieved
MS4.1	Prototype data analysis services completed	31-May-20	Achieved
MS4.2	Data analysis services accessible through EOSC	31-May-22	Pending
MS5.1	Simulation codes in PaNData Software Catalog	31-May-19	Achieved
MS5.2	Demonstration of simulation services	31-May-20	Achieved
MS5.3	VINYL Software release	31-May-22	Achieved
MS5.4	Validation of simulation services	30-Nov-22	
MS6.1	Implementation of AAI integration at the level of the Identity providers	30-Nov-21	Achieved
MS6.2	First release of PaNOSC services	31-May-20	Achieved
MS6.3	Second release of PaNOSC services, data and resources	30-Nov-21	Achieved
MS7.1	Stakeholder database ready	31-May-19	Achieved
MS7.2	First Sustainability Plan	30-Nov-22	
MS8.1	Joint WP4 & 8 plan	31-May-19	Achieved
MS8.2	Joint WP5 & 8 plan	31-May-19	Achieved
MS8.3	pan-learning.org up running	29-Feb-20	Achieved
MS8.4	Jupyter integrated with e-learning platform	31-May-21	Achieved
MS8.5	e-learning virtual facilities	30-Nov-21	Achieved
MS9.1	PaNOSC's Website Ready	31-May-19	Achieved

Please note that deliverable D8.3 (Teaching material for users) has been delayed to September 2022 (~ 3.5 months delay) to match the planned WP8's summer school, as they are linked – as additional teaching material is being developed for the school.

Next steps

Work has already started on planning the next hybrid meetings (face-to-face encouraged but enabling remote participation), with dates in June and October being considered (ELI being the hosting organisation)I. These events will be joint events with ExPaNDS, further cementing the good collaboration between the two projects.

PaNOSC aims to continue delivering for the rest of the project as it has been doing so far, with special focus on

services that add value to the PaN community and thus become sustainable:

- The VISA PaN Portal for data analysis
- The Search API within the PaN Portal
- AAI with UmbrellaID for the PaN Portal, e-learning platform and Search API
- Increasing the amount and usefulness of training content in the e-learning platform
- Implement the new data policy at as many RIs and instruments as possible, maintaining the alignment whenever possible, amongst facilities
- Increase awareness of PaNOSC in the PaN Community by gathering more scientific use cases.
- Increase awareness on the funders about the needs for funding to maintain PaNOSC outputs viable and up to date in the future.

However, several challenges remain:

- Achieve the production stage for the PaNOSC developments
- The interaction and technical integration with the European Open Science Cloud (EOSC)
- Engagement of certain stakeholders (e.g. users) in open science in general and with the tools developed by the project in particular.
- Collaboration with other Science Cluster projects.

In order to address these challenges:

- PaNOSC project members are working hard to ensure that key services are in production stage before the project ends
- PaNOSC is publishing real use cases and engaging with the community
- PaNOSC is collaborating with ExPaNDS
- PaNOSC representatives are regularly attending EOSC events, participating in EOSC Future, and liaising with the other cluster projects.