

Sustainability Policy Report

Document Control Information

Settings	Value
Document Identifier:	D1.7
Project Title:	ExPaNDS
Work Package:	WP1
Document Authors:	Mirjam van Daalen (PSI), Valentina Piffer (PSI), Patrick Fuhrmann (DESY), Sophie Servan (DESY)
Responsible Partner:	PSI
Doc. Issue:	1
Dissemination level:	Public
Date:	18/03/2020

Abstract

This document describes the mechanisms and policies the ExPaNDS collaboration is envisioning to make the achievements of their activities available after the end of the project funding period.

Licence

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit creativecommons.org/licenses/by/4.0/ or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.



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	Comment	Author/Partner
0.1	03.03.2020	First Draft	Valentina Piffer, Mirjam van Daalen / PSI
0.2	11.03.2020	Internal review	Patrick Fuhrmann / DESY
0.3	12.03.2020	Second draft	Valentina Piffer, Mirjam van Daalen / PSI
0.4	16.03.2020	Final version ready for review	Patrick Fuhrmann, Sophie Servan / DESY
0.5	17.03.2020	Internal review	Daniel Salvat / ALBA
1	18.03.2020	Final version	Sophie Servan / DESY

EXECUTIVE SUMMARY

ExPaNDS offers a unique opportunity to produce a sustainable ecosystem of EOSC enabled Photon and Neutron (PaN) standards, data sources and services as it harnesses disparate activities to build on and develops existing initiatives and tools within national PaN Research Infrastructures (RIs).

By making PaN data and PaN data services federated and easily accessible through EOSC portals, marketplaces and regulations, the ExPaNDS project makes them more valuable and more utilized, and with that increases the probability of them to be sustained after the project's lifetime.

ExPaNDS coordinates and builds on existing networks and infrastructures. It will bring together the various sections of Photon and Neutron data management, currently existing within national or facility based catalogues and analysis services, under a single EOSC service umbrella. Adoption and sustainability of an EOSC federated model will be promoted through extensive training in data management, service delivery and service management. ExPaNDS will participate in the EOSC governance to ensure its services are well integrated and aligned to EOSC objectives.

During the project's lifetime, all ExPaNDS services are maintained and operated by partner facilities. The goal is to make those services indispensable by improving their scalability, maintainability and visibility.



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

1. Introduction

A considerable challenge to all projects is the question of sustaining its precious results. The focus of ExPaNDS is to coordinate activities to enable national Photon and Neutron Research Infrastructures (PaN RIs) to make the majority of their data 'open' following FAIR principles and to harmonise their efforts to make their data catalogues and data analysis services accessible through the European Open Science Cloud (EOSC).

ExPaNDS' RIs curate many tens of petabytes of data annually from a large variety of experiments. Individual facilities maintain their own data catalogues that provide access to experimental data sets, metadata and documentation. These catalogues include curated metadata, search indexes and download mechanisms. Any scientist can use these catalogues to search for experimental data and request access to the data if still under embargo or download them if embargo periods have expired. What is currently not available is the integration of these catalogues into a consolidated and federated resource, which would enable a researcher to explore a broader set of data. Moreover, the ExPaNDS community has a rich collection of scientific software encompassing data reduction and analysis, referenced to by the PaN software catalogue¹.

Enabling our organisations and users to access services from the EOSC is an ongoing effort that provides great scientific opportunities. Expanding the services to be available through a cross-European solution will not only build on existing strengths available within the RIs, it will also create enhanced and sustainable user accessibility. Making data available to new users via an Open Data policy provides opportunities, through the EOSC, to broaden the exploitation of the data and the range of science delivered. In doing so, ExPaNDS will act as a facilitator of the EOSC and will help to create a culture of Open Science in the numerous domains that are served by the Photon and Neutron user communities and thus propagate Open Science much more broadly. Data analysis involves large amounts of time, great effort and is of high value once achieved. The federated data services provided by ExPaNDS will reduce unnecessary repetition of research and enable further advances in science and technological innovation through these shared resources.

Through the EOSC, ExPaNDS offers an opportunity to overcome the existing fragmented data management and exploitation across national RIs and so harmonise, optimise and advance the use of data produced at our sites. The EOSC will enable a coherent set of services of data services within a common data policy framework at European RIs to give scientists better tools to fully exploit their data and facilitate the use of Open Data. As demonstrated by regular PaN user surveys (2012, 2014, 2016), scientists are increasingly working across several different RIs to perform their research. Harmonisation of metadata, interoperability of services and standardisation have become of the utmost importance. This is the primary goal of ExPaNDS. Ensuring that data from different RIs are made interoperable, not only simplifies and improves the user experience, it also opens up access to new groups of academic and industrial users, with no or little, previous experience in Photon or Neutron techniques.

The project aligns with the EOSC strategy to federate existing research and scientific infrastructures into a European wide platform of cloud-based services for Open Science, as set out by the European Commission. PaN RIs, as national facilities with user-focused services and a user community of more

¹ <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 857641.

than 30'000, are service providers with a high relevance for the EOSC. Through ExPaNDS, coherent FAIR data services will be enabled to the scientific users of PaN RIs, universities and even industry.

It is essential that ExPaNDS gathers feedback and cooperates with the EOSC governance bodies

- to ensure its services are well integrated and aligned to EOSC objectives,
- to improve the EOSC and develop standard relationships and interconnections between scientific publications, PaN scientific datasets, experimental reports, instruments and authors (via ORCID).

To that end, ExPaNDS is taking an active role in the EOSC stakeholder forum and other bodies such as the different technical Working Groups (WG) of the EOSC Executive Board on Architecture, FAIR, Sustainability, Rules of Participation, Landscaping, Skills and Training. These WGs are part of the EOSC Governance structure, ensuring a community-sourced approach to the current challenges of the EOSC.

Well-defined “Rules of Participation” in EOSC, securing the integration of only high quality services are key to the success of the planned federated infrastructure and are currently worked out by the corresponding EOSC WG.

The sustainability of the ExPaNDS project is directly connected to and influenced by the outcome of the “Sustainability WG” of the EOSC EB, which can provide us with precious input on which we can build upon. Concretely, we expect the WG to deliver a set of recommendations concerning the implementation of an operational, scalable and sustainable EOSC.

A Tinman Report, “Solutions for a sustainable EOSC”, was produced by the “Sustainability WG” in December 2019, exploring possible means for sustaining the EOSC beyond its initial phase which terminates at the end of 2020. It considers the financing model, legal vehicle, governance structure as well as the regulatory and policy environment of the EOSC with a first iteration to establish a Minimum Viable EOSC (MVE) addressing the needs of publicly funded researchers exploiting openly available data. Subsequent iterations expand the EOSC to address more sophisticated use cases and a wider user base including the public sector and industry.

2. Sustainability Policies

ExPaNDS consists of six work packages, where WP3 and WP4 form the technical outward facing part of the project, delivering federated data catalogue (WP3) and data analysis (WP4) services from the national RIs and making them available as services through the EOSC. They are supported by WP1 Project Management and Sustainability, WP2 Enabling FAIR data for national RIs, WP5 training and WP6 outreach.

1. WP1: Management and sustainability

WP1, Management and Sustainability, implements the project’s governance and management structure, undertaking overall project administration to ensure successful achievement of the project’s objectives. It addresses sustainability of the EOSC provided services, supporting the technical WPs in making their products sustainable.



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

WP1 closely engages with EOSC governance structures to ensure representation and sustainability for the national RIs, and will have direct involvement with other EOSC stakeholders in the EOSC Liaison Group. Throughout the ExPaNDS programme of work the project will liaise with other EOSC projects and activities, seizing the opportunities offered by the community to build upon.

On a more pragmatic approach, the management tools and mechanisms themselves are being developed taking FAIR into account, which will enhance the ease and use of the project, supporting the activities, service providers and users. For example, WP1 makes sure the project's key deliverables are published in community-used platforms like Zenodo to make them findable in the long-term.

2. WP2: Enabling FAIR data for national RIs

WP2, Enabling FAIR data for national RIs, extends and deepens the adoption and use of FAIR data principles within the Photon and Neutron community to allow publication and access of national RI data and services within the EOSC.

Ensuring such large quantities of data to be *Findable, Accessible, Interoperable* and *Reproducible* will provide a wealth of instantly available knowledge and processes, resulting in making scientists and researchers work simpler. They will be able to start any piece of work from a scientifically higher level of knowledge rather than having to create their own data prior to their research activities commencing at an RI.

Within this WP, ExPaNDS will make sure that all facilities involved will adopt the FAIR principles and with that increase the chance of their data to become sustainable. This will be done for example by implementing a **common data policy framework** to be adopted by all ExPaNDS facilities and providing consistent **DMP templates** for instruments, to be used for each new experiment to make the produced data as FAIR as possible.

Additionally, by feeding into existing **standardisation initiatives**, like the Research Data Alliance (RDA), ExPaNDS definitions and specifications will be adopted and extended by the communities beyond the end of the project's lifetime.

3. WP3: Data catalogue services

WP3, EOSC data catalogue services, enables the implementation of harmonised policies, practices and guidelines coordinated in WP2, as well as provide catalogue services for WP4 (EOSC data analysis services). It facilitates federated access to metadata, experimental data and related information.

Concretely ExPaNDS proposes to standardise and link all the relevant PaN RI catalogues to ensure that the user community has access to both the raw data they collect, which is linked to their research session at the various national RIs and to the relevant peer reviewed articles, produced as a direct result of their usage.

For a detailed technical architecture of the ExPaNDS project on how to interface and federate current and future services, locally provided by national PaN infrastructures, to a high level scientific service, accessible via state-of-the-art cloud mechanisms, please refer to ExPaNDS D.1.6 "General Architecture description in relation to the EOSC services" (DOI: [10.5281/zenodo.3697703](https://doi.org/10.5281/zenodo.3697703)).



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

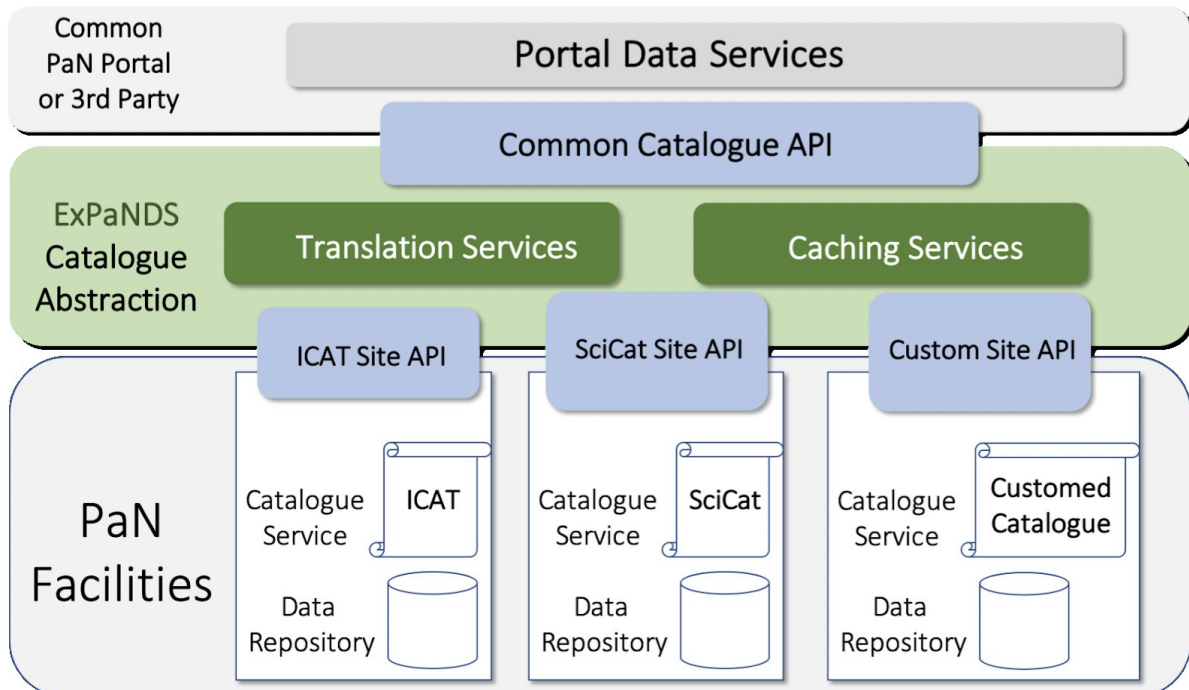


Fig. 1: ExPaNDS architecture - Catalogues services (Search)

Many national RIs have already catalogues of experiment data that have their own lifecycle, such as for example metadata catalogues like ICAT and SciCat (see Fig. 1). WP3 will coordinate with these EOSC-ready initiatives to make services available long-term to all users by providing best practices and cross-site support.

For instance, it is paramount that we develop a **common ontology** to fully integrate all the elements of the catalogues as well as a roadmap for the back-end architecture and functionalities. We will also develop a powerful taxonomy strategy in line with the requirement of the EOSC user community.

By building on existing links with communities such as **PaNOSC**, standards such as the **NeXus meta data format** and services as provided by OpenAIRE, this WP will enable and coordinate the adoption of agreed standards at national science community levels.



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

4. WP4: Data analysis services

WP4, EOSC data analysis services, exploit the deliverables provided by WP2 and WP3 that provide FAIR-ready data through established metadata catalogues. The focus of this work package is to coordinate, adapt and align existing data analysis services at national RIs with the new EOSC services available through the EOSC Service Catalogue and EOSC Portal.

Similarly important for FAIRness is to work towards interoperability of Data Analysis platforms amongst project partners in ExPaNDS, to guarantee future reproducibility of scientific results.

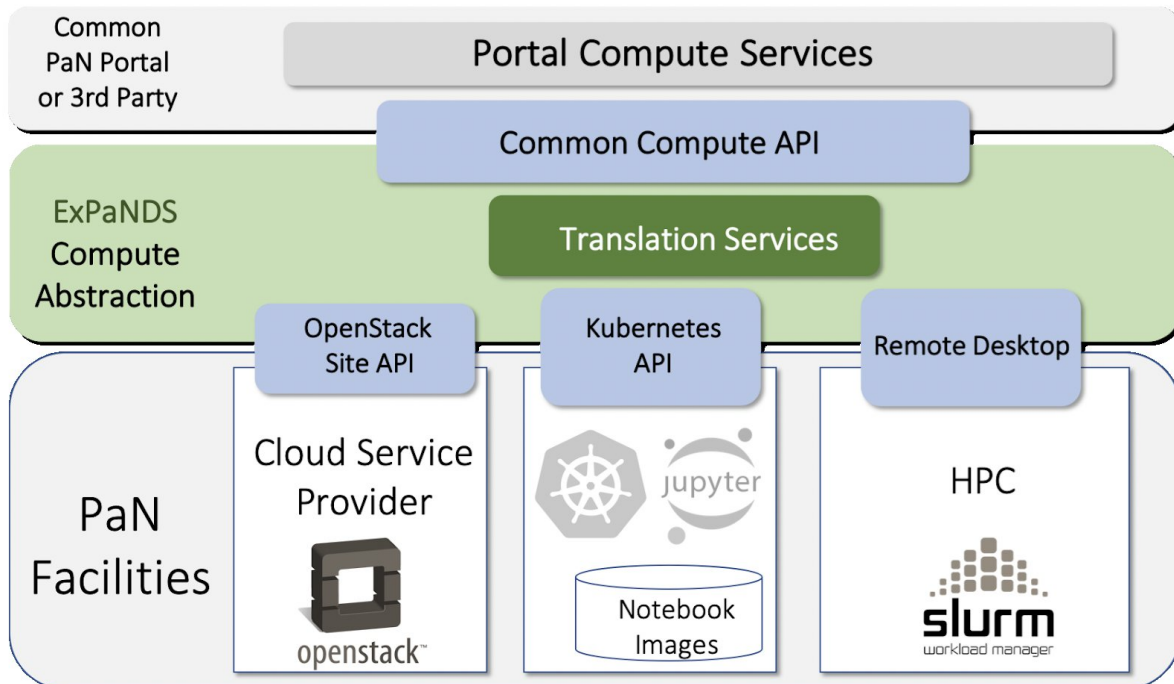


Fig. 2: ExPaNDS architecture - Data analysis services

The ExPaNDS services, as shown in figures 1 and 2 are maintained and operated by the ExPaNDS facilities, during the project's lifetime. In order to make ExPaNDS products available for long-term use, it is essential that they are built upon existing national services for data catalogues and data analysis as well as EOSC core services, whenever possible.

The PaN RIs will make sure that the related software/pipelines are production ready, stable, useful, open source and maintained. ExPaNDS will prepare for those prerequisites by ensuring the following:

- Data services need to be **reliable and attractive to all facility users**.
- Data services needs to be **maintained and financed by the facility itself**, guaranteeing sustainability after the end of ExPaNDS and PaNOSC.
- Know-how must be kept in the PaN RIs by **training** people at the facilities and **continuously documenting** these services (see WP5).

Additionally, we will make use of **as many services as possible from already existing e-infrastructures**, like EGI and EUDAT **and initiatives**, like the EOSC Hub and INDIGO-DataCloud. Furthermore we will link



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

to solutions demonstrated by other Horizon 2020 initiatives, such as CALIPSOplus (JRA2: <http://www.calipsoplus.eu/jra2-daas/>), the DataLake from the ESCAPE cluster and the “Quality of Service” model specified by the eXtreme DataCloud project.

We will also reuse the existing PaN-data software catalogue service instead of creating a new implementation, which will significantly increase the probability for it to be more sustainable. The cost of making the different analysis software accessible as wrapped solutions (in containers or via virtual machines) via the software catalogue will be shared between the facilities. This makes for an economy of scale which will encourage further uptake.

Finally, the ExPaNDS services (Fig 1 and 2) for the European PaN Community will leverage the distributed architecture depicted in Fig 3 (below) and will be offered to the **wider EOSC community**.

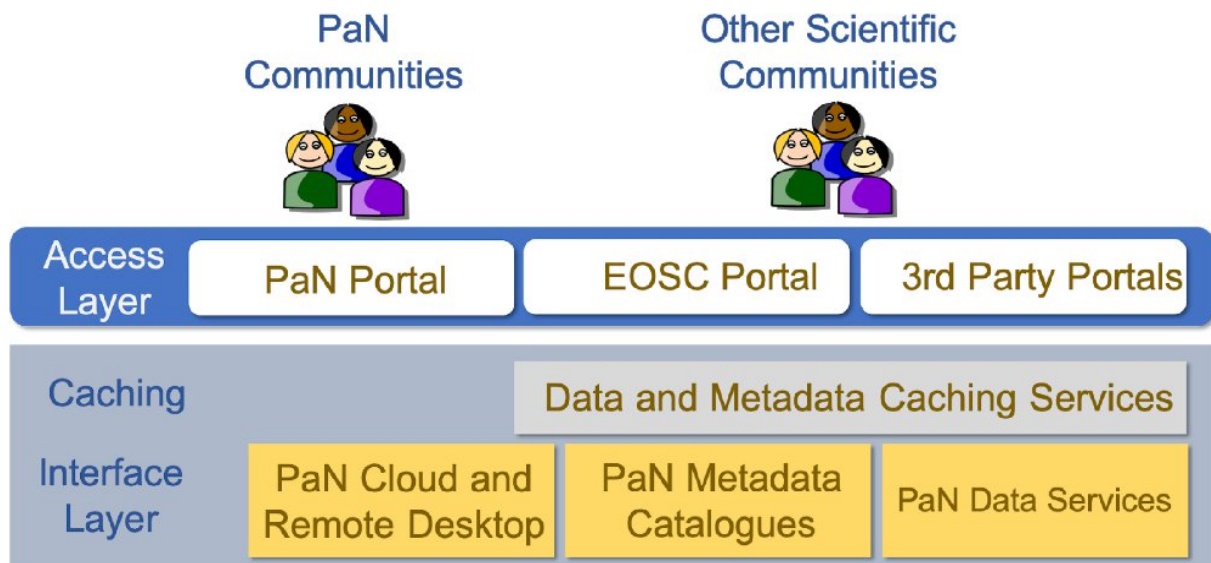


Fig. 3 Accessing ExPaNDS Services in EOSC. Photon and Neutron communities and other EOSC user communities will be able to access the ExPaNDS services through the PaN portal and, possibly, via the EOSC portal.

EOSC will be a key instrument to facilitate access to scientific services, lower barriers to integrate and compose services and promote the usage of services between adjacent communities. To achieve this for our services, we will use **EOSC standards** wherever possible and will define interoperability guidelines that allow to identify EOSC compliant services. These services will offer well-established and documented interfaces for usage and integration, based on well-known standard or APIs, facilitating their exploitation and the combined usage of more EOSC services. This activity is currently led by the EOSC-hub project that proposed a reference EOSC Technical Architecture² that includes functions, interfaces, APIs and standards as technical concepts, with the final aim of fostering interoperability and, ultimately, service composability. The adoption of standard interfaces is also relevant for horizontal or core services as we can chose different providers for those services, like EOSC-hub or even industry.

² EOSC-hub D10.4 EOSC Hub Technical Architecture and standards roadmap v2: <https://www.eosc-hub.eu/deliverable/d104-eosc-hub-technical-architecture-and-standards-roadmap-v2-under-ec-review>



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

5. WP5: Training

WP5, Training activities through EOSC platforms, will organise workshops and deliver training materials through the e-learning platforms made available on the EOSC. Training will be organised in cooperation with EOSC related activities. Training will address the following: service provision, user training in FAIR principles, data stewardship, data management and data analysis services integrated into the EOSC. Input for training will be provided by all WPs, in particular by WP3 and WP4 in regards to technical and user interface related information.

Training will foster **faster adoption of best practices** and facilitate the **uptake of ExPaNDS open source software** by an enlarged number of Photon and Neutron users. Extensive training in the areas of data management, service delivery and service management will also stimulate the sustainability of an EOSC federated model.

6. WP6: Outreach

WP6, Dissemination and Outreach, supports and promotes the ExPaNDS initiative in Europe and beyond and fosters Open Data and FAIR data principles among users and RIs.

The WP6 outreach activities are essential to **increase the customer basis** and to **attract more RIs to contribute to the already existing ExPaNDS and PaNOSC federated service portfolio**.

3. Conclusions and Outlook

The sustainability task within WP1 will make sure the policies described above are properly implemented by the WPs. The following KPIs can hint our level of success in making ExPaNDS sustainable:

- The take-up in provisioning of ExPaNDS services by national PaN RIs in support of their science programme,
- The successful integration of national PaN RIs data catalogues into the EOSC, measuring the number of users accessing the catalogues ,
- The successful integration of national PaN RIs data analysis services into the EOSC, measuring the number of users processing data.

Finally, our work will be aligned to the findings of the "Sustainability WG" of the EOSC Executive Board , of PaNOSC and of the other INFRAEOSC-5B projects on making all our components and FAIR policies sustained.



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