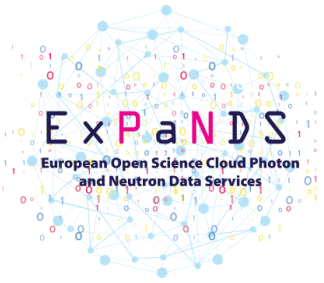
Project Data Management Plan



Document Control Information

Settings	Value
Document Identifier:	D1.10
Project Title:	ExPaNDS
Work Package:	WP1
Document Author:	Sophie Servan (DESY)
Responsible Partner:	DESY
Doc. Issue:	5
Dissemination level:	Public
Date:	14/02/2020

Abstract

This Data Management Plan (DMP) describes the data management life cycle for the data to be collected, processed and generated by the ExPaNDS project. Like all Horizon 2020 projects, the first version of the DMP is submitted by month 6 of the project but is intended as a *living document*. Information will be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur.

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 WP
1	22/01/2020	Template and main input	Sophie Servan / DESY WP1
2	23/01/2020	First review	Patrick Fuhrmann / DESY WP1 Alejandra Gonzalez Beltran / UKRI WP2 Heike Goerzig / HZB WP2 Kat Roarty / DLS WP6
3	27/01/2020	Review	Alun Ashton / PSI WP3
4	03/02/2020	Final draft ready for review	Sophie Servan / DESY WP1
5	14/02/2020	Final version after review	Patrick Fuhrmann / DESY WP1 Andy Götz / ESRF PaNOSC Sophie Servan / DESY WP1

Table of Contents

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

Executive summary

One of ExPaNDS main objectives is to promote FAIR (findable, accessible, interoperable and reusable) data in national Photon and Neutron Research Infrastructures (PaN RIs), e.g. through the harmonisation and generalisation of the use of DMPs for the experiments carried out in the facilities. ExPaNDS also takes part in the Open Research Data Pilot, which aims to make the research data generated by H2020 projects accessible with as few restrictions as possible, while at the same time protecting sensitive data from inappropriate access.

The way we deal with the project's data must then be an example of FAIRness and openness. That is why we plan to publish all our deliverables in the ExPaNDS community created on Zenodo (zenodo.org/communities/expands) and to integrate our software in the PaN software catalogue (panosc.eu/services/pan-software-catalogue) which will be further developed by PaNOSC.

Survey and personal data, which are the only sensitive data in ExPaNDS, will be dealt with in compliance with the General Data Protection Regulation (GDPR).

1 Introduction

A Data Management Plan is a key element of good data management as it is the initial tool to implement our FAIR data policy. The following paragraph (Data summary) is about FAIR principles applied to the data collected or generated by ExPaNDS. Then, the DMP addresses resources, security and ethics.

This document follows the EC – H2020 FAIR DMP template provided in the EC portal (ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm).

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

2 Data summary

ExPaNDS provides services for scientists who produce or process scientific data. It will therefore generate documents (text, presentation, video), websites, demonstrators and source code. It will also generate simulated or re-use existing data sets but not generate scientific data like instruments do.

Table 1: Generated and collected data in ExPaNDS

Data type	Origin	Expected size	Contents
Text documents	PC, collaboration space	100s Megabytes	Deliverables, meeting minutes, documentation
Presentations	PC, collaboration space	100s Megabytes	Training material, dissemination material, presentations at conferences
Videos	PC	100s Megabytes	Training material, dissemination material
Metadata schemas & example files	GitHub, NeXus Format	10s Megabytes	Adopted ontologies for the metadata catalogues
Source code	GitHub, several ExPaNDS facilities	10s Megabytes	Extension & integration of existing catalogues and services
Simulated data sets	TBD	TBD	Test data sets for analysis services to be implemented in EOSC
Re-used data sets	Several ExPaNDS facilities	10s Gigabytes	

In the case of existing data sets being re-used, the data policy and the infrastructure of the hosting partner facility will have to be complied with.

Concerning the software, most ExPaNDS developments will be extending the functionality of existing software, for which the pre-existing licences will also be complied with.

The documents and source code produced by ExPaNDS will be useful to national PaN RIs and to ESFRIs, particularly through the collaboration work with the PaNOSC project.

2.1 Making data findable, including provisions for metadata

The naming convention for the project documents is defined in the project's Quality Assurance Plan (Deliverable 1.2). All project deliverables will be uploaded to Zenodo (see 2.2) and by that will get a DOI assigned. They will be findable using the Zenodo search feature. The software and the associated reference data sets will also be referenced in the PaN software catalogue.

The reference data sets will use the metadata standards developed in the project, which will be based, consistently with PaNOSC, on the NeXus format standards (www.nexusformat.org). They will be referenced using the facility generated persistent identifiers. Indeed, each facility will be

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

encouraged to provide DOIs for the data they produce, so that their origin can be easily traceable (which would not necessarily be the case with a Zenodo DOI).

In order to allow detailed tracking of data produced in ExPaNDS, we will apply versioning of all entities as follows:

- The project documents will get version numbers, as described in the projects Quality Assurance Plan.
- The data sets versions will follow the data policy of hosting facilities.
- The different versions of the source code will use the Zenodo release management.

The combined use of GitHub and Zenodo to strictly version the different releases of software is a good practice that will be promoted at the different facilities of ExPaNDS.

2.2 Making data openly accessible

All public project deliverables will be uploaded to Zenodo where ExPaNDS has a community (zenodo.org/communities/expands). Only a few deliverables will not be public because they contain sensitive data (Collaboration Agreement, ethics and human resources allocation) and they will only be shared with the project contributors in the private collaboration space hosted by Diamond Light Source. The re-used datasets will be available in the contributing RI's own repository.

For the project documents, only widely used or open source software will be needed for access (e.g. PDF, MS suite, LibreOffice suite). The code source, GIT tool set and software catalogue will be accessible via a web browser. The reference data will be accessible through the different RI's repositories which store it and through the PaN software catalogue.

Training and documentation on how to apply the ExPaNDS software on our reference data-sets will be provided as part of the ExPaNDS deliverables.

Given their size, uploading the deliverables to Zenodo does not require previous arrangement. The arrangements for the PaN software catalogue are and will be discussed in real-time with PaNOSC.

There is no restriction on use for all public ExPaNDS deliverables, which will be under Creative Commons licences (see 2.4). The access to confidential deliverables will be provided by the SharePoint collaboration space through nominative accounts for project participants.

Anonymous access to public deliverables is guaranteed. For confidential deliverables, the identity will be ascertained using the SharePoint's credentials.

2.3 Making data interoperable

ExPaNDS does not produce scientific data per se. However, the projects deliverables are shared and can be re-used by other EOSC-related projects for efficient collaboration and good practices.

The development of a common ontology for PaN scientists is one of the objectives of the project, in collaboration with PaNOSC. The contribution to developing the NeXus metadata standard will promote the interoperability of data produced in the PaN community. Some ExPaNDS partners are

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

part of the Nexus International Advisory Committee (NIAC, www.nexusformat.org/NIAC) and the participation of all of them will be encouraged.

2.4 Increase data re-use (through clarifying licenses)

The use of Creative Commons licences (CC) BY 4.0 for all public project deliverables will be implemented. Concerning the source code for software development, we will encourage the projects we contribute to to choose a Creative Common licence.

Depending on their origin, the reference data sets will be made available as soon as possible. The public deliverables will be made available as soon as they are published on the EU portal. There is no restriction on the re-use of the projects' public deliverables after the end of the project.

The reference data sets will benefit from the 10 years guaranteed availability of their hosting facility. All data uploaded to Zenodo will benefit from its sustainability (see 3).

ExPaNDS Data quality assurance processes are described in a dedicated document "Quality Assurance Plan" (Deliverable 1.2).

3 Allocation of resources

Zenodo is hosted by CERN and paid for by the EU. GitHub is a public resource. The PaN software catalogue is hosted by ILL (Institut Laue-Langevin, www.ill.eu/about-the-ill/) and is maintained by the community. Human resources dedicated to FAIR in ExPaNDS are included in its Project Grant.

WP leaders are responsible for the data management of their own products. WP1 is responsible for the general compliance of the data management with respect to this DMP.

4 Data security

Data security for all documents uploaded to Zenodo naturally relies on Zenodo's security measures. Documents for long-term storage will be in PDF/A format (en.wikipedia.org/wiki/PDF/A). Data security for confidential deliverables relies on SharePoint and the DLS infrastructure.

For the development phases, data security relies on each partner's facility IT infrastructure.

5 Ethical aspects

The handling of the Protection of Personal Data (POPD) requirements in ExPaNDS are described in a dedicated deliverable (D7.1).

Some surveys have been and will be made in the framework of ExPaNDS. Personal data collected for them will be deleted after anonymisation or aggregation. Other personal data, like the project's "who's who" document or the list of attendees to training, workshops and conference events, will be stored in the collaboration space with restricted access to the project's contributors.

Consent will be obtained before any public dissemination.

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