

# ExPaNDS Vision and Roadmap

Settings	Value
Document Identifier:	D6.04
Project Title:	ExPaNDS
Work Package:	WP6
Document Author:	Isabelle Boscaro-Clarke, Francois Cesmat, Kat Roarty
Responsible Partner:	Diamond Light Source
Doc. Issue:	4
Dissemination level:	Public
Date:	18/12/2020



## 2019

Over 20 European Photon and Neutron (PaN) facilities generating dozens of petabytes of data every year.

Over 25,000 users for the photon and over 5,000 researchers for the neutron communities.

Thousands of journal articles published by the users every year.

Most facilities have a different experimental data policies.

No centralised or federated data repository.

No centralised or federated publications catalogue.



## 2020

General Architecture in relation to the EOSC services.

Sustainability policy report.

Annual report on training activities, workshops held, materials published and investigations performed.

Draft extended data policy framework for PaN Research infrastructures (RIs).

FAIR workshops held on 1<sup>st</sup> and 2<sup>nd</sup> October 2020.

Data Management Plan for ExPaNDS.

Draft recommendations for FAIR PaN Data Management.

1<sup>st</sup> PaN EOSC Symposium and Joint Annual Meeting with PaNOSC - November 2020.

Report on status, gap analysis and roadmap towards harmonised and federated metadata catalogues for EU national PaN RIs.

## 2021-2022

2<sup>nd</sup> and 3<sup>rd</sup> PaN EOSC Symposium and Joint Annual Meeting with PaNOSC.

Use case for Life sciences:

- Collaboration with EOSC*life* & FREYA for ligand screening;
- Explore AI for automation of classification and mining;
- User meeting for life science community to engage their views on better digitising all data analysis workflows;
- Deliver Life sciences use case.

1<sup>st</sup> ExPaNDS/PaNOSC Symposium for librarians and data policy facility staff.

Series of workshop for facility users to make them FAIR aware and work on data management planning.

Joint ExPaNDS/PaNOSC learning portal to map training available for facility users.

Engage with publishers to add data fields for publications.

## 2022-2023

**National RIs policy change:** FAIR principles fully integrated with the facilities experimental data policy and demonstrating adoption of FAIR principles.

**Impact on alignment of National RI data catalogues:**

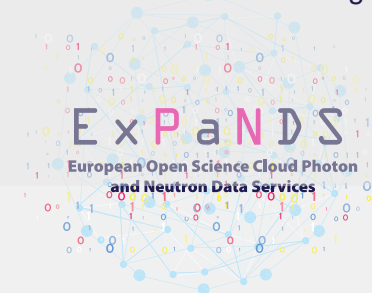
1. A common API for all the data catalogues.
2. Alignment of data catalogues with a common ontology and using agreed PIDs.
3. An agreed common taxonomy for classification of datasets & publications.

A federated data catalogue for the EOSC service (using the common API) – EOSC Component.

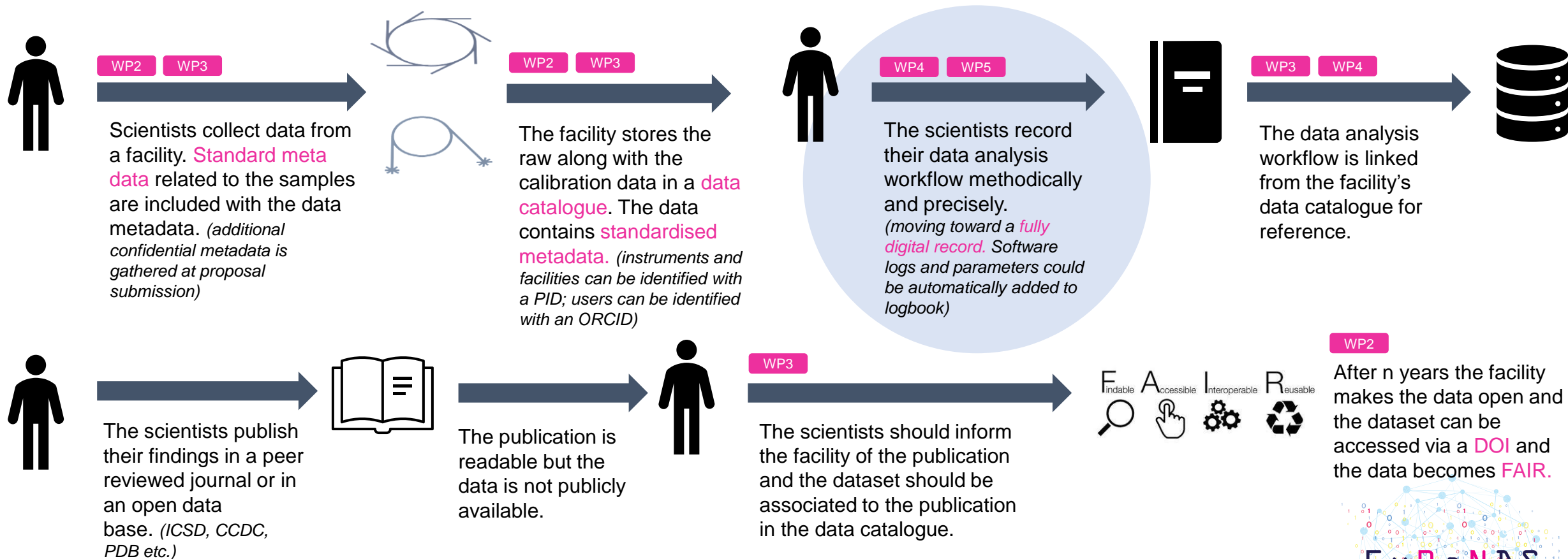
Moving toward an EOSC analysis services for national RIs within the EOSC-hub.

Working with a subset of facility users being trained and experienced at cataloguing their experimental data and describing their data analysis workflow with the aim of expanding to the rest of the user community – beyond ExPaNDS.

ExPaNDS must enable each national RIs to generate and associate data DOIs for the data sets they generate. The DOIs must be associated with any publicly available data set so that they can be referenced in science literature and other catalogues.



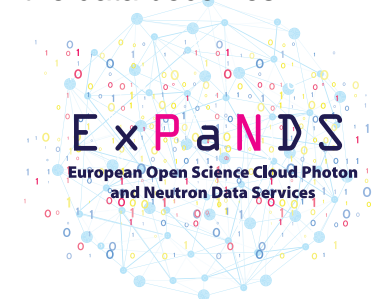
# Use case 1: Experiment from data collection to publication – the facility publishes the data after n years



WP1 WP6 Driving the vision, engagement and dissemination of the ExPaNDS grant project to the community



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



**Workflow of ExPaNDs Data Management**

```

graph LR
    subgraph Stage1 [Data Collection and Storage]
        S1[Scientists collect data from a facility. Standard meta data related to the samples are included with the data metadata. (additional confidential metadata is gathered at proposal submission)]
        S2[The facility stores the raw data along with the calibration data in a data catalogue. The data contains standardised metadata. (instruments and facilities can be identified with a PID; users can be identified with an ORCID)]
        S1 -- WP2, WP3 --> S2
    end

    subgraph Stage2 [Data Analysis]
        S3[The scientists records their data analysis workflow methodically and precisely. (moving toward a fully digital record. Software logs and parameters could be automatically added to logbook)]
        S2 -- WP4, WP5 --> S3
    end

    subgraph Stage3 [Data Publication]
        S4[The scientists publish their findings in a peer reviewed journal, the article contains link to the data set(s).]
        S5[The scientists also publish at the same time or before the raw data and the complete data analysis workflow via the facility's data catalogue (DOI provided by the facility) making the data FAIR.]
        S3 -- WP3 --> S4
        S3 -- WP3 --> S5
        S4 -- WP3 --> S5
    end

    S5 -- WP3 --> FAIR[FAIR: Findable, Accessible, Interoperable, Reusable]
    FAIR -- WP3 --> S6[The scientists should inform the facility of the publication. The dataset and publications are associated by the facility on both the data catalogue and facility publications database.]
    S6 -- WP3 --> S7[Expands to engage with Publishers to add data fields for Publications.]
    S7 -- WP3 --> S5
  
```

**WP2 WP3**

Scientists collect data from a facility. **Standard meta data** related to the samples are included with the data metadata. *(additional confidential metadata is gathered at proposal submission)*

**WP2 WP3**

The facility stores the raw data along with the calibration data in a **data catalogue**. The data contains **standardised metadata**. *(instruments and facilities can be identified with a PID; users can be identified with an ORCID)*

**WP4 WP5**

The scientists records their data analysis workflow methodically and precisely. *(moving toward a **fully digital record**. Software logs and parameters could be automatically added to logbook)*

**WP3 WP4**

The data analysis workflow is linked from the facility's data catalogue for reference.

**WP3**

The scientists publish their findings in a peer reviewed journal, the article contains link to the data set(s).

**WP3**

Expands to engage with Publishers to add data fields for Publications.

**WP2 WP3**

The scientists also publish at the same time or before the raw data and the complete data analysis workflow via the facility's data catalogue (**DOI provided by the facility**) making the data FAIR.

**WP3**

**FAIR**

**Findable Accessible Interoperable Reusable**

The scientists should inform the facility of the publication. The dataset and publications are associated by the facility on both the data catalogue and facility publications database.

**WP3**

Expands to engage with Publishers to add data fields for Publications.

**WP1 WP6**

Driving the vision, engagement and dissemination of the ExPaNDs grant project to the community

**ExPaNDs**

European Open Science Cloud Photon and Neutron Data Services