



Achieving a Photon and Neutron community federated cloud in EOSC

2nd November 2020

Enabling our facilities to produce FAIR data

Brian Matthews,
UKRI-STFC

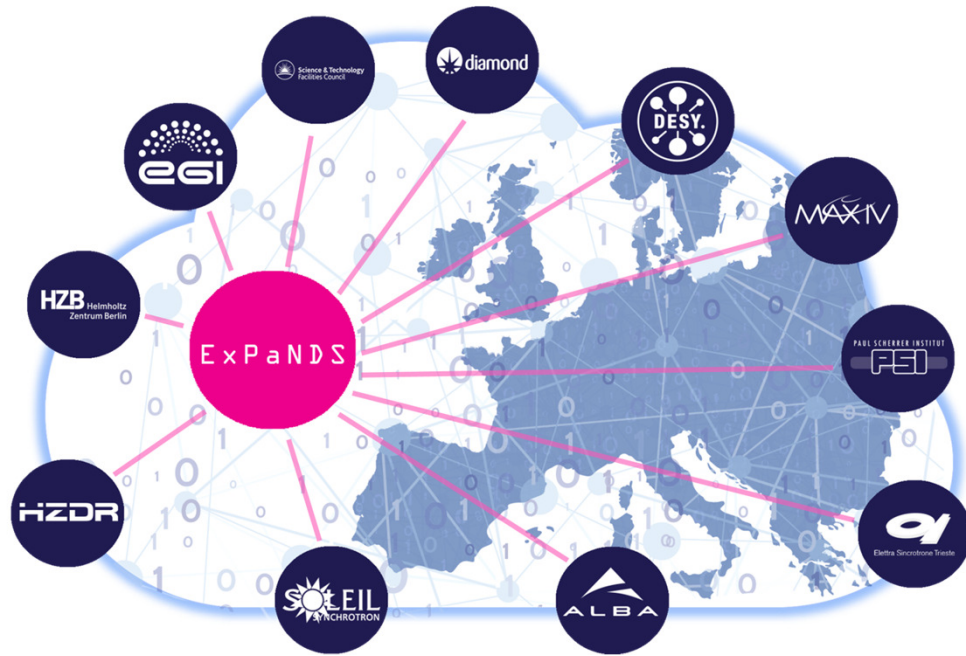


These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641

ExPaNDS WP2: Enabling FAIR data PaNOSC WP2: Data Policy and Stewardship

- Review and recommend the policies, practises, standards and tools which would develop best practise for FAIR data generation and use in the National Photon and Neutron RIs.
 - In the policies of the RIs
 - In the data-generation, collection and analysis process
 - In Data Management Planning
- Raising awareness and competence in FAIR data of our scientific communities.

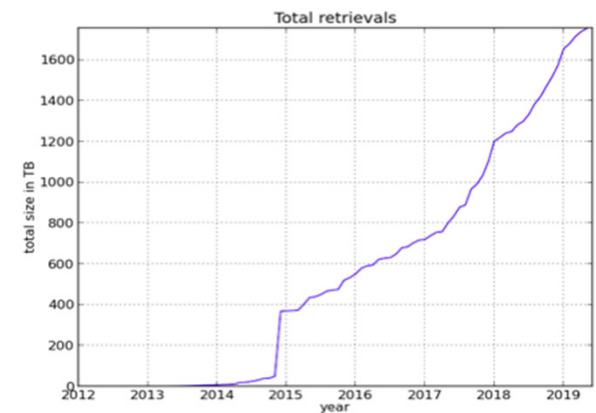
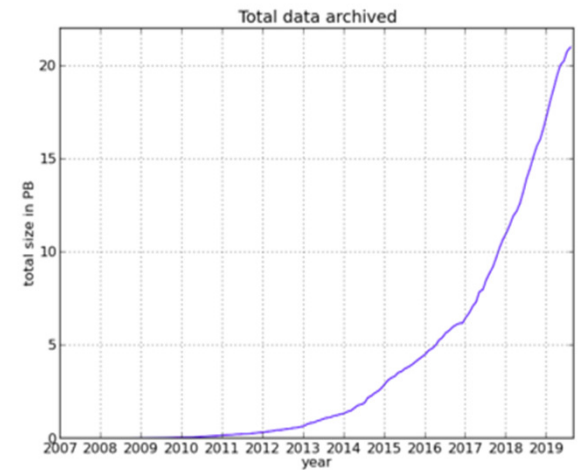
To guide services to support FAIRness



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641

What's the problem?

- P&N experiments across disciplines
 - No single discipline approach
- Data getting ever larger and more complex
 - Hard to move and handle
 - Hard to Process
- Maximise the Science Value
 - Facilities are expensive!
- P&N Facilities can support users to (re-)use data
 - Common data lifecycle
 - Common infrastructure



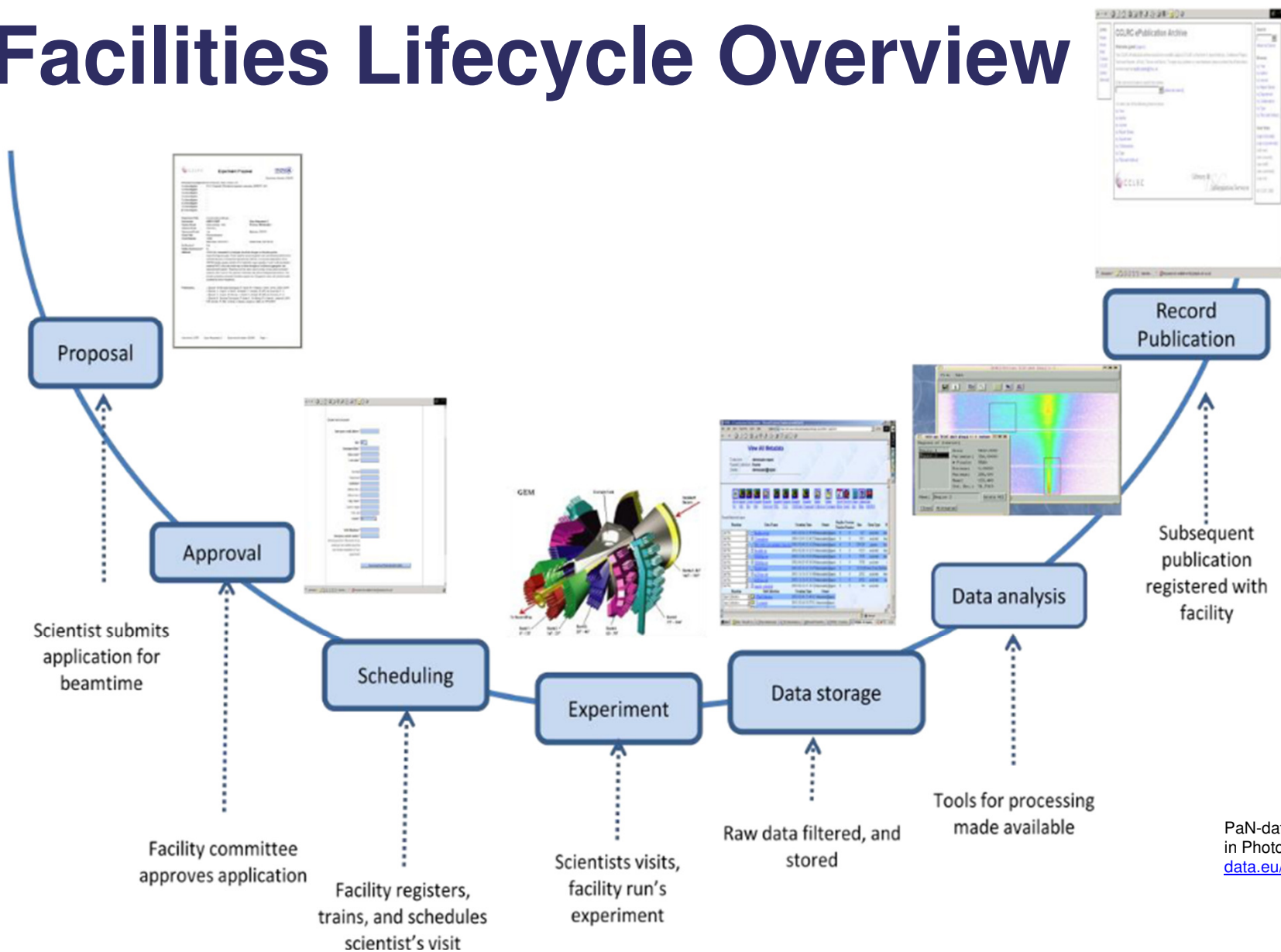
Daniel Salvat, Alba



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641



Facilities Lifecycle Overview



P&N Science is multi-discipline

P&N Facilities can support users to (re-)use data

- **Common infrastructure**
- **Common data lifecycle**

**Make this FAIR!
Facilities have a good foundation**

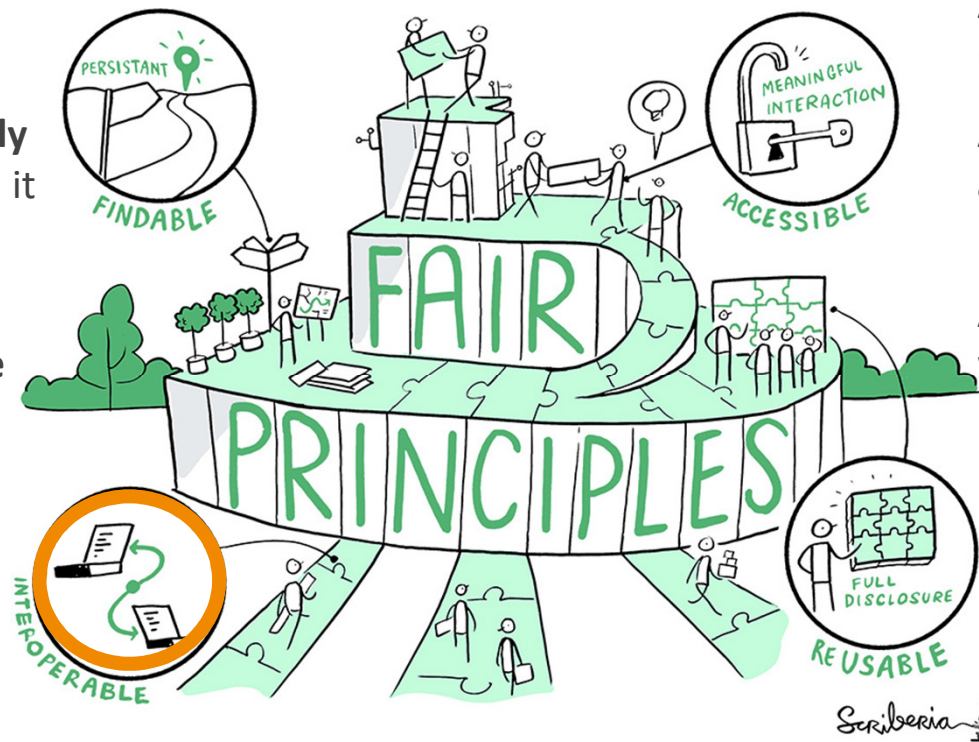
PaN-data Open Data Infrastructure. Model of the data continuum in Photon and Neutron Facilities. Deliverable 6.1, 2012 <http://pan-data.eu/sites/pan-data.eu/files/PaNdataODI-D6.1.pdf>

F1. (meta)data are assigned a **globally unique and persistent identifier**

F2. data are described with **rich metadata** (defined by R1 below)

F3. metadata clearly and **explicitly include the identifier** of the data it describes

F4. (meta)data are registered or indexed in a **searchable resource**



I1. (meta)data use a **formal, accessible, shared, and broadly applicable language for knowledge representation.**

I2. (meta)data use **vocabularies** that follow FAIR principles

A1. (meta)data are **retrievable** by their **identifier** using a **standardized communications protocol**

A1.1 the **protocol is open, free, and universally implementable**

A1.2 the protocol allows for an **authentication and authorization** procedure, where necessary

A2. **metadata are accessible**, even when the data are no longer available

R1. meta(data) are **richly described** with a plurality of accurate and relevant attributes

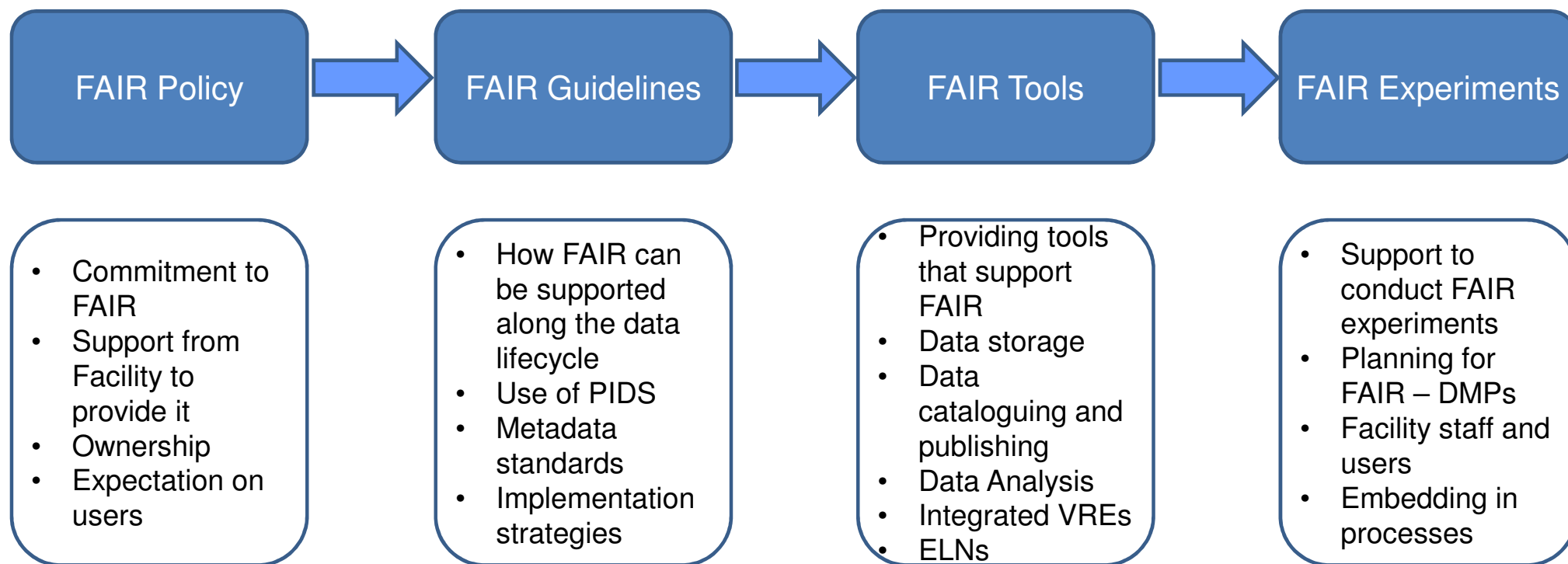
R1.1. (meta)data are released with a clear and accessible **data usage license**

R1.2. (meta)data are associated with **detailed provenance**

The Turing Way Community, & Scriberia. (2020, March 3). Illustrations from the Turing Way book dashes. Zenodo. <http://doi.org/10.5281/zenodo.3695300>

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

Steps towards FAIR Facilities



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641



Data policy in P&N Facilities

Photon and Neutron RIs have a long track record on policy
PaN-data Europe common policy framework February 2011

Neutrons:

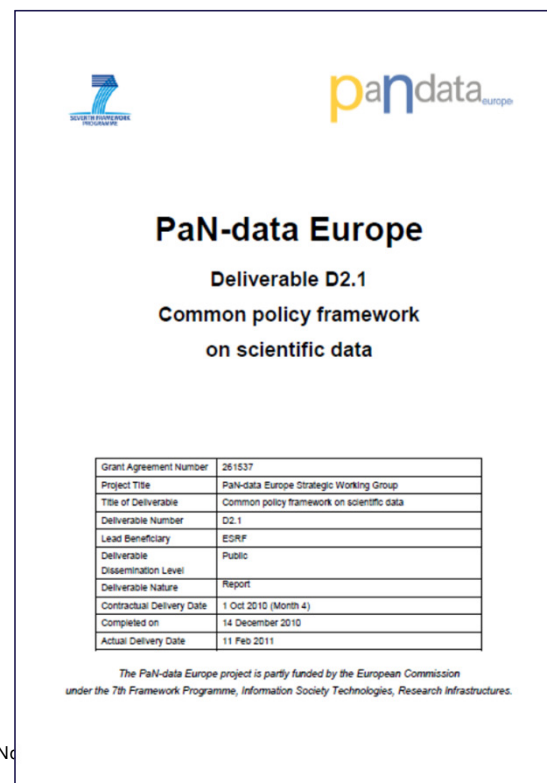
ILL- PanData data policy - 2012
ISIS-PanData data policy - 2012

Photons:

Elettra- PanData data policy - 2013
ESRF-PanData data policy - November 2015
MAXIV-PanData data policy - 2015
HZB-PanData data policy - June 2016
HZDR-PanData data policy - June 2016
PSI-PanData data policy - August 2016
EUXFEL-PanData data policy - August 2017
ALBA-PanData data policy - July 2017
DESY-PanData data policy - August 2017
SOLEIL PanData data policy - 2018
Diamond PanData data policy - March 2019
FELIX PanData data policy - 2019

Set some key principles

- Open access to raw data and metadata
- Curation of raw data supported by the facility
- Data catalogue to make data accessible
- Embargo periods for exclusive access to experimental data



PaN-data Europe
Deliverable D2.1
Common policy framework
on scientific data

Grant Agreement Number	261537
Project Title	PaN-data Europe Strategic Working Group
Title of Deliverable	Common policy framework on scientific data
Deliverable Number	D2.1
Lead Beneficiary	ESRF
Dissemination Level	Public
Deliverable Nature	Report
Contractual Delivery Date	1 Oct 2010 (Month 4)
Completed on	14 December 2010
Actual Delivery Date	11 Feb 2011

The PaN-data Europe project is partly funded by the European Commission under the 7th Framework Programme, Information Society Technologies, Research Infrastructures.



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No.



Towards FAIR policies for Photon and Neutron RIs

- ExPaNDS and PaNOSC have been working together to revise data policies frameworks in the light the of FAIR data principles
 - PaNOSC D2.1: PaNOSC data policy framework
 - A model policy for adaptation and adoption by PaNOSC Partners
 - [May 2020: https://doi.org/10.5281/zenodo.3826039](https://doi.org/10.5281/zenodo.3826039)
 - ExPaNDS D2.1: Draft Extended Data Policy Framework for Photon and Neutron RIs
 - 18th September 2020: <https://doi.org/10.5281/zenodo.4014810>
 - Guidance on adopting a FAIR data policy for national RIs
 - Taking into account FAIRsFAIR's recommendations on Data Policy
 - Key Policy Elements within a PaN RI Data Policy Framework
 - 30 data policy framework elements
- Aim to have a common approach to data policy across all P&N RIs
 - Value in a compatible approach in different facilities
 - Easier for users to move around, easier to combine data



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agree

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

paNOSC
photon and neutron
open science cloud

Deliverable: D2.1 - PaNOSC data policy framework

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

Page | 1

ExPaNDS
European Open Science Cloud Photon
and Neutron Data Services

D2.1: Draft Extended Data Policy Framework for Photon and Neutron RIs

Document Control Information

Settings	Value
Document Identifier:	D2.1
Project Title:	ExPaNDS
Work Package:	WP2
Work Package Lead:	UKRI
Deliverable Lead:	PSI
Document Author(s):	Brian Matthews (UKRI), Abigail McBirnie (UKRI), Andrei Vukolov (Elettra), Alun W Ashton (PSI), Stephen Collins (DLS), Sylvie Du Gracq Ramos (DLS), Brigitte Gagey (SOLEIL), Alejandra Gonzalez-Beltran (UKRI), Maria Johansson (Lund University), Rolf Krahl (HZB), Majid Cunsy (SOLEIL), Mirjam van Daalen (PSI)
Document Contributor(s):	Andy Gotz (ESRF), Uwe Konrad (HZDR), Simon Lambert (UKRI), Daniel Salvat (ALBA), Sophie Servan (DESY)
Doc. Version:	1.0
Dissemination level:	Public
Date:	18/09/2020

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

Date: 18/09/2020 1 / 74 DOI: 10.5281/zenodo.4014811



Adding FAIR to Policy

“RIs’ data policies should enable the experimental data in scope to be FAIR”

This gives an implicit commitment to uphold the FAIR principles.

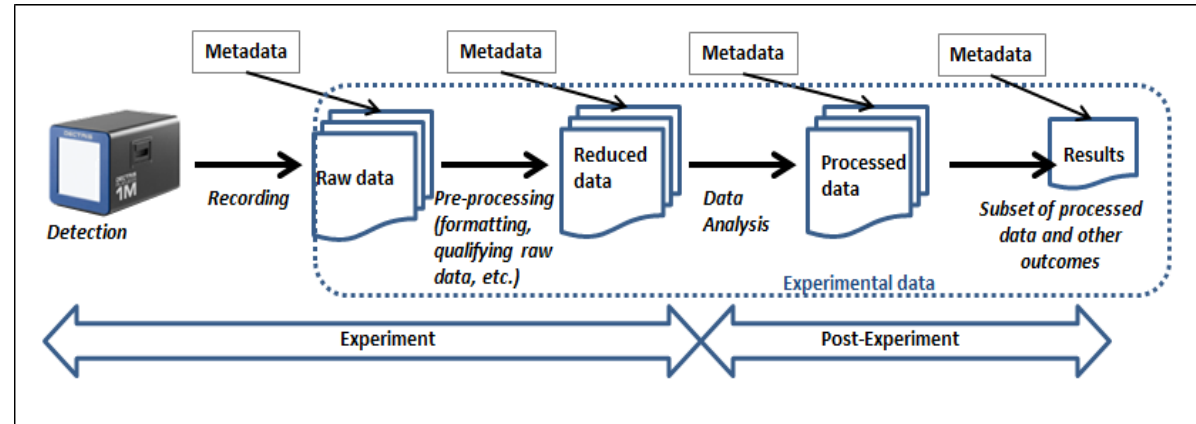
Leads to other policy principles

- *RIs should specify the grounds for restricting access to data (A1.2)*
- *In the event that data are deleted, the facility should retain a “digital footprint” of the data (A2)*
- *The RI’s data policy should specify a licence under which the data are made available (R1.1)*
- *The RI’s data policy should include commitments to enabling FAIR data which include [PIDs and Collecting sufficient Metadata]*

And also requirements for implementations conforming to policy



These projects have received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641



Simplified illustration of Classes of Experimental Data in the Science Life Cycle (from the Soleil Data Policy)

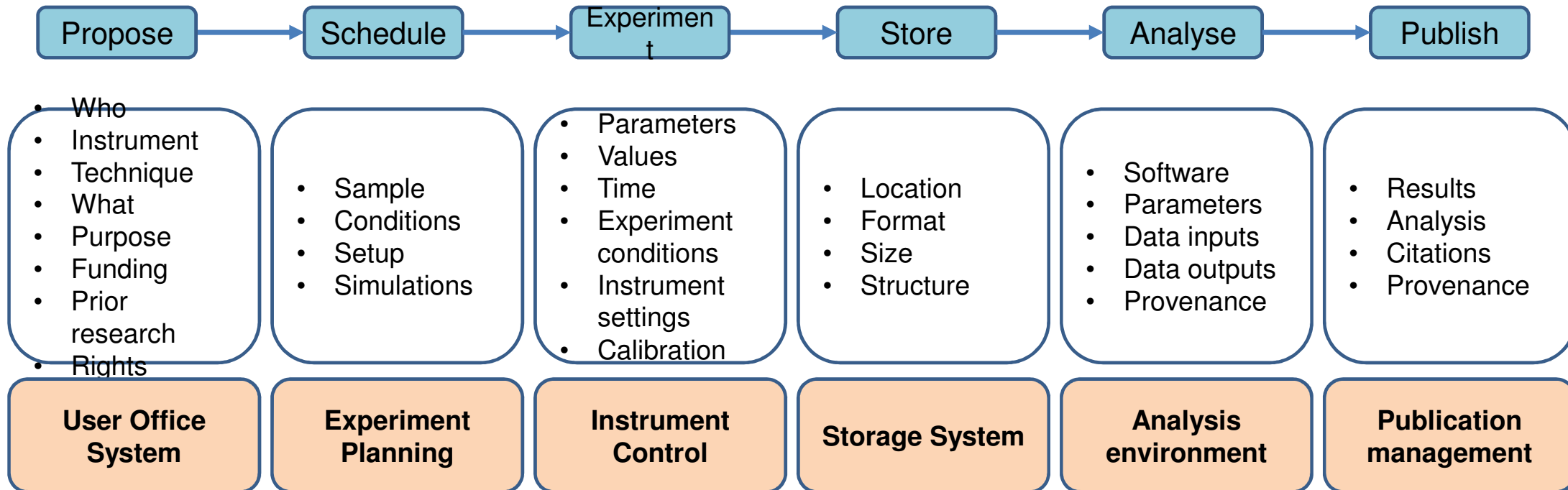
“Data should be FAIR when it leaves the Facility”

Can be assessed via FAIR Metrics



FAIR Guidelines: FAIR at every step

PaN-data Open Data Infrastructure. Model of the data continuum in Photon and Neutron Facilities. Deliverable 6.1, 2012 <http://pan-data.eu/sites/pan-data.eu/files/PaNdataODI-D6.1.pdf>



Collect, Connect, Curate



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641



FAIR Guidelines: Resource Identification

Persistent Identifier (PID) Services

- Purpose
- Scope
- Technology
- Governance
- Metadata
- Cost
- Uptake

What are the best choices for Facilities?

EOSC PID Policy

A survey of PID services is available in: FREYA project. D3.1 Survey of Current PID Services Landscape

https://www.project-freya.eu/en/deliverables/freya_d3-1.pdf



Publication



PURL



Data

ARK (Archival Resource Key)



PURL



People



RESEARCHERID



Organisation



Project



Instrument



Persistent Identification of Instruments WG

Sample



Software



RDA/FORCE11 Software Source Code Identification WG



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641



FAIR Guidelines: Findability

- Metadata catalogues
 - Searchable resource where data, and associated metadata can be registered or indexed
 - New use cases being considered
- Search for humans: graphical user interfaces
- Search for machines: common Application Programming Interface
- Integration with EOSC services
- Integration with catalogues in other domains

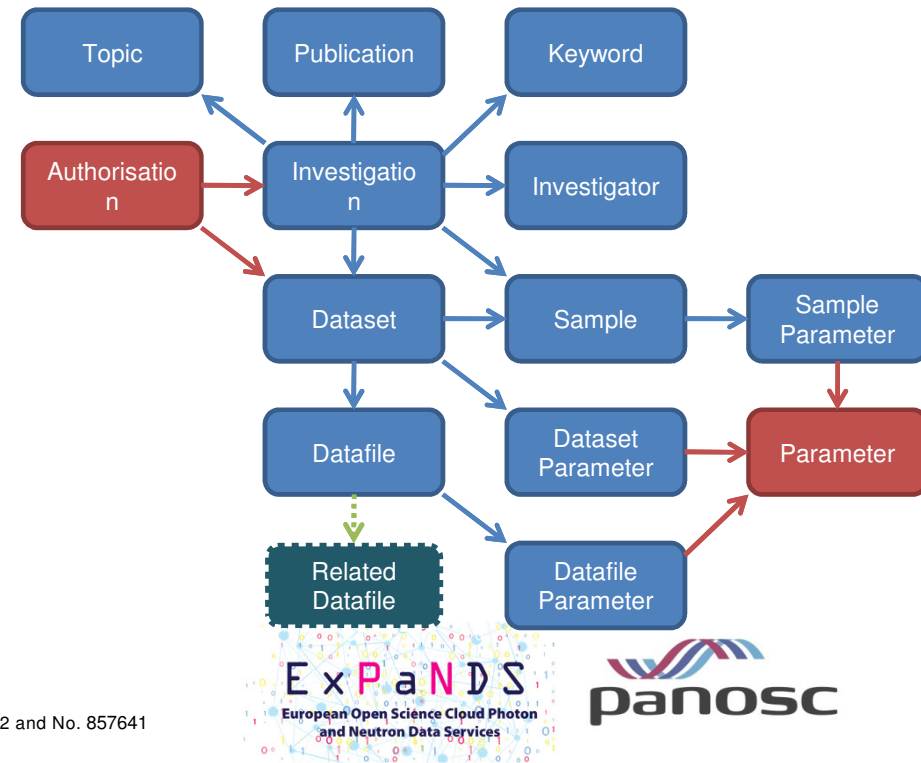


F1. (meta)data are assigned a **globally unique and persistent identifier**

F2. data are described with **rich metadata** (defined by R1 below)

F3. metadata clearly and **explicitly include the identifier** of the data it describes

F4. (meta)data are registered or indexed in a **searchable resource**



FAIR Guidelines: Interoperability



Data Meaning & Context
Terminologies establishing the meaning and context for the data

Semantics

Common vocabularies and mappings

WP2 & WP3 - work on vocabularies framework & ontologies



Data Topology
Data models, data schemas, shapes

Structure

Common data schemas and validation processes

Use of NeXuS and its application definitions at facilities



Data Language
Data formats used for data representation

Syntax

Common data formats for different types of data and metadata

Use of HDF5 format and metadata catalogues



Data Exchange
through computers, applications, networks and web services

System

Common web services and Application Programming Interfaces (APIs) for data discovery, data access and data exchange

WP3 - implementation of common search API



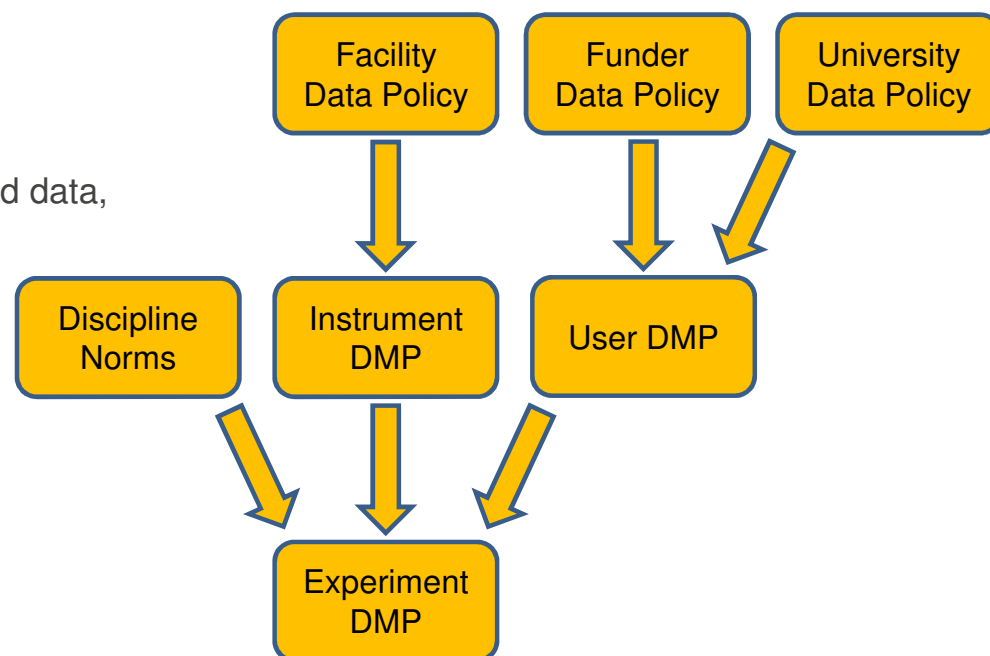
Alejandra Gonzalez-Beltran

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641



Bringing FAIR to the Experiment

- Data Policy applies at the facilities level
 - This needs to be made happen for each experiment
- For each experiment
 - Data/Metadata to be collected, data storage, connections to derived data, software used
 - Sounds like Hard Work!!!
 - Data Management Planning
- DMP for an experiment needs to be done in context:
 - The DMP for its instrument and technique
 - The community norms for the discipline
 - The DMP of the user's institution and funder
- The DMP should be **active**
 - Help steer the collection of metadata in the experiment
 - Work with for example an Electronic Notebook



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641



FAIR Experiment: Data Management Planning

Plan the many aspects of data and metadata generation, preservation, and analysis at the outset

- Information on data and data format
 - Types of data generated, Volumes of data, File formats
 - Collection processes
 - Software used, Analysed data
 - Quality control
 - Metadata content and format
 - Metadata items collected.
 - Metadata standards
 - How collected?
 - Policies for access, sharing, and re-use
 - Obligations from funders
 - Specific ethical/privacy and IPR issues
 - Data Publication
 - Digital Object Identifiers
 - Long-term storage and data management
 - Where is the data going to land?
 - Especially derived data
 - Costs
-
- **Pros:** can really assist in the allocation of resources and generation of reusable data
 - Expected data volumes Identify analysis routes and workflows Identify and fix bottlenecks ahead of time
 - **Cons:** Users may be agnostic or less convinced.
 - Extra work for proposal writing, Additional bureaucracy for access, Poor past experiences with DMPs

DMPs need careful consideration and presentation in the P&N community



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641



Summary

- P&N Facilities well-placed to assist their diverse user communities to produce FAIR data
 - However a lot of work required to cover the range of user communities and disciplines.
- Guidelines for FAIR data and metadata implementation
 - Under preparation in both ExPaNDS and PaNOSC
- Guidelines for PIDs and DMPs scheduled
- Feeding into service development in the projects

Need to bring User Communities along



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641





Thankyou

Brian Matthews

Brian.Matthews@stfc.ac.uk

<https://expands.eu>

Thanks to: Abigail McBirnie, Daniel Salvat, Alejandra Gonzalez-Beltran, Andy Gotz, Jonathan Taylor



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852 and No. 857641