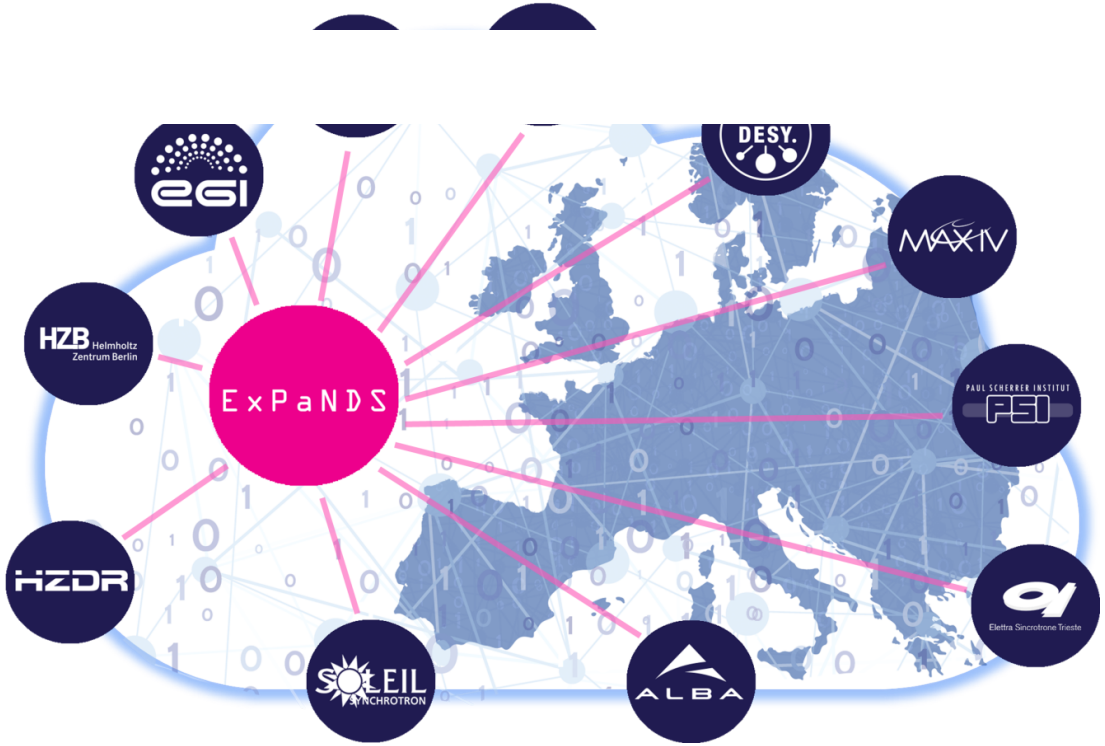
10-11 November 2020

PaNOSC & ExPaNDS Annual Meeting

WP2 & WP2: Enabling our facilities to produce FAIR data

Andy Gotz, Brian Matthews,
UKRI-STFC

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.



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

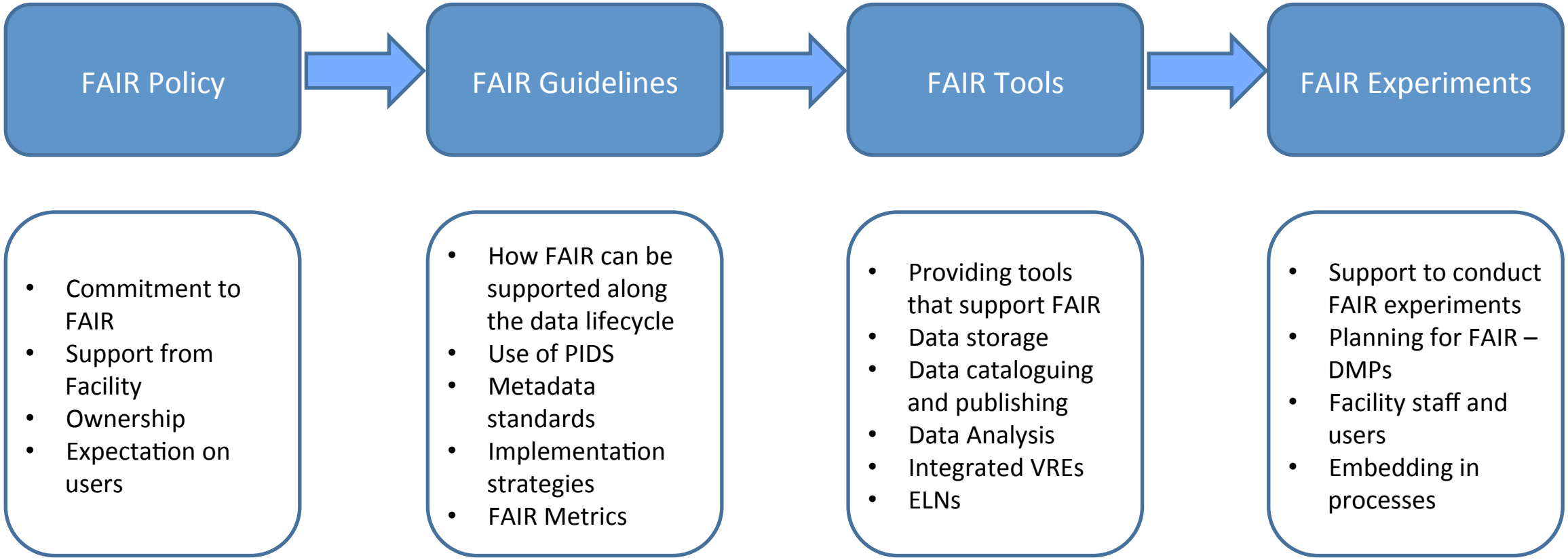
paNOSC

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

panosc photon and neutron open science cloud

ExPaNDS European Open Science Cloud Photon and Neutron Data Services

Steps towards FAIR Facilities



PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

ExPaNDS WP2: Work Tasks

2.1 Alignment policies & practices:

- Reviewing and revising data policies for FAIR data in the EOSC
- Aligning with similar work in PaNOSC

2.2 Data management planning:

- Considering how FAIR data can be collected during the facilities research lifecycle
- Using Data Management Planning to capture metadata and allow the creation of FAIR data for an experiment
- Active DMP as a mechanism for enabling and automating FAIR data creation and use in the lifecycle

2.3 Mainstreaming of standards:

- Consider how each of the 13 FAIR principles apply to Facilities Science
- Review how they are currently being applied
- Recommend best practises and standards so that they can be satisfied
- Promote standards across the facilities.

2.4 Persistent identifier infrastructure:

- Best practises for assigning and linking PIDs into a “PID Graph”
- Investigate emerging PIDs – e.g. Organisations, Instruments

2.5 QA & certification schemes:

- Consider emerging FAIR Data Maturity models and how they might apply to P&N RIs
- Self-assessment of ExPaNDS RIs

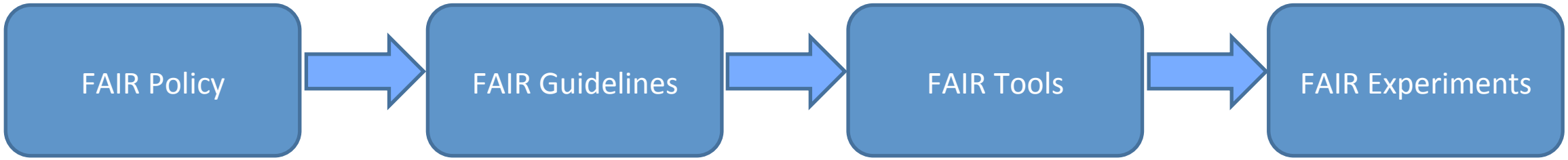
2.6 Uptake of FAIR data practices:

- Advocacy and awareness of FAIR data practise to ExPaNDS stakeholders
- Senior management, Instrument Scientists, User Researchers
- Competencies of Data Stewardship

PaNOSC and ExPaNDS projects have received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Steps towards FAIR Facilities

D2.6



- Commitment to FAIR
- Support from Facility
- Ownership
- Expectation on users

T2.1

- How FAIR can be supported along the data lifecycle
- Use of PIDS
- Metadata standards
- Implementation strategies
- FAIR Metrics

T2.3, T2.4, T2.5 WP3

- Providing tools that support FAIR
- Data storage
- Data cataloguing and publishing
- Data Analysis
- Integrated VREs
- ELNs

WP3 WP4

- Support to conduct FAIR experiments
- Planning for FAIR – DMPs
- Facility staff and users
- Embedding in processes

T2.2

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

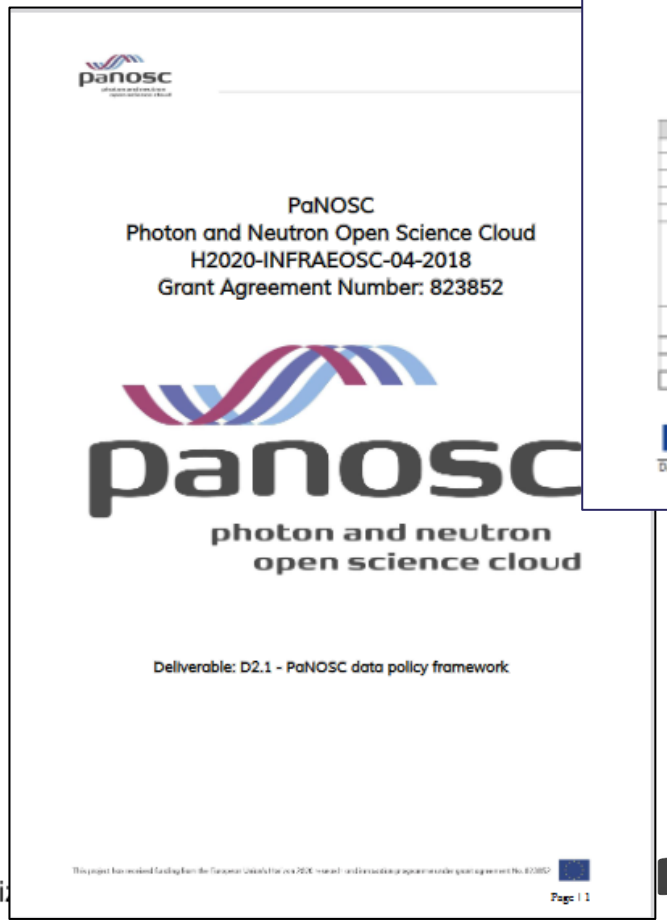
ExPaNDS WP2: Deliverables

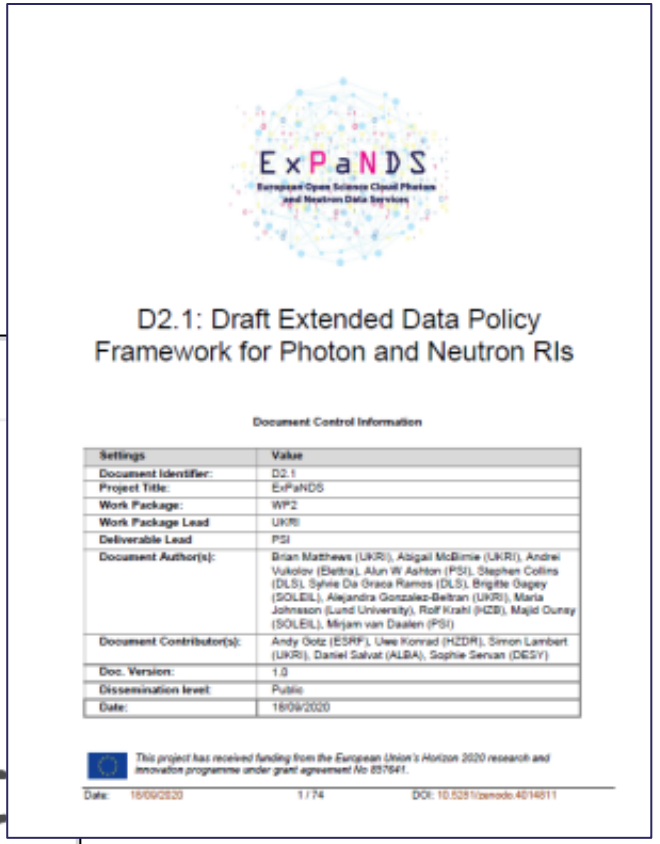
Deliverable		Month	Task	Status
D2.1	**Draft extended data policy framework for Photon and Neutron RIs**	**12**	**T2.1**	**Done**
D2.2	**Draft recommendations for FAIR Photon and Neutron Data Management**	**15**	**T2.3**	**In Progress**
D2.3	Final data policy framework for Photon and Neutron RIs	24	T2.1	Next
D2.4	DMPs for Photon and Neutron RIs	27	T2.2	Started working with PaNOSC
D2.5	Advanced infrastructure for PIDs in Photon and Neutron RIs	30	T2.4	About to start
D2.6	Self-evaluation of Photon and Neutron RIs for FAIR data certification	39	T2.5	Not started
D2.7	Final Recommendations for FAIR Photon and Neutron Data Management	36	T2.3	Not started
D2.8	Active DMPs for Photon and Neutron RIs	39	T2.2	Not started
D2.9	Report on promotion of FAIR data within Photon and Neutron RIs	42	T2.6	Started – need input

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

D2.1: Towards FAIR policies for Photon and Neutron RIs

- ExPaNDS and PaNOSC have been working together to revise data policies frameworks in the light of the 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
  - 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





105 views

82 downloads

PaNOSC and ExPaNDS projects have received funding from the European Union's Horiz
and innovation programme under grant agreements 823852 and 857641, respectively.

Adding FAIR to Policy

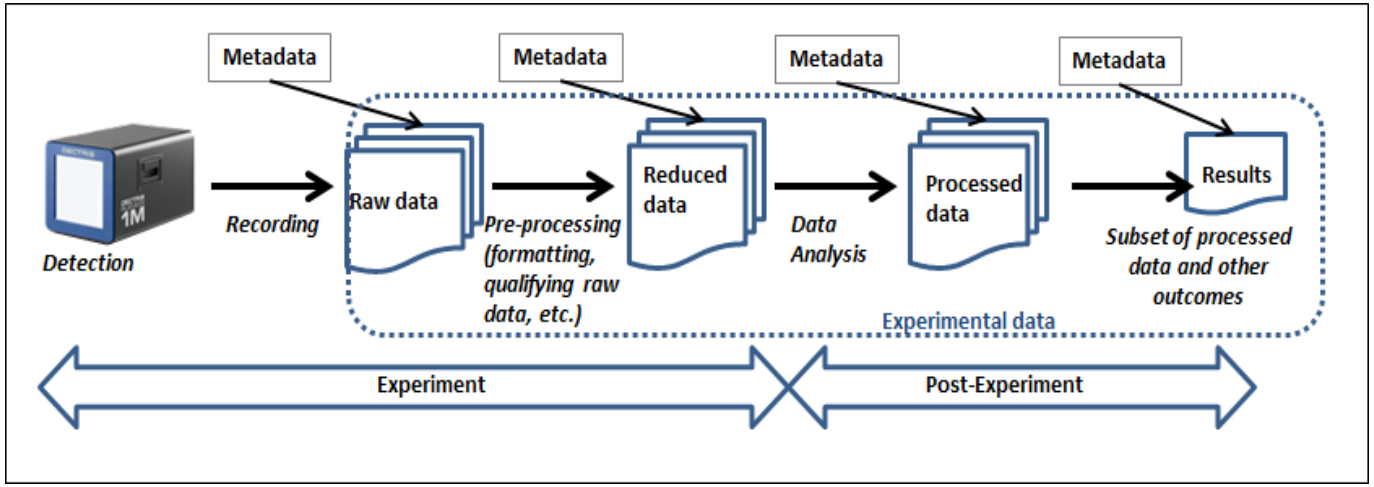
"RIs' data policies should enable the experimental data in scope to be FAIR"

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



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

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

T2.1: Policy Next Steps

- Consult on recommendations with ExPands Facilities
- Revise and extend guidelines
- Convergence with PaNOSC Model policy
- Final Policy recommendations: August 2021

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

D2.2: Draft recommendations for FAIR Photon and Neutron Data Management

- In advanced draft
  - Lead by Daniel Salvat

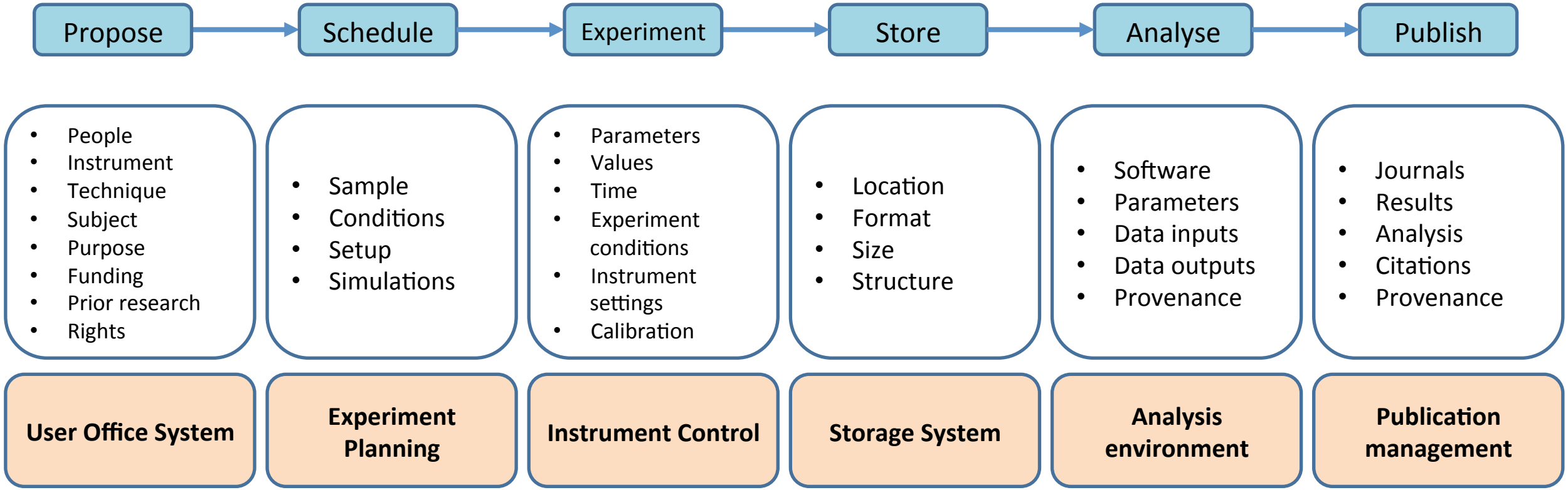
- A detailed analysis of the steps of Experimental process and what is needed for FAIR-ness at each stage

- Towards recommendations for Metadata

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

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

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

FAIR Guidelines: Findability

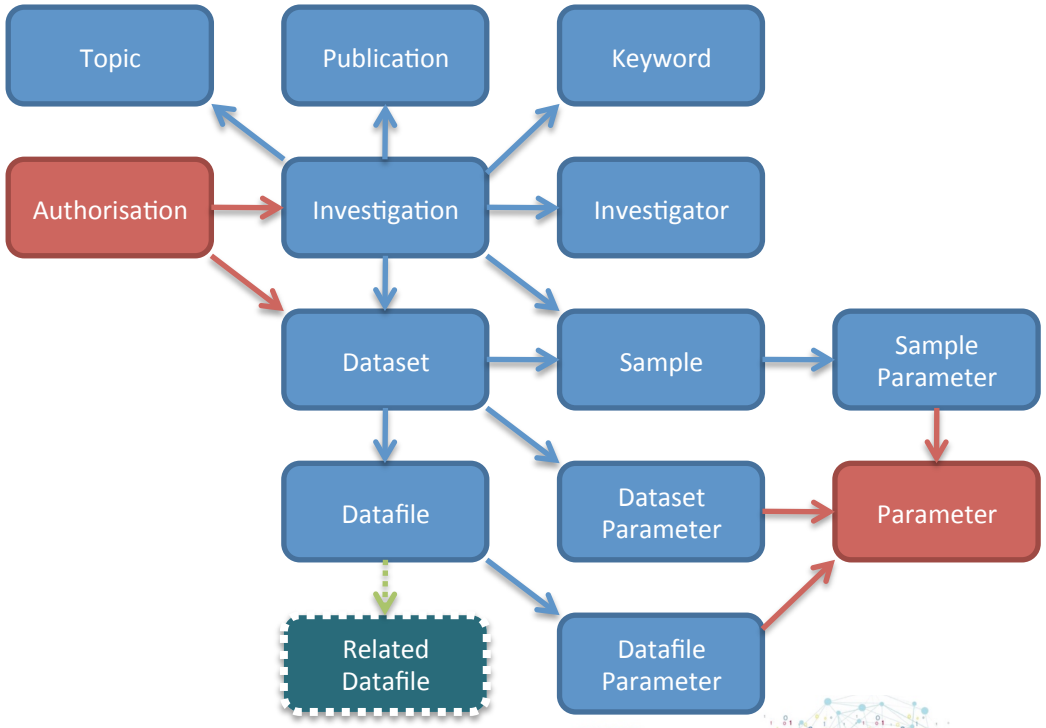
- Metadata catalogues
  - Searchable resource where data, and associated metadata can be registered or indexed
  - New use cases being considered
  - Search for humans: graphical user interfaces
  - Search for machines: common Application Programming Interface
  - Integration with EOSC services
  - Integration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



PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Accessibility

- Metadata catalogues
  - Authentication and authorisation
  - DOI for datasets
  - Landing pages with metadata
    - Methods to access data if open
  - Are identifiers needed at other levels?
    - DOIs for visits?
    - DOIs for datafiles?

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

Alejandra Gonzalez-Beltran

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

panosc photon and neutron open science cloud

ExPaNDS European Open Science Cloud Photon and Neutron Data Services

Reusability

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**

R1.3. (meta)data meet **domain-relevant community standards**

Your data

- Rich descriptions, provenance, standards

Others' data

- Data usage license
- All the above

Facilities / Institutional Policies

- Embargo periods
- Licenses

Ashton, Alun, Da Graca Ramos, Silvia, Matthews, Brian, Salvat, Daniel, & Sander, Knut. (2019, December 19). ExPaNDS Data Landscaping Survey. Zenodo. http://doi.org/10.5281/zenodo.3673811

Gotz, Andy, Perrin, Jean-Francois, Fangohr, Hans, Salvat, Daniel, Gliksohn, Florian, Markvardsen, Anders, et al. (2020). PaNOSC FAIR Research Data Policy framework. Zenodo. doi: https://doi.org/10.5281/zenodo.3862701

Matthews, Brian. Draft extended data policy framework for Photon and Neutron RIs. (Zenodo, 2020). doi: https://doi.org/10.5281/zenodo.4014811

Alejandra Gonzalez-Beltran

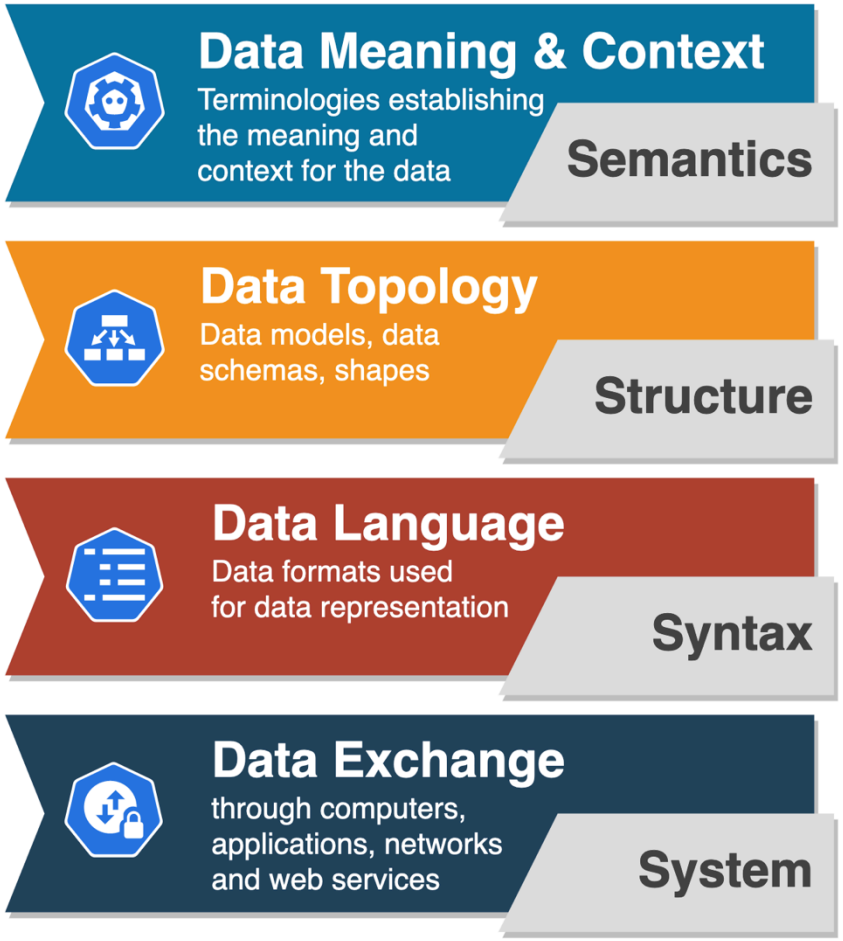
PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

FAIR Guidelines: Interoperability

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

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

I3. (meta)data include **qualified references** to other (meta)data



Common vocabularies and mappings	WP2 & WP3 - work on vocabularies framework & ontologies
Common data schemas and validation processes	Use of NeXuS and its application definitions at facilities
Common data formats for different types of data and metadata	Use of HDF5 format and metadata catalogues
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

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

T2.4: 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	doi, PURL, Handle.Net® Registry
Data	ARK (Archival Resource Key), Handle.Net® Registry, PURL, doi
People	ORCID, RESEARCHERID, isni
Organisation	ROR, GRID Global Research Identifier Database, isni
Project	RAiD Research Activity Identifier
Instrument	RDA Research Data Alliance — Persistent Identification of Instruments WG
Sample	InChITRUST, RRID, IGSN
Software	doi, RDA Research Data Alliance — RDA/FORCE11 Software Source Code Identification WG

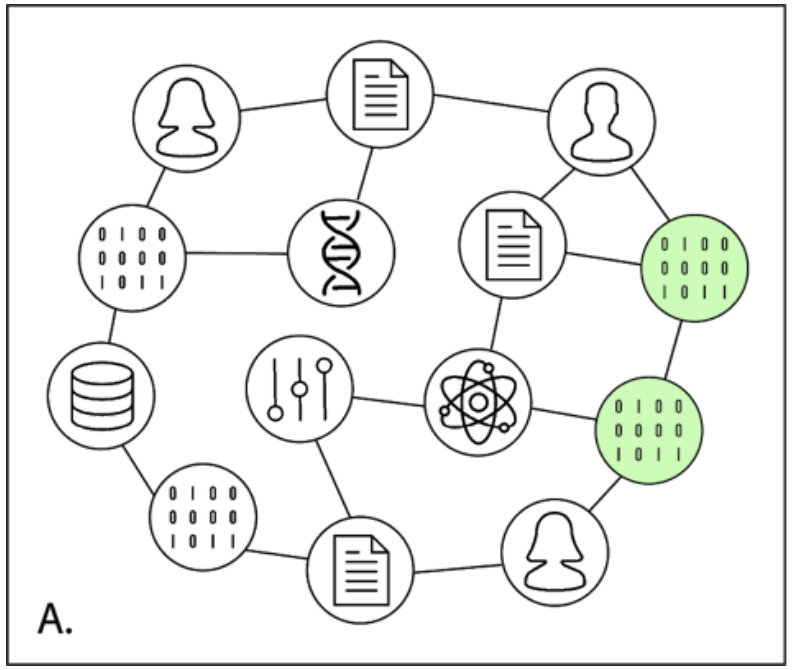
PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

photon and neutron open science cloud

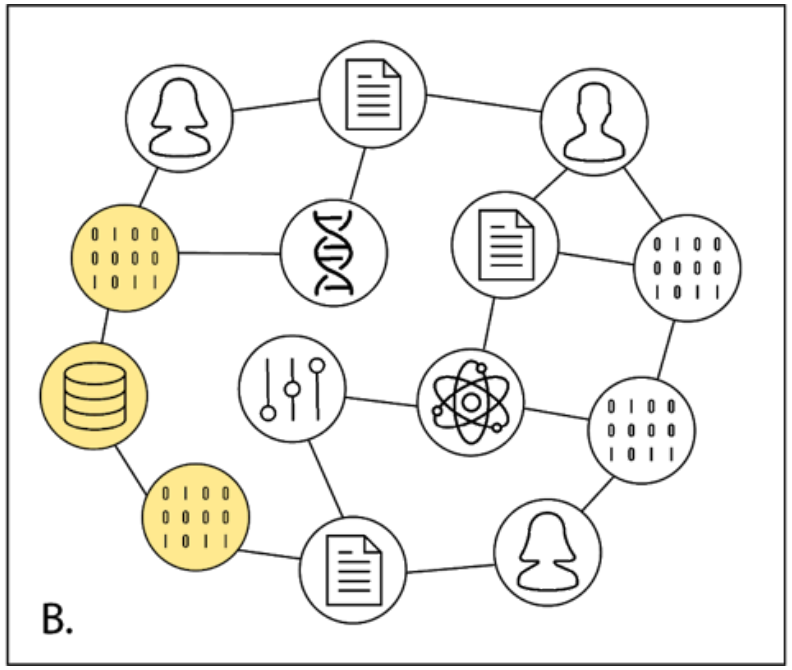
ExPaNDS European Open Science Cloud Photon and Neutron Data Services

Resource Identification Graph

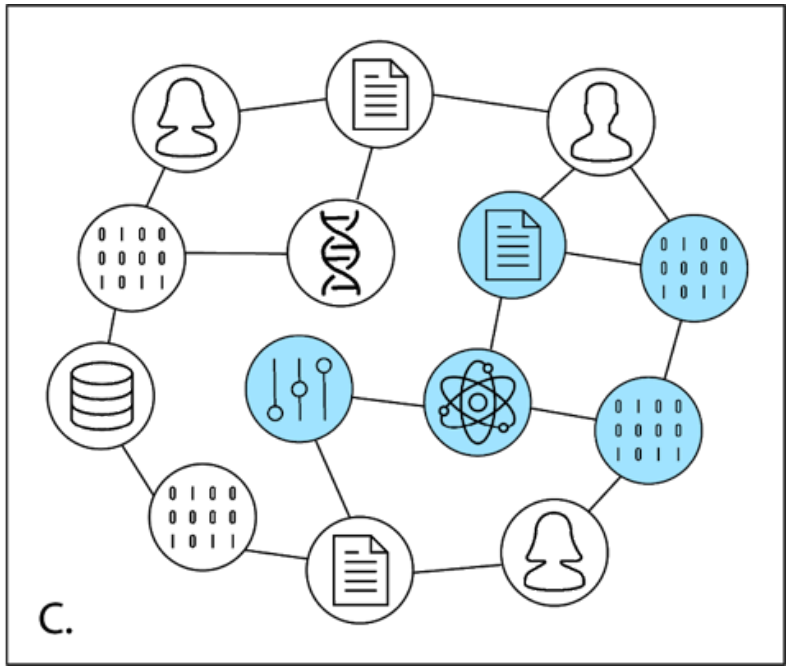
PIDs + metadata support linkages between resources



Software ←→ Software



Dataset ←→ Dataset



All digital objects associated with a dataset

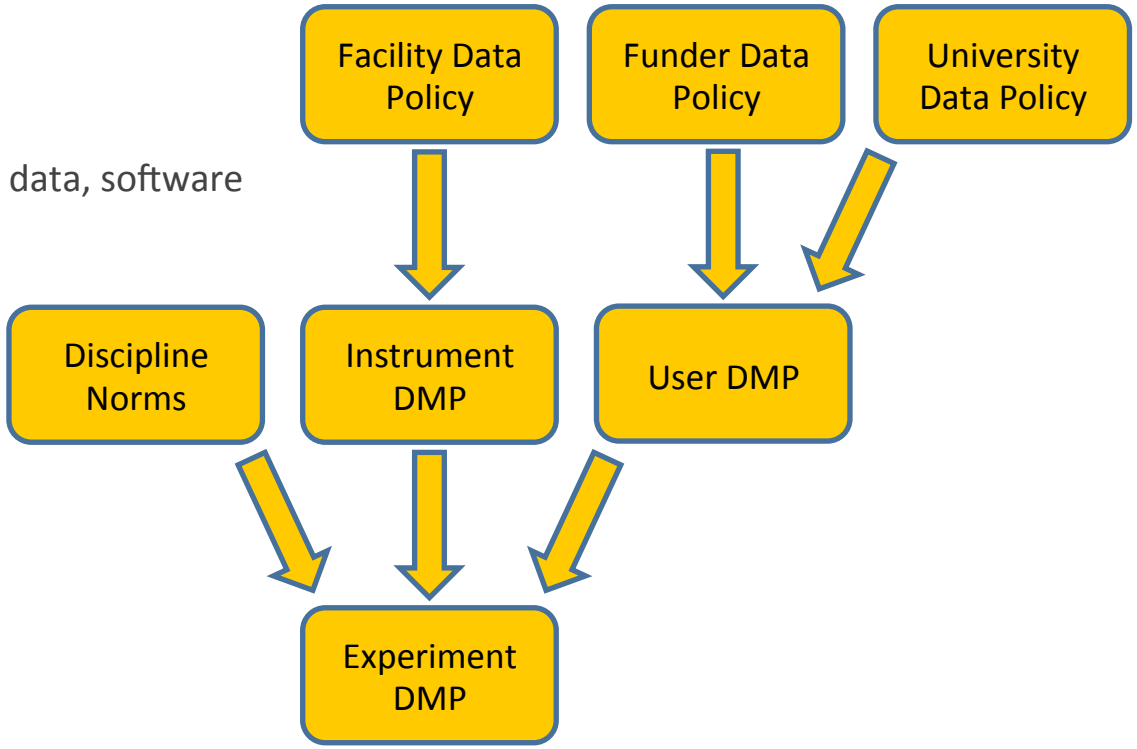
Alejandra Gonzalez-Beltran

Introducing the PID Graph - March 28, 2019 by Martin Fenner and Amir Aryani
https://doi.org/10.5438/jwvf-8a66

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

T2.2: 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



PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

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

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

KPIs for WP2

WP2 Enabling FAIR data for EU PaN National Ris

Level of FAIR maturity of participating facilities (% - self-assessment)

# of facilities data policy explicitly mentioning FAIR

# of facilities data policy mentioning DMPs

# of facilities offering training related to FAIR data

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Questions For Discussion

- Making the WPs work more closely together

  Key objectives

  - Harmonised Data Policy (guidelines and model policy)
  - Promoting standards (e.g. NEXUS, Metadata and Ontologies, PIDs)
  - DMPs (guidelines and template)
  - FAIR Metrics for Facilities

- Advocacy for FAIR data
  - who should we influence and how?
  - Senior managers - LEAPS/LENS
  - Users
  - How can we maximise our impact ?

- Take up
  - Challenges in implementing a FAIR Data Policy
  - Who is using DMPs and challenges in getting them used.

- Working with standards bodies

- Use cases and Success Stories
  - how do we gather use cases for added value ?

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

PaNOSC & ExPaNDS Annual Meeting

Thankyou

Brian Matthews

Brian.Matthews@stfc.ac.uk

https://expands.eu

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

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.