PaNOSC Overview

PaNOSC

Photon and Neutron Open Science Cloud

WP2 - Data Policy + Stewardship

Andy Götz

15/01/2019

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

WP2 Objectives

1. Definition and **harmonisation** of PaN specific **data policies**
2. Definition and **adoption** of **common open standards** for interoperability.
3. **Registering** with and citing of these **standards** by standards bodies and **publishers**.
4. **Stewardship** of data handled by the involved research infrastructures **according** to the **FAIR** principles.
5. **Citing** of PaN **data repositories** and data descriptors by **publishers**.
6. Produce **guidelines** for **best practices** based on experience of those PaN partners who already have Open Data data policies.
7. Develop **guidelines** for dealing with typical PaN issues like **huge data sets** will be dealt with by exploring **data reduction** and **compression** schemes which reduce the burden on the data infrastructure.

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

WP2 Tasks

1. Task 2.1: **Lessons learned and FAIR Definitions (M1-M6)**
   Leader: CERIC-ERIC. *Contributors: ESRF, ILL, XFEL.EU, ESS, ELI*

2. Task 2.2: **Updated PaNOSC Data Policy framework (M6-M18)**
   Leader: ESS. *Contributors: ESRF, ILL, XFEL.EU, ESS, ELI*

3. Task 2.3: **Approve Data Policy framework (M9-M36)**
   Leader: CERIC-ERIC. *Contributors: ESRF, ILL, XFEL.EU, ESS, ELI*

4. Task 2.4: **Create Best Practices Guidelines (M1-M24)**
   Leader: ESRF. *Contributors: ESS, ELI, XFEL.EU,CERIC-ERIC*

5. Task 2.5: **Implement DMP template (M12-M36)**
   ***Leader*: *ESS*.** *Contributors: ILL, CERIC-ERIC*

6. Task 2.6: **Validation of Data Policy implementation (M12-M36)**
   Leader: CERIC-ERIC. *Contributors: ELI*

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

WP2 Deliverables

1. **Deliverable 2.1 PaNOSC data policy framework updated**
   M18, ESRF (R, PU)
2. **Deliverable 2.2 DMP Template for facility users published**
   M36, ESS (R, PU)
3. **Deliverable 2.3 Guidelines on best practices implementing the PaNOSC data policy framework published.**
   M24, ESRF (R, PU)
4. **Deliverable 2.4 Integration of the policy in the User Access and facility information systems**
   M36, CERIC (R, DEC)

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

WP2 Data Policy – led by ESRF

Before PaNOSC (2018)

	ILL	ESRF	CERIC	XFEL	ELI	ESS
Data Policy	2011	2016	2014 (3/8)	2017	In Progress	2017

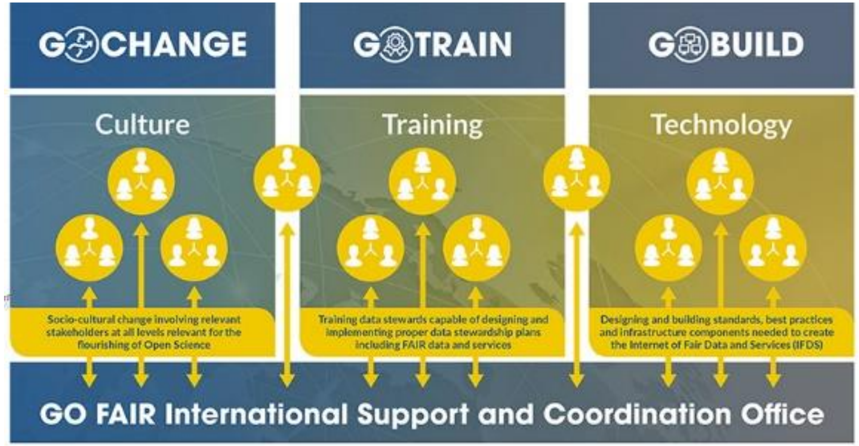
After PaNOSC (2023)

	ILL	ESRF	CERIC	XFEL	ELI	ESS
Effort (PMs)	10	17	12	3	20	14
Common Framework Data Policy	2011	2016	2019	2017	2019	2017
Data Archiving	YES	YES	YES	YES	YES	YES
DOIs	YES	YES	YES	YES	YES	YES
Open Data	YES	YES	YES	YES	YES	YES
DMP templates	YES	YES	YES	YES	YES	YES

FAIR organisations + data repositories

1. GO-FAIR
2. OpenAire-Advanced
3. R3data

GO FAIR


GO CHANGE
GO TRAIN
GO BUILD
Culture
Training
Technology
Socio-cultural change involving relevant stakeholders at all levels relevant for the flourishing of Open Science
Training data stewards capable of designing and implementing proper data stewardship plans including FAIR data and services
Designing and building standards, best practices and infrastructure components needed to create the Internet of Fair Data and Services (IFDS)
GO FAIR International Support and Coordination Office

OpenAIRE | ADVANCE

The Open Science Pillar of EOSC

re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

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

Scientific standards bodies

International Union of
CRYSTALLOGRAPHY

- **IUCr Committee on Data (CommDat) advises on:**

  Raw data and its metadata preservation and their digital object identifiers

  Data mining within individual and across two or more databases

  Data and software development

  Data and instrumentation

  Data policy drivers as received from policy makers (*e.g.* funding agencies)

  Data type domains (discrete *versus* diffuse, *i.e.* continuum scattering)

  Data and eScience

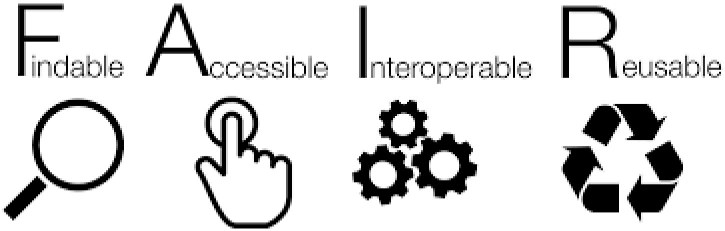
  Data and data publishing [*IUCrData*; recommendation of editors for *IUCrData*; linking of data to articles in IUCr publications; new article categories involving data]

  Data repositories

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

FAIR principles

To be **Findable**:



F1. (meta)data are assigned a globally unique and eternally persiste
F2. data are described with rich metadata.
F3. (meta)data are registered or indexed in a searchable resource.
F4. metadata specify the data identifier.

To be **Accessible**:

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.

To be **Interoperable**:

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.

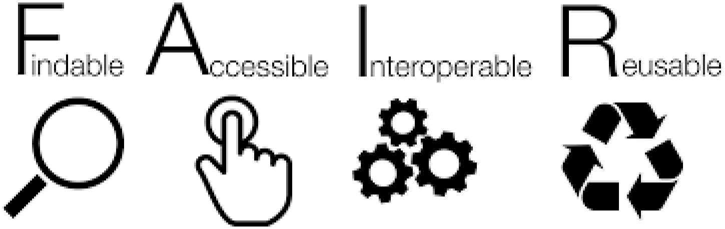
To be **Re-usable**:

R1. meta(data) have 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 their provenance.
R1.3. (meta)data meet domain-relevant community standards.

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

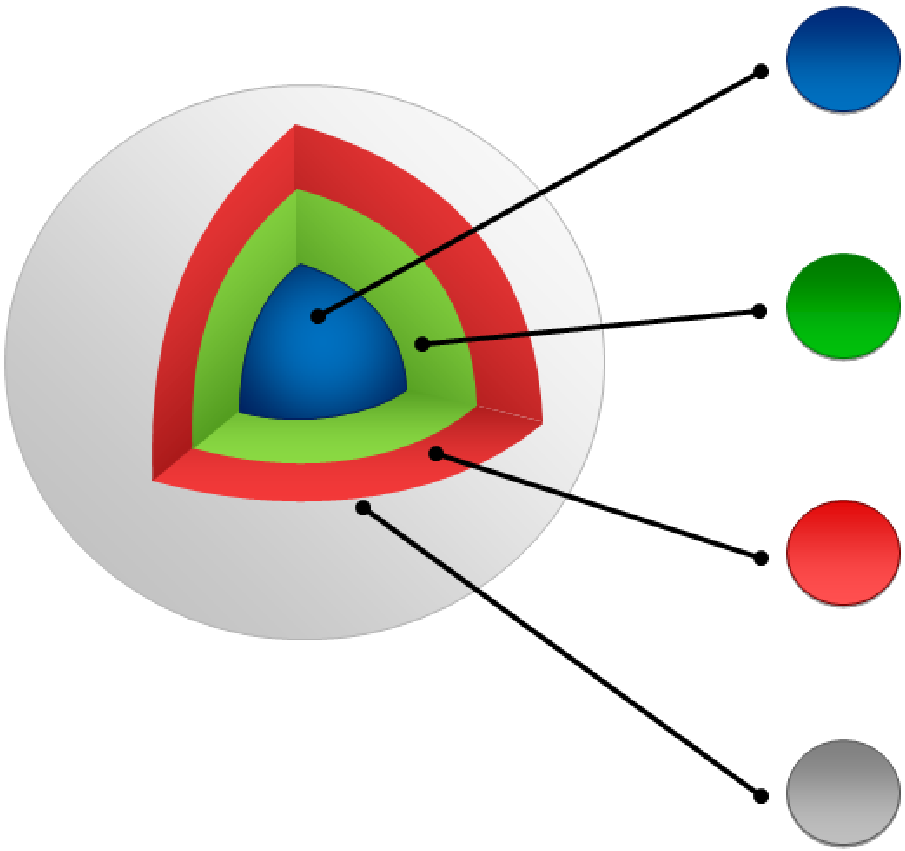
How FAIR are our data policies and data ?

- Compare current **Data Policies** with latest FAIR principles
- Consult with **FAIR experts** (GO FAIR, OpenAire, B2FIND)
- Define **KPI**s for FAIR data and measure
- **Train** scientists how to create + use FAIR data



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

FAIR Data Objects



DATA

The core bits

At its most basic level, data is a bitstream or binary sequence. For data to have meaning and to be FAIR, it needs to be represented in standard formats and be accompanied by Persistent Identifiers (PIDs), metadata and code. These layers of meaning enrich the data and enable reuse.

IDENTIFIERS

Persistent and unique (PIDs)

Data should be assigned a unique and persistent identifier such as a DOI or URN. This enables stable links to the object and supports citation and reuse to be tracked. Identifiers should also be applied to other related concepts such as the data authors (ORCIDs), projects (RAIDs), funders and associated research resources (RRIDs).

STANDARDS & CODE

Open, documented formats

Data should be represented in common and ideally open file formats. This enables others to reuse the data as the format is in widespread use and software is available to read the files. Open and well-documented formats are easier to preserve . Data also need to be accompanied by the code use to process and analyse the data.

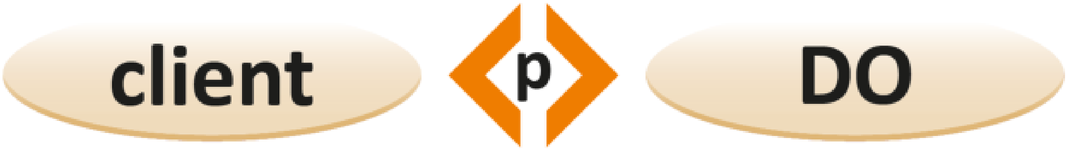
METADATA

Contextual documentation

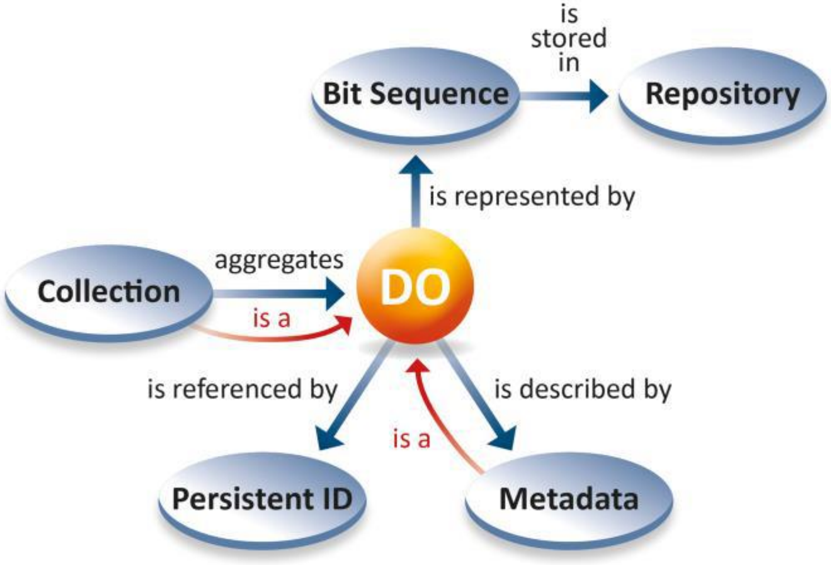
In order for data to be assessable and reusable, it should be accompanied by sufficient metadata and documentation. Basic metadata will enable data discovery, but much richer information and provenance is required to understand how, why, when and by whom the data were created. To enable the broadest reuse, data should be accompanied by a 'plurality of relevant attributes' and a clear and accessible data usage license.

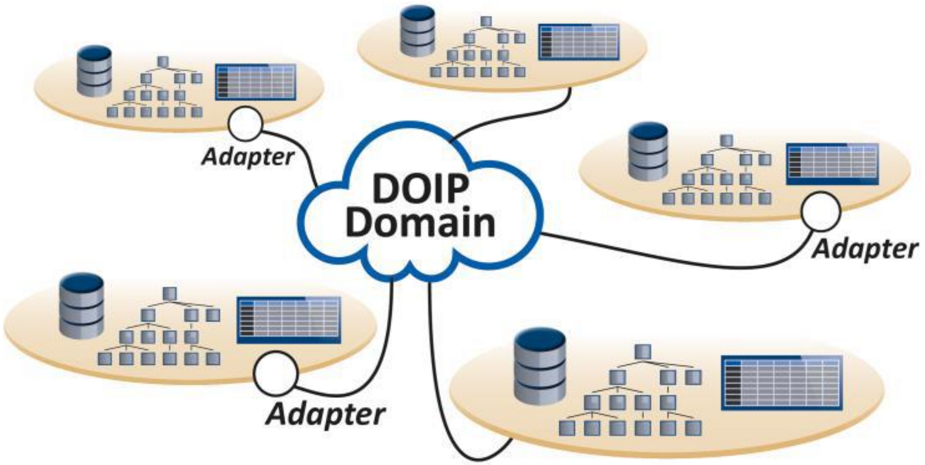
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852

FAIR Data Objects – how far should we go ?



Digital Object Interface Protocol





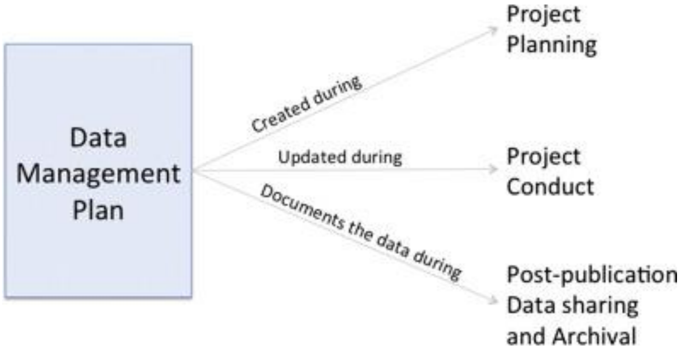
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852

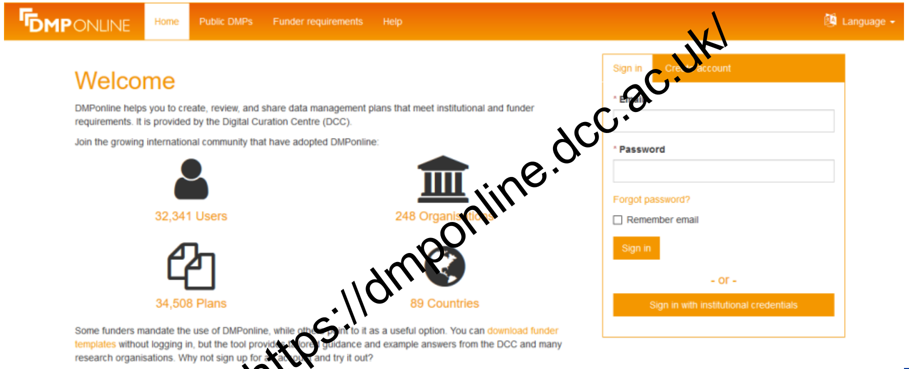
Data Management Plans (DMP)

A data management plan or DMP is a formal document that outlines how data are to be handled both during a research project, and after the project is completed.

A number of online services exist which we plan to re-use

Adapt / customize existing services to help users fill in DMPs





Welcome

DMPonline helps you to create, review, and share data management plans that meet institutional and funder requirements. It is provided by the Digital Curation Centre (DCC).

Join the growing international community that have adopted DMPonline:

32,341 Users

248 Organisations

34,508 Plans

89 Countries

Some funders mandate the use of DMPonline, while others point to it as a useful option. You can download funder templates without logging in, but the tool provides tailored guidance and example answers from the DCC and many research organisations. Why not sign up for an account and try it out?

* Email

* Password

Forgot password?

□ Remember email

Sign in

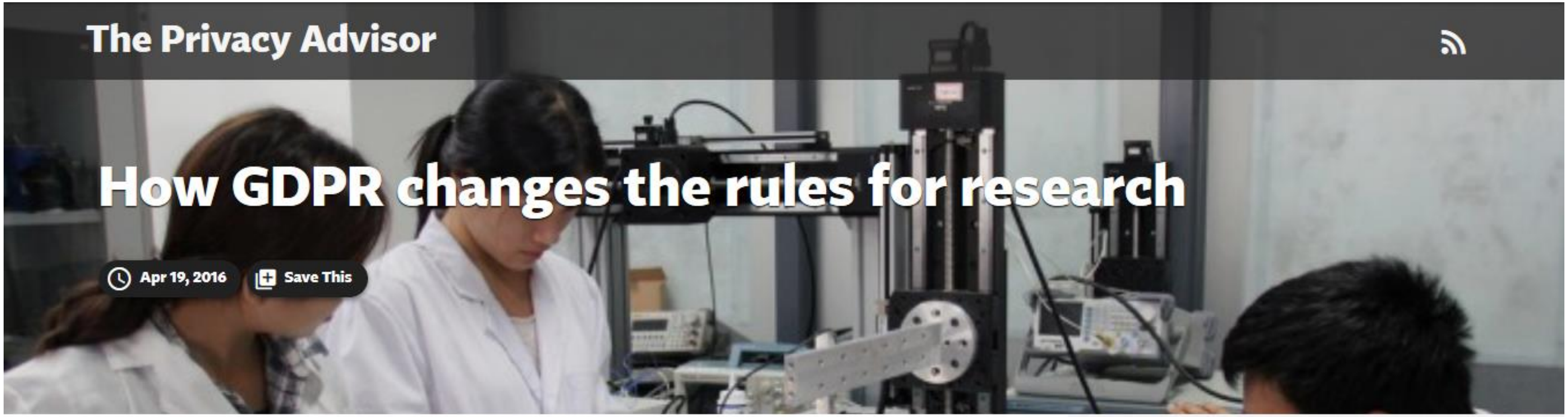
- or -

Sign in with institutional credentials

https://dmponline.dcc.ac.uk/

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

GDPR – how does it affect scientific data


The Privacy Advisor

How GDPR changes the rules for research

Apr 19, 2016

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

Key Performance Indicators (KPIs)

- Number of datasets cited in publications
- Number of publications NOT citing datasets
- Number of datasets really re-used
- Number of DMPs generated by PaNOSC DMP tool
- Number of data policies adopting the PaNOSC framework

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

Open Questions

- How to reference users profiles e.g. Orcid ?
- How to include additional material e.g. videos, files, … ?
- How to publish reduced / analysed data ?
- How to generalize the use of e-logbooks as rich metadata ?
- Should the embargo be shorter / longer ?
- Should we propose a different licence ?

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

Conclusion

- **FAIR principles** are the **de facto standard** for **Open Data**
- **PaNdata/PaNOSC Data Policies** needs to be **updated + adopted**
- **Large body** of **expertise** out there to **consult** and **re-use**
- Need to **train scientists** on **FAIR data**
- Need to **demonstrate Open Data use + re-use**

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