

# WP2 Data policy and Stewardship

**5 November, 2019**

**Author: Andy Götz (ESRF)**



# Talk overview

- Objectives, Tasks, Deliverables, KPIs
- Input from FAIR organisations
- Updating the Data Policy framework
- Best practices for implementing Data Policy
- Data Management Plan templates
- Spending profile



## 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.



## 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*



## 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)



# Key Performance Indicators

- **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**



## 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 DP adopted	2011	2016	2019	2017	2019	2017
Implementation of Data Policy	2017	2020	2021	2017	2021	2019
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



# GDPR compliancy

- **Answer from Datacite:**

*Hello Andy -*

*Please see our [privacy policy](#) which outlines the DataCite's GDPR compliance with various services. In addition, the onus is on the member to comply with GDPR when submitting metadata:*

*By providing DataCite with personal data which was provided to the Member by a natural person(s), including Member staff (the "origin party"), the Member guarantees that:*

*-- the Member collected and processed the data in accordance with applicable law, including the General Data Protection Regulation (GDPR);*

*the Member acquired the origin party's informed consent to share the data with DataCite;*

*-- the Member acquired the origin party's consent for the data to be transferred to Germany for processing.*

*-- The Member further agrees that it will maintain appropriate mechanisms to ensure that it will provide natural person(s) whose personal data it provides to DataCite with a means to have access to, to correct, and to delete such data and understands that the burden is on the Member to communicate such corrections or deletions to DataCite.*





# FAIR Organisations – January 2019

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



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



# FAIR Organisations – November 2019

→ ESRF chosen as one of 10 repositories to be FAIR certified



→ Would be great to share and learn from other clusters



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



## D6.1

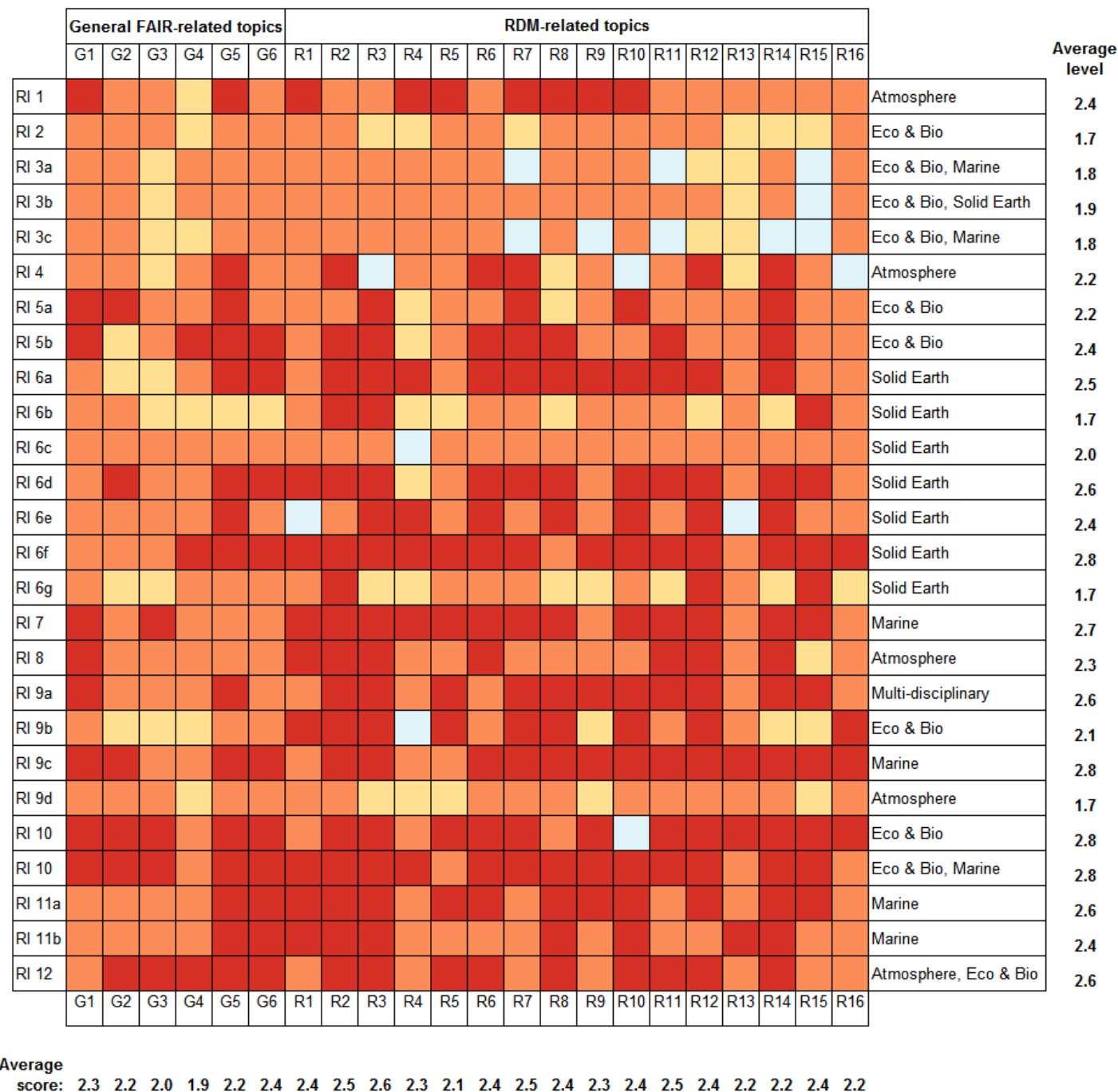
# “Inventory & gap analysis of FAIR training materials”

by Margaret Hellström  
(2/10/2019)



This project has received funding from the European U

RI current knowledge level



# Updating PaN Data Policy framework

Policy	Description	ILL	ESRF	XFEL	ESS	CERIC	ELI
<b>Based on PaN-data policy</b>	<a href="#">PaN-data policy deliverable 2.1 of EU FP7 project PaN-data Europe</a>	YES	YES	YES	YES	YES	-
<b>Date</b>	The date when the latest version of the DP was released	July 2017	November 2015	June 2017	February 2019	February 2019	-
<b>Versioned</b>	The DP have a version number	NO	NO	NO	NO	NO	-
<b>DOI</b>	The DP has a DOI	NO	NO	NO	NO	NO	-
<b>FAIR principles</b>	The DP mentions the FAIR principles explicitly	NO	NO	NO	NO	NO	-
<b>GDPR compliant</b>	Specific mention / treatment of GDPR	NO	NO	NO	NO	NO	-
<b>Introduction</b>	The DP has an introduction motivating why a DP is necessary	YES	NO	YES	NO	YES	-
<b>General principles</b>	Does the DP have a section on General Principles	YES	YES	YES	YES	YES	
<b>Definitions</b>	The DP has clear definitions for the principal concepts e.g. PI, raw data	YES 20 terms	YES 11 terms	YES 22 terms	YES 10 terms	YES 10 terms	-



# Updating PaN Data Policy framework

Policy	Description	ILL	ESRF	XFEL	ESS	CERIC
<b>Metadata</b>	Types of metadata identified in the DP	only metadata from instruments, excludes notebooks	metadata captured automatically y instruments			instrument, context of the experiment, experimental conditions
<b>Raw data</b>	List of raw data stored	All data resulting from non industrial experiment, including test, evaluation				data collected from measurements performed at CERIC – ERIC Partners Facilities
<b>Processed data</b>	Does the DP cover processed data	Yes, if at the ILL	mentions derived data	yes	yes	not mentioned
<b>Data format</b>	The DP's preferred data format for data	NONE	NONE	NONE	NONE	Nexus/HDF5
<b>Metadata archiving</b>	How long metadata are archived for (GDPR compliant)	Forever	Forever	5-10y	>10y	Not defined
<b>Raw data archiving</b>	How long raw data are archived for	Forever	10y	5-10y	>10y	Not defined

# Updating PaN Data Policy framework

Policy	Description	ILL	ESRF	XFEL	ESS	CERIC
<b>High-level metadata</b>	How are high-level metadata treated i.e. Title, Authors, Instrument, ...	Public right after the experiment.	Public when experiment starts	Public directly after experiment	not in policy	not in policy
<b>DOI</b>	What are DOIs defined for	raw data + processed if they exist on the archive storage	raw data	data sets (blurry)	raw data files	raw data + metadata
<b>Proposals</b>	What access to proposals is foreseen in DP	Proposal team (which includes local contact,	PI, experimental team, local beamline staff, admins	proposal team, experiment team, members of review panel, XFEL staff, those authorized by PI	not defined	experimental team and staff



# Updating PaN Data Policy framework

Policy	Description	ILL	ESRF	XFEL	ESS	CERIC
<b>Good practices</b>	What good practices are listed in the DP (link to details)	Very few, but need to cite the DOI in every related publication)	Completeness of metadata	Completeness of metadata, linking results of data analysis to the raw data	none defined	Completeness of metadata, facility help to capture of metadata, collaborations between PI and researchers, linking results of data analysis to the raw data
<b>Publication</b>	How to cite data	Citing DOI is mandatory	Citing DOI is mandatory	Citing PID of data	DOI	unique identifier
<b>Proprietary data</b>	How industrial/commercial data are treated in the DP	Separate DP from the general one	explicitly excluded, metadata can be stored	explicitly excluded from DP	not defined will be added later	an agreement must be established



# Updating PaN Data Policy framework

Policy	Description	ILL	ESRF	XFEL	ESS	CERIC
<b>Metadata catalogue</b>	What metadata catalogue is mentioned in the DP	<a href="https://data.ill.eu">https://data.ill.eu</a>	<a href="https://data.esrf.fr">https://data.esrf.fr</a>		none	none
<b>Embargo period</b>	The length of the embargo period	3 to 5 years	3 years	3 years	3 years	not defined
<b>Extending embargo period</b>	Conditions under which the embargo period can be extended	A request could be made to the science director.	on request of PI	defined	defined	defined
<b>Principal Investigator</b>	Role of the PI	Could grant access to data to other scientist. Could also "open" the data at any time.	can extend embargo, mint dois	defined	defined	defined
<b>E-logbook</b>	Role of E-logbooks	E-logbook is part of the metadata	not defined	not defined	not defined	defined





# Updating PaN Data Policy framework

Policy	Description	ILL	ESRF	XFEL	ESS	CERIC
<b>Ownership of IP</b>	Who owns the IP produced from the data	The person who produce the results form the data.	the team who generate the data from which the ip is derived	determined by the contractual obligations of the person(s) performing the analysis	defined for processed data	defined
<b>Enforcement of DP</b>	How is the DP enforced	The acceptance of the DP is mandatory for the submission of a proposal. Raw data are preserved automatically. The difficult part is the citation of DOIs in related articles.	automatic procedures as part of experiments, trust		not defined	
<b>Process of accepting DP</b>	Describes the process how the DP was accepted	3 years of internal discussions before reaching agreement with the ILL scientits. Dp has been presented to our shareholders and accepted.	presentations to EXPD, discussions with Directors, endorsement by SAC and Council, accepted by Users at proposal submission	Upon registration in User Portal and for subsequent version upon first login after change is made	acceptance by ESS council; for Users default agreement at proposal submission;	
<b>Process of accepting changes to DP</b>	Describes the change process for the DP	Discussions take place in an internal group of scientific representatives + User office and IT presented to the board of directors for validation.	Not defined		In scope of ESS structure	
<b>Review of DP</b>	Who reviewed the DP	Not defined	No-one		all ESS internal stakeholders	

# Updating PaN Data Policy framework

Policy	Description	ILL	ESRF	XFEL	ESS	CERIC
<b>ORCID ID</b>	How ORCID ID's are mentioned in the DP	Mentioned only in the privacy policy	no		no	no
<b>Uploading of results</b>	Can processed results be uploaded to the DP	yes	no		yes	yes
<b>Licence</b>	Licence under which data are made available	CC BY 4.0	CC BY 4.0	CC BY	not defined	not defined
<b>Anonymous access</b>	Is anonymous access allowed	no	yes	no	no	no
<b>Implementation of DP</b>	How complete is the DP for all data produced	Complete, but can always be improved	50% of beamlines	raw, processed results	raw processed, analysed data are discussed	
<b>Date of completion of DP implementation</b>	When will the DP be completely implemented	2012	2020	fairly complete	from day 1	in progress
<b>Does your DP include DOIs for non scientific data</b>		no	yes e.g. documents			

# Data Policy implementation BEST PRACTICES

- **Input being gathered in google doc:**

<https://docs.google.com/document/d/1TvlkRYgxduADzDeeNjsyReVDtyibnGpp5oH9cFiXpfM/edit?usp=sharing>

- **Next step** is to write them up collectively as a **Handbook of Best Practices** in a collective sprint session e.g. Open Science Training Handbook (<https://open-science-training-handbook.gitbook.io/book/>)
- **FAIR WG** has expressed interest in this document



# Data Management Plan Templates

- **For whom and when?**
- **Vision:**
  - provide a service to fill in DMP templates for users who require a DMP for their funding agency
  - Provide a DMP service for partners who (will) require them
- **Use an existing tool e.g. DMPTool or DMP-online**
- **Configure it based on needs of PaNOSC partners**



## Open issues

- **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 ?**



# Spending profile

- **WP2 is underspending (25%), we have identified new resources**
- **Need to increase effort over next 6 months to produce updated PaNOSC Data Policy framework Deliverable D2.1 in M18**
- **Another increase in effort for DMP tool**
- **Need more regular meetings + engagement from partners in WP**
- **WP2 ends in M36**



# Thank you

**[andy.gotz@esrf.fr](mailto:andy.gotz@esrf.fr)**

