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 Da Graca Ramos (DLS), Brigitte Gagey (SOLEIL), Alejandra Gonzalez-Beltran (UKRI), Maria Johnsson (Lund University), Rolf Krahl (HZB), Majid Ounsy (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

ExPaNDS

Abstract

Research Infrastructures within the Photon and Neutron science community have a significant track record in publishing and implementing data policies to support the access and use of experimental data generated by the user. Recent developments within European science policy, in particular the drive towards supporting FAIR data in the European Open Science Cloud has meant that it is timely to review and revise these data policies. The PaNOSC project has developed a new *PaNOSC Data Policy Framework* with the contribution of ExPaNDS partners. The aim of this *Deliverable 2.1: Draft extended data policy framework for photon and neutron research infrastructures* is to reflect upon and extend that work. We review the FAIR data policy landscape at European and national levels, consider the current state of data policy adoption and implementation at ExPaNDS partner facilities, and examine existing FAIR ecosystem data policy recommendations, in particular, from the *Turning FAIR into reality* report and the recent *FAIRsFAIR Deliverable 3.3: Policy enhancement recommendations* In response, we make twenty-six recommendations of our own that serve to translate these recommendations to the local level of photon and neutron research infrastructures. As the report's key output, we present thirty elements that should inform a data policy framework for photon and neutron research infrastructures. These focus on enabling FAIR data and are those elements about which RIs **need to make choices** — both on the level of commitments that they themselves are prepared to make as well as with regards to the obligations that they will place on users. We conclude with some next steps that will inform the final version of this deliverable: *D2.3 Final data policy for photon and neutron research*.

Document Log

Version	Date	Comment	Author/Partner
Complete Draft	21/08/2020	Internal review version	Brian Matthews (UKRI), Abigail McBirnie (UKRI)
1.0	18/09/2020	Final version for submission	Brian Matthews (UKRI), Abigail McBirnie (UKRI)

Abbreviations and acronyms

ANR	French National Research Agency
BBSRC	Biotechnology and Biosciences Research Council
CC	Creative Commons
CERIC-ERIC	Central European Research Infrastructure Consortium
CNRS	French National Centre for Scientific Research
CoSO	Committee for Open Science
DESY	Deutsches Elektronen-Synchrotron
DFG	German Research Foundation
DLS	Diamond Light Source
DMP	Data Management Plan
EC	European Commission
EGI	European Grid Infrastructure Foundation
EOSC	European Open Science Cloud
EPSRC	Engineering and Physical Sciences Research Council
ERC	European Research Council
ESRF	European Synchrotron Radiation Facility
ESS	European Spallation Source
EU	European Union
ExPaNDS	European Open Science Cloud (EOSC) Photon and Neutron Data Service
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
HZB	Helmholtz-Zentrum Berlin
HZDR	Helmholtz-Zentrum Dresden-Rossendorf
ICDI	Italian Computing and Data Infrastructure
ILL	Institut Laue-Langevin
INIST	Institute for Scientific and Technical Information
INRAE	French National Research Institute for Agriculture, Food and Environment
IPR	Intellectual Property Rights
NERC	Natural and Environmental Research Council
ORD Pilot	Open Research Data Pilot
OSPP	Open Science Policy Platform
PaN	Photon and Neutron
PaNOSC	The Photon and Neutron Open Science Cloud

PCIs	Practical Commitments for Implementation
PI	Principal Investigator
PID	Persistent Identifier
PSI	Paul Scherrer Institute
R&D	Research and Development
RD&I	Research, Development, and Innovation
RI	Research Infrastructure
RoPs	Rules of Participation
SNSF	Swiss National Science Foundation
STFC	Science and Technology Facilities Council
swissuniversities	the umbrella organization of Swiss universities
TFiR	Turning FAIR into Reality report
UKRI	UK Research and Innovation
WG	Working Group
WP	Work Package

Executive Summary

This report, *ExPaNDS Deliverable 2.1: Draft extended data policy framework for photon and neutron research infrastructures*, considers how data policies for Research Infrastructures providing Photon and Neutron facilities should be framed, in particular in the light of supporting open science and FAIR data. It begins with an overview of FAIR research data policy. After exploring key European-level initiatives, we consider the national data policy landscapes of the seven ExPaNDS partner countries, noting key similarities and differences between these as well as to what extent they reflect higher-level European policy aspirations. We also review relevant background and context specific to data policy for photon and neutron research infrastructures and summarize the findings of the *ExPaNDS Data Policy Landscape Survey* (December 2019), which capture the present state of data policy adoption and implementation in ExPaNDS partner research infrastructures.

ExPaNDS was heavily involved in the development of the new *PaNOSC Data Policy Framework* (May 2020). Our aim in our deliverable here is to expand on that work, reflecting on what common elements a data policy for photon and neutron research infrastructure (RI) framework should address, especially in seeking to enable FAIR. We consider existing policy recommendations to develop a wider FAIR ecosystem in the European Open Science Cloud, in particular, the recommendations of the 2018 *Turning FAIR into reality* report and the recent *FAIRsFAIR Deliverable 3.3: Policy enhancement recommendations*. In response to the latter, we make twenty-six recommendations of our own. These serve to translate and relate the FAIRsFAIR recommendations to the local level of photon and neutron research infrastructures.

As the report's key output, we present thirty elements that should inform a data policy framework for photon and neutron research infrastructures. In particular, these elements focus on enabling FAIR data. In essence, they are those elements about which research infrastructures **need to make choices** — both on the level of commitments that they themselves are prepared to make as well as with regards to the obligations that they will place on users.

Elements of a Data Policy Framework

The thirty data policy framework elements are as follows:

1. A RI should openly publish a data policy, including the period in which the policy is in force and when it is planned to be reviewed. A PID should be used to refer to the published version of the policy.
2. The data policy should seek to clarify the ownership and access to data collected at a facility.
3. The data policy should specify the extent to which the facility will supply the user with support to access and analyse the data, within the experiment and beyond for subsequent analysis, to maximize the opportunity for the user to develop their science.
4. The data policy should specify the obligations on users in the subsequent use of the data.
5. The data policy should specify the actions the facility should undertake to ensure that experimental data is made available and reusable to maximize the scientific impact of the experiment.

6. The data policy should be sufficiently flexible to accommodate the data policies of RIs' national funders and users' funders.

7. A RI's data policy should comply with their national research funder's policy.

8. RIs should seek to align their data policies, within the constraints of divergent national funder policies and legal frameworks.

9. RIs should consider making their policies machine-readable.

10. RIs should seek to harmonize the terminology used, if possible, and use common data policy elements.

11. The RI's data policy should apply to data which are generated, stored, and analysed using the facility's resources (e.g. instruments, compute infrastructure, software, staff).

12. The RI's data policy should cover the classes of experimental data:

13. The RI's data policy should seek to maximize the scientific impact of experiments through enabling the validation of research results and maximizing opportunities for reuse.

14. The policy should specify the retention policy for each class of experimental data, with a minimum retention period and criteria for deletion. As this includes auxiliary data, this also includes software and tools.

15. In the event that data are deleted, the facility should retain a "digital footprint" of the data. This could constitute a (metadata) record of their essential characteristics or a method to allow the reconstruction of the data. This should allow as much as possible the validation of published research results.

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

17. The RI's data policy should specify a licence under which the data are made available.

18. The RIs should acknowledge the application of the relevant national legislation under the GDPR framework in the handling of personal and sensitive data.

19. RIs should specify the grounds for restricting access to data.

20. RIs should specify the time limit for which users are allowed exclusive control on the use of experimental outputs. This should also specify who can access the data, who can determine who should be given access rights, and the appeals process established to alter the embargo period.

21. The RI's data policy should include commitments to enabling FAIR data.

22. RIs should specify the rights and responsibilities of particular classes of actors involved in the experimental process.

23. The RI's policy should consider the extent to which it commits providing infrastructure to support the retention and distribution of FAIR data,

24. The policy should specify the requirements on users to participate in data management planning activities.

25. The RI's policy should specify whether the PI is responsible for preparing a DMP.

26. The RI's Policy should promote the recognition and citation of the use of facilities:

27. RIs should have regular audits on compliance to FAIR data.

28. Compliance to policy by users may be monitored and checked.

29. Users may be requested to report on compliance for previous experiments when applications for further access to the facility are received. Non-compliance may be a contributing factor in the refusal of further access.

30. Changes or termination to the data policy will be given in sufficient time for PIs to take alternative action to provide alternative provision to comply with their funders' data policies.

Table of Contents

1. Introduction

1.1 Overview of the Deliverable

In this document, ExPaNDS presents initial work to develop a new data policy framework for partner Research Infrastructures (RIs),[1] taking into account the current national and European research data policy landscape, the present state of data policy adoption within ExPaNDS facilities, and recommendations that are emerging in relation to the European Open Science Cloud (EOSC) on adapting data policy to make data FAIR (Findable, Accessible, Interoperable, Reusable).

This document represents the first deliverable on ExPaNDS work package (WP) 2 task 2.1:

> Relevant Research Infrastructures have a variety of data policies and practices, typically building on the PaNdata Common Policy Framework (2011), and on later activities in CALIPSOplus. This task will review current data policies and revise this framework within the policy recommendation of the EOSC and FAIR data principles. Further factors on data policy, for example, IPR and data licensing, commercial data, and sensitive data [e.g. General Data Protection Regulation (GDPR)] will also be considered. The task will work closely with PaNOSC, participating in a policy workshop and other consultation exercises.[2]

This document is the initial draft outcome, and gives the background and context of data policy within RIs, and a draft set of principles and recommendations to frame RIs data policy development.   These principles and recommendations will be used as a basis for further discussion with facilities senior staff, policy makers and practitioners to develop a final set of policy recommendations.   In this first draft, we place particular attention to the FAIR data principles; the other factors (e.g. restrictions and requirements on licensing and sensitive data) are touched upon, and will be considered further in the final version.

1.2 Document Content

Beyond this introduction (**section 1**), this document comprises six additional sections, which between them cover a range of topics of relevance to the development of a data policy framework for national photon and neutron RIs:

Section 2 overviews the European and national research data policy landscapes that are influencing and driving the development of policy in ExPaNDS RIs.

---

[1] The ExPaNDS partners are: Deutsches Elektronen-Synchrotron (DESY), Paul Scherrer Institute (PSI), Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Diamond Light Source (DLS), MAX IV, Elettra, ALBA, SOLEIL, Helmholtz-Zentrum Berlin (HZB), UK Research and Innovation (UKRI), Science and Technology Facilities Council (STFC), ISIS Neutron and Muon Source, and European Grid Infrastructure Foundation (EGI). ExPaNDS (2020). Partners. https://expands.eu/partners/

[2] ExPaNDS (2018). ExPaNDS European Open Science Cloud Photon and Neutron Data Services [proposal].

Section 3 presents the current landscape of data policies within the ExPaNDS partner RIs, including background on previous common approaches to data policy and changes in the facilities science context and also the results of a survey of partners considering the common approaches and features and noting any divergences.

Section 4 describes the development of the PaNOSC model data policy framework.

Section 5 considers the influences on data policy that arise from the requirements of policy to encourage and enable RIs to produce FAIR data. This leads to a series of recommendations on how RIs' data policies should be framed.

Section 6 draws on these recommendations and previous sections to present proposals for general principles and high-level recommendations for inclusion in data policies for ExPaNDS facilities. These proposals serve as an initial data policy framework.

Section 7 discusses the next stages of the work on developing a data policy framework in ExPaNDS.

2. Research Data Policy Landscape

2.1 European Level Influences

A number of factors influence the European research data policy landscape. While the national initiatives of Member States contribute much of the detail present in this landscape (see section 2.2), several wider, European level influences act to shape the overarching approach to the development of research data policy within Europe. Open Science policy is a major influencer as evident in EC research data policy guidance, the EOSC Rules of Participation (RoPs),[3] the Turning FAIR into Reality (TFiR) report,[4] and the two mandates of the Open Science Policy Platform (OSPP).[5] Taken together, these drivers play a significant role in shaping the research data policy landscape in Europe.

2.1.1 EC research data policy guidance

The EC seeks to make the outputs of the research it funds 'open' — that is, available online, long term, and at no extra cost to researchers, industry, and the public. The potential benefits of this approach are numerous: ready access to scientific information such as publications and datasets allows others to build on previous research, avoids unnecessary duplication, speeds up innovation, and improves transparency of the scientific process. In line with the principle of open

---

[3] EOSC RoP Executive Board Working Group (2020). European Open Science Cloud rules of participation version 0.2. https://repository.eoscsecretariat.eu/index.php/s/QWd7tZ7xSWJsesn#pdfviewer
[4] EC Expert Group on FAIR data (2018). Turning FAIR into reality. https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf
[5] EC (2020). Open Science Policy Platform. https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-policy-platform

access,[6] the ERC promotes the retention, documentation, and sharing of research data as best practice.

To encourage open access and reuse of research data, the ERC established the [Open Research Data Pilot (ORD Pilot)](7) in connection with the Horizon 2020 funding programme. A key expectation of the ORD Pilot is that data are made not just open but also FAIR. The value of open data increases significantly when they are FAIR; for example, humans and machines can quickly find, access and share FAIR data while the associated contextual and provenance-related information enables these data to be repurposed outside of the context in which they were originally generated. ERC projects participating in the ORD Pilot are also expected to employ Data Management Plans (DMPs), which set out details of how the research data produced by the projects will be collected, processed, stored, and made FAIR. The EC intends to draw on the outcomes of the ORD Pilot to further develop its policy on Open Science.

2.1.2 EOSC Rules of Participation

The EOSC plays a crucial role in enabling Open Science in Europe. In relation to data, the EOSC has three main objectives:

1. to increase the value of research data by making them more readily available
2. to reduce scientific data management costs
3. to ensure adequate protection of personal and sensitive information in line with European Union (EU) legislation such as GDPR.[8]

The [EOSC Rules of Participation](9) (RoPs) are designed to support these objectives. Currently in draft form, revised RoPs, intended to apply to the EOSC after 2020, are expected later this year.[10]

The RoPs are a minimal set of rules around rights, obligations and accountability that govern transactions by users, providers, and operators in the context of the EOSC. In turn, the RoPs reflect broader policy decisions on what is needed to constitute a minimal viable EOSC. The focus is on common requirements that can and will apply irrespective of Europe's very heterogeneous landscape of research infrastructures, services, and providers.[11] In relation to FAIR, two examples

---

[6] Note that the EC's definition of open access incorporates both publications **and** data: "Open access (OA) refers to the practice of providing online access to scientific information that is free of charge to the end-user and reusable… In the context of research and innovation, 'scientific information' can mean: 1. peer-reviewed scientific research articles (published in scholarly journals), or 2. research data (data underlying publications, curated data and/or raw data)." EC (2020). Horizon 2020 online manual: Open access. https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm

[7] ERC (2017). Guidelines on implementation of open access to scientific publications and research data. https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf

[8] EC (2018). Prompting an EOSC in practice. https://ec.europa.eu/info/sites/info/files/prompting_an_eosc_in_practice.pdf

[9] EOSC RoP Executive Board Working Group (2020). European Open Science Cloud rules of participation version 0.2. https://repository.eoscsecretariat.eu/index.php/s/QWd7tZ7xSWJsesn#pdfviewer

[10] EOSCsecretariat (2020). Rules of participation working group. https://www.eoscsecretariat.eu/working-groups/rules-participation-working-group

[11] Ibid.

of such requirements include the (second draft) [EOSC Persistent Identifier (PID) policy](#)[12] (e.g. to support findability) and the integration of a generic authentication and authorisation infrastructure (e.g. to support accessibility).

2.1.3 Turning FAIR into Reality report

The 2018 EC report, [Turning FAIR into Reality](#) (TFiR), produced by the EC FAIR Data Expert Group (Simon Hodson, Chair), "focuses on actions needed in terms of research culture and technology to ensure data, code and other research outputs are made FAIR".[13] In particular, the report examines what is needed to make data FAIR in the context of the EOSC. Although acknowledging that FAIR does not equate with open, TFiR is written from the perspective of the wider shift towards Open Science, with the recognition that, certainly where publicly-funded research is concerned, research data should be 'as open as possible and as closed as necessary'. TFiR makes twenty-seven recommendations in total: fifteen priority recommendations, considered the initial and essential steps towards implementing FAIR, and twelve supporting recommendations, designed to follow up on the priority recommendations with more detail and specifics. A set of actions aimed at specific stakeholder groups accompanies each recommendation.

Three priority and eight supporting recommendations have associated actions aimed specifically at policy makers. Relevant 'priority' actions for data policy include: clearly defining and articulating the difference between FAIR and open, balancing openness and protection appropriately while accommodating both, ensuring policies and DMPs are machine-readable and actionable, and employing common standards (e.g. metadata, semantics).[14] Indexing policies in a policy registry, tailoring policy for specific contexts, the use of embargo periods, and the requirement to deposit data in certified repositories are just a few examples of the TFiR 'supporting' actions relevant to data policy.

2.1.4 Open Science Policy Platform

The Open Science Policy Platform (OSPP) is a group that was tasked with advising the EC on how to develop Open Science policy. Established by the EC Directorate-General for Research and Innovation, the OSPP recently completed its four-year mandate, which ran from 2016 to 2020. The group comprised a range of nominated organisations and institutions associated with or representing scholarly research in Europe.

---

[12] EOSC FAIR Working Group and EOSC Architecture Working Group (2020). Second draft persistent identifier (PID) policy for the European Open Science Cloud (EOSC). https://doi.org/10.5281/zenodo.3780423
[13] EC Expert Group on FAIR data (2018). Turning FAIR into reality.
https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf
[14] Note that the phrase 'common standards' is used here in a general sense, i.e. to refer not only to common metadata standards, but also to common semantics, which may be supported by vocabularies, taxonomies, and ontologies.

The OSPP's first mandate (May 2016 – May 2018) gave rise to a set of [Recommendations for Open Science](#).[15] These target key stakeholder groups represented by the OSPP and focus on publicly-funded research. The set encompasses five 'general principle' recommendations and a number of 'prioritized' recommendations. The latter relate to eight priority focus areas drawn from the [European Open Science Agenda](#).[16] FAIR data is one priority; examples of others include the EOSC, research integrity, next-generation metrics, skills, and citizen science.

The OSPP's second mandate (May 2018 – May 2020) built on the first mandate recommendations to produce a [report](#),[17] which sets out Practical Commitments for Implementation (PCIs). PCIs aim to address the cultural changes needed for Open Science: "A PCI is a realistic and affordable action that a stakeholder (or a representative) has the will and jurisdiction to implement in relation to a particular aspect or recommendation in Open Science."[18] As with the recommendations, the PCIs align with the eight European Open Science priority focus areas. Several of the PCIs, although aimed at the national and international levels of European Member States, funders, or policy making organisations, are also relevant to the development of community policy (e.g. such as that of the Photon and Neutron (PaN) community) and local research data policy. Examples include needs for policy alignment, an agreed framework for data management, and evidence-based policy development that incorporates monitoring of policy outcomes in relation to policy intentions.

2.2 National Initiatives and Motivations

Although driven heavily by European-level policy making, national research data policy varies across Europe,[19] depending on country-level motivations and needs. ExPaNDS partners include RIs from seven countries: France, Germany, Italy, Spain, Sweden, Switzerland, and the United Kingdom. The sections below overview the current state of research data policy and policy making in each of these. In particular, the summaries focus on capturing the special features of each country, including the key players involved in shaping research data policy and the formal legislation, initiatives, and policies that drive national recommendations and actions around research data.

---

[15] EC (2017). Open Science Policy Platform recommendations. https://ec.europa.eu/research/openscience/pdf/integrated_advice_opspp_recommendations.pdf

[16] EU Presidency, hosted by the Government of the Netherlands (2016). Amsterdam call for action on Open Science. file:///C:/Users/zrs94363/Downloads/amsterdam-call-for-action-on-open-science.pdf

[17] EC (2020). Progress on Open Science: Towards a shared research knowledge system. Final report of the Open Science Policy Platform. https://ec.europa.eu/research/openscience/pdf/ec_rtd_ospp-final-report.pdf#view=fit&pagemode=none

[18] Ibid.

[19] SPARC Europe, in collaboration with the Digital Curation Centre (2017). An analysis of open data and Open Science policies in Europe, May 2017. https://sparceurope.org/new-sparc-europe-report-analyses-open-data-open-science-policies-europe/

2.2.1 France

Following the 2016 [French Law for a Digital Republic](#)[20] (in French), which encourages public establishments to make their data open and reusable, France published its [National Plan for Open Science](#)[21] in July 2018. One of the ambitions of this national plan is "to ensure that data produced by government-funded research in France are gradually structured to comply with the FAIR Data Principles … and that they are preserved and, whenever possible, open to all." As part of this plan, the Committee for Open Science ([CoSO](#)) was created: among its actions, the CoSO promotes [Recommendations for Services in a FAIR Data Ecosystem](#)[22] on how existing data infrastructures can evolve and collaborate to provide services that support the implementation of the FAIR data principles, particularly in the context of building the EOSC.

The French National Research Agency (ANR) funds project-based research carried out by public operators cooperating with each other or with private companies. The ANR has a strong [Open Science policy](#),[23] which is fully aligned with the French National Plan for Open Science, and promotes open access to publications, contributes to open research data wherever possible, and develops a coordinated approach at national, European, and international levels. From 2019, the ANR requires any funded project to produce a DMP summarizing what datasets will be produced, how they will evolve, how and where they will be preserved, and how they will be FAIR. The ANR launched a [FLASH call](#)[24] in 2019 "to accelerate the structuring of the French scientific community to promote the FAIR Data Principles and open up data". Twenty-five projects were selected, relating to the certification of data repositories as well as to the implementation of new tools, standards, or the deployment of networks of actors.

Following the publication of the National Plan for Open Science, the French National Centre for Scientific Research (CNRS), the main research institution in France, adopted a strong [Roadmap for Open Science](#)[25] in November 2019. Data produced by CNRS researchers or deriving from CNRS resources must, as far as possible, be made accessible and reusable, according to the FAIR principles. The CNRS announced an infrastructure charter to broaden the application of FAIR principles to all disciplines and as a commitment for infrastructures to adopt FAIR practices and quality standards and to release data policies jointly-drafted with the scientific communities using them.

---

[20] République Française (2016). LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique. https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000033202746&categorieLien=id

[21] Ministère de l'Enseignement Supérieur de la Recherche et de l'Innovation (2018). National Plan for Open Science. https://cache.media.enseignementsup-recherche.gouv.fr/file/Recherche/50/1/SO_A4_2018_EN_01_leger_982501.pdf

[22] Committee for Open Science (2019). Recommendations for services in a FAIR data ecosystem. https://www.ouvrirlascience.fr/recommendations-for-services-in-a-fair-data-ecosystem-2/

[23] ANR (2013). Open Science policy. https://anr.fr/en/anrs-role-in-research/values-and-commitments/open-science/

[24] ANR (2019). FLASH CALL Open Science: Research practices and open research data. https://anr.fr/en/call-for-proposals-details/call/flash-call-open-science-research-practices-and-open-research-data/

[25] CNRS (2019). CNRS roadmap for Open Science. https://www.science-ouverte.cnrs.fr/wp-content/uploads/2019/11/CNRS_Roadmap_Open_Science_18nov2019.pdf

Accessible by any member of the Higher Education and research community as well as its French or foreign partners, the OPIDoR DMP tool[26] (in French) provided by the CNRS Institute for Scientific and Technical Information (INIST) guides researchers through the drafting and implementation into practice of data or software management plans. It integrates all known funder DMP models: H2020, ANR, public funders such as the French National Research Institute for Agriculture, Food, and Environment (INRAE) and private ones such as Institut Pasteur, a non-profit foundation, whose mission is to help prevent and treat diseases, mainly those of infectious origin.

The French state is also encouraging RIs to adopt data policies. SOLEIL adopted its own Data Policy[27] in October 2018 based upon the PaNData policy framework[28] and its application at other facilities.

2.2.2 Germany

Research freedom is a fundamental right protected by the German Constitution. As a result, there exists little legislation in Germany that regulates norms internal to science. Policies are predominantly set out by scientific organizations and funding bodies, rather than by the state.

The Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) is the central, independent research funding organization in Germany. In response to a case of scientific misconduct, in 1997, the DFG formulated the Proposals for Safeguarding Good Scientific Practice.[29] These recommendations are concerned mainly with the assurance of integrity in the sciences. Recommendation 7 states that "primary data as the basis for publications shall be securely stored for ten years in a durable form in the institution of their origin."

The concept of open access to scientific results created through publicly funded research has been widely discussed in the scientific community. In this context, the Alliance of German Science Organisations formulated the Principles for the Handling of Research Data,[30] through which it "supports the long-term preservation of, and the principle of open access to, data from publicly funded research." This is substantiated by the DFG in the Leitlinien zum Umgang mit Forschungsdaten[31] (in German) in the context of their funding regulations.

---

[26] CNRS Institute for Scientific and Technical Information (n.d.). DMP OPIDoR. https://dmp.opidor.fr/
[27] SOLEIL (2018). SOLEIL data management policy. https://www.synchrotron-soleil.fr/en/files/soleil-data-management-policy
[28] PaN-data Europe (2011). Common policy framework on scientific data. https://doi.org/10.5281/zenodo.3738497
[29] Deutsche Forschungsgemeinschaft (1998). Proposals for safeguarding good scientific practice. https://web.archive.org/web/20110519131326/http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf
[30] Alliance of German Science Organisations (2010). Principles for the handling of research data. https://doi.org/10.2312/ALLIANZOA.035
[31] Deutsche Forschungsgemeinschaft (2015). Leitlinien zum Umgang mit Forschungsdaten. http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf

In 2019, the DFG revised its recommendations and replaced them with the [Guidelines for Safeguarding Good Research Practice](#).[32] This code of conduct requests researchers to document all information relevant to the production of research results, to provide public access to research results as a rule, and to archive research data and results that have been made publicly available. The publication of research data should be made in accordance with the FAIR Data Principles. Higher education and research institutions are requested to implement these guidelines in a legally binding manner in order to be eligible for funding by DFG.

2.2.3 Italy

As a part of European data landscape, Italy supports most European initiatives in the area of data protection and stewardship. European policy strives to implement the FAIR data model and to distribute openly accessible data without restrictions. In relation to the protection of personal data, GDPR implements strict rules of anonymity designed to protect human-related research data and results. The Agenzia per l'Italia Digitale oversees data-related activities at the highest level from a legal point of view.[33] Governmental support is focused mostly on indirect funding and auditing, rather than on direct curation of scientific data. Auditing and direct funding institutions such as Fondazione Bruno Kessler and PONgovernance support research according to their organizational policies and strategic plans,[34,35] not data policies.

The Italian Computing and Data Infrastructure (ICDI) works on PaN activities as part of the GARR Consortium.[36,37] In fact, ICDI yields the regulation of data stewardship to the RIs. For datasets and direct data stewardship, the Consiglio Nazionale delle Ricerche provides sample dataset definitions and ontologies.[38] At a local level, the national RIs prepare and implement their own data policies and maintain compliance to the generalized European norms. On the one hand, this is an advantage because it eases the self-regulating activities and enables data policies to be written "by scientists, for scientists". On the other hand, however, it makes the data policy landscape at the RI level very complex. The main complexity problems are differences in terminology and bunches of legacy regulations. Harmonization of this landscape is the prime task for Italian data governance initiatives. However, the absence of more prescriptive regulation eases the implementation of European policies and the FAIR principles.

---

[32] Deutsche Forschungsgemeinschaft (2019). Guidelines for safeguarding good research practice. Code of conduct. https://doi.org/10.5281/zenodo.3923602

[33] Agenzia per l'Italia Digitale (n.d.). About us. https://www.agid.gov.it/en/agency/about-us

[34] Fondazione Bruno Kessler (2018). Piano Strategico. https://www.fbk.eu/wp-content/uploads/2018/12/Piano-Strategico-FBK_2018.11.16.pdf

[35] PON Governance (2020). PON Governance e Capacità Istituzionale 2014–2020 Home. http://www.pongovernance1420.gov.it/en/

[36] Italian Computing and Data Infrastructure (n.d.). ICID. https://www.icdi.it/en/

[37] Consortium GARR (n.d.). Chi siamo. https://www.garr.it/it/chi-siamo

[38] Consiglio Nazionale delle Ricerche (2010). About. http://data.cnr.it/site/about

2.2.4 Spain

In June 2011, the Spanish Government issued the first [Science, Technology and Innovation (STI) Act 14/2011](#)[39] informing about the creation of the Scientific, Technological and Innovation Policy Council, whose primary purpose is to work together with the Ministry of Science and Innovation to implement the 2017–2020 [State Plan for Scientific and Technical Research and Innovation](#)[40] (in Spanish). Before the creation of the Council, any strategy referred to Scientific and Technical Research was included in the National Research and Development (R&D) Plans (see [Plans I+D+i Anteriores,](#)[41] in Spanish).

[Article 37 "Open Access dissemination"](#)[42] (in Spanish) of the Act (14/2011) mandates that the public workers of the scientific, technical and innovation system shall drive forward the creation of their own or shared, new open access repositories for the publications produced by their researchers. They will also enable new systems to link these repositories with other, similar national and international initiatives. Results from publicly funded activities and accepted for publication in research journals will be made public, in a digital version, in an open access repository and may be used by public administrations in their assessment procedures.

The State Plan for Scientific and Technical Research and Innovation envisages open access to data and metadata, aligning with the Act (14/2011) and the recommendations of the EC regarding open data and Open Science. To promote research data access, funded R&D projects are able to attach DMPs related to the data produced at scientific institutions and stored in their repositories. The data are made public once their release date is reached. Other aspects, such as data protection or confidentiality, are carefully respected during the process of making the data open.

The State Plan sets out a series of objectives for achieving the main goal of open access to data:

- Reinforce the scientific and technological capabilities of the Spanish scientific, technical, and innovation system by promoting the development of scientific careers in both the public and private sectors.
- Through funding, enhance the Spanish leadership role in the scientific international landscape.
- Activate private funding investment in Research, Development, and Innovation (RD&I) and in the technological capability of the Spanish production system.

[39] Gobierno de España (2011). Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación. https://www.boe.es/eli/es/l/2011/06/01/14/con
[40] Gobierno de España (2017). Plan Estatal de Investigación Científica y Técnica y de Innovación 2017-2020. https://www.ciencia.gob.es/portal/site/MICINN/menuitem.7eeac5cd345b4f34f09dfd1001432ea0/?vgnextoid=83b192b9036c2210VgnVCM1000001d04140aRCRD
[41] Gobierno de España (various). Planes Nacionales I+D+i Anteriores. https://www.ciencia.gob.es/portal/site/MICINN/menuitem.29451c2ac1391f1febebed1001432ea0/?vgnextoid=79dbec05f2a7d210VgnVCM1000001d04140aRCRD
[42] Gobierno de España (2011). Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación. Artículo 37. Difusión en acceso abierto. https://www.boe.es/eli/es/l/2011/06/01/14/con#a37

- Promote the potential and impact of research and innovation for the benefit of societal challenges.
- Promote an open and responsible RD&I model supported by the participation of society by promoting the adoption of the principles of open access to (FAIR) research data by scientific communities and research institutions.
- Enhance coordination, synergies, and efficient implementation of RD&I funding policies at regional, national, and European levels.[43]

2.2.5 Sweden

Since 2017, the Swedish Research Council has worked in consultation with the National Library of Sweden, the Swedish National Archives, and higher education institutions to coordinate and increase open access to research data. The goal is to fully implement the transition to open access to research data by 2026. In 2018, the Swedish Research Council published a government-commissioned report[44] (in Swedish) that sets out criteria for assessing how data resulting from publicly-funded research meet the FAIR principles. Research funders such as the Swedish Research Council, Riksbankens Jubileumsfond, and Formas have introduced requirements for open data and DMPs.[45,46,47,48] RIs, including MAX IV, having established research data policies at the local level, are working intensively on setting up good research data management practices and support.

Sweden engages with a range of activities related to the EOSC and participates in EOSC-Nordic (one of the '5b projects', of which ExPaNDS is another). Swedish RIs also collaborate internationally on several research data projects (e.g. MAX IV with ExPaNDS, the National Bioinformatics Infrastructure Sweden with ELIXIR, and the Integrated Carbon Observation System with ENVRI-FAIR). As part of Sweden's aim to open access to research data by 2026, the Swedish Research Council hopes to release a national tool for DMPs in 2020. A national project focused on a high-capacity data storage solution is also ongoing, aimed at meeting an urgent need for many universities in Sweden. Future planned investments include training in research data management to raise the competency of researchers and research support staff.

---

[43] Gobierno de España (2017). Plan Estatal de Investigación Científica y Técnica y de Innovación 2017-2020. https://www.ciencia.gob.es/portal/site/MICINN/menuitem.7eeac5cd345b4f34f09dfd1001432ea0/?vgnextoid=83b192b9036c2210VgnVCM1000001d04140aRCRD

[44] Vetenskapsrådet (2018). Kriterier för FAIR forskningsdata. https://s3.amazonaws.com/helpscout.net/docs/assets/5673e406c6979143615582d5/attachments/5c5adc74042863543ccca279/Kriterier-FAIR-forskningsdata_VR_2018.pdf

[45] Swedish Research Council (2019). Producing a data management plan. https://www.vr.se/english/applying-for-funding/requirements-terms-and-conditions/producing-a-data-management-plan.html

[46] Riksbankens Jubileumsfond (2019). Guidelines for open access. https://www.rj.se/en/About-RJ/Work-methods/guidelines-for-open-access/

[47] Riksbankens Jubileumsfond (2019). Data management plan for RJ. https://www.rj.se/en/About-RJ/Work-methods/guidelines-for-open-access/data-management-plan-for-rj/

[48] FORMAS (2020). Open access to research results and data. https://www.rj.se/en/About-RJ/Work-methods/guidelines-for-open-access/data-management-plan-for-rj/

2.2.6 Switzerland

The Swiss National Science Foundation (SNSF) and the umbrella organization of Swiss universities (swissuniversities) dominate Swiss funded research. They decided to make sharing data from funded projects mandatory in 2016. Since October 2017, most funding schemes have required researchers to include DMPs in funding applications. The SNSF also expects that data generated by funded projects are publicly accessible in digital databases, provided there are no legal, ethical, copyright, or other issues. More broadly, the SNSF has an internationally-aligned perspective on goals and aspirations for open research data.

A joint SNSF/swissuniversities-commissioned study[49] conducted in 2018 concluded that motivations for and concerns around data sharing and data reuse are similar for the Swiss community and other scientific communities. Sharing and reuse behaviour differs between research disciplines. In particular, while they share equally in general repositories and smaller disciplinary repositories (which exist in greater numbers), researchers prefer to use disciplinary repositories if they want to reuse data. Overall, about a third of the Swiss research community shares data in repositories. The main reason for not sharing was researchers' plans to publish their results first. Also, many participants claimed to have a different concept of 'data'; while the study tried to define terms carefully, apparently there is a need for more discipline-specific information and discussion of the topic.

The SNSF policy on Open Research Data outlines:

> The SNSF values research data sharing as a fundamental contribution to the impact, transparency and reproducibility of scientific research. In addition to being carefully curated and stored, the SNSF believes research data should be shared as openly as possible.
>
> The SNSF therefore expects all its funded researchers
>
> - to store the research data they have worked on and produced during the course of their research work,
> - to share these data with other researchers, unless they are bound by legal, ethical, copyright, confidentiality or other clauses, and
> - to deposit their data and metadata onto existing public repositories in formats that anyone can find, access and reuse without restriction.

---

[49] von der Heyde, M. (2019). Open research data: Landscape and cost analysis of data repositories currently used by the Swiss research community, and requirements for the future [Report to the SNSF]. https://doi.org/10.5281/zenodo.2643460

> Research data is collected, observed or generated factual material that is commonly accepted in the scientific community as necessary to document and validate research findings.[50]

The SNSF also provides guidelines on FAIR data principles and data repositories to facilitate an open research data landscape.[51]

2.2.7 United Kingdom

The UK government considers data and data-related services as both vital to existing UK infrastructure and a priority for future development. The UK's stance on research data is clear: "data produced by publicly-funded researchers should be regarded as a public good".[52] Alongside the aim of open research data as the norm sits the additional aspiration that data acquired through the research process should be made FAIR by default. There is still some way to go with the latter: while the data policies of major UK funders embrace the concept of open data, FAIR currently receives considerably less policy attention. As well, although UK research data policies cover similar ground, for example, metadata, preservation and access, and DMPs, a notable lack of harmonization persists, with major differences evident in terminology, scope, and expectations (i.e. even after accounting for varying disciplinary norms). Despite their policy differences, however, there is a general consensus amongst UK research funders that data policy implementation – let alone, compliance – remains a considerable and pressing challenge.[53,54] More broadly, though, an "examination of a range of policies from research funders and related agencies overseas does not suggest that in general UK policies lag much behind, if at all, in terms of the policy requirements laid on researchers."[55]

The UK data policy landscape is complex, shaped by national research councils, research charities, government departments, research institutions, and universities. As the national research funding body, UK Research and Innovation (UKRI) has a role in encouraging the harmonization of research data policy in the UK. While their uptake in university data policy remains mixed, the principles set out in the Concordat on Open Research Data[56] (2016) and the

---

[50] Swiss National Science Foundation (n.d.). SNSF policy on open research data. http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx#SNSF%20policy%20on%20Open%20Research%20Data

[51] Swiss National Science Foundation (n.d.). Open research data. http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx

[52] Open Research Data Taskforce with Michael Jubb (2017). Research data infrastructures in the UK landscape report. https://www.universitiesuk.ac.uk/policy-and-analysis/research-policy/open-science/Documents/ORDTF%20report%20nr%201%20final%2030%2006%202017.pdf

[53] Open Research Data Taskforce (2018). Realising the potential: Final report of the Open Research Data Taskforce. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/775006/Realising-the-potential-ORDTF-July-2018.pdf

[54] Open Research Data Taskforce with Michael Jubb (2017). Research data infrastructures in the UK landscape report. https://www.universitiesuk.ac.uk/policy-and-analysis/research-policy/open-science/Documents/ORDTF%20report%20nr%201%20final%2030%2006%202017.pdf

[55] Ibid.

[56] Higher Education Funding Council for England, Research Councils UK, Universities UK, and Wellcome (2016). Concordat on open research data. https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf/

Common Principles on Data Policy[57] (first published in 2011, with guidance notes[58] updated in 2015 and 2018) are reflected in the policies of the individual research councils as well as in the data policies of major charitable research funders such as the Wellcome Trust, Cancer Research UK, and the Royal Society. In relation to PaN research data, the research data policies and related guidance of the Engineering and Physical Sciences Research Council (EPSRC), Biotechnology and Biosciences Research Council (BBSRC), Science and Technology Facilities Council (STFC) and the Wellcome Trust — all major funders of research undertaken at Diamond Light Source and ISIS, the UK's PaN RIs — set out a clear expectation that research data will be made openly available in a timely manner. [59,60,61,62,63]

While none of the UK's major national and charitable research policies yet stipulate specifically that data must be made FAIR, some policies do echo aspects of the FAIR principles. Examples include the Wellcome Trust's policy to ensure that, regarding research data, "other research can verify it, build on it and use it to advance knowledge",[64] the Royal Society's attention to "intelligent openness",[65] and the Natural and Environmental Research Council's (NERC) guidance on the inclusion of provenance metadata "to allow others to effectively re-use … data".[66] As FAIR continues to gain traction in the European and international research data landscape, the UK's participation in related projects (such as ExPaNDS and FAIRsFAIR) should help to raise awareness and increasingly encourage UK research data policy to incorporate FAIR principles.

2.3 Reflections on European and National Policy Making

Several similarities emerge from the research data policy landscapes of ExPaNDS partner countries. These commonalities at the national level reflect back the key themes found in European level policy making. This said, some notable differences exist at the national level, illustrating how policy can differ when implemented in the context of national planning and needs. Supported by increased awareness of the cultural and technological changes required, the rapid transition to open and FAIR data looks set to continue in Europe.

---

[57] UKRI (2020). Common principles on data policy. https://www.ukri.org/funding/information-for-award-holders/data-policy/common-principles-on-data-policy/

[58] UKRI (2018). Guidance on best practice in the management of research data. https://www.ukri.org/files/legacy/documents/rcukcommonprinciplesondatapolicy-pdf/

[59] EPSRC (2011). EPSRC policy framework on research data. https://epsrc.ukri.org/about/standards/researchdata/

[60] BBSRC (2017). BBSRC data sharing policy. https://bbsrc.ukri.org/documents/data-sharing-policy-pdf/

[61] BBSRC (2016). Safeguarding good scientific practice. https://bbsrc.ukri.org/documents/safeguarding-good-scientific-practice/

[62] STFC (2016). Scientific data policy. https://stfc.ukri.org/about-us/our-purpose-and-priorities/requesting-information-from-uk-research-and-innovation/scientific-data-policy/

[63] Wellcome (2017). Data, software and materials management and sharing policy. https://wellcome.ac.uk/grant-funding/guidance/data-software-materials-management-and-sharing-policy

[64] Ibid.

[65] The Royal Society (2012). Science as an open enterprise: Open data for Open Science. https://royalsociety.org/~/media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

[66] NERC (2019). NERC data policy – guidance notes. https://nerc.ukri.org/research/sites/data/policy/datapolicy-guidance/

2.3.1 Key themes and similarities

Across the research data policy landscapes of ExPaNDS partner countries, several similar themes emerge. All ExPaNDS partner countries have formal laws, acts, or plans aimed at making data, especially those which derive from publicly-funded research, open, reusable, and aligned with the FAIR principles. Mandated DMPs feature heavily in national research data landscapes, with many countries also developing DMP tools or similar support.

While responsibility for research data policy differs in terms of whether it is centrally or locally delegated, some degree of harmonization is generally sought (even if not yet achieved) amongst the key players such as government, funders, scientific organisations, universities, and RIs who operate within each national research landscape. At the country-level, there is also considerable evidence of concerted efforts to address the changes to research culture needed for the transition to FAIR and open data. In practice, this effort often takes the form of training and support around data management for research communities and researchers themselves. Some technical projects also aim to deliver related services such as data storage and repositories.

In all cases, these themes and implementations reflect back European-level policy initiatives and ambitions such as those set out in TFiR, the EOSC RoPs, and the OSPP recommendations and PICs. Furthermore, in some cases, it can even be argued that national implementations around FAIR and open data are further ahead than European-level initiatives. The EC Horizon 2020 ODR Pilot offers a case in point: many national stances on open data have moved well beyond what could be deemed a pilot stage.

2.3.2 Differences

Despite the similarities highlighted above, some notable differences can be found between ExPaNDS partner national research data policy landscapes. A key difference is the role of the state and how devolved or not policy making is from the centre. For example, in both Germany and Italy, there is little regulation by the state, and not even necessarily by a central funding body. Instead, responsibility for policy making and regulation sits at the local level. This difference in the delegation of responsibility affects who has the key roles for data policy making in a country. It also impacts policy harmonization efforts, sometimes even where strong national guidance is in place. The UK provides an example of this situation; although national research data policy is supported through UKRI (i.e. a central funder), considerable variety persists across the data policies of other funders and universities.

Policy attention to FAIR also differs from country to country, especially at the local level. In Germany, universities and research institutions must implement guidelines, which include the FAIR principles, in a legally binding manner to be eligible for funding, while many of the other countries stop short of legal measures, choosing instead to 'encourage' or 'require'. And, in the case of the latter, the more strongly-worded 'require', it is notable that considerable gaps remain around monitoring and compliance. As noted above, all ExPaNDS countries are working in some way to promote the research culture needed for FAIR and open data. This said, efforts do differ

from country to country, with some being more well-defined and centrally organized and supported than others.

2.3.3 Future direction

One point that stands out across the national data landscapes is just how much policy has moved towards open data and the FAIR principles specifically within the last five years. This represents a rapid pace of change (or at least, the ambition to change) that clearly aligns with similarly rapid European level developments around policy making in this area.

In future, given the direction of research data policy in Europe, all ExPaNDS partner countries are likely to move increasingly to make data FAIR as well as open. To this end, use of repositories will grow as will the recognition in practice that research culture and data management training both play significant roles. There are also signs that, once the EOSC is better established, it, too, will feature increasingly in national policy landscapes. Given the speed with which research data policy around FAIR and open has developed at both the European and national levels, the key question for the future is how well and how quickly this policy can be implemented successfully in practice across Europe.

3. Current Data Policy Landscape in National RIs

3.1 Background and Context

The landscape in which PaN RIs operate has changed significantly over the last decade. Experiments at RIs now generate much larger and more complex data, and these data need to be managed effectively and efficiently. In this regard, data policies are a key tool in providing guidance to both users and facilities on research data management. A common data policy framework supports RIs in building their local data policies, while at the same time encourages harmonisation across RIs around principles such as FAIR and legislation such as GDPR that have emerged in recent years as new features of the research data policy landscape.

3.1.1 A changed landscape for RIs

Over the last decade, the working landscape of PaN facilities has changed completely. RIs have transformed from being experiment-focused facilities to being data producers with high data flow, which varies widely between facilities' instruments and techniques (from, for example, several megabytes per scan in ptychography to dozens of gigabytes per acquisition in hi-resolution crystallography). The experiments themselves have also changed. From operations supporting experiments which produced relatively limited data, requiring specialized processing tools and human interpretation by the experimenter, the experiments have become serial test cases producing very large amounts of data. In general, this change is closely related to new detectors arising from the development of the special vision techniques reconstructing the images and structural models from redundant data recorded in special settings of the experiment, including multi-parametric recording in controllable lighting conditions. On the one hand, such test cases

require extremely high repeatability of the settings, and, on the other, produce a lot of redundant and high-density context-dependent data. Data curation becomes very important as the experimental context must be provided to make any further analysis reproducible. In the case of data, if integrity is kept, the analysis could be reproduced on any other infrastructure compatible with the existing software. As a reaction to changes in scientific practices, separated data storage, integrity control, and stewardship facilities are required within existing RIs. All the above has led to the need for refreshing of RIs data management policies.

3.1.2 Role of data policy frameworks for PaN RIs

A data policy framework provides guidance for RIs on how they should build their own data policies to comply with community aims and known restrictions, and to ensure data quality. From this point of view, the framework should be considered as a simplified skeletal model of future policy: without a determination of the constraints and prescriptions, it gives propositions about how the end beneficiary would obtain the benefits (for example, immediate reusability for FAIR-compliant metadata) from the use of the facility. Limiting the set of possible implementations, the policy framework also indicates key points for harmonization of relations between RIs. In addition, the hidden task of the policy framework is to set limits that allow avoiding misuse cases for data.

A data policy is used as an instrument to guide facilities staff and users on the rights and obligations of facilities and their user communities with respect to the management and use of data arising from the RI. In the past, the data policies of most RIs were dedicated to the regulation of relationships between the scientists working for the RI and external investigators that guide the experiments and handle the experimental data. In fact, implementing the experiments, RIs should also comply with national regulations, especially in cases when personal data are in the processing pipeline.

Funders, publishers and other beneficiaries may completely define the data policy landscape of the given RI. In fact, when most of the existing policies were being developed, there were no frameworks or regulations to guide the development. This made the data policy landscape complex but enlightened the existing general prescriptions which had arrived from the practice. For example, almost all member RIs have an embargo period for access to the produced data in their policies. Also, the necessity to categorize the data by derivation criterion was elaborated. However, obligations placed on RIs by funding, publishing agreements, and management produced many aspects in which the data stewardship regulations yielded to the external acts and a lot of differences in parts dedicated to the relationships between data producers and users. Now the ExPaNDS project is striving to provide the framework that can inform the partner RIs implementing the data policies how to harmonize funder prescriptions, legal regulations, interpersonal relationships, and working practices from the legalistic point of view. Special attention should be paid here to the drive from science policy makers towards the adoption of open science, in particular the FAIR data principles, and also evolving international legislation, for example, GDPR. The intercontinental legal issues also should be under consideration of the project because of huge differences in areas of patent and publicly accessible scientific data between EU and US, for example, the norms dedicated to localization of sensitive data.

Development of open-access software and data produced new challenges for the PaN community. Data have become redistributable resources that could be used in a reproducible

way. However, the amount of the data produced today is so large that even the data already declared openly accessible cannot be treated and exposed to the users from a single RI. The huge amount of existing data requires establishing stewardship initiatives across the countries and institutions. The EOSC initiative gathers scientists to develop almost all aspects of open data distribution, and forms a driver to reconsider data policy at this moment.

3.1.3 The development of data policy frameworks in PaNOSC and ExPaNDS

The stages of data policy development can be tracked for decades within Photon and Neutron RIs. In particular, the previous projects, PaN-data Europe[67] and PaNdata Open Data Infrastructure,[68] developed an influential data policy framework document, the PaNData Policy Framework,[69] thus encapsulating the early recognition that PaN RIs will need to alter their practices. This has been the basis of the adoption of data policy by facilities over the last ten years, and forms the basis of the further refinement of data policy currently being undertaken within ExPaNDS and its sister project, PaNOSC.[70]

There is a difference in focus in the work of PaNOSC and ExPaNDS. The PaNOSC data policy framework and data management guidelines focus on the data ownership and practical aspects of curation more than on the governance landscape and stewardship (including handling of law issues in corresponding areas). The PaNData D2.1: Policy Framework and its extended derivation PaNOSC Deliverable 2.1 - PaNOSC data policy framework close the extremely important part of the development giving general directions and initial definition of the problems which ExPaNDS expands on. The PaNData policy is an attempt to gather the facilities, not the data, but the PaNOSC policy describes the importance and applicability of FAIR and metadata curation. These frameworks defined the data cataloguing principles and basics of metadata stewardship with the essential principle of authentication and authorization for the users who place the request to access the data. The data catalogue is indicated in PaNOSC documentation as the entity that is planned to be built. Thus PaNOSC does not regulate the stewardship principles on the gathering and analysis levels but it does set the principles. The experiment proposal is declared as the basis of cataloguing, so contractual obligations and overall access rights regulation are yielded to the relationship between RI and Principal Investigator (PI), especially in case of a proprietary experiment, and the PaNOSC Policy Framework indicated the difference between publicly funded experiments producing openly accessible data and proprietary research from a legalistic point of view..

ExPaNDS, aims to make a common self-regulated environment for different aspects of work done by PaN scientists, including activity governance, data stewardship, publishing, and training. The ExPaNDS Data Landscaping Survey describes the state of data stewardship evolution by the end of 2019. From a very general point of view, the main emphasis is made inside ExPaNDS on,

---

[67] PaN-data-Europe project (2010-2011) co-funded by the partners and the European Commission under the 7th Framework programme Grant Agreement RI-261537

[68] PaNdata Open Data Infrastructure project (2011-2014) co-funded by the partners and the European Commission under the 7th Framework programme Grant Agreement RI-283556

[69] PaN-data Europe (2011). Common policy framework on scientific data. https://doi.org/10.5281/zenodo.3738497

[70] Photon and Neutron Open Science Cloud (PaNOSC) project (2018-2022) funded by the European Commission under the H2020-EU.1.4.1.1. programme Grant Agreement 823852 www.panosc.eu

firstly, open access to the metadata, and, secondly, to the data. Open access policies are the natural extension of the data policy evolution because public and international funders are interested in keeping the data flow transparent. The emphasis in newly developed data policies is given to reproducibility of the data and experimental contexts published by the partner RIs of ExPaNDS, and thus the FAIR data model has become central for all PaN data workflows. Thus the aim of this task is to collaborate and converge with PaNOSC, refining these principles with the philosophy of openly accessible data and cataloguing possibilities developed through work of PaNOSC. Ideally the data policy frameworks should be mutually ratified.

3.1.4 Challenges for data policy frameworks

New challenges have arisen with the evolution of national and international data stewardship legislation, especially in relation to the protection of personal data. But in fact, most countries do not have a standardized set of procedures to prove human-related anonymity for the given data. Also, data transfer processes are now the subject of regulations. Because of high reproducibility and the huge number of details needed to support experimental context, copying is inseparable from integrity verification. Storage formats with integrated structural naming like NEXUS/HDF and hash table-based versioning systems made the RIs able to create data repositories allowing end-users to obtain a proof of integrity from any downloaded piece of data.

The endorsement of the FAIR principles for data is also a major challenge. For the implementation of the FAIR principles in practice, the most challenging task is to handle derived data obtained from the existing ones. The first step to deal with this task was made when the term 'raw data' was defined within the PaNOSC Data Policy Framework; however, even developing guidelines on how to maintain reproducibility for data analysis and derivation is a non-trivial task, requiring a huge amount of work. The project ExPaNDS strives to implement the glossary, definitions and information interchange guidelines for member RIs, taking advice from FAIRsFAIR[71] as one of the most important initiatives aiming to implement openly accessible data into practice for Europe in the EOSC. This data policy framework gathers the guides for scientists who participate in data sharing using the FAIR model, Open Science policies, and cloud computing.

3.2 ExPaNDS Landscaping Survey

ExPaNDS WP2 and WP3 jointly conducted a landscape analysis survey[72] in December 2019 in order to establish a baseline on the current state of the 10 participating RIs on FAIR data policies and data management practices. This activity can be compared with a similar survey[73] undertaken within the PaNOSC project. This baseline can be used to steer the subsequent activity of the project and to assess its impact. In this report, we focus on the responses to the survey relevant to data policy and provide a summary of the key findings.

---

[71] FAIRsFAIR: Fostering FAIR Data Practices in Europe (2019-2022) funded by the European Commission under the H2020-EU.1.4.1.1. programme Grant Agreement 823852 www.fairsfair.eu
[72] Ashton, A., Da Graca Ramos, S., Matthews, B. et al. (2019). ExPaNDS data landscaping survey. https://doi.org/10.5281/zenodo.3673811
[73] PaNOSC (2019). WP2 data policy and stewardship. https://github.com/panosc-eu/panosc/tree/master/Work%20Packages/WP2%20Data%20Policy%20and%20stewardship

3.2.1 Existing data policies

The national research facilities represented in ExPaNDS have been actively developing data policy. All the facilities report that they either have a data policy in place, or else they are in the process of agreeing a policy in their institution. All have now taken the 2011 PaN-Data Policy framework document into account. Some have had policies in place for some years which have gone through several revisions. The current state of data policies at the 10 participating Research Infrastructures is as in the following table.

ExPaNDS

RI	Does your organisation have a data policy (DP) defined?	Year of first adoption of a DP based on PaN-data	When was the last version of the DP released?	Web address of policy
DESY	In progress	2017		DESY endorsed the PanData Policy as applicable to Photon Science In 2017. DESY is in the process of implementing the policy as part of the beam-time application process.
PSI	Yes	2016	30/11/2016	https://www.psi.ch/en/science/psi-data-policy
DLS	Yes	2019	April 2019	https://www.diamond.ac.uk/Users/Policy-Documents/Policies/Experimental-Data-Management-Pol.html
UKRI-ISIS	Yes	2012	March 2020[74]	https://www.isis.stfc.ac.uk/Pages/Data-Policy.aspx
SOLEIL	Yes	2018	02/10/2018	https://www.synchrotron-soleil.fr/en/file/11308/download?token=96KVRymM
CELLS-ALBA	Yes	2017	01/07/2017	https://www.cells.es/en/users/call-information-1/bases/2017_07_data_policy_alba_approved-cr.pdf
HZB	Yes	2016	19/01/2017	https://hz-b.de/datapolicy
HZDR	Yes	2016	26/06/2018	https://www.hzdr.de/db/Cms?pOid=57725
MAX IV	Yes	2015	19/01/2017	https://www.maxiv.lu.se/wp-content/plugins/alfresco-plugin/ajax/downloadFile.php?object_id=d606d8dc-4593-413f-8261-b1fa7dc46184
ELETTRA	Yes	2013	01/02/2020	https://vuo.elettra.eu/vuo/cgi-bin/download-tm4.py?frm_user_id=8707&frm_iddocumenttype=14&frm_iddocument=492382&frm_hash=167f09e730ec9ae648ad6831d39e983719bede89

---

[74] The Data Policy of UKRI-ISIS was under revision at the time of the survey in December 2019. We have updated the response to include the revised version.

3.2.2 General principles underlying RI data policies

The data policies generally include advice and guidelines on principles and practices behind the policy. All 9 facilities that have a published policy include general principles, with between 2 and 17 clauses and definitions identified.

The development of data policies in this community is influenced by the PaNdata policy framework, which predates the elaboration of the FAIR data principles as a specific set of guidelines, although it is strongly aligned with the principles of Open Science. Consequently, there is an explicit treatment of the FAIR data principles in only 3 data policies.

Where FAIR is mentioned, the motivations for collecting FAIR are included in the policy. Several institutes report that while they do not explicitly mention FAIR, they do cover the motivations and benefits of Open Science.

Further good data management practices are generally explicitly covered in policies. Some data policies include information on the processes involved in the enactment of the policy; however, some would see that this would be out of scope of the policy.

3.2.3 Elements addressed by RI data policies

All of the data policies cover the raw data generated by the facilities' instruments and give explicit definitions of what is covered by 'raw data'. Most policies go further to consider the data resulting from subsequent processing and analysis. This typically would clarify ownership and encourage users to manage processed data in an open manner; processed data may not be covered by the policy as ownership and responsibility for these data can be seen to change during the scientific process.

Similarly all policies cover the metadata collected to support the publication of data covered by the policy, while half of the policies give explicit consideration to PIDs being used to support the stable reference and publication of experimental data.

Policies can also cover how other information may be stored and released. In particular, the experimental proposal is an area that is often considered. This gives information on the experimental context, which is important for subsequent reuse, and is thus valuable to be made openly available; however, users will often consider that it contains sensitive intellectual property covering their research plans. Thus, facilities give the experimental proposal special consideration, with an explicit statement of how and when the proposal is made public and how it is used to seed metadata.

RIs are also mindful of needs of users to be able to benefit from the use of experimental data within their own research, and thus 7 facilities specify an embargo period of between 3 and 5 years with access restricted to the experimental team.

The relationship of the institutional data policy to the user's DMP is emerging as an important factor in the application of policy. Some data policies now give explicit consideration to DMPs, although these still represent a minority.

Data licencing is considered in 7 of the policies, with others reporting this is an aspect which is being developed (with one now including a licence in a recent policy update) with the adoption of Creative Commons (CC) licences.

Data protection (e.g. GDPR) is another area which has become more prominent since the development of the PaNdata policy framework, and data policies are now beginning to cite it explicitly. Note that data protection would also typically be covered within the policies and general terms and conditions of the use of the facility.

4. The PaNOSC Data Policy Framework

4.1 Development of the PaNOSC Policy Framework

This section focuses on the development of the [PaNOSC Data Policy Framework](#)[75] document. The discussion reviews the aims of PaNOSC's data policy framework task, as well as the approach and methods used. In particular, the considerable contribution of ExPaNDS is noted.

4.1.1 Aims

The PaNOSC Data Policy Framework represents deliverable D2.1 for the PaNOSC project. The document largely seeks to fulfil the aims of PaNOSC WP2, task 2.2:

> Based on the existing PaNData policy, create a new PaN data policy framework that all facility specific data policies* should adhere to. Ensure that the data policy framework is aligned with EOSC activities on data policy harmonization. The aim of the policy [framework] is to ensure that FAIR principles are applied as broadly as possible.[76]

Note that the PaNOSC proposal (March 2018) was submitted and approved before the ExPaNDS project (proposal submitted November 2018) was proposed. As such, formally, the PaNOSC policy framework task refers only to the PaNOSC partners, which represent pan-European research infrastructures.[77] It is important to recognize, however, that the same scientist or research group from the PaN community may well conduct experiments at both pan-European and national facilities. Thus, to provide a consistent and compatible environment for facility users, it is helpful for data policy making at both types of RI to align. To this end, ExPaNDS has a strong interest in the development of the PaNOSC data policy framework.

---

[75] Gotz, A., Perrin, J., Fanghor, H. et al. (2020). PaNOSC data policy framework. https://doi.org/10.5281/zenodo.3862701
[76] PaNOSC proposal (2018).
[77] The PaNOSC partners are: European Synchrotron Radiation Facility (ESRF), Central European Research Infrastructure Consortium (CERIC-ERIC), ELI Delivery Consortium, European Spallation Source (ESS), European Grid Infrastructure Foundation (EGI), European XFEL, and Institut Laue-Langevin (ILL). PaNOSC (2020). Partners. https://www.panosc.eu/partners/

4.1.2 Approach and methods

The PaNOSC data policy framework derives from a three-phase development approach:

1. Phase one resulted in a first draft document, which was based on the 2011 PaNdata policy framework[78] for the PaN community. This first draft was prepared during the breakout sessions of PaNOSC WP2 at the project's first Annual Meeting in November 2019.

2. Phase two consisted of a series of ten review meetings, during which experts from both PaNOSC and ExPaNDS worked together to review and enhance the draft's contents.

3. The third, and final, phase focused on the evaluation of the new policy framework in relation to the RDA FAIR Data Maturity Model. This led to further revisions of the data policy framework. As in phase two, the work was undertaken in a series of meetings attended by both PaNOSC and ExPaNDS colleagues.[79]

As should be evident from the approach outlined above, ExPaNDS partners made a significant contribution to the development of the PaNOSC Data Policy Framework. ExPaNDS is keen for our contribution to be noted, even though the new policy framework is the formal deliverable of another project. Likewise, we would propose to continue to collaborate with PaNOSC as we further develop the data policy framework in ExPaNDS and seek to engage with national PaN facilities regarding its usefulness. This joint ExPaNDS/PaNOSC effort provides a good example, not only of ongoing collaboration between PaN RIs on key concerns, but also of synchronization of effort across two Horizon 2020 projects.

4.2 Overview of the PaNOSC Data Policy Framework

This section overviews the PaNOSC Data Policy Framework, providing an abstract and brief summaries of the document's five sections and two appendices. For ease of reference, information about the online location of the document is also supplied.

4.2.1 Abstract

While the data policy framework document itself does not incorporate an abstract, one is available on the document's online landing page:

> This paper presents the new photon and neutron research data policy framework based on the previous PaNData policy (https://doi.org/10.5281/zenodo.3738497) applicable to all photon and neutron facilities and scientific research data in general. The data policy framework is strongly aligned with the FAIR principles. The aim of the policy is to ensure that the FAIR principles are applied in research data policies. This deliverable has been prepared by the EOSC projects PaNOSC (https://panosc.eu) and ExPaNDS

---

[78] PaN-data Europe (2011). Common policy framework on scientific data. https://doi.org/10.5281/zenodo.3738497

[79] Gotz, A., Perrin, J., Fanghor, H. et al. (2020). PaNOSC data policy framework. https://doi.org/10.5281/zenodo.3862701

(https://expands.eu) together to ensure harmonisation [*sic*] of the updated data policies for the photon and neutron communities.[80]

4.2.2 Document sections

The Data Policy Framework comprises five main sections:

1. The **introduction** section provides background and sets out the reasons behind the development of the new data policy framework. The original Common Policy Framework on Scientific Data, published in 2011 and subsequently endorsed by PaN RIs across Europe, has played a key role in shaping data policy for individual RIs over the last nine years. However, during this time, the FAIR principles have emerged and gained considerable traction, especially in relation to European policy making and research data-related practice. As such, there is a need for a revised PaN data policy framework that incorporates FAIR.

2. Section two sets out a few brief **recommendations** for PaN facilities in relation to the suggested uptake of the new data policy framework and summarizes the minimum steps a facility needs to take to incorporate and adapt the policy framework for local use. The need to retain the FAIR elements of the policy framework is emphasized.

3. Section three presents a **generic scientific data management policy** – in essence, a model policy. Content covered includes definitions, general principles, the use of PIDs, raw and processed data and associated metadata, auxiliary data, and results. 4. A set of nine '**best practice**' statements follows the model policy section. Many of these focus on metadata and on the retention of contextual information about the experiment.

5. The final section of the framework focuses on **termination of custodianship** — "the event that a facility decides to stop acting as a custodian and/or maintainer and provider of the metadata catalogue" related to experimental data. In particular, steps for informing principal investigators are set out.[81]

4.2.3 Associated appendices

The Data Policy Framework also includes two appendices:

1. Appendix 1 provides detailed implementation notes.

2. Appendix 2 presents an analysis of the policy framework in relation to the RDA FAIR Data Maturity Model. The results of the analysis are collected in a table and illustrated using radar charts.[82]

---

[80] Gotz, A., Perrin, J., Fanghor, H. et al. (2020). PaNOSC FAIR research data policy framework [landing page in Zenodo]. https://zenodo.org/record/3862701
[81] Gotz, A., Perrin, J., Fanghor, H. et al. (2020). PaNOSC data policy framework. https://doi.org/10.5281/zenodo.3862701
[82] Ibid.

4.2.4 Online access to the policy framework document

The PaNOSC Data Policy Framework is an online document published as a pdf through the Zenodo Open Science platform. The document is persistently available at: https://doi.org/10.5281/zenodo.3862701.[83] Appendix A includes a link to the deliverable, accompanied by a screenshot of the document's front cover page.

5. Analysis of Data Policy Framework in Relation to FAIR

5.1 Revising Data Policy in Light of FAIR

A prime motivation for revisiting data policies is to consider how they might be modified to consider and encourage appropriate practices for FAIR data by both the RI and user community. The data management community as a whole has been considering the implications of FAIR data on data policy in general. In this section, we consider some of these recommendations and reflect on how they can be applied to the context of PaN facilities as represented in ExPaNDS. Of particular relevance is the work which has been undertaken in the context of the EOSC, and thus we focus on the recommendations of two EOSC actions, in particular: the TFiR report and the work of the FAIRsFAIR project on best practices for FAIR data policy.

5.2 TFiR Actions for Policymakers

Alongside each of its 27 recommendations to support FAIR, TFiR presents a number of actions, with each action aimed at particular stakeholders. One such stakeholder is 'policymakers'. While RIs primarily represent another stakeholder group identified in TFiR ('institutions', i.e. research performing organisations), in the case of data policy specifically, one can argue that RIs do act as 'policymakers'. This is especially the case when, like the ExPaNDS and PaNOSC partner facilities, they are seeking to develop an overarching data policy framework for the PaN community.

Several TFiR actions with 'policymakers' as associated stakeholders are relevant for the development of FAIR data policy frameworks:

- The relationship between FAIR and Open should be clarified and well-articulated as the concepts are often wrongly conflated. (Recommendation 1: Define FAIR for implementation, Action 1.3)
- Policies must assert that the FAIR principles should be applied to research data, to metadata, to code, to DMPs and to other relevant digital objects, as well as to policies themselves. (Recommendation 16: Apply FAIR broadly, Action 16.1)

---

[83] Ibid.

- The greatest potential reuse comes when data are both FAIR and Open. Steps should be taken to ensure coherence across data policy, emphasising both concepts… (Recommendation 17: Align and harmonise FAIR and Open data policy, Action 17.1)
- Policies should be versioned, indexed and semantically annotated in a policy registry to enable broad reuse within the FAIR data ecosystem. (Recommendation 17: Align and harmonise FAIR and Open data policy, Action 17.3)
- … policies may explicitly allow a reasonable embargo period to facilitate the right of first use of the data creators. Embargoes should be short (e.g. c. six months to two years) based on the prevailing culture in the given research community. (Recommendation 17: Align and harmonise FAIR and Open data policy, Action 17.5)
- Policies should require an explicit and justified statement when (publicly-funded) data cannot be Open and a proportionate and discriminating course of action should be followed to ensure maximum appropriate data accessibility, rather than allowing a wholesale opt-out from the mandate for Open data. (Recommendation 17: Align and harmonise FAIR and Open data policy, Action 17.6)
- … refine appraisal and selection criteria and to improve guidance and processes on what to keep and make FAIR and what not to keep. (Recommendation 19: Select and prioritise FAIR Digital Objects, Action 19.1)
- Policy should require data deposit in certified repositories and specify support mechanisms (e.g. incentives, structural funding and/or funding for deposit fees, and training) to enable compliance. (Recommendation 20: Deposit in Trusted Digital Repositories, Action 20.1)
- The results of monitoring processes should be used to inform and iterate data policy. (Recommendation 25: Implement FAIR metrics to monitor uptake, Action 25.5)[84]

Although these specific recommendations are aimed at policymakers, i.e. the role that RIs adopt when they create data policy, they are not the only TFiR recommendations aimed at policy making. FAIRsFAIR has produced an analysis of the latter in relation to policy enhancement, and we consider these in detail in the next section.

5.3 FAIRsFAIR Practical Recommendations for Policy Enhancement

The FAIRsFAIR project has prepared a series of practical recommendations for policy enhancement to support the realization of a FAIR ecosystem, based on the recommendations of the TFiR report. FAIRsFAIR WP3: FAIR Data Policy and Practice "carried out an analysis of the current data policy landscape at various levels (national, funder, publisher, institutional) to provide a snapshot of the situation in 2019 and to identify policy elements that support or hinder FAIR data practice…". A key aim for FAIRsFAIR is to amplify existing policy recommendations wherever possible rather than to duplicate what has already been done. In this respect, the initial set of

---

[84] EC Expert Group on FAIR data (2018). Turning FAIR into reality. https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf

recommendations builds upon recommendations made by a number of initiatives including EOSC-hub, EOSCpilot, RDA Europe, OpenAIRE, FREYA, and FAIRsFAIR.[85,86,87,88,89,90]

We discuss the application of these recommendations to the ExPaNDS PaN RIs, using the results of the landscape assessment and the joint work on draft policy framework recommendations with PaNOSC. We note that recommendations such as those in FAIRsFAIR arise from consulting largely with researchers in research performing organisations (typically universities or research institutes) and the organisations which support those researchers (e.g. funders, research libraries, publishers). The PaN facilities are in an unusual position in that ecosystem: the facility sets up the measurement workflows at its instruments, has control over the data curation, the data is typically stored in its archives, and it controls the access to the data and the retention times, while users and user funding are typically from outside the RI. As a consequence, not all of the FAIRsFAIR recommendations fit the facilities context well. Thus we consider carefully for each recommendation whether it makes sense in our case.

The recommendations[91] are presented below under each of the three stages outlined by Turning FAIR into Reality. [92]

- **Define** - concepts for FAIR Digital objects and the ecosystem

- **Implement** - culture, technology and skills for FAIR practice

- **Embed and Sustain** - incentives, metrics and investment

We relate the recommendations back to the TFiR recommendations, thus giving an assessment of the extent to which the ExPaNDS community is working towards satisfying these recommendations. Further, we include recommendations on how the recommendations from FAIRsFAIR should inform the ExPaNDS Data Policy Framework.

---

[85] EOSC-hub (2019). D2.8 - First data policy recommendations. https://www.eosc-hub.eu/deliverable/d28-first-data-policy-recommendations-approved-ec
[86] EOSCpilot (2019). EOSCpilot's 9 recommendations for EOSC policy. https://eoscpilot.eu/news/eoscpilots-9-recommendations-eosc-policy
[87] Research Data Alliance (2017). RDA Europe. https://www.rd-alliance.org/rda-europe
[88] OpenAIRE (2020). OpenAIRE. https://www.openaire.eu/
[89] FREYA (2020). Welcome to FREYA. https://www.project-freya.eu/en
[90] Bangert, D., Hermans, E., van Horik, R. et al. (2019). Recommendations for services in a FAIR data ecosystem. https://doi.org/10.5281/zenodo.3585742
[91] Davidson, J., Grootveld, M., Whyte, A. et al. (2020). D3.3 policy enhancement recommendations. https://doi.org/10.5281/zenodo.3686901
[92] EC Expert Group on FAIR data (2018). Turning FAIR into reality. https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf

ExPaNDS

5.3.1 Define

#	FAIRsFAIR recommendation	TFiR action	ExPaNDS commentary	ExPaNDS recommendation
1	Provide practical guidance to researchers and data stewards on how to implement FAIR within different domains – specifically on how to describe data using appropriate metadata standards, data tags and ontologies. Commitments are needed from all stakeholders to support and meet training needs relating to Open Science - for both researchers and data stewards.	Action 16.3	This is not a topic that is directly addressed by the RIs Data Policy documents. Nevertheless, data policies are increasingly reflecting a commitment to open data, which should evolve to FAIR. The Policy Framework should recommend the collection and publication of sufficiently rich metadata. Recommendations and guidance on appropriate metadata is again, not a topic for policy, but instead represent part of implementation of policy — in ExPaNDS, in tasks 2.3 and 3.2. Given that the policy should recommend FAIR data, then in order to achieve this, it would be appropriate to support FAIR data by adopting an advocacy and training programme for staff and users.	**Recommendation 1.1**: The Policy Framework **should** give an explicit commitment to FAIR Data *at the point of leaving the facility*. **Recommendation 1.2:** The Policy Framework should recommend the collection and publication of sufficiently rich metadata to support the delivery of FAIR data, at the point that data "leaves the facility".
2	Cooperate with relevant initiatives to support funding bodies to characterise and, where needed, enhance policies to align with FAIR	Action 1.3	ExPaNDS RIs work within a complex policy landscape. They are subject to policy drivers from a number of different sources:	**Recommendation 2.1:** The Policy Framework should be adaptable to accommodate national and other relevant funder policies.

| principles - either explicitly or implicitly. | | National funder policy. National RIs are working primarily within their national context, who both directly fund the RI and also fund their user communities. Thus the national policy is likely to be the primary driver for an RI. International funder policy. RIs also will often have agreements with funders in other countries and also transnational funders, in particular for ExPaNDS RIs, the EC. Other funder policy. Other funders will also work with ExPaNDS RIs, particularly charitable and commercial organisations. User supplied policy. Users are typically from other institutions (universities) and will be subject to their own institutional policies. Domain norms. ExPaNDS RIs are cross-disciplinary and discipline communities will have their own norms on best practice for data, metadata, and publication. Thus ExPaNDS RIs need to have a policy framework which is sufficiently | |

			broad and flexible to accommodate these demands.	
3	Support policy makers to ensure that they include the dates of validity for their policies as well as any planned review dates.	Action 3.2	ExPaNDS RIs typically include the date of validity (see the Landscape Survey).	**Recommendation 3.1:** The Policy Framework should specify that the policy includes dates of validity and planned review dates.
4	Building on the work of other initiatives (FAIRsharing, EOSCpilot, RDA), agree on a common set of FAIR policy elements and work with stakeholders to employ them to describe their policies. The emphasis should be on describing those policy elements that may be considered 'rules' rather than simply suggested good practice to support machine-actionability	Action 3.2	While not a direct concern of ExPaNDS RIs, they should nevertheless track efforts to build a common set of FAIR policy elements, and frame their policy using this terminology.[93] ExPaNDS RIs have not as yet issued policies in machine-readable format. Further use cases and examples would be needed to guide how machine-actionable policies can be used in practice, and so, the RIs should track emerging experience and use machine-actionable policy as appropriate.	**Recommendation 4.1:** The Policy Framework should use emerging set of FAIR policy elements to describe RIs' policies. **Recommendation 4.2:** The Policy Framework should propose that the policy should be machine-readable as an option.
5	PIDs should be assigned to clearly versioned policies. These PIDs should be included in the metadata records in registries such as	Action 17.3	ExPaNDS RIs are typically not issuing PIDs to policies, and not registering policies in policy registries.	**Recommendation 5.1:** The Policy Framework should propose that RIs' policies have a PID in their own right.

---

[93] See for example the Data Policy Standardisation and Implementation Interest Group of the Research Data Alliance (RDA) https://www.rd-alliance.org/groups/data-policy-standardisation-and-implementation-ig, although the focus of this group is on Publisher Policies.

	FAIRsharing.org or other policy registry services (such as those envisaged by EOSCpilot)[94]			**Recommendation 5.2:** The Policy Framework should propose that RIs might consider registering their policies in a policy registry. In particular, they should register their policy in a registry if it were to be mandated or recommended by the EOSC Rules of Participation.
6	Working with research communities to define data outputs, policymakers should adopt standard descriptions to ensure that definitions provide clarity on the range of outputs that should be considered and what might be considered "FAIR enough".	Action 17.1	ExPaNDS RI's policies include definitions of those outputs which are in the scope of the policy (e.g. "raw data", "processed data", "publications"). These definitions vary, and so, within a common policy framework, there should be a consensus view on the definitions and scope of policy. Current policies do not consider FAIR explicitly, and should be modified to provide support to be sufficiently FAIR for the community.	**Recommendation 6.1:** The Policy Framework should include a consensus view on the definitions of outputs within a common glossary. This common glossary should take into account glossaries within the EOSC ecosystem. **Recommendation 6.2:** The Policy Framework should specify which outputs are in the scope of the policy. **Recommendation 6.3:** The Policy Framework should consider how it should be modified to deliver FAIR outputs at the point of leaving the RI. **Recommendation 6.4:** The Policy Framework should propose that RIs should undertake FAIR audits and assessment, publishing their current state of FAIRness.

---

[94] Mack, L. and Papadopoulou, E. (2018). EOSCpilot D3.4 Open Science policy registry. https://eoscpilot.eu/content/d34-open-science-policy-registry

7	Standardised exceptions for not sharing data should be developed and promoted in associated policy guidance.	Action 1.3 & 17.6	RIs' policies will specify that they are restricted to publicly funded research and that research funded from other institutions should be excluded. There should also be explicit restrictions on grounds of personal confidentiality, law enforcement, and national security. Further, RIs also specify embargos, restricting access to the experimental team for a limited period.	**Recommendation 7.1:** The Policy Framework should give the explicit scope of the policy to publicly-funded research. **Recommendation 7.2:** The Policy Framework should allow outputs to be embargoed to the experimental team for a limited period of time. **Recommendation 7.3:** The Policy Framework should specify the body that adjudicates on setting restrictions to data sharing.
8	Standard exceptions should be added to metadata schemas used by repositories for consistency.	Action 1.3 & 17.6	RIs' policies will typically identify groups of users who are counted as exceptions to open science obligations, typically including non-public users.	**Recommendation 8.1:** Administrative metadata should include information on access control and embargoes.
9	Working with relevant stakeholders, support adoption of rights and licensing documentation schemas for different types of research outputs as they are defined.[95]	Action 17.1	Data licencing was not explicitly considered in the PaNdata policy framework and is variably covered in the existing data policies of RIs.	**Recommendation 9.1:** The Policy Framework should propose that data is licenced, including a recommended open licence.
10	Provide mechanisms to enable searching for data by license type in repositories.	Action 17.1	This is an implementation issue out of scope of the policy framework. The	

---

[95] As recommended in Implementation Action 4.1 of EOSCpilot D3.6 Final Policy Recommendations https://www.eoscpilot.eu/content/d36-final-policy-recommendations

			policy can recommend that outputs can be searched flexibly.	
11	Provide legal guidance on choosing appropriate licenses during active stage of research and for assessing the compatibility of different license types when reusing multiple data outputs.	Action 17.1	RIs can provide advice on appropriate licenses for derived products or related outputs arising from the use of facilities, but cannot mandate their use. RIs would also be aware of their legal liability in providing formal advice.	
12	Working collaboratively, define and require standardised Data Accessibility Statements.	Action 1.3	RIs should cooperate with journals that specify in their data accessibility statements that they require the accessibility of experimental data. But as the policy should cover the availability of those data, then there is nothing to be added to the policy.	
13	Provide support to repositories and data stewards to develop tombstone metadata records that are maintained - even when data is no longer available - and to ensure that these metadata records are referenced in Data Availability Statements.	Action 20.1	RIs endeavour to maintain records of experiments indefinitely, and by providing PIDs to experiments and data, will commit to maintaining that record, thus providing tombstoning.	**Recommendation 13.1:** The Policy Framework should specify the minimal length of time that outputs will be maintained. **Recommendation 13.2:** The Policy Framework should specify that RIs should maintain and publish a record of experiments, even if the outputs are no longer available.

ExPaNDS

5.3.2 Implement

#	FAIRsFAIR recommendation	TFiR action	ExPaNDS commentary	ExPaNDS recommendation
14	Working with all stakeholders, ensure that data management planning is supported across the entire research lifecycle so that data can be "born FAIR" and kept "FAIR enough" over time. Require updating of DMPs over the research lifecycle leading to comprehensive, high-quality end stage DMPs that are included in end-stage reporting.	Action 5.2	The aim of the revised policy framework is to ensure that as far as is practical and under the control of the facility, research outputs are FAIR at the point of leaving the facility. This would need to be within an experimental lifecycle which ensures FAIRness; this is considered in detail in ExPaNDS task 2.3 and its deliverables. Polices currently do not generally support the use of DMPs in RIs' processes. There will need to be further investigation and advocacy to guide facilities on the appropriate application of data management planning which is acceptable to research communities. DMPs are considered in detail in ExPaNDS task 2.3.	**Recommendation 14.1:** The Policy Framework should propose that a data management planning activity is undertaken with experimental teams to ensure that the data is as FAIR as possible.
15	Policies and related guidance should emphasise that data management planning and sharing data supports research integrity goals, enhances data quality and contributes to reproducibility and transparency.	Action 5.1	Advocacy and training activities will need to be undertaken to support the policies within RIs.	**Recommendation 15.1:** While out of scope directly of the Policy Framework, it should be framed within material providing motivations and training for FAIR research.

16	Support researchers to assess the potential risks, benefits and associated costs to enable the sharing of FAIR data as they draft their DMP.	Action 5.2	Advocacy and training activities will need to be undertaken to support the policies within RIs, which should include risks and costs.	**Recommendation 16.1:** While out of scope directly of the Policy Framework, the Policy Framework should be framed within material providing guidance on risks and costs of FAIR research.
17	RDM support should place an emphasis on selecting which data to make and keep FAIR as well as advising on where data should be deposited.	Action 20.1	RIs' standard processes will select which outputs to keep within an experimental workflow at a facility. Facilities should provide advice and assistance to users on data management planning and implementation activities. Where data and derived products are deposited after the experiment, this is out of the facilities direct control, although facilities can make their expectations explicit.	**Recommendation 17.1:** The Policy Framework should specify which outputs are managed by the facility. **Recommendations 17.2:** The Policy Framework can propose guidance on derived products from experiments.
18	Where resources allow, RPOs [Research Performing Organisations] should provide domain specific RDM support locally (research group, faculty/department). Where local support isn't feasible, the development of shared domain-specific resources should be supported and maintained with resources provided by all stakeholders.	Action 5.3	While outside the scope of policy, RIs can provide support to experimental teams whilst they are within the facility, including how to handle research outputs to make them FAIR.	

| 19 | Building upon previous work on defining cost types work with funding bodies and research performing organisations to implement these in new grant applications. RPOs should monitor and review RDM costings over the life of the project and beyond to assess the effectiveness of current cost models. | Action 5.2 & 18.3 | The cost of maintaining research outputs is of prime concern to ExPaNDS RIs, in particular, balancing the cost of long-term preservation of large volumes of data against the benefits of reproducible and reusable research to support better and new science. Costs are further considered in #21 below. | |

5.3.3 Embed and sustain

#	FAIRsFAIR recommendation	TFiR action	ExPaNDS commentary	ExPaNDS recommendation
20	Provide guidance on how to cite a broader range of research outputs including data and software, as well as actors and enablers such as data managers, data stewards, funding bodies, research infrastructures and organisations.	Action 26.2	RIs' policies will specify how to cite experimental data within research publications. Citation of other research outputs (e.g. software) and associated resources (e.g. instruments) are subject of active interest (e.g. ExPaNDS WP2, task 2.4) within the RI community, to provide context for reproducible research and also to further assess the impact of the use of facilities.	**Recommendation 20.1:** The Policy Framework should recommend how experiments are cited. **Recommendation 20.2**: The Policy Framework should recommend that PIDs and metadata are provided for associated research outputs (e.g. software) and resources (e.g. instruments), so that they can be citable.
21	Working collaboratively on carefully scoped pilots, funding bodies, RPOs and repositories should assess and report on the	Action 18.3	While out of scope for the Policy Framework, RIs are working within carefully managed budgets and will routinely monitor costs, including the	**Recommendation 21.1**: The policy should recognise that cost constraints may limit the capacity of RIs to maintain access to FAIR data indefinitely.

	costs of making and keeping data FAIR to build up a picture of how the costs might change over time and to leading to the development of sustainable funding models.		additional costs of maintaining FAIR data, and attempt to assess the benefits.	
22	Support stakeholders to consider compliance monitoring across the FAIR ecosystem using identifiers and knowledge graphs. An emphasis should be placed on rewarding good practice but, where necessary, the introduction of penalties for non-compliance should be considered.	Action 26.4	While out of scope of the Policy Framework, RIs routinely track the impact arising from the use of facilities, through use in publications and other research results, and therefore have an interest in the use of "knowledge graphs".	

6. Key Policy Elements within a PaN RI Data Policy Framework

6.1 Overview

A policy framework might be defined as:

> *"A policy framework is [a] document that sets out a set of procedures or goals, which might be used in negotiation or decision-making to guide a more detailed set of policies, or to guide ongoing maintenance of an organization's policies"*[96]

A data policy framework aims to provide a supporting structure around which a data policy can be built.[97] Three components come together to form this supporting structure:

1. Principles

   ➢ Principles are propositions that set out fundamental beliefs, behaviours or requirements.
   ➢ In the case of the data policy framework, what is the point of the framework? What is being proposed and why?

2. Underpinning ideas

   ➢ Underpinning ideas represent the approach and key concepts that serve as a basis or foundation.
   ➢ These need to be relevant to the context in which the framework will operate to ensure that the framework is grounded in and applicable to a 'real' situation.
   ➢ In the case of the data policy framework, in what context will the framework sit and how will it be used in practice?

3. Elements

   ➢ Elements are the essential or characteristic parts.
   ➢ A framework can only capture limited information about 'real life practice'. The focus is on those elements that are likely to play the most significant roles.
   ➢ In terms of data policy, what elements must the framework address, and which can it afford to leave out?

A key point is that a framework is a supporting structure designed to leave flexibility for a range of different implementations to be 'built' around the framework, and requiring deviations from the principles to be justified. In the case of a data policy framework, this means a number of different data policies.

---

[96] Wikipedia (2020). Policy framework. https://en.wikipedia.org/wiki/Policy_framework
[97] Cambridge Dictionary (2020). Framework. https://dictionary.cambridge.org/dictionary/english/framework

In the rest of this section, we discuss some key principles and elements that should be considered within a RI's data policy. These elements are those in which RIs need to make choices on the level of commitments which they are prepared to make and as well as the obligations that they would have on users. In particular, we have taken into account the needs of enabling FAIR data in the light of the recommendations above. We have been guided on the structure of the principles by a [CODATA note](#)[98] on best practice for research data management policies, which gives a useful categorization of the topics that a data policy should consider.

6.2 General Drivers and Principles

RIs wish to maximize their scientific value and their broader societal impact, while at the same time serving the requirements of their user communities. The construction and operation of user facilities are significant long-term investments for public sector research, and therefore, those facilities are obliged to seek the maximum return for the public expenditure. This includes the selection of user experiments that have the maximum science value from the allocation of instrument time. This has been traditionally measured in terms of high-impact publications and other research outputs (e.g. patents and products).

Facilities science has become more data intensive, with the volume and complexity of research data increasing as beam intensity increases, and more and more sensitive detectors have been developed. Thus, the value of the experiment is increasingly encapsulated in those data. At the same time, the volume and complexity has meant that user communities have found it increasingly challenging to store and process those data. Consequently, facilities themselves have taken a role to help manage the data directly, with data storage capacity, specialized analysis software, and often, significant compute power.

This additional responsibility of user facilities has meant that RIs have needed to clarify the rights and responsibilities of their facilities and their user communities, and thus the need has emerged for data policies.

1. RIs should openly publish a data policy, including the period in which the policy is in force and when it is planned to be reviewed. A PID should be used to refer to the published version of the policy.

The RIs' data policies should seek to address the following objectives.

2. Clarify the ownership and access to data collected at a facility.

---

3. Specify the extent to which the facility will supply the user with support to access and analyse the data, within the experiment and beyond for subsequent analysis, to maximize the opportunity for the user to develop their science.

4. Specify the obligations on users in the subsequent use of the data.

Further, the requirement of RIs to maximize the scientific value of the use of the facilities forms a strong motivation for making experimental data available for re-analysis and reuse. As a rare and specialized source of data, the potential for reuse would require specialized expertise. It is also a reasonable expectation that the user should be in the best position to exploit the experimental data results. Nevertheless it remains of value for the data to be made available for others to reanalyse and validate the results, and to reuse the data within their own lines of research. Thus the policy should further the following additional goal.

5. Specify the actions the RI should undertake to ensure that experimental data is made available and reusable to maximize the scientific impact of the experiment.

Facilities are usually not funders of research, but rather most experiments have a dual funding regime, where the core facilities' funders resource the operations and staff while the users are supported by grants from other funders. For a particular experiment, the data policies of both funders need to be respected.

6. The data policy should be sufficiently flexible to accommodate the data policies of RIs' national funders and users' funders.

In ExPaNDS, we are focussing on national RIs, funded via major national public sector research bodies.

7. A RI's data policy should comply with their national research funders' policy.

One feature of the PaN community is that there is a shared user community. The extent that PaN facilities share a common user community has been demonstrated in the PaNData-ODI project[99] and remains the case. Users will use instruments at different facilities, taking advantage of the different characteristics of instruments and the different capabilities of neutron and synchrotron sources. If different facilities have different approaches to managing and sharing data, this forms a barrier to the integration and sharing of data, where the user can bring data from different experiments, and then publish the data in a reliable and consistent manner.

8. RIs should seek to align their data policies, within the constraints of divergent national funder policies and legal frameworks.

The aim of this data policy framework and that of PaNOSC is to provide a common framework to guide the development of compatible data policies across facilities.

Further, machine-readable policies are recommended within the TFiR report to make interpreting policies easier for machine-to-machine access to data. This is not widely done currently, and should be left as an option for future refinement.

> 9. RIs should consider making their policies machine-readable

Further, if facilities services are to be included within the EOSC ecosystem, the policy should also take into account the EOSC Rules of Participation and other requirements, such as on-boarding into the EOSC service registry.

6.3 Scope

The data policy should define the scope of its coverage and what it excludes.

6.3.1 Definitions

RIs' data policy scope include definitions of terms to ensure that there is clarity in the scope and coverage of the policy. This should include definitions of fundamental concepts. For ease of understanding and harmonizing of policies, it would be beneficial if RIs were to harmonize the definitions used as much as is possible.

> 10. RIs should seek to harmonize the terminology used, if possible, and use common data policy elements.

ExPaNDS is developing a Glossary to share fundamental concepts, developed via a community consensus on the definitions of key terms; we give a selection of these in Appendix B. This Glossary should also refer to the wider definitions of terms as given in, for example, the emerging EOSC Glossary.[100] Fundamental to defining the scope of data policy is the classification of the data in scope. This has not as yet gained consensus. Consequently, we give as an illustrative example the definitions of data as defined in the SOLEIL data policy[101] as representative of the definitions given in data policy.

---

[100] EOSC Glossary Interest Group (2020). First intermediate version of the EOSC glossary released. https://www.eoscsecretariat.eu/eosc-glossary
[101] Gagey, B. (ed.) (2018). SOLEIL data management policy. https://www.synchrotron-soleil.fr/en/file/11308/
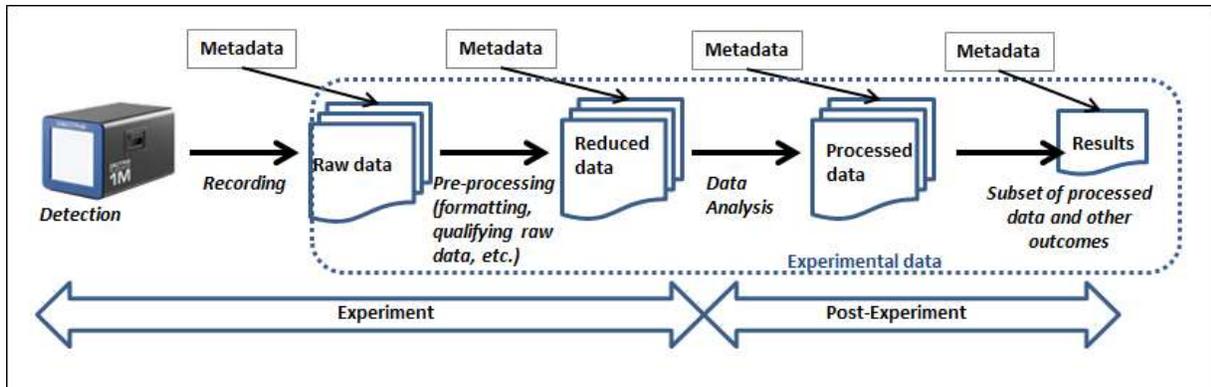
Figure 1: Simplified illustration of Classes of Experimental Data in the Science Life Cycle
(from the Soleil Data Policy)

2.1. The term experimental data, see Figure 1, pertains to data collected from experiments performed on instruments. This definition includes (but is not limited to) data that are created automatically or manually by facility specific software and/or facility staff expertise to facilitate subsequent analysis of the experimental data.

2.2. The term raw data, see Figure 1, pertains to the experimental data that is recorded during experiments, as produced by the detection system, and cannot be derived from other persistent data.

2.3. The term reduced data, see Figure 1, pertains to the experimental data that is derived from raw data through pre-processing during experiments including (but not limited to) formatting and qualifying raw data and helping to decide on the continuation of the experiment.

2.4. The term processed data, see Figure 1, pertains to the experimental data that is derived from raw data along the analysis steps.

2.5.The term results, see Figure 1, pertains to a subset of processed data and other outcomes arising from the analysis of experimental data, excluding publications based on such analysis and intellectual property (IP) rights.

2.6. The term metadata, see Figure 1, describes information pertaining to data collected from instruments, including (but not limited to) the context of the experiment, the experimental team, experimental conditions and other logistical information.

Figure 2: Definitions of data classes as given in the Soleil Data Policy

Different experiments may "leave the facility" at different stages, so the scope of the policy may apply differently for different experiments. A further additional data class is identified in the PaNOSC data policy framework for the wide variety of data resources which provide experimental context.

The term auxiliary data refers to data that provide contextual information regarding the experiment and its datasets but which are collected outside the context of the experiment conducted at the research facility, such as information about the sample images, provenance and preparation, data processing scripts, processing environment information such as software tools and versions used, etc.

Figure 3: Definition of Auxiliary Data from the PaNOSC Data Policy Framework

It is recognized that there is currently no community consensus on these definitions. Therefore for the purpose of this document, we shall use the term 'experimental data' to refer to all data related to the experiment using the facility's instrument, including auxiliary data items. We recognize that 'experimental data' has a number of subclasses, including raw, reduced, processed, and auxiliary, which are typically related to each other as in Figure 1 above. However, we shall not give at this stage a precise definition.

6.3.2. Data within the scope of the policy

Facilities should specify the scope of the data policy. The facilities experiment is typically within a wider scientific process that the user is undertaking and other research actions are taken and data generated and analysed which are outside the direct influence of the facility. These would have other IPR conditions and are subject to different data policies. Thus the RI can only assert that

11. The RI's data policy should apply to data which are generated, stored, and analysed using the facility's resources (e.g. instruments, compute infrastructure, software, staff).

In particular, this should apply to the data generated directly from instruments and reduced using facilities' compute and software resources.

12. The RI's data policy should cover the classes of experimental data :

- Raw data generated directly from the use of instruments and stored on facilities' storage resources.
- Reduced data generated from raw data produced using facilities' compute and software resources.
- Processed data generated from the use of instruments and produced using facilities' compute and software resources.

- Auxiliary data to provide contextual information, in so far as it is owned by the facility, processed using facilities resources, or submitted by users to provide supplementary information.
- Metadata used to describe, contextualize, and constrain the data resources above.

Note that this does not constitute a commitment to handle all these classes of data in the same way, or to keep them equally for the same time. The policy should provide statements on their treatment within the policy, if they are subject to different regulation.

Facilities should consider, however, what policy actions can be taken to maximize the scientific value of the data resources. For example, facilities should consider how to maintain access to data that directly underpin or substantiate published research findings and are required for validation.

13. The RI's data policy should seek to maximize the scientific impact of experiments through enabling the validation of research results and maximizing opportunities for reuse.

However, resource constraints (e.g. costs, ongoing storage capacity) are likely to limit the ability of facilities to commit to maintaining access to all data indefinitely, and so the policy should indicate the criteria for the selection of experimental data.

14. The policy should specify the retention policy for each class of experimental data, with a minimum retention period and criteria for deletion. As this includes auxiliary data, this also includes software and tools.

15. In the event that data are deleted, the facility should retain a "digital footprint" of the data. This could constitute a (metadata) record of their essential characteristics or a method to allow the reconstruction of the data. The facility should support as much as possible the provenance and validation of published research results.

Once the experiment is complete, then subsequent research actions are out of the facilities' control. However, facilities can request that users continue to keep derived results data available and reusable.

6.4 Data Sharing and Necessary Restrictions

Funder policy for publicly funded research within Europe now largely supports the aim of maximizing scientific impact by releasing research results, in particular, research data, as openly, widely, and as early as possible. Further, to encourage the use of data by third-parties, data should be interoperable with other data and software and reusable as widely as possible. Thus this policy framework recommends that the data should be FAIR "at the point of leaving the facility".

16. RIs' data policies should enable the experimental data in scope to be FAIR:

- Support the ongoing **findability** of experimental data and their associated discovery metadata to uniquely identify experimental data to as wide a spectrum of users as possible.
- Support the ongoing availability of data and associated administrative metadata to allow users, including new users, to **access** experimental data.
- Support the presentation of data and provide sufficient contextual metadata and supporting auxiliary data to maximize the opportunities for **interoperability** of experimental data with other data sources and with third-party software.
- Support the presentation of data and provide sufficient contextual metadata and supporting auxiliary data to maximize the opportunities for **reuse** of experimental data in novel research contexts.

Thus the policy should clearly refer to FAIR as opposed to open, which can have a broader and vaguer interpretation. However, FAIR does not mean open; facilities should be clear on the nature and extent of restrictions that are applied to data, which should be specified in a data licence. In particular:

17. The RI's data policy should specify a licence under which the data are made available.

18. The RIs should acknowledge the application of the relevant national legislation under the GDPR framework in the handling of personal and sensitive data.

19. RIs should specify the grounds for restricting access to data. Typical grounds would be:

- data arising from experiments which are not publicly funded (typically, expressly excluded from data sharing)
- restrictions applied by reasons of national security or prevention of criminality
- access to personal sensitive data.

Further, facilities experiments are a joint enterprise between the facility and a user group. While most facilities would assert rights to experimental outputs, nevertheless the subject and conduct of the experiment is the result of the contribution of the researcher, who has a wider research goal in conducting the experiment. The goals of the researcher to further their personal research agenda should be reflected in allowing "first use" of the experimental results to further their research objectives, by providing restricted access for a time-limited period.

20. RIs should specify the time limit (an "embargo period") for which users are allowed exclusive control on the use of experimental outputs. This should also specify who can access the data (e.g. facilities staff), who can determine who should be given access rights, and the appeals process established to alter the embargo period.

6.5 Recommendations on Enabling FAIR Data

Additionally, we propose that data policies should include specific commitments in the way they will handle experimental data which would enable the production of FAIR data.

21. The RI's data policy should include commitments to enabling FAIR data which may include:

- The RI should provide the globally unique identification of experimental data via the association of appropriate globally unique PID that conforms to the EOSC PID Policy.[102]
- The RI should annotate data with sufficient rich metadata in conformance to publicly available community and domain standards.
- The RI should support standard protocols for accessing data.
- The RI should provide data in formats conformant to publicly available standards.
- The RI should provide sufficient contextual metadata and auxiliary data.
- The RI should provide access to experimental data and associated metadata via human and machine-readable interfaces.

Note that these are not implementation decisions brought into the policy, but rather form an explicit commitment that subsequent implementation decisions will be guided by the FAIR principles. This is inherent in the commitment to enable FAIR data, which implies an adherence to the principles given in the commonly accepted definition of FAIR data.[103]

6.6 Summary of Responsibilities

The data policy should outline the rights and responsibilities, with respect to the data policy, of the actors involved.

22. RIs should specify the rights and responsibilities of particular classes of actors involved in the experimental process.

A proposed set of core actors are given below with recommendations on their rights and obligations under a facility data policy.

---

[102] Current version: EOSC FAIR Working Group and EOSC Architecture Working Group (2020). Second draft persistent identifier (PID) policy for the European Open Science Cloud (EOSC). https://doi.org/10.5281/zenodo.3780423

[103] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 3, 160018. https://doi.org/10.1038/sdata.2016.1

Actor	Definition	Rights	Responsibilities
RI	Large-scale centre owning and providing access to specialized instruments and other resources (including staff) for research purposes.	• Develop data policy and conditions on the access to facilities' resources and outputs.	• To maximize the scientific impact of the use of its resources for its user community and the wider research community. • To respect the data policy requirements of funders and users. • To provide resources for the long-term stewardship and sharing of experimental data as outlined in the data policy.
Facilities support staff	Staff employed at the facility to support research. This includes for example: User Office Staff, Instrument Scientists, Computing Staff.	• Access Experimental Data and metadata, and modify it with additional metadata for data curation and data sharing purposes, and to improve facilities' processes and performance.	• To respect the data sharing restrictions on experimental data • To maintain the long-term access and stewardship • To maintain FAIRness of data as is practicable.
PI	The main proposer of an experiment, taking the decision making for the conduct of the experiment and main liaison with the facility	• To steer and control the collection of experimental data. • To determine who has access to experimental data during the embargo period.	• To agree with the data policy of the RI • To ensure that data management planning for the experiment is completed and followed.

Experimental Team	The PI and any other persons to whom the PI assigns access rights for the conduct and analysis of the experiment	• Access to the experimental data • Add to the experimental data from additional runs and subsequent processing actions	• To comply with the RI Data Policy and data management planning for the experiment. • To provide accurate information to maintain the FAIRness of experimental data
Data re-users	Third parties accessing the experimental data for further scientific purposes.	• Access to metadata describing experiments as soon as is practical after the experiment. • Access to the experimental data after any embargo period.	• Ethical use of the data. • Acknowledgement and citation of the RI and experimental team.

6.7 Availability of Infrastructure and Responsibility for Costs

Facilities should recognize that supporting a FAIR data policy comes with the provision of infrastructure to support the retention and distribution of FAIR data. Thus the data policy should commit that the facility should support the provision of infrastructure as far as the coverage of the data policy for the facility specifies. This would then commit the facility to identifying how resources might be allocated to cover these costs, within the reasonable funding limitations available.

The infrastructure should thus provide:

23. The RI's policy should consider the extent to which it commits providing infrastructure to support the retention and distribution of FAIR data, for example:

   • a storage and curation service to keep experimental data for the specified retention periods.
   • a data discovery service to keep experimental data or its record findable.
   • a data and metadata access and movement service to allow users (including new users) to interrogate the experimental context and access experimental data.

The policy should also specify which infrastructure and costs would reasonably be expected to be incurred by users. Note that this commitment does not bind the facility to a specific implementation strategy.

6.8 Data Management Planning Requirements

DMPs are not necessarily standard within facilities. However, if data are to be well-managed and curated and also made FAIR, there is a need for users to cooperate with facilities staff to estimate the storage and computational needs of the experiment, and to assist in providing accurate metadata. This will guide the facility to provide computational resources the experimental team might need as well as document the data for future FAIR use. Thus:

24. The policy should specify the requirements on users to participate in data management planning activities. This might include:

   - Providing accurate information on the experiment for inclusion in the experimental metadata.
   - Providing estimates on the storage and computation requirements for data storage and data processing.
   - Providing additional experimental metadata, using tools provided by the facilities (e.g. via electronic laboratory notebooks).
   - Specify software needed by the experimental team to process the data.

This may include preparing a DMP – if so, this should be specified in the policy.

25. The RI's policy should specify whether the PI is responsible for preparing a DMP.

6.9 Recognition and Reward for Data

Data policy should encourage or specify how the use of experimental data should be recognized and cited.

26. The RI's Policy should promote the recognition and citation of the use of facilities. Specifically it should:

   - specify that use of experimental data should be acknowledged, including within citations.
   - encourage the citation of experimental data in publications by the experimental team and also re-users.
   - encourage re-users to contact the experimental team to express their interest in the experiment.
   - encourage the citation of software and instruments supplied by the facility.

The data policy should be presented to users within a context which promotes FAIR data and is supported with training material to ease the collection and exploitation of FAIR data. This is outside the scope of the policy itself, but should make its acceptance and use more straightforward to users.

6.10 Reporting Requirements, Compliance Monitoring, and Any Possible Sanctions

Data policies should indicate how compliance will be monitored, what reporting is required, and what sanctions may be imposed.

27. RIs should have regular audits on compliance to FAIR data.

28. Compliance to policy by users may be monitored and checked.

29. Users may be requested to report on compliance for previous experiments when applications for further access to the facility are received. Non-compliance may be a contributing factor in the refusal of further access.

The policy should also cover changes of circumstances or policy, for example arising from unforeseen restrictions on future budget or the continuity of service of the facility.

30. Changes or termination to the data policy will be given in sufficient time for PIs to take alternative action to provide alternative provision to comply with their funders' data policies.

7. Conclusion

7.1 Summary

This document brings together ExPaNDS' work to date on developing a data policy framework for PaN RIs. An overview of European and national level research data policy in relation to FAIR (section 2) coupled with consideration of the changed working landscape of RIs and the current state of data policies within ExPaNDS partner facilities (section 3) highlights not only the complexity of the national PaN RI policy environment but also the need to update PaN data policy to incorporate newer features of the research data policy landscape such as FAIR and GDPR.

A result of joint PaNOSC/ExPaNDS effort, the new PaNOSC data policy framework presents a generic example data policy (section 4). Stepping back to reflect on this new data policy framework, we argue that a FAIR data policy framework should be framed within a higher level of abstraction and reflect a deeper consideration of how FAIR is made an integral part of policy, which would then steer any model policy recommendation. We explore recommendations from the TFiR report and the FAIRsFAIR project to consider how a PaN data policy framework could incorporate FAIR effectively. As an outcome of this work, ExPaNDS responds with our own set of recommendations around FAIR data policy for PaN RIs (section 5).

We also set out key elements that PaN RIs should consider in relation to their data policies (section 6). These elements range from underlying general principles and the scope of the data policy, to more specific concerns such as data sharing, making data FAIR, data management

planning, and policy compliance and monitoring. It is not our intent to be prescriptive on any of these points. Instead, we emphasize that these are all elements around which RIs need to make policy choices, both in relation to the level of commitment RIs are prepared to make and also on the requirements they will place on users.

7.2 Key Messages

Several key messages emerge from our review of areas relevant to data policy frameworks for RIs:

1. National RIs work in a complex research data policy landscape. They need to consider not only national funder requirements, but also high level European policy directives. A review of the national research data policy landscapes of ExPaNDS partners indicates that these landscapes do reflect key European research data ambitions, for example, such as FAIR. However, despite this, harmonization of research data policy at the national level remains a work in progress for many countries. As such, in practice, RIs need to choose with which funders at the national level to most align their data policies. And the reality is that these funders are likely to be the ones most involved in funding the work undertaken at the facilities — not necessarily those who are most aligned with FAIR or any other particular initiative.

2. A harmonized approach to data policy across all PaN RIs is desirable, especially in relation to the user-experience given that scientists in the PaN community may move between facilities to undertake experiments. Certainly, RIs all face similar challenges around managing the complexity and volume of the data generated by their facilities, complying with ethical and legislative requirements, engaging with FAIR and the EOSC, and making meaningful and useful data policy. Landscaping surveys undertaken by both ExPaNDS and PaNOSC further reinforce this point: all PaN RIs do currently have data policies, and these tend to address issues that are of common concern, such as embargo periods, licensing, and the use of metadata and PIDs.

3. A data policy framework aims to provide a supporting structure around which a data policy can be built, and PaNOSC and ExPaNDS have taken different approaches. Working jointly with ExPaNDS, PaNOSC delivered a new data policy framework in May 2020. The approach of the PaNOSC framework has been to propose a generic PaN data policy, accompanied by detailed implementation notes, and thus provides a model policy for facilities to adapt. Based on initial feedback and discussion with ExPaNDS partners, we have taken the opportunity in the present deliverable to reflect on what a framework should be, what elements it should include, and how it should be presented. Thus we have framed the policy framework as a set of principles and options as guidance for the choices which RIs have to make in setting a FAIR data policy. We see these as complementary approaches which should converge on a common view.

4. There is a large body of work which has tried to capture and codify best practice around good data management and FAIR data practices. This is now being brought into the guidance emerging within the EOSC programme, for example, in the TFiR report and in

the FAIRsFAIR project. When we seek to apply these guidelines, in our case into RIs supporting PaN science, we need to carefully consider how they might apply into specific institutional contexts. PaN RIs are unusual institutions, representing neither funders nor higher education institutions performing research which are the frequent subject of these recommendations. PaN RIs have specific established infrastructures and practices and are sensitive to the needs of their user communities. These factors must be recognized in the interpretation of general guidelines into the ExPaNDS context.

5. Taking into consideration all the inputs from the previous sections, we can draw a set of general principles. These form the *elements around which RIs need to make policy choices*. In many respects, these reflect the existing best practices in current data policies and are also largely reflected in the PaNOSC data policy framework deliverable.

7.3 Next Steps

On the basis of key findings from this deliverable, ExPaNDS WP2 will proceed with several next steps:

1. Over the next twelve months, we will undertake an ExPaNDS partner consultation on the new PaNOSC data policy framework document to collect initial feedback from national RIs. We have already piloted two consultations (with ISIS and Elettra) and will report on these, along with the consultations with the remaining ExPaNDS partner facilities, in August 2021 as part of our final version data policy framework deliverable. Appendix C sets out the consultation aims, approach, and methods and presents the list of questions designed to frame the discussion at the consultation meetings. Based on our experience with the pilot consultations, we are considering adjusting this list of questions slightly to reflect feedback from colleagues.

2. In the coming months, we will continue with our ongoing work on developing an ExPaNDS glossary of terms relevant to the policy and framework activity we are undertaking in WP2. As section 6.2.1 highlights, definitions play a role in data policy; however, in relation to PaN data policy, as was evident during the development of the PaNOSC data policy framework, there is sometimes debate around what a particular term means and how exactly it should be defined. As a result, ExPaNDS has been devoting specific effort to the discussions needed to come to agreement on key definitions. To date, we have agreed nineteen terms and definitions. Appendix B presents these and also provides more detail on our glossary approach. While nineteen may not seem a large number, experience has shown us that it can take a significant amount of community discussion to agree even a single definition.

Definitions are often specific to a community, in ExPaNDS' case, to PaN RIs. This said, where a term is more widely used, for example, 'metadata' or 'raw data', it is important that community definitions reflect general understandings of the term. To this end, ExPaNDS WP2 is also liaising with the work of the EOSC Secretariat Glossary Working Group (WG). The WG recently released the first intermediate version of the EOSC

glossary.[104] ExPaNDS WP2 note, for example, that our definitions of both 'raw data' and 'metadata' (see Appendix B) echo closely the EOSC glossary definitions for these terms.

Coordinating with PaNOSC, ExPaNDS intends to present the results of steps 1 and 2 above in a final version deliverable focused on data policy: *D2.3 Final Data Policy Framework for Photon and Neutron RIs* (expected August 2021). We will also link our data policy framework activity to other aspects of ExPaNDS WP2 tasks, particularly tasks 2.2 (data management planning), 2.3 (mainstreaming of standards for data management), 2.4 (PID infrastructure), and 2.5 (quality assurance and certification schemes for data repositories). In turn, these tasks feed into and impact the tasks of other ExPaNDS work packages, especially WP3 and WP4.

---

[104] EOSC Glossary Interest Group (2020). First intermediate version of the EOSC glossary released. https://www.eoscsecretariat.eu/eosc-glossary

References

Alliance of German Science Organisations (2010). Principles for the handling of research data. https://doi.org/10.2312/ALLIANZOA.035

ANR (2013). Open Science policy. https://anr.fr/en/anrs-role-in-research/values-and-commitments/open-science/

ANR (2019). FLASH CALL Open Science: Research practices and open research data. https://anr.fr/en/call-for-proposals-details/call/flash-call-open-science-research-practices-and-open-research-data/

Ashton, A., Da Graca Ramos, S., Matthews, B. et al. (2019). ExPaNDS data landscaping survey. https://doi.org/10.5281/zenodo.3673811

Bangert, D., Hermans, E., van Horik, R. et al. (2019). Recommendations for services in a FAIR data ecosystem. https://doi.org/10.5281/zenodo.3585742

Battaglia, S., Beagrie, N., Cavalli, V. et al. (2019). EOSCpilot D3.6 final policy recommendations. https://www.eoscpilot.eu/content/d36-final-policy-recommendations

BBSRC (2016). Safeguarding good scientific practice. https://bbsrc.ukri.org/documents/safeguarding-good-scientific-practice/

BBSRC (2017). BBSRC data sharing policy. https://bbsrc.ukri.org/documents/data-sharing-policy-pdf/

Bicarregui, J., Matthews, B. and Schluenzen, F. (2015). PaNdata: Open data infrastructure for photon and neutron sources, *Synchrotron Radiation News*, 28:2, 30-35. https://doi.org/10.1080/08940886.2015.1013418

Cambridge Dictionary (2020). Framework. https://dictionary.cambridge.org/dictionary/english/framework

CNRS (2019). CNRS roadmap for Open Science. https://www.science-ouverte.cnrs.fr/wp-content/uploads/2019/11/CNRS_Roadmap_Open_Science_18nov2019.pdf

CNRS (n.d.). DMP OPIDoR. https://dmp.opidor.fr/

Committee for Open Science (2019). Recommendations for services in a FAIR data ecosystem. https://www.ouvrirlascience.fr/recommendations-for-services-in-a-fair-data-ecosystem-2/

Davidson, J., Grootveld, M., Whyte, A. et al. (2020). D3.3 policy enhancement recommendations. https://doi.org/10.5281/zenodo.3686901

Deutsche Forschungsgemeinschaft (1998). Proposals for safeguarding good scientific practice. https://web.archive.org/web/20110519131326/http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf

Deutsche Forschungsgemeinschaft (2015). Leitlinien zum Umgang mit Forschungsdaten. http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf

Deutsche Forschungsgemeinschaft (2019). Guidelines for safeguarding good research practice. Code of conduct. https://doi.org/10.5281/zenodo.3923602

EC (2017). Open Science Policy Platform recommendations. https://ec.europa.eu/research/openscience/pdf/integrated_advice_opspp_recommendations.pdf

EC (2018). Prompting an EOSC in practice. https://ec.europa.eu/info/sites/info/files/prompting_an_eosc_in_practice.pdf

EC (2020). Horizon 2020 online manual: Open access. https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm

EC (2020). Progress on Open Science: Towards a shared research knowledge system. Final report of the Open Science Policy Platform. https://ec.europa.eu/research/openscience/pdf/ec_rtd_ospp-final-report.pdf#view=fit&pagemode=none

EC Expert Group on FAIR data (2018). Turning FAIR into reality. https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf

EOSC FAIR Working Group and EOSC Architecture Working Group (2020). Second draft persistent identifier (PID) policy for the European Open Science Cloud (EOSC). https://doi.org/10.5281/zenodo.3780423

EOSC Glossary Interest Group (2020). First intermediate version of the EOSC glossary released. https://www.eoscsecretariat.eu/eosc-glossary

EOSC RoP Executive Board Working Group (2020). European Open Science Cloud rules of participation version 0.2. https://repository.eoscsecretariat.eu/index.php/s/QWd7tZ7xSWJsesn#pdfviewer

EOSC-hub (2019). D2.8 - First data policy recommendations. https://www.eosc-hub.eu/deliverable/d28-first-data-policy-recommendations-approved-ec

EOSCpilot (2019). EOSCpilot's 9 recommendations for EOSC policy.
https://eoscpilot.eu/news/eoscpilots-9-recommendations-eosc-policy

EPSRC (2011). EPSRC policy framework on research data.
https://epsrc.ukri.org/about/standards/researchdata/

ERC (2017). Guidelines on implementation of open access to scientific publications and research data. https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf

EU Presidency, hosted by the Government of the Netherlands (2016). Amsterdam call for action on Open Science. file:///C:/Users/zrs94363/Downloads/amsterdam-call-for-action-on-open-science.pdf

ExPaNDS (2018). ExPaNDS European Open Science Cloud Photon and Neutron Data Services [proposal].

Fondazione Bruno Kessler (2018). Piano Strategico. https://www.fbk.eu/wp-content/uploads/2018/12/Piano-Strategico-FBK_2018.11.16.pdf

FORMAS (2020). Open access to research results and data. https://www.rj.se/en/About-RJ/Work-methods/guidelines-for-open-access/data-management-plan-for-rj/

Gagey, B. (ed.) (2018). SOLEIL data management policy. https://www.synchrotron-soleil.fr/en/file/11308/

Gobierno de España (2011). Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación. https://www.boe.es/eli/es/l/2011/06/01/14/con

Gobierno de España (2011). Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación. Artículo 37. Difusión en acceso abierto.
https://www.boe.es/eli/es/l/2011/06/01/14/con#a37

Gobierno de España (2017). Plan Estatal de Investigación Científica y Técnica y de Innovación 2017-2020.
https://www.ciencia.gob.es/portal/site/MICINN/menuitem.7eeac5cd345b4f34f09dfd1001432ea0/?vgnextoid=83b192b9036c2210VgnVCM1000001d04140aRCRD

Gobierno de España (2017). Plan Estatal de Investigación Científica y Técnica y de Innovación 2017-2020.
https://www.ciencia.gob.es/portal/site/MICINN/menuitem.7eeac5cd345b4f34f09dfd1001432ea0/?vgnextoid=83b192b9036c2210VgnVCM1000001d04140aRCRD

Gobierno de España (various). Planes Nacionales I+D+i Anteriores.
https://www.ciencia.gob.es/portal/site/MICINN/menuitem.29451c2ac1391f1febebed1001432ea0/?vgnextoid=79dbec05f2a7d210VgnVCM1000001d04140aRCRD

Gotz, A., Perrin, J., Fanghor, H. et al. (2020). PaNOSC data policy framework.
https://doi.org/10.5281/zenodo.3862701

Gotz, A., Perrin, J., Fanghor, H. et al. (2020). PaNOSC FAIR research data policy framework [landing page in Zenodo]. https://zenodo.org/record/3862701

Higher Education Funding Council for England, Research Councils UK, Universities UK, and Wellcome (2016). Concordat on open research data.
https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf/

Hodson, S., and Molloy, L. (2015). Current best practice for research data management policies. http://doi.org/10.5281/zenodo.27872

Mack, L. and Papadopoulou, E. (2018). EOSCpilot D3.4 Open Science policy registry.
https://eoscpilot.eu/content/d34-open-science-policy-registry

Ministère de l'Enseignement Supérieur de la Recherche et de l'Innovation (2018). National Plan for Open Science. https://cache.media.enseignementsup-recherche.gouv.fr/file/Recherche/50/1/SO_A4_2018_EN_01_leger_982501.pdf

NERC (2019). NERC data policy – guidance notes.
https://nerc.ukri.org/research/sites/data/policy/datapolicy-guidance/

Open Research Data Taskforce (2018). Realising the potential: Final report of the Open Research Data Taskforce.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/775006/Realising-the-potential-ORDTF-July-2018.pdf

Open Research Data Taskforce with Michael Jubb (2017). Research data infrastructures in the UK landscape report. https://www.universitiesuk.ac.uk/policy-and-analysis/research-policy/open-science/Documents/ORDTF%20report%20nr%201%20final%2030%2006%202017.pdf

PaN-data Europe (2011). Common policy framework on scientific data.
https://doi.org/10.5281/zenodo.3738497

PaNOSC (2019). WP2 data policy and stewardship. https://github.com/panosc-eu/panosc/tree/master/Work%20Packages/WP2%20Data%20Policy%20and%20stewardship

PaNOSC (2018). The Photon and Neutron Open Science Cloud (PaNOSC) [proposal].

République Française (2016). LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique. https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000033202746&categorieLien=id

Riksbankens Jubileumsfond (2019). Data management plan for RJ. https://www.rj.se/en/About-RJ/Work-methods/guidelines-for-open-access/data-management-plan-for-rj/

Riksbankens Jubileumsfond (2019). Guidelines for open access. https://www.rj.se/en/About-RJ/Work-methods/guidelines-for-open-access/

SOLEIL (2018). SOLEIL data management policy. https://www.synchrotron-soleil.fr/en/files/soleil-data-management-policy

SPARC Europe, in collaboration with the Digital Curation Centre (2017). An analysis of open data and Open Science policies in Europe, May 2017. https://sparceurope.org/new-sparc-europe-report-analyses-open-data-open-science-policies-europe/

STFC (2016). Scientific data policy. https://stfc.ukri.org/about-us/our-purpose-and-priorities/requesting-information-from-uk-research-and-innovation/scientific-data-policy/

Swedish Research Council (2019). Producing a data management plan. https://www.vr.se/english/applying-for-funding/requirements-terms-and-conditions/producing-a-data-management-plan.html

Swiss National Science Foundation (n.d.). SNSF policy on open research data. http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx#SNSF%20policy%20on%20Open%20Research%20Data

Swiss National Science Foundation (n.d.). Open research data. http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx

The Royal Society (2012). Science as an open enterprise: Open data for Open Science. https://royalsociety.org/~/media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

UKRI (2018). Guidance on best practice in the management of research data. https://www.ukri.org/files/legacy/documents/rcukcommonprinciplesondatapolicy-pdf/

UKRI (2020). Common principles on data policy. https://www.ukri.org/funding/information-for-award-holders/data-policy/common-principles-on-data-policy/

Vetenskapsrådet (2018). Kriterier för FAIR forskningsdata. https://s3.amazonaws.com/helpscout.net/docs/assets/5673e406c6979143615582d5/attachments/5c5adc74042863543ccca279/Kriterier-FAIR-forskningsdata_VR_2018.pdf

von der Heyde, M. (2019). Open research data: Landscape and cost analysis of data repositories currently used by the Swiss research community, and requirements for the future [Report to the SNSF]. https://doi.org/10.5281/zenodo.2643460

Wellcome (2017). Data, software and materials management and sharing policy. https://wellcome.ac.uk/grant-funding/guidance/data-software-materials-management-and-sharing-policy

Wikipedia (2020). Policy framework. https://en.wikipedia.org/wiki/Policy_framework

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 3, 160018. https://doi.org/10.1038/sdata.2016.1

Appendix A: PaNOSC Data Policy Framework

The PaNOSC Data Policy Framework document is available online:
https://doi.org/10.5281/zenodo.3862701

PaNOSC
Photon and Neutron Open Science Cloud
H2020-INFRAEOSC-04-2018
Grant Agreement Number: 823852
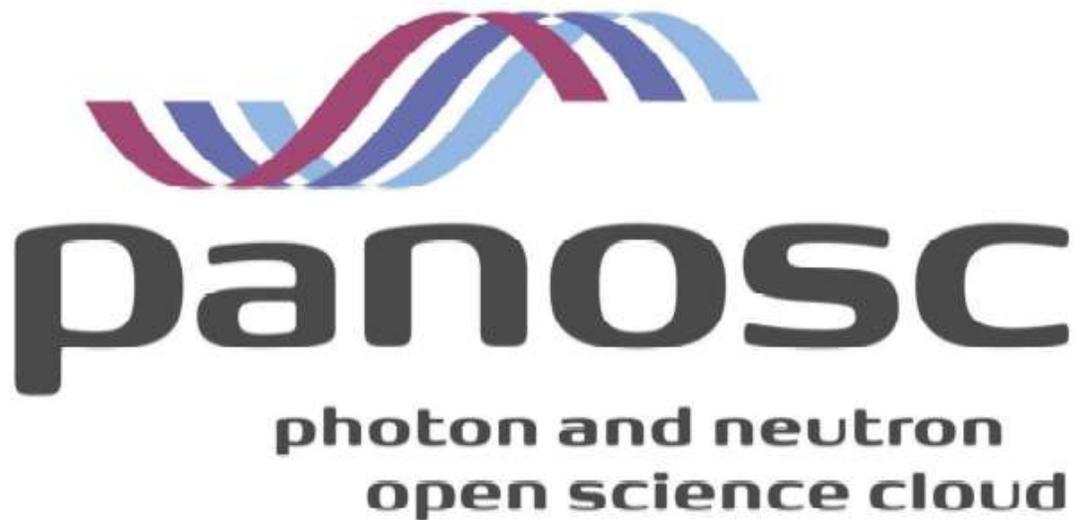
Deliverable: D2.1 - PaNOSC data policy framework

Appendix B: Work to Date on the Glossary

In relation to its reconsideration of the 'Definitions' section of the PaNOSC Data Policy Framework, ExPaNDS work package 2 has to date agreed nineteen terms and definitions for inclusion in the ExPaNDS *FAIR Data Management at National Research Infrastructures Common Glossary of Terms*:

General Terms

1. The term **public research** refers to publicly funded research that has been allocated access to facility resources through a peer-review process or through instrument time allocated to facility employees and their collaborators and which is intended to lead to publication(s). (WP2)
2. The term **proprietary research** refers to research done through purchased (commercial) access to the research facility, subject to local jurisdiction. (WP2)
3. The term **open access** means digital, online, free of charge, and free of most copyright and licensing restrictions.
4. The term **authentication** refers to the process of verifying the identity of a user, process, or device, often as a prerequisite to allowing access to resources in an information system.
5. The term **authorization** refers to the process of verifying that a requested action or service is approved for a specific entity.
6. The term **long term preservation** means continued access to digital materials, or at least to the information contained in them, indefinitely.
7. The term **medium term preservation** means continued access to digital materials beyond changes in technology for a defined period of time to be determined by the facility, but not indefinitely.
8. The term **embargo period** refers to a period of time during which access to meta/data is restricted to those with appropriate authorization, normally the experimental team.
9. The term **best practice** refers to a technique or methodology that, through experience and research, has proven to reliably lead to a desired result.
10. The term **research infrastructure** (RI) refers to a publicly-funded European or national photon and/or neutron facility.
11. The term **facility** refers to the part of the research infrastructure aggregating instruments, equipment, hardware, software and other resources necessary to perform experiments and to curate associated meta/data.
12. The term **Users** refers to those making use of the research infrastructure facility, including scientists, engineers and students from academia, research councils and charitable institutions, researchers from commercial and non-commercial organizations, and facility staff.
13. The term **internal user** refers to a User who has authorization from the Research Infrastructure or the Principal Investigator to perform operations with the data (or operations that can affect the data).
14. The term **external user** refers to an authenticated user who is not allowed by the Research Infrastructure or Principal Investigator to perform operations with the data.
15. The **Principal Investigator** is the User who is responsible for control and stewardship over the entire Research including but not limited to: the experiment, the Experimental Team and

other involved persons, software, objectives, access rights, ethics, data and metadata processing, and legal and other communications with the Research Infrastructure.

16. The term **experimental team** includes the persons authenticated and authorized to use the facility instruments and to perform all operations needed to collect, support, curate and process raw data and all associated data, including metadata.

17. The term **metadata** refers to data about other data that define and describe the characteristics of that other data, including, for example, their content, structure, provenance and/or the rights attached to the data. There are many types of metadata, e.g. administrative, descriptive, structural, etc.

18. The term **raw data** refers to the data collected from experiments performed on facility instruments that are considered as the source for any further analysis / processing.

19. The term **reference data** refers to data collected about instrument alignment or during an instrument calibration run, often against a reference sample.

As section 7 (next steps) highlights, work on the glossary is ongoing. Ninety-nine terms and definitions remain out of an initial list of 214. Note that both of these numbers include 'duplicate terms', for example, where the term is the same but the definition differs.

Appendix C: Consultation Aims, Methods, and Framing Questions

PaNOSC data policy consultation with ExPaNDS partners

Consultation aims:

1. To offer an opportunity to capture any changes in the current data policy situation across ExPaNDS partners (i.e. since the Dec 2019 Landscape survey and analysis of data policy and practise for ExPaNDS partners)
2. To identify current data policy elements that support/hinder FAIR data practice
3. To identify similarities/differences (and reasons for these) between ExPaNDS partners' data policies
4. To compare ExPaNDS partners' current data policies with the proposed PaNOSC data policy
5. To capture information on 'outside FAIR' matters (e.g. data retention/deletion, GDPR) that could have in/direct effects on FAIR data policy
6. To understand what data policy aspects seem to have the most influence on scientists'/users' practice
7. To explore which aspects of data policy are most likely to positively/negatively influence scientist/user behaviour

Method and approach

To support ExPaNDS WP2 deliverable *D2.3: Final data policy framework for Photon and Neutron RIs* (Aug 2021), we intend to consult with all ExPaNDS partners to gather feedback on the content, FAIRness, and implications (e.g. for the facility, for users, for practice) of the PaNOSC Data Policy Framework. The aims of the consultation are set out in more detail below.

The consultation process will take place over the next twelve months and will involve a meeting with each facility, followed up by formal write up of the meeting discussion, reviewed and checked for accuracy by the facilities. The meetings will take the format of semi-structured interviews: a set of framing questions is designed to help the meeting conversations to cover similar ground.

Questions to frame the meeting discussions:

- Policy sections ➜ Do you agree with the sections of the PaNOSC Data Policy, e.g. Definitions? General principles?
- Specific policy elements ➜ We would like to draw your attention to particular elements of the policy:
  - Use of Persistent Identifiers
  - Differentiation between raw and processed data
  - Embargos and their length

ExPaNDS

- Are these elements part of your current data policy/practice? Do you foresee that these elements will pose any particular challenges for your facility?
- Impact of FAIR on the policy ➜ Do you feel that the policy engages with all aspects of FAIR adequately? If you compare the PaNOSC policy with your current data policy, can you distinguish where changes may have been made to accommodate FAIR principles?
- Consequences of legal frameworks ➜ Are the policy elements focused on legal frameworks (e.g. GDPR, licensing) sufficient and appropriate?
- Implementing the policy in practice (bear in mind that the policy is a 'best practice' document, i.e. some aspects may yet be aspirational for facilities):
    - Are there any aspects of the policy with which you would disagree or that you consider unhelpful?
    - Are there any aspects that would form a barrier to users or that you would find difficult/impossible to implement?
    - Are there any tools/roles that you would need in place to implement the policy?
    - Based on your understanding of the policy, are there elements that you consider mandatory/optional?
- Missing elements ➜ Is there anything missing from the policy that you would like to see included?