Proposal name	**PaNOSC**
Proposal full name	**Photon and Neutron Open Science Cloud**
H2020 Call	**INFRAEOSC-04-2018**
Coordinator	**Andrew Götz (andy.gotz@esrf.fr)**
Coordinating organisation	**ESRF**
Written by	**Florian Gliksohn, Andy Götz, Alessandro Olivo, Dario Roccella, Jon Taylor, Jean-Francois Perrin**
Deliverable Name	**Internal document for WP 2, Task 2.1**
Reviewed by	
Reviewed by	

List of participants

Participant No	Participant organisation name	Country
1	European Synchrotron Radiation Facility (ESRF)	France
2	Institut Laue-Langevin (ILL)	France
3	European XFEL (XFEL.EU)	Germany
4	The European Spallation Source (ESS)	Sweden
5	Extreme Light Infrastructure Delivery Consortium (ELI-DC)	Belgium
6	Central European Research Infrastructure Consortium (CERIC-ERIC)	Italy
7	EGI Foundation (EGI.eu)	Netherlands

Contents

General questions

1. **Who was the main driver within your Organization to adopt a Data Policy (management, IT, researchers, etc.?)**

- **ESRF**:

  The main drivers were the Directors and Heads of Division. They were assisted by the two data policy experts (R.Dimper + A.Götz) who had participated in the writing of the PaNdata data policy framework. Together with other IT experts they estimated the costs of implementing the data policy and provided background information on what other sites (ILL and ISIS) had done to implement their DP. The proposed DP was presented to the ESRF scientists before being presented and discussed by the SAC (Scientific Advisory Council) and finally endorsed by the Council. Some ESRF scientists were openly in favour of the DP. Most of them had no strong opinion for or against while only one or two expressed strong doubts about Open Data policy either because they thought it not technically / financially feasible to store all raw data, or they doubted the usefulness of the DP.

- **ILL:**

  The Data Policy project was initiated by the IT in the scope of the PaNData-Europe project in 2008 and was largely supported by the Directors during the 3 years of internal discussion that were necessary to obtain an acceptable consensus for all parties. It was officially adopted in Sept 2011 after validation by our Scientific Council.

  A working group (composed of instrument control, user office, scientific computing, IT and representatives of scientific groups) was then set up to discuss, steer and monitor its implementation. This working group has also the responsibility to propose policy modifications. The first policy revision, driven by this group, took place in July 2017.

- **EuXFEL:**

  The introduction of Scientific Data Policy was driven by the IT and Data Management group with the strong support of Scientific Director in their division. The draft of the policy was presented to the Management Board where main directions were endorsed. Then it was sent for comments to the leading Instrument Scientists and technical group leaders and also legal aspects were checked. After corrections were implemented the final draft was accepted by Management Board, then presented to Scientific Advisory Committee and finally accepted by European XFEL Council on 30 June 2017 which was last day before entering operation phase of the facility

- **ESS**:

  The main drivers for ESS to develop and formally agree a policy for scientific data early in the construction phase was the head of the Data Management and Software Centre (Jonathan Taylor) The Head of the Data Management Group (Tobias Richter) and the Director for science (Andreas Schreyer). The process to develop the policy and gain council approval was undertaken in 2017. The process took approximately 12 months. There was considerable support from the ESS Director General John Wormersley.

- **CERIC-ERIC**:

The Executive Director of CERIC-ERIC.

- **ELI:**

The ELI Facilities will start to enter into initial operations in 2019 and gradually open up their capabilities to the user community in the coming years. They will be operated by an ELI ERIC in the process of being established. A Data Policy will be adopted by the General Assembly of the ELI ERIC shortly after establishment. There is therefore an institutional push to adopt such a policy, even more so since the data policy is mentioned as one of the statutory data policies in the draft version of the ELI ERIC Statutes. There is also a strong expectation and request within the organization for such policy from scientists and staff in charge of data management as clear data policy principles are needed as guiding elements in the on-going communication with prospective users and design and implementation of the data infrastructure. PaNOSC from that standpoint adds to the pressure in a positive way, creating a favourable environment for ELI to develop the policy. It is foreseen that a working group, involving IT and data management staff, scientists, staff in charge of instruments) will be involved, as well as the Scientific Advisory Committee prior to approval by the ELI ERIC General Assembly.

2. **Which was the main reason/benefits for adopting a Data Policy?**

- **ESRF**:

The reasons the ESRF adopted a DP are multiple. They were motivated by the fact that the ESRF (like other RIs) produces huge quantities of data which need to be managed and curated in order to provide services for data and allow scientists to profit from them fully. Without a DP many fundamental issues like ownership, embargo, sharing, archiving, open data etc. were not possible. The DP allowed us to define these. The DP was motivated by the need to define and collect metadata and raw data in well defined formats. The changing scientific publishing landscape which requires data to be made available and citable was an additional motivation for the DP. Without the DP the burden is on the users to store and curate the data. Another motivation for the DP was the growing volumes of data produced which makes it more and more difficult to export data and therefore requiring them to be kept on site. This would be difficult without a DP because the ownership is otherwise undefined. A strong motivation of the DP was to make data openly available after an embargo period to increase its usefulness.

- **ILL:**

The initial driver was the reuse of data: ILL was archiving experimental data since its first run in 1973, the IT was taking care to migrate the data files for every technology change (e.g. from IBM tapes to LTO), the cost of preservation as always been relatively important for a limited number of requests to access legacy data.

As soon as we started the implementation, other excellent motivations for this work, such as the improvement of the service quality for our users, became obvious.

- **EuXFEL:**

The main reasons to introduce the Scientific Data Policy was to impose coherent approach to data management across all instruments and to allow defining obligations and rights with respect to data for the facility and facility users who had to accept it upon the registration in the User Portal. The policy defined the basis for implementation of data management services and it turned out to be extremely useful.

- **ESS**:

  The key reason to develop a policy for scientific data early in the ESS construction phase (some 5 years before beam on target) was to set the policy framework in place to assist future developments in scientific computing. I.e. One can use the existence of a data policy as a lever for developing scientific computing in a way that is commensurate with empowering open data for the ESS scientific community. An important but certainly not the only reason for the ESS data policy was to maintain compliance with the core EU ambition for Open data.

- **CERIC-ERIC**:

  The data policy is necessary to be compliant with H2020 funding. Also, CERIC committed to the ORDP for the data generated in the ACCELERATE project. CERIC believes that open data will benefit researchers and institutions, increasing the visibility, enhancing collaborations and allowing a better use of resources in general.

- **ELI**:

  For an emerging infrastructure like ELI entering into operations, there is simply an expectation, both from users and funding agencies, that experimental data will be made available and comply with the FAIR principles. In other terms, not having a data policy is considered a failure. In this context, the data policy is a necessary framework, because it addresses a number of critical issues that organise the relation with visiting users and the user community in general when it comes to data (ownership, embargo, access to data, storage, curation, etc.). It is also an internal driver pushing ELI to look at data also in terms of services leveraging the value of experimental data. Naturally, this perspective has a direct impact on technological choices made around data management.

3. **To write yours, did you use a template, a management platform or an existing policy? If yes, please specify.**

- **ESRF**:

  ESRF based their data policy on the PaNdata data policy framework. As ESRF had actively participated in the writing of the PaNdata DP (ESRF was WP leader for the deliverable) it had strong knowledge of the contents. The PaNdata DP was then modified based on the input from the ESRF management, discussions with scientists and SAC.

- **ILL:**

  The work done during during the PaNData EU project with our colleagues from the other EU analytical facilities was the basis for the internal discussion. With the strong competition that exists between these user facilities to attract the best scientific team and the fear that existed, at this time, to lose some users because of such “Open Data” regulation, the fact that it was a common and de facto standard framework was extremely important.

- **EuXFEL:**

European XFEL based the scientific data policy on PaNdata recommendations and followed majority of modifications made by ESRF and ILL.

- **ESS**:

  The ESS policy for scientific data is based upon the proforma policy created by the PaNdata project. An initial comparison map was made of existing data policies from european research infrastructure This document and the PaNdata proforma was used to develop a policy for ESS.

- **CERIC-ERIC**:

  Yes. We used the PaN-data data policy guidelines and incorporated elements of other existing policies (ALBA synchrotron, Elettra Sincrotrone Trieste, EuXFEL, ESS, ESRF, ILL).

- **ELI**:

  ELI will use the PaN-data guidelines as initial reference, analyse the data policies collected within the framework of PaNOSC's WP2 and build on the work of this work package.

4. **Were all departments/persons affected by the policy consulted?**

- **ESRF**:

  The effort of implementing the DP at the ESRF involved the beamline scientists, the beamline control system staff, the data analysis scientists, the IT staff. Two staff positions are dedicated to data management to implement the data catalogue and manage the data curation.

- **ILL:**

  A large part of the organisation was involved in the discussions, and more specially the Direction, the instrument responsibles, the User office, IT, the legal office and the scientific council. The internal discussions took 3 years to reach an agreement.

- **EuXFEL:**

  Various scientific and technical groups were consulted after the main directions were accepted by Management Board. Substantial support was given by the legal office.

- **ESS**:

  Implementation of the data policy from a technical perspective falls within the remit if the data management and software centre. DMSC was the driver for the development of the policy details DMSC staff were involved in development from the initial stage. The DMSC scientific and technical advisory panel were consulted for advice. For broader stakeholder engagement the policy was presented to the ESS scientific advisory council for discussion before being presented to the ESS council for approval

- **CERIC-ERIC**:

CERIC-ERIC is a consortium offering access to 9 facilities in Europe. Our Partner Facilities were consulted and the final word will be given in June 2019 by the General Assembly. Users were not consulted so far, this may happen at a later stage.

- **ELI**:

Policies, within the context of the future management system of ELI ERIC, are short high-level documents. It is expected that, as such, the data policy will be complemented by more detailed regulatory documents describing the practicalities of the principles enshrined in the policy document. The policy document will be approved by the General Assembly and will contain the core principles and strategic objectives of ELI in terms of data management and access to data. It is expected that the Scientific Advisory Committee of ELI ERIC will be invited to comment on the proposal. Complementary management and regulatory documents will be developed with the likely involvement of researchers and operators (scientific directorate, beamline scientists), of the staff involved in control systems and IT management and, possibly, of user representatives.

5. **Before the adoption of your current data policy, was its compliance with legal and regulatory aspects assessed?**

- **ESRF**:

The ESRF DP was submitted to the ESRF legal counsel for checking. She did not make any changes. This was before GDPR.

- **ILL:**

The initial version of the ILL data policy and especially the question of the protection of data was discussed with a layer specialist of the IPR questions and like for any policy was checked by our legal office.

- **EuXFEL:**

Yes. Amendments were introduced based on the Legal Office advise, especially personal data protection and liability aspects.

- **ESS**:

The ESS legal team were involved in the development of the policy text to ensure compliance with legislation and latterly GDPR.

- **CERIC-ERIC**:

No, it wasn’t checked by a lawyer so far.

- **ELI**:

Yes, it is planned that a legal assessment will be performed.

6. **Does your Policy address all the data produced and related metadata? If not, which kind of data does it exclude (Personal data, sensitive data, etc.)?**

- **ESRF**:

  The ESRF DP only excludes data produced by proprietary (commercial) research. All data from public research, including the CRG beamlines, are covered by the ESRF DP. Analysed data is currently excluded from the DP.

- **ILL:**

  The first revision of the data policy, published in July 2017, also address reduced data and more generally all scientific data resulting from the analysis of the raw data and stored on the ILL IT infrastructure. This revision also take into account data generated from CRG instruments (Collaborating Research Groups instruments are instruments managed on ILL beamlines by third party organisation https://www.ill.eu/fr/users-en/instruments/crgs/). Only data resulting from proprietary research are excluded from this text.

- **EuXFEL:**

  The Scientific Data Policy excludes data produced by proprietary research. It applies to all scientific data generated at European XFEL instruments including those contributed by third party organizations and User Consortia.

- **ESS**:

  The ESS policy specifically excludes data from proprietary use of ESS beamlines / instruments. Meta data that constitutes sensitive data is not implicitly included or excluded. From the ESS perspective this aspect falls within other policies set by the organisation.

- **CERIC-ERIC**:

  Our policy can be applied to all the data produced and relative metadata. Personal or sensitive data will not be disclosed.

- **ELI**:

  Similar to other PaNOSC partner organisations, it is planned that our policy will apply to all data generated by ELI instruments and related metadata, It will address data from proprietary research and sensitive data, that will not be disclosed.

7. **Was the personnel of your organization trained on how to apply the Data Policy?**

- **ESRF**:

  ESRF dedicated two positions to data management. The DP were presented to staff but (so far) users are not trained how to use the data portal.

- **ILL:**

  The personnel was not trained, but support exists (data@ill.eu) to reply to data management related questions.

- **EuXFEL:**

  The Scientific Data Policy is the first point in the data acquisition and data management training provided to instrument scientists.

- **ESS**:

  DMSC has specific positions for data management. It is the intention of ESS to train users and staff in certain aspects of data management and aspects that directly pertain to our data policy (such as the SciCat data catalogue)

- **CERIC-ERIC**:

- **ELI:**

  No such training has yet been planned, but including data policy aspects in the compulsory user training is being considered.

8. **Does the policy include a review cycle?**

- **ESRF**:

  Not yet. This is one thing we would like to introduce as part of PaNOSC.

- **ILL:**

  The current policy does not include a formal review process or cycle**.** Nevertheless it has already been reviewed when it has clearly appeared to the stakeholders that it was necessary. This review was simplified by the fact that the reationales were well understood by all the parties, this will not necessarily be the case with a defined time scale review.

- **EuXFEL:**

  The Policy does not define review cycles. However, it allows within certain limits for modifications of storage periods on different levels of storage systems according to the experience and available resources.

- **ESS**:

  No specific timescale is included in the policy for review of the policy (and implementation in itself)

- **CERIC-ERIC**:

Not yet. We have just agreed on a final version that still needs to be adopted by the General Assembly. Training will be necessary during the implementation.

- **ELI**:

  Regular policy assessment and review are considered good management practice. A review cycle will therefore be proposed as part of the data policy submitted for approval to the ELI ERIC General Assembly.

9. **If you used a template or model, did you have a need to change some standard definitions? If yes, why?**

- **ESRF**:

  The definition of proprietary data was added to the PaNdata DP framework. We needed to add the definition of a session to the DP for DOIs.

- **ILL:**

  There was no major changes but some definitions had to be adapted to the "language" of the ILL and its users, as an example we use the wording "main proposer" instead of PI. The revision also adopted a less formal and more practical approach in order to be more easily understood by users and personnel. It was mainly rewritten by a scientist whereas the initial one was mainly written by managers.

- **EuXFEL:**

  The definition of various data types was introduced. A separate paragraph on warranty and liability regarding scientific data was introduced.

- **ESS**:

  The template was used to develop the overall concept of the data policy rather than a direct copy paste of text.

- **CERIC-ERIC**:

  We needed to include the definition of 'Partner Facility', due to the particular nature of CERIC-ERIC, see table "PaNOSC definitions for data Policy, cell D5).

- **ELI**:

  Not applicable.

10. **Have you defined one or more standard format for the raw data? If yes, which one/s?**

- **ESRF**:

HDF5 is the preferred data format but we are a long way from having all beamlines producing HDF5.

- **ILL:**

  The standard data format at ILL is NEXUS and it is in place for almost all instruments, exceptions only exist for instruments when it exists a strong community standard (e.g. root format for the nuclear physics community). Nevertheless this standard is not defined formally in the Data Policy document.

- **EuXFEL:**

  The policy document does not name any specific data format. In practice, the only format supported across the facility is HDF5.

- **ESS**:

  Nexus & HDF5 are preferred

- **CERIC-ERIC**:

  Yes, it was decided that HDF5 may fit the needs of all the partners.

- **ELI**:

  Not yet decided, but Nexus and HDF5 are preferred.

11. **Which were the considerations behind the choice of the embargo period?**

- **ESRF**:

  The ESRF embargo period of 3 years is based on the length of a PhD.

- **ILL:**

  The ILL formal embargo period of 3 years is based on the standard length of a PhD. There is also a possible extension of 2 years when no one is requesting access to the data, this period came from the discussion that took place in 2010 and was put in place to avoid having to face too many extension requests from the users. In practise, after 7 years of implementation only 1 request was received by the scientific director. It is difficult to know if they very limited number of extension requests is due to this mechanism, the feedback of the other facilities that did not implement it will be extremely interesting.

- **EuXFEL:**

  Length of a PhD project and following the recommendation of PaNdata.

- **ESS**:

  3 years embargo was chosen to match the average length of a PhD project.

- **CERIC-ERIC**:

We chose 3 years of embargo period, which is the standard duration for a PhD degree, and it is a reasonable period in which all data should have led to a publication.

- **ELI**:

  3-year embargo currently being considered. It is considered to be a reasonable period of time and is the average duration of a PhD.

12. **Was your data policy reviewed since it was adopted? If yes, how was/were the revised version/s improved, in comparison to the first one adopted?**

- **ESRF**:

  The ESRF DP has not been revised since the first adoption in 2016. It will be revised during the PaNOSC project.

- **ILL:**

  The Data policy was reviewed in 2017 (see previous sections), the main drivers were to :

  1) Handle the reduced data and other derived data in preparation of the set up of data analysis services.
  2) Take into account the CRG instruments
  3) Create a more accessible text for the users (too few people had a general understanding of the policy). For instance, the term PID was replaced by DOI.

- **EuXFEL:**

  The Scientific Data Policy has not been reviewed yet. There is ongoing attempt to redefine retention periods of different data types.

- **ESS**:

  Our DP has not been reviewed since adoption by council in 2018

- **CERIC-ERIC**:

  It has not been adopted by our General Assembly yet.

- **ELI**:

  ELI ERIC's DP has not been adopted yet.

13. **Do you allow the embargo to be extended and how do you manage this?**

- **ESRF**:

The ESRF DP allows the embargo period to be extended by the PI (Principal Investigator) on demand to the ESRF Scientific Directors. We have had one request so far but we have not defined the workflow for the implementation yet.

- **ILL:**

  In the current policy, the non disclosure period is extended to 5 years is no request has been received to access the data.

- **EuXFEL:**

  Any PI that wishes to extend the embargo period might submit a written request, specifying the reasons for the proposed prolongation, to the management board of European XFEL GmbH, which decides on the request at its own discretion. In exceptional circumstances, data can be made openly accessible during the embargo period if the PI informs the European XFEL GmbH to do so and subject to its own discretion.

- **ESS**:

  Yes by application to ESS

- **CERIC-ERIC:**

- **ELI:**

  It is not established if such option will be regulated in the Data Policy.

14. **What data services do you provide as part of the DP?**

- **ESRF**:

  The implementation of the DP provides the following data services: (1) well defined metadata in hdf5 file and in metadata catalogue, DOI for sessions + on demand, raw data long term archiving, web portal (https://data.esrf.fr), search engine for metadata, download service. The following services will be added as part of PaNOSC - Jupyter notebooks, container service for data analysis programs, globus download service.

- **ILL:**

  The services that were put in place for the implementation of the Data policy are:
  - Better management of the data files and repository
  - Access to the data from any ILL computer directly available from the desktop of the users.
  - Data portal, with search engine and user self management of the access (ACLs), including termination of the embargo period.
  - Internet access to the data through SFTP service
  - Generation of DOIs
  - Elogbook available on the instruments

- **EuXFEL:**

The following services deal with scientific data and are compliant with the policy:

- User Portal
- Metadata Catalogue service
- Automatic data acquisition service
- Data calibration service
- DOI generation
- Data processing on site
- Data archiving
- Data export services (ftp, globus)

- **ESS**:

  Specific services are not discussed in the DP. Services developed at ESS should be compliant with the policy but not defined in the policy itself - this allows some flexibility in service provision and alteration thereof. That is to say one can change downstream policies, procedures and rules without changing the governing policy itself. Which from our perspective is a more flexible approach.

- **CERIC-ERIC**:

- **ELI**:

  Similar to ESS, it is not anticipated that data management services will be discussed in the data policy, though data services will obviously have to be compliant with it. This being said, it is anticipated that the data policy proposal will be submitted to the General Assembly for approval together with some background information on the data services foreseen at ELI and an accompanying implementation roadmap.

Lessons Learned

This chapter will be completed once all the answers have been received and compared

- **ESRF**:

  This is a non-exhaustive list of lessons learned while working on the data policy at the ESRF:

  1. Implementing a Data Policy is a long process, especially when it is being implemented on an existing installation where the implementation has to be retro-fitted to the running installation and habits of scientists, engineers and scientists need to be changed. At ESRF we started working on the Data Policy 10 years ago (2009) with the writing of the PaNdata data policy framework. Implementation on beamlines started in 2016.
  2. Support of upper and top management are essential to get the data policy accepted and implemented
  3. An initial hurdle was the feasibility and cost of storing all raw data for the ESRF (hundreds of petabytes over 10 years) but we started off by discussing metadata and metadata policy. Because metadata is much less and there is no cost or feasibility issue it unblocked the discussion. It would be an option for all sites. Every site must have a metadata data policy. This also ensures that when there is budget to store raw data the data are already well organised with metadata.

4. A top down approach for deciding that we need a data policy reduced the discussion time. It was considered the right approach at the ESRF because policy is the prerogative of the management and not the scientists or users.
5. A majority of beamline scientists at the ESRF know nothing or very little about data management concepts like PID i.e. DOI. They needed explaining and will need more training.
6. There were no ready to use solutions with all features we needed for the metadata and data catalogue. We found icat was the closest to what we needed because it had a data model which mimicked our proposal and scientific data flow. However it suffered from lack of widespread adoption and an active user community. We invested in extending icat.
7. We identified the need for an electronic logbook in order for scientists who were not part of the original experiment to understand the experiment and produced data. We therefore developed hired someone to develop one. We did not find any of the Open Source solution which fitted our needs. One requirement was to have a modern web UI.
8. GDPR can be a pain because there is no clear directive for scientific institutes coming from the standards bodies. In fact the standards bodies and EU projects have not been of much practical value so far. We hope this will evolve with the EOSC.
9. The PaNdata policy is implicitly FAIR but we have not mentioned FAIR explicitly in our data policy. We need to do this but how is still an open question. We are working on this in PaNOSC.
10. We did not find any practical guidance on how to present landing pages for DOIs. We use a dynamic web page built out of the datacite metadata. We are not sure this is the best solution. Sometimes datacite is down. Dynamic pages are not indexed as well. Datacite search engines are not user friendly. Here is an example of our landing page: https://doi.esrf.fr/10.15151/ESRF-ES-135816585
11. It is essential to setup a contract with datacite in order to mint DOIs.

- ILL:

- EuXFEL:

- ESS:

- CERIC-ERIC:

Best Practices

- PAN Data Best Practices:

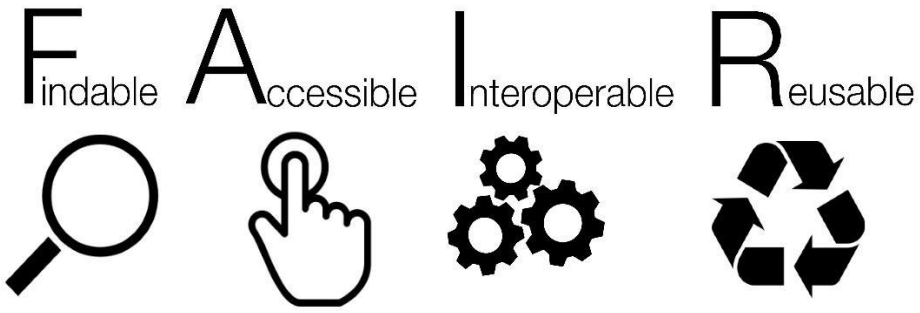
The experimental team is encouraged to ensure that experiments metadata are as complete as possible, as this will enhance the possibilities for them to search for, retrieve and interpret their own data in the future.
Each facility undertakes to provide means for the capture of such metadata items that are not automatically captured by an instrument, in order to facilitate recording the fullest possible description of the raw data.
Researchers who aim to carry out analyses of raw data and metadata which are openly accessible should, where possible, contact the original Principal Investigator to inform them and suggest a collaboration if appropriate.
Researchers must acknowledge the source of the data and cite its unique identifier and any publications linked to the same raw data.
Principal Investigators and researchers who carry out analyses of raw data and metadata are encouraged to link the results of these analyses with the raw data / metadata using the facilities provided by the on-line catalogue. Furthermore, they are encouraged to make such results openly accessible.

Open Questions

1. Do we need a specified chapter regarding the ownership of the data? Is it enough to define this in the data policy?
2. Is it necessary to define "process data"? Maybe we could consider them as results? Otherwise it is needed to well-define the concept to avoid misunderstanding
3. Could make sense to introduce concept of "experiment outside the proposal system"?
4. Is it really necessary to define a subset of data and metadata types of data? (as calibrated, calibration, alignment etc.)?
5. Could make sense well-define who is a registered user avoiding misunderstanding?
6. Is it necessary to define group and sub-group of authorized experiment participants (as proposal managers, proposal team, experiment team)?
7. Is it really necessary to differentiate between peer-reviewed and in-house experiments?
8. Maybe we could introduce personal data in the definition section in order to recipe the GDPR?
9. Could we introduce the DOI definition in the definition section? This is a keyword in the FAIR principles.
10. Should we introduce the definition of curation?

Appendix 1 - Definition of FAIR

In 2016, the 'FAIR Guiding Principles for scientific data management and stewardship[1]' were published in Scientific Data. The authors intended to provide guidelines to improve the findability, accessibility, interoperability, and reuse of digital assets. The principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.



Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process.

F1. (Meta)data are assigned a globally unique and persistent identifier
F2. Data are described with rich metadata (defined by R1 below)
F3. Metadata clearly and explicitly include the identifier of the data they describe
F4. (Meta)data are registered or indexed in a searchable resource

Accessible

[1] https://www.nature.com/articles/sdata201618

Once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation.
A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
A1.1 The protocol is open, free, and universally implementable
A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
A2. Metadata are accessible, even when the data are no longer available

Interoperable
The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.
I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (Meta)data use vocabularies that follow FAIR principles
I3. (Meta)data include qualified references to other (meta)data

Reusable
The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.
R1. Meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (Meta)data are released with a clear and accessible data usage license
R1.2. (Meta)data are associated with detailed provenance
R1.3. (Meta)data meet domain-relevant community standards

The principles refer to three types of entities: data (or any digital object), metadata (information about that digital object), and infrastructure. For instance, principle F4 defines that both metadata and data are registered or indexed in a searchable resource (the infrastructure component).

References

[1] https://doi.org/10.1038/sdata.2016.18 - “The FAIR Guiding Principles for scientific data management and stewardship” by MD.Wilkinson et. al.

[2] https://doi.org/10.1080/08940886.2019.1608119 - “ESRF Data Policy, Storage, and Services” by R. Dimper et. al.