Deliverables

Deliverable Number	**D2.8**
Deliverable Title	**1st update of DMP**
Lead Beneficiary	**PSI**
Type	**ORDP: Open Research Data Pilot**
Dissemination Level	**Public**
Due date of delivery	**Month 24**

Deliverable report

D2.8

Update of the Data Management Plan (DMP) based on Data Policies of the different facilities (month 24)

WP2

User tools for access and data management

deliverable report

PROJECT DETAILS

PROJECT ACRONYM
CALIPSOplus

GRANT AGREEMENT NO:
xxxxxxxx 730872

START DATE
01/05/2017

PROJECT TITLE
User tools for access and data management

CALL Horizon 2020-H2020-INFRAIA-2016-2017
INFRAIA-01-2016-2017: Convenient Access to Light Sources Open to Innovation, Science and to the World (CALIPSOplus)

DELIVERABLE DETAILS

WORK PACKAGE ID
WP2

WORK PACKAGE TITLE
WP2

WORK PACKAGE LEADER

Cecilia Blasetti (Elettra)

Mirjam van Daalen (PSI)

DELIVERABLE ID
D2.8

ESTIMATED INDICATIVE PERSON/MONTHS

0.1

EXPECTED DATE

Month 24 - 30/04/2019

DELIVERABLE TITLE

Update of the Data Management Plan (DMP)

DELIVERABLE DESCRIPTION

Data Management Plan DMP describing data curation

PERSON RESPONSIBLE FOR THE DELIVERABLE

Mirjam van Daalen (PSI)

NATURE

[X] R - Report

[ ] P - Prototype

[ ] D - Demonstrator

[ ] O - Other

DISSEMINATION LEVEL

[X] P - Public

[ ] PP - Restricted to other programme participants & EC:

_____

[ ] RE - Restricted to a group:

_____

[ ] CO - Confidential, only for members of the consortium

REPORT DETAILS

VERSION

1

STATUS

[ ] Template
[ ] Draft
[X] Final
[ ] Released to EC

DATE

30/04/2019

NUMBER OF PAGES

11

DELIVERABLE REPORT AUTHOR(S)

Mirjam van Daalen (PSI)

Valentina Piffer (PSI)

FOR MORE INFO PLEASE CONTACT

Mirjam van Daalen (PSI)
mirjam.vandaalen@psi.ch
0041/56/3105674

TABLE OF CONTENTS

PREMISE

The present deliverable is the first update of the Data Management plan for NA1, i.e. WP2 "User tools for access and data management". The existing data policies of facilities (mostly based on the framework created within PaNdata Europe Deliverable D2.1) were the basis for the first version of the CALIPSOplus Data Management Plan D2.2. Since the first version of the DMP D2.2, delivered in Month 6 of the project, the situation at the CALIPSOplus partners regarding the implementation of data policies changed and is described in deliverable 2.6 "Harmonisation of data policies".

The present Data Management Plan makes research data generated in the frame of the CALIPSOplus project findable, accessible, interoperable and reusable (FAIR). The DMP makes sure that the data are soundly managed.

The present DMP follows the guidelines on FAIR Data Management in Horizon 2020 given by the European Commission (Version 3.0; July 2016).

DESCRIPTION OF WORK

WP2: User tools for access and data management

Introduction

The document presents an updated overview on the Data Management Plan for data generated within the framework of the CALIPSOplus project. Data Management Plans are part of the Open Data Research Pilot of the European Commission. As the data policies are the basis for the DMP this document is strongly linked to D2.6 "Harmonisation of data policies roadmap", both documents are based on the survey done in March 2019 on data policies and data management plans amongst CALIPSOplus partners.

SURVEY amongst CALIPSOplus partners

The March 2019 survey was done in order to map the status of the data policies at the different partner facilities and to see how the facilities engage in research data management (i.e. metadata catalogues, data life cycle management, DMPs) on the national level. Further, questions regarding the integration of the federated identity management tool (UmbrellaID) were part of the survey. The survey was divided into two sections as follows:

SECTION RELATED TO INTERNAL DATA POLICY AND DATA MANAGEMENT PLAN

1. Does your facility have a data policy? Is it compatible with PanData and compliant with the GDPR?
2. Is your data policy aligned with the F.A.I.R. data principles?
3. What is the facility's strategy and / or Metadata management?
4. Does your facility have a Metadata catalogue? If so, what kind?
5. Have you validated and implemented a data management plan?
6. Please, describe the data life cycle process

SECTION RELATED TO INTEGRATION WITH NA1 WP2 Task 2.3

1. Is the facility's user office connected to Umbrella?
2. How is it available?

The outcome of the survey is listed in the tables on the following pages.

	CALIPSOplus partners	Country	Data Policy	F.A.I.R.	GDPR compliant	Implementation date	Metadata Catalogue	Federated identity Management UmbrellaID
1	HELMHOLTZ-ZENTRUM DRESDEN-ROSSENDORF EV Germany	Germany	PaNdata	✓	✓	01.05.2018	RODARE	✓
2	ANKARA UNIVERSITESI	Turkey	PaNdata	✓	✓	In preparation	Not yet in place	NO
3	AARHUS UNIVERSITET	Denmark	No data policy				Not planned yet	NO
4	ALBA - CONSORCIO PARA LA CONSTRUCCION EQUIPAMIENTO Y EXPLOTACION DEL LABORATORIO DE LUZ DE SINCROTRON	Spain	PaNdata	✓	✓	07.2017	iCat	✓
5	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE	France	No data policy				Not planned yet	NO
6	STIFTUNG DEUTSCHES ELEKTRONEN-SYNCHROTRON DESY	Germany	PaNdata	NO	✓	25.06.2015	Gamma Portal DESY	✓
7	DIAMOND LIGHT SOURCE LIMITED	United Kingdom	PaNdata	✓	✓	01.03.2019	iCat	✓
8	ELETTRA - SINCROTRONE TRIESTE SCPA	Italy	PaNdata	✓	✓	01.12.2014	Other	✓
9	EUROPEAN X-RAY FREE-ELECTRON LASER FACILITY GMBH (XFEL)	Germany	PaNdata	NO	✓	30.06.2017	myMdC	NO
10	HELMHOLTZ-ZENTRUM BERLIN FUR MATERIALIEN UND ENERGIE GMBH	Germany	PaNdata	✓	✓	16.06.2016	iCat	NO
11	ISTITUTO NAZIONALE DI FISICA NUCLEARE	Italy	No data policy				Not yet in place	NO
12	INSTALLATION EUROPEENNE DE RAYONNEMENT SYNCHROTRON	France	PaNdata	✓	✓	2015 - 2020	iCat	✓
13	KARLSRUHER INSTITUT FUER TECHNOLOGIE	Germany	No user operation yet				EUDAT	✓
14	MAX IV - LUNDS UNIVERSITET	Sweden	Data Policy partly compatible with PanData	NO	N/A	19.01.2017 (experimental)	SciCat + ISByP	NO

15	PAUL SCHERRER INSTITUT	Switzerland	PaNdata	NO	✓	08.2016	SciCat	✓
16	FELIX - STICHTING KATHOLIEKE UNIVERSITEIT	Netherlands	PaNdata	✓	✓	2019	Not planned yet	✓
17	SYNCHROTRON-LIGHT FOR EXPERIMENTAL SCIENCE AND APPLICATIONS IN THE MIDDLE EAST	Jordan	2 beamlines in operation / PaNData policy in preparation through BEATS H2020				In the context of the BEATS project, metadata catalogue will be ready within the next three years.	NO
18	Société Civile Synchrotron SOLEIL	France	PaNdata	✓	✓	10.2018	Not yet in place	✓
19	SOLARIS-UNIWERSYTET JAGIELLONSKI	Poland	PaNdata	NO	✓	N/A	Not yet in place	NO

Table 1. Overview of the survey results

Table 1 shows that fourteen (14) out of nineteen (19) CALIPSOplus partners have implemented a data policy based on the PaNData framework for scientific data management at photon and neutron facilities (PaNData Europe D2.1). The remaining five facilities are in the process of planning to have a PanData based policy within the project period of CALIPSOPlus. These are smaller facilities with a limited amount of users, thus the need for a data policy is less urgent. The PaNdata framework policy is a very good basis to handle the data that are generated in the frame of the CALIPSOplus project findable, accessible, interoperable and reusable (FAIR).

	CALIPSOplus partners	**Country**	**Does your facility have a metadata catalogue? If so, what kind?**
1	HELMHOLTZ-ZENTRUM DRESDEN-ROSSENDORF EV Germany	Germany	RODARE
2	ANKARA UNIVERSITESI	Turkey	Not yet in place They are planning to have an Online-Catalogue, which will be prepared in accordance with ICAT.
3	AARHUS UNIVERSITET	Denmark	Not planned yet At the ASTRID2 facility (Aarhus University), there are six operational beam lines where each of the specialised experiments has its own individual method of data collection and analysis procedures. All data are backed up onto the University servers. We do not currently have a data management policy at ASTRID2, but if one were to be applied in the future it would be defined by the University.
4	ALBA - CONSORCIO PARA LA CONSTRUCCION, EQUIPAMIENTO Y EXPLOTACION DEL LABORATORIO DE LUZ E SINCROTRON	Spain	iCat
5	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE	France	Not planned yet
6	STIFTUNG DEUTSCHES ELEKTRONEN-SYNCHROTRON DESY	Germany	Gamma Portal (DESY)
7	DIAMOND LIGHT SOURCE LIMITED	United Kingdom	iCat iCAT holds light metadata throughout the facility but ISPyB, which is being rolled out facility-wide, holds rich metadata. How much metadata will be held for FAIR data consumption is TBD.
8	ELETTRA - SINCROTRONE TRIESTE SCPA	Italy	Elettra doesn't have a catalogue like iCat in production (even if it has deployments of it in testing). It has an database in its digital user office (VUO) that contains metadata relevant to the experiment the datasets. These metadata connect the proposal with the actual data but miss machine metadata that are often in the files. The intention is to fully implement and deploy an iCat catalog with data ingested from the VUO. Elettra will implement a fully-fledged iCat catalog and a DMP in the timeframe of the ExPaNDs and PANOSC projects (within 2020).
9	EUXFEL - EUROPEAN X-RAY FREE-ELECTRON LASER FACILITY GMBH	Germany	myMdC
10	HELMHOLTZ-ZENTRUM BERLIN FUR MATERIALIEN UND ENERGIE GMBH	Germany	iCat
11	ISTITUTO NAZIONALE DI FISICA NUCLEARE	Italy	Not yet in place
12	ESRF - INSTALLATION EUROPEENNE DE RAYONNEMENT SYNCHROTRON	France	iCat
13	KIT - KARLSRUHER INSTITUT FUER TECHNOLOGIE	Germany	Not planned yet (as no external user operation) - EUDAT
14	MAX IV - LUNDS UNIVERSITET	Sweden	SciCat + ISByP

15	PSI - PAUL SCHERRER INSTITUT	Switzerland	SciCat
16	FELIX - STICHTING KATHOLIEKE UNIVERSITEIT	Netherlands	Not planned yet
17	SESAME - SYNCHROTRON-LIGHT FOR EXPERIMENTAL SCIENCE AND APPLICATIONS IN THE MIDDLE EAST	Jordan	In the context of the BEATS project, metadata catalogue will be ready within the next three years. The kind of metadata catalogue is not deceided on yet.
18	Société Civile Synchrotron SOLEIL	France	Not yet in place
19	SOLARIS - UNIWERSYTET JAGIELLONSKI	Poland	Not yet in place

Table 2. Implementation of metadata catalogue software

Table 2 shows that 10 out of 19 partners of CALIPSOplus have a metadata catalogue in place. The kind of software used for the metadata catologues at each facility varies. In comparison to D2.2 the amount of implemented metadata catalogues at the partner facilities increased from 6 to 10, clearly showing the trend of more accurated and better-organised data management during the CALIPSOplus project.

As mentioned in D2.2 of CALIPSOplus, the implementation of a Data Policy can start only with the availability of a metadata catalogue software. Such a software manages the metadata of raw and derived data taken at experiment facilities (i.e. partners of CALIPSOplus). In a metadata catalogue different types of metadata are saved: 1) *administrative metadata:* data management lifecycle, ownership, file catalog and 2) *scientific metadata:* describing the sample, the beamline and experiment as well as parameters relevant for data analysis.

A data catalogue software enables management of the data lifecycle, i.e. from data aquisition to data analysis and eventual deletion of the data. The data can be linked to proposals and samples, to publications (DOI, PID) and can be migrated to and from long-term storage on tape.

A metadata catalogue helps keep track of the data provenance (i.e. the steps leading to the final results) and it allows to check scientific integrity (checksum of data). It allows to find data based on the metadata (i.e. the users own data and handles open access to data). In the long term: metadata catalogues will help to automate standardised analysis workflows and support the standardisation of data formats.

In a future step, the national metadata catalogues will be federated using cloud-based services connecting to the European Open Science Cloud (EOSC). For the CALIPSOplus facilities this will be realised through projects such as European Photon and Neutron Data Services (ExPaNDS) and Photon and Neutron Open Science Cloud (PaNOSC).

To have a better overview on the data management strategy each facility, the survey asked for a short description. A description per facility is given below:

1. HELMHOLTZ-ZENTRUM DRESDEN-ROSSENDORF EV Germany:

   We differentiate between administrative metadata (proposal, title, abstract, users) and physical metadata (experiment parameters). Administrative metadata are imported from the user portal GATE into ICAT before the begin of the experiment. Experimental metadata are collected (automatically, whenever possible) during the experiment and stored directly with the measured raw data. Physical metadata that are considered relevant for the search are additionally stored in ICAT.

2. ANKARA UNIVERSITESI - TARLA:

   Access to Raw Data and Metadata obtained from an experiment is restricted to the Experimental Team for an embargo period of three (3) years after the end of the experiment. Thereafter, the data will become openly accessible. Any PI that wishes data to retain restricted access for a period longer than three (3) years will have this possibility on a yearly basis on a maximum prolongation of two (2) years. For longer extension a written request shall be submitted, specifying the reasons for the proposed prolongation, to the head of the corresponding TARLA division who decides on the request, all at their sole discretion. Data can always be made openly accessible earlier on simple request of the PI. In exceptional circumstances the head of the corresponding TARLA division can grant access to official committees at any time for the purpose of verifying data integrity.

3. AARHUS UNIVERSITET:

   There is no defined strategy for data management. At the facility, there are six operational beam lines where each of the specialised experiments has its own individual method of data collection and analysis procedures. All data are backed up onto the University servers. We do not currently have a data management policy at ASTRID2, but if one were to be applied in the future it would be defined by the University.

4. ALBA - CONSORCIO PARA LA CONSTRUCCION, EQUIPAMIENTO Y EXPLOTACION DEL LABORATORIO DE LUZ E SINCROTRON:

   Data and metadata is collected by the control system during data acquisition at each beamline. Metadata is saved in iCat catalog. Data is archived. A data_portal webpage gives access to metadata and the possibility of downloading data. Data will be kept in custody by ALBA for a minimum of 5 years, applying an initial embargo period of 3 years during which access to data will be restricted.

5. CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE: no information provided.

6. STIFTUNG DEUTSCHES ELEKTRONEN-SYNCHROTRON DESY:

   It is planned to implement a data policy largely following the recommendations of the PaN-data Europe project (D2.1) on data ownership, data curation, data archiving and open access to data. The main points are: DESY acts as a custodian of the experimental primary data and metadata, data and metadata become open access after an embargo period, a searchable catalogue gives access to the data and metadata.

7. DIAMOND LIGHT SOURCE LIMITED:

   Data will be stored by Diamond for a minimum period of 30 days from the date of its creation. Following this initial storage period, diamond will create a single archive copy of the Experimental data for long-term access. If more resilient long-term storage is required, this will be the User's responsibility. Diamond currently

   - endevours rich metadata at the level of the file;
   - is rolling out ISPyB as an information management system across all experiment facilities;
   - uses the iCat framework but without rich metadata due to scaleabilit issues.

8. ELETTRA - SINCROTRONE TRIESTE SCPA:

   Elettra has a strategy for data and metadata management. Elettra has been an active participant in relevant EU projects (PaNdata (ODI & Europe), CALIPSOplus, ExPaNDs, and PANOSC as a partner facility of CERIC-ERIC). It is an early adopter of a data policy (2014) and has an expertise in cataloging systems and databases. It aims at following the FAIR principle and at the moment is defining a strategy for sustainability (including funds for hardware.

9. EUXFEL - EUROPEAN X-RAY FREE-ELECTRON LASER FACILITY GMBH:

   Data strategy is based on Scientific Data Policy document. Raw data are kept in the archive for 5 years, striving for 10. Access to data and analysis facility is provided to the experiment team. After a 3-year embargo period, access to data is open. Metadata catalogue maintains information on who has access to the data; any modifications are managed there and propagated to low level services.

   Users are encouraged to deposit data analysis routines together with data.

10. HELMHOLTZ-ZENTRUM BERLIN FUR MATERIALIEN UND ENERGIE GMBH:

    HZB differentiates between administrative metadata (proposal, title, abstract, users) and physical metadata (experiment parameters). Administrative metadata are imported from the user portal GATE into ICAT before the begin of the experiment. Experimental metadata are collected (automatically, whenever possible) during the experiment and stored directly with the measured raw data. Physical metadata that are considered relevant for the search are additionally stored in ICAT.

11. ISTITUTO NAZIONALE DI FISICA NUCLEARE:

    All raw data and the associated metadata obtained by Users as a result of publically funded access to the research facility will become open access after a reasonable embargo period. An on-line catalogue will allow linking experimental data to experimental proposals. Access to raw data and metadata will be provided by the facility via a F.A.I.R. on-line catalogue. All information used to identify the data sets must be controlled and approved by the user.

12. ESRF - INSTALLATION EUROPEENNE DE RAYONNEMENT SYNCHROTRON:

    Pre-defined metadata are collected automatically on beamlines during experiments and then stored in a metadata catalogue forever. Raw data are curated and stored in archive for 10 years. Metadata definitions follow the Nexus standard as close as possible.

13. KIT - KARLSRUHER INSTITUT FUER TECHNOLOGIE:

    Strategy is being developed. KIT is investigating the EUDAT services for data management and are harvesting metadata of the accelerator. Prototype is in operation.

14. LUNDS UNIVERSITET:

    Max IV is funded for 5 years to upgrade the Data Storage and Management in a project called (DataStaMP). Data will be curated for many years and in addition extra services will be implemented to increase the value of the stored data and improve access and tools for researchers.

15. PSI - PAUL SCHERRER INSTITUT:

    PSI is developing the SciCat Metadata catalogue in collaboration with ESS. This catalogue is rolled out beamline by beamline and will be fully compatible with iCat.

16. FELIX - STICHTING KATHOLIEKE UNIVERSITEIT:

    FELIX aims to store data and metadata in a reliable and reusable way. We will develop tools and protocols based on the PanData best practice experience and align to the F.A.I.R. principles.

17. SESAME - SYNCHROTRON-LIGHT FOR EXPERIMENTAL SCIENCE AND APPLICATIONS IN THE MIDDLE EAST:

    SESAME is still developing their own strategy. Most likely, they will keep both of the experimental raw data and metadata for each experiment for the long term (a minimum of 5 years) on a best-effort basis.

18. Société Civile Synchrotron SOLEIL:

    SOLEIL strategy is to encourage its Users and Scientists to produce F.A.I.R. data & to facilitate this with appropriate data management services, in coordination with EXPANDS (EOSC Photon and Neutron Data Services Project) activities and by taking into account results of other collaborative works or projects as JRA2. SOLEIL requires a DMP for each proposal; will provide Users with automatically collected metadata for all experiments carried out on its beamlines; will act as a custodian for raw data and associated metadata - read-only, long-term curation of up to five years (striving for 10 years), DOI, made open access (CC-BY license) after an embargo period, available via a searchable on-line catalog; aims at providing means for reduction and/or processing of raw data.

19. SOLARIS - UNIWERSYTET JAGIELLONSKI

    SOLARIS follows the following steps in regards to their strategy on data and metadata:

    a. Provide access to Digital User Office via Umbrella or eduGAIN.
    b. Implement F.A.I.R. Data Management Principles.
    c. Adopt ESRF solutions on metadata tagging based on ICAT.
    d. Provide data storage mechanism based on HDF5 or NeXus.

	CALIPSOplus partners	Country	Data Management Plan
1	HELMHOLTZ-ZENTRUM DRESDEN-ROSSENDORF EV Germany	Germany	Yes
2	ANKARA UNIVERSITESI	Turkey	No
3	AARHUS UNIVERSITET	Denmark	No
4	ALBA - CONSORCIO PARA LA CONSTRUCCION, EQUIPAMIENTO Y EXPLOTACION DEL LABORATORIO DE LUZ DE SINCROTRON	Spain	No
5	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE	France	No
6	STIFTUNG DEUTSCHES ELEKTRONEN-SYNCHROTRON DESY	Germany	No
7	DIAMOND LIGHT SOURCE LIMITED	United Kingdom	Yes
8	ELETTRA - SINCROTRONE TRIESTE SCPA	Italy	No
9	EUROPEAN X-RAY FREE-ELECTRON LASER FACILITY GMBH	Germany	No
10	HELMHOLTZ-ZENTRUM BERLIN FUR MATERIALIEN UND ENERGIE GMBH	Germany	No
11	ISTITUTO NAZIONALE DI FISICA NUCLEARE	Italy	No
12	INSTALLATION EUROPEENNE DE RAYONNEMENT SYNCHROTRON	France	No
13	KARLSRUHER INSTITUT FUER TECHNOLOGIE	Germany	No
14	LUNDS UNIVERSITET	Sweden	No
15	PAUL SCHERRER INSTITUT	Switzerland	Yes
16	FELIX - STICHTING KATHOLIEKE UNIVERSITEIT	Netherlands	Not yet* *planning to fully implement in 2019
17	SYNCHROTRON-LIGHT FOR EXPERIMENTAL SCIENCE AND APPLICATIONS IN THE MIDDLE EAST	Jordan	No user operation yet* *planning to implement within one year (BEATS project)
18	Société Civile Synchrotron SOLEIL	France	No
19	SOLARIS - UNIWERSYTET JAGIELLONSKI	Poland	No

Table 3. Data management plan

Implementation of Data Management Plan for CALIPSOplus

As is clear from the survey that 14 out of 19 partners have a data policy in place based on the PaNdata data policy (Deliverable D2.1. of PaNdata Europe FP7 project in 2011). The PaNData data policy frame work defines (long term) goals concerning data storage, life cycle management, data access and ownership. Implementation these data policies can only be realised using a metadata catalogue. This implies that the implementation of the data policies can only start with the availability of metadata catalogue software.

Ten (10) out of nineteen (19) facilities do have a metadata catalogue in place now (Table 2), which is a clear progress with regard to the status 1.5 years ago during the last survey done amongst CALIPSOplus partners (see D2.2). Several different metadata catalogue softwares are used (iCAT, SciCAT, ISPyB). Implementation of the Data Policies is done step by step (i.e. role out from beamline to beamline) at all facilities. This stepwise process implies that a DMP will only be complete once these processes at the different facilities have been finished.

From Table 3 it is clear that only three (3) out of nineteen (19) partners have a DMP in place on the national level, here the was no increase with respect to the last survey done for D2.2.

Overall the survey shows that the partner facilities are moving towards a more advanced harmonised data management structure, by putting in place data policies, metadata catalogues, data analysis infrastructures and storage capacities. All accessible via the UmbrellaID.

F.A.I.R. Data management at a glance: DMP components to be covered

1. **Data summary:** Within this project **data** are **collected** during the experiments at the facilities. The data are collected by users that received transnational access money from the CALIPSOplus project. Most data are collected in the HDF5 or Nexus **format**. Data are open access after the embargo period of 3-5 years and can be re-used by third parties. Data origin is from the experimental station of the CALIPSOplus partner large scale facilities.

2. The different types of metadata catalogues planned to be used by the CALIPSOplus partners (mainly ICAT and SciCat; see Table 2) make data findable, accessable, interoparable and reusable **(FAIR data),** with standards for metadata creation.

2.2. The timeline in which **data** become **accessible** is defined by the embargo period (3-5 years) in the specific facility data policy (see examples of data policies in ANNEX 2).

2.3 The **interoperability of the data** is guaranteed by the metadata catalogue softwares in place.

3. The implementation of the data policies at the different facilities and with this the putting in place of metadata catalogue software is an ongoing process, in which we see progress since the publication of D2.2. Therefore the **allocation of recources**, i.e. the cost of making our data FAIR and the costs for long term preservation of data will be descibed in detail in the next update of our DMP which is due in month 36 of the project.

Conclusions

The stepwise implementation of the Data Policies (i.e. role out from beamline to beamline) at all facilities implies that a DMP will only be complete once these processes at the different facilities have been finished. The present version of the DMP reflects the status as it is now for the different partners of CALIPSOplus and shows a clear improvement in the amount of data policies and metadata catalogues that are in place now at the CALIPSOplus partner facilities. In the next update of the DMP in month 36, again clear progress in role out of this process will be visible. In the meantime the NA1 (i.e. WP2) team is working on even more harmonisation in this area striving towards the creation of an optimal DMP for CALIPSOplus.

ATTACHMENTS

ANNEX 1 SURVEY

ANNEX 2 DATA POLICIES RECEIVED

ANNEX 1

SURVEY RESPONSES

Survey

As part of **NA1 WP2 Task 2.3 Deliverables 2.8 and 2.9**

Good Data Management is a prerequisite for Data integration.

The purpose of this survey is to:

- ✓ **learn about the status of your data policy as input for the harmonization of data policies roadmaps (NA1 WP2 Task 2.3)**
- ✓ **learn more about how your facility engages in research data management;**
- ✓ **learn about the degree of implementation of the facility's data management policies;**

Name of Institution: _____CELLS / ALBA SINCROTRON_____

Date of submission 22 / 03 / 2019

Please, note: Deadline for submission is April 1st 2019 h 12:00 PM

In the questions below, please specify your institution's level of engagement

0 – No activity

1 – Information gathering, explorative

2 – Specification and development

3 – Partly implemented

4 – Fully implemented

1	2	3	4

Section related to INTERNAL Data Policy and Data Management Plan of your facility

1. Does your facility have a data policy? **YES**

1	2	3	4

2. If yes, is it compatible with PanData? **YES**

    a. Date of implementation: **July 2017**

1	2	3	4

3. Is your data policy compliant with the GDPR and the implementation of the policy?

    YES

1	2	3	5

4. Is your data policy aligned with the F.A.I.R. data principles?

    YES

5. What is the facility's strategy on data and/ or Metadata management?

Please, describe

Data and metadata is collected by the control system during data acquisition at each beamline. Metadata is saved in iCat catalog. Data is archived. A data_portal webpage gives access to metadata and the possibility of downloading data. Data will be kept in custody by ALBA for a minimum of 5 years, applying an initial embargo period of 3 years during which access to data will be restricted.

6. Does your facility have a Metadata catalogue? **YES**

a. If yes, what kind?

iCat

b. If not, when is your facility planning to implement one?

Please, specify

There is an iCat server running already for development and test purposes.

1	2	3	4

7. Have you validated and implemented a data management plan?

NO

a. If yes, date of implementation: _____

b. If no, when are you planning to fully implement it: **not decided yet**

1	2	3	4

8. Please, describe the data life cycle process inherited from your data policy, based on your Metadata catalogue storage and processes:

See point 5 above

9. Please, attach **a copy of your facility's data policy and data management plan**, where applicable.

Section related to integration with NA1 WP 2 Task 2.3 - CALIPSOPlus PROJECT

1. Is the facility's user office connected to UMBRELLA? **YES**

1	2	3	4

2. How is it available?

In ALBA's user office login page, the user can select to login using his/her umbrella login id

Survey

As part of **NA1 WP2 Task 2.3 Deliverables 2.8 and 2.9**

Good Data Management is a prerequisite for Data integration.

The purpose of this survey is to:

- ✓ **learn about the status of your data policy as input for the harmonization of data policies roadmaps (NA1 WP2 Task 2.3)**
- ✓ **learn more about how your facility engages in research data management;**
- ✓ **learn about the degree of implementation of the facility's data management policies;**

Name of Institution: _____ASTRID2, Aarhus University_____

Date of submission 27 / 3 / 2019

Please, note: Deadline for submission is April 1st 2019 h 12:00 PM

In the questions below, please specify your institution's level of engagement

0 – No activity

1 – Information gathering, explorative

2 – Specification and development

3 – Partly implemented

4 – Fully implemented

1	2	3	4

Section related to INTERNAL Data Policy and Data Management Plan of your facility

1. Does your facility have a data policy? ~~YES~~ ___ NO _ ✓ _

1	2	3	4

2. If yes, is it compatible with PanData? YES ___ NO __

   a. Date of implementation: _____

1	2	3	4

3. Is your data policy compliant with the GDPR and the implementation of the policy?

   YES ___ NO __

1	2	3	5

4. Is your data policy aligned with the F.A.I.R. data principles?

   YES ___ NO __

5. What is the facility's strategy on data and/ or Metadata management?

Please, describe

There is no defined strategy for data management at the facility. See comments at the end._____

_____

6. Does your facility have a Metadata catalogue? ~~YES~~ ___ NO ✓

a. If yes, what kind?

iCat _____ SciCat _____ Other _____

b. If not, when is your facility planning to implement one?

Please, specify

At this time there are no plans to implement a metadata catalogue.

_____

1	2	3	4

7. Have you validated and implemented a data management plan?

~~YES~~ ___ NO _✓_

a. If yes, date of implementation: _____

b. If no, when are you planning to fully implement it: ___No current plans___

1	2	3	4

8. Please, describe the data life cycle process inherited from your data policy, based on your Metadata catalogue storage and processes:

_____

_____

_____
_____

9. Please, attach **a copy of your facility's data policy and data management plan**, where applicable.

Section related to integration with NA1 WP 2 Task 2.3 - CALIPSOPlus PROJECT

1. Is the facility's user office connected to UMBRELLA? ~~YES~~ ___ NO ✓

1	2	3	4

2. How is it available?

_____
_____
_____
_____
_____

A more general comment:
At the ASTRID2 facility (Aarhus University) we have six operational beam lines where each of the specialised experiments has its own individual method of data collection and analysis procedures. All data are backed up onto the University servers. We do not currently have a data management policy at ASTRID2, but if one were to be applied in the future it would be defined by the University.

Von: Jean-Michel Ortega
An: Piffer Valentina (PSI)
Betreff: Re: [na1] Survey Reminder
Datum: Dienstag, 26. März 2019 14:52:16

Dear Valentina,

As for CLIo we have no data policy...sorry

Best Regards

Jean-Michel

De: "Piffer Valentina (PSI)" <valentina.piffer@psi.ch>
À: na1@calipsoplus.eu, "jra2@calipsoplus.eu" <jra2@calipsoplus.eu>
Envoyé: Mardi 26 Mars 2019 08:46:58
Objet: [na1] Survey Reminder

Dear NA1 and JRA2 Teams!

First of all, I'd like to thank the team members who have already filled out and submitted the survey we sent out last week. To all other team members, we would like to kindly remind you that the deadline is **next Monday, April 1st**. It is crucial we all respect this deadline in order to submit the deliverables on time.

Thank you so much for your kind attention and precious collaboration.

Have a great day!

Valentina

_____

Paul Scherrer Institut
Valentina Piffer
CALIPSOPlus Project Assistant

OVGA/415
Forschungsstrasse 111
5232 Villigen PSI
Schweiz

Telefon:+41 56 310 4360
E-Mail: valentina.piffer@psi.ch

www.psi.ch

Von: Piffer Valentina (PSI)
Gesendet: Dienstag, 19. März 2019 13:34

An: 'jra2@calipsoplus.eu' <jra2@calipsoplus.eu>; 'na1@calipsoplus.eu' <na1@calipsoplus.eu>
Cc: Piffer Valentina (PSI) <valentina.piffer@psi.ch>
Betreff: WP2 NA1 -
Priorität: Hoch

Dear NA1 and JRA2 Team Members,

Please, find attached a questionnaire that we kindly ask you to fill out and **return to us by April 1st 2019**. Aim of this survey is to gather information on the status of your facility's data policy, data catalogue, data management plans and UmbrellaID, in order to complete and submit deliverables 2.8 and 2.9 of WP 2 NA1.

Should you feel that you are not the person directly responsible for the topics , please feel free to forward the questionnaire to the expert within your facility.

Thank you very much for the collaboration.

We look forward to receiving your information.

Kind regards,

Valentina

Paul Scherrer Institut
Valentina Piffer
CALIPSOPlus Project Assistant

OVGA/415
Forschungsstrasse 111
5232 Villigen PSI
Schweiz

Telefon:+41 56 310 4360
E-Mail: valentina.piffer@psi.ch

www.psi.ch

Survey

As part of **NA1 WP2 Task 2.3 Deliverables 2.8 and 2.9**

Good Data Management is a prerequisite for Data integration.

The purpose of this survey is to:

- ✓ **learn about the status of your data policy as input for the harmonization of data policies roadmaps (NA1 WP2 Task 2.3)**
- ✓ **learn more about how your facility engages in research data management;**
- ✓ **learn about the degree of implementation of the facility's data management policies;**

Name of Institution: DESY_____

Date of submission 27 / 03 / 2019

Please, note: Deadline for submission is April 1st 2019 h 12:00 PM

In the questions below, please specify your institution's level of engagement

0 – No activity

1 – Information gathering, explorative

2 – Specification and development

3 – Partly implemented

4 – Fully implemented

1	2	3	4

Section related to INTERNAL Data Policy and Data Management Plan of your facility

1. Does your facility have a data policy? YES 4 NO __

1	2	3	4

2. If yes, is it compatible with PanData? YES ___ NO x

    a. Date of implementation: 25.6.2015

1	2	3	4

3. Is your data policy compliant with the GDPR and the implementation of the policy?

    YES x NO

1	2	3	5

4. Is your data policy aligned with the F.A.I.R. data principles?

    YES ___ NO 3

5. What is the facility's strategy on data and/ or Metadata management?

Please, describe
It is planned to implement a data policy largely following the recommendations of the PaN-data Europe project (D2.1) on data ownership, data curation, data archiving and open access to data. The main points are: DESY acts as a custodian of the experimental primary data and metadata, data and metadata become open access after an embargo period, a searchable catalogue gives access to the data and metadata.

6. Does your facility have a Metadata catalogue? YES x NO __

a. If yes, what kind?

iCat _____ SciCat _____ Other Gamma Portal (DESY)

b. If not, when is your facility planning to implement one?

Please, specify

_____

_____

1	2	3	4

7. Have you validated and implemented a data management plan?

YES ___ NO x

a. If yes, date of implementation: _____
b. If no, when are you planning to fully implement it: DESY has a concept for data management, archiving, download, etc. without having an explicit data management plan

1	2	3	4

8. Please, describe the data life cycle process inherited from your data policy, based on your Metadata catalogue storage and processes:

Experimental data are centrally stored and archived using an authorization scheme based on data from the digital user office. After data have been analyzed they are moved to a second-level storage from where they can be re-staged, if needed. A metadata catalogue allows for file discoveries and downloads through a web portal. The portal supports simple data management.

9. Please, attach **a copy of your facility's data policy and data management plan**, where applicable.

Section related to integration with NA1 WP 2 Task 2.3 - CALIPSOPlus PROJECT

1. Is the facility's user office connected to UMBRELLA? YES x NO __

1	2	3	4

2. How is it available?

   If users who are authenticated by Umbrella connect to the DESY user office, they are automatically forwarded, if their local account has been matched with the Umbrella account before. Otherwise they may match their local accounts with the Umbrella account.

Survey

As part of **NA1 WP2 Task 2.3 Deliverables 2.8 and 2.9**

Good Data Management is a prerequisite for Data integration.

The purpose of this survey is to:

- ✓ **learn about the status of your data policy as input for the harmonization of data policies roadmaps (NA1 WP2 Task 2.3)**
- ✓ **learn more about how your facility engages in research data management;**
- ✓ **learn about the degree of implementation of the facility's data management policies;**

Name of Institution: Diamond Light Source

Date of submission 26/03/2019

Please, note: Deadline for submission is April 1st 2019 h 12:00 PM

In the questions below, please specify your institution's level of engagement

0 – No activity

1 – Information gathering, explorative

2 – Specification and development

3 – Partly implemented

4 – Fully implemented

1	2	3	4

Section related to INTERNAL Data Policy and Data Management Plan of your facility

1. Does your facility have a data policy? YES ✓ NO __

1	2	3	[4]

2. If yes, is it compatible with PanData? YES ✓ NO __

   a. Date of implementation: 1/03/2019

1	[2]	3	4

3. Is your data policy compliant with the GDPR and the implementation of the policy?

   YES ✓ NO __

1	2	3	[4]

4. Is your data policy aligned with the F.A.I.R. data principles?

   YES ✓ NO __

5. What is the facility's strategy on data and/ or Metadata management?

Please, describe

Data will be stored by Diamond for a minimum period of 30 days from the date of its creation. Following this initial storage period, Diamond will create a single archive copy of the Experimental Data for long-term access. If more resilient long-term storage is required, this is the Users' responsibility. Diamond currently 1) endeavours rich metadata at the level of the file, 2) is rolling out ISPyB as an information management system across all experiment facilities, 3) uses the ICAT framework but without rich metadata due to scaleability issues.

6. Does your facility have a Metadata catalogue? YES ✓ NO __

a. If yes, what kind?

iCat ✓ SciCat _____ Other ISPyB

b. If not, when is your facility planning to implement one?

Please, specify

ICAT holds light metadata throughout the facility but ISPyB, which is being rolled out facility wide holds rich metadata. How much metadata will be held for FAIR data consumption is TBD

1	[2]	3	4

7. Have you validated and implemented a data management plan?

YES ✓ NO __

a. If yes, date of implementation: 1/1/2007

b. If no, when are you planning to fully implement it: _____

1	2	[3]	4

8. Please, describe the data life cycle process inherited from your data policy, based on your Metadata catalogue storage and processes:

see point 5

9. Please, attach **a copy of your facility's data policy and data management plan**, where applicable.

https://www.diamond.ac.uk/Users/Policy-Documents/Policies/Experimental-Data-Management-Pol.html

Section related to integration with NA1 WP 2 Task 2.3 - CALIPSOPlus PROJECT

1. Is the facility's user office connected to UMBRELLA? YES ___ NO X

1	2	3	4

2. How is it available?

The facility user office will connected to the UK Access Management Federation, which includes Umbrella, and is currently in Alpha testing.

Survey

As part of **NA1 WP2 Task 2.3 Deliverables 2.8 and 2.9**

Good Data Management is a prerequisite for Data integration.

The purpose of this survey is to:

- ✓ **learn about the status of your data policy as input for the harmonization of data policies roadmaps (NA1 WP2 Task 2.3)**
- ✓ **learn more about how your facility engages in research data management;**
- ✓ **learn about the degree of implementation of the facility's data management policies;**

Name of Institution: ELETTRA SINCROTRONE TRIESTE

Date of submission 01 / 04 / 2019

Please, note: Deadline for submission is April 1st 2019 h 12:00 PM

In the questions below, please specify your institution's level of engagement

0 – No activity

1 – Information gathering, explorative

2 – Specification and development

3 – Partly implemented

4 – Fully implemented

1	2	3	4

Section related to INTERNAL Data Policy and Data Management Plan of your facility

1. Does your facility have a data policy? YES

1	2	**3**	4

2. If yes, is it compatible with PanData? YES

    a. Date of implementation: 01/12/2014

1	2	**3**	4

3. Is your data policy compliant with the GDPR and the implementation of the policy?

YES

1	2	3	5

4. Is your data policy aligned with the F.A.I.R. data principles?

YES

5. What is the facility's strategy on data and/ or Metadata management?

Elettra has a strategy for data and metadata management. Elettra has been an active participant in relevant EU projects (PaNdata (ODI & Europe), CALIPSOplus, ExPaNDs, and PANOSC as a partner facility of CERIC-ERIC). It is an early adopter of a data policy (2014) and has an expertise in cataloging systems and databases. It aims at following the FAIR principle and at the moment is defining a strategy for sustainability (including funds for hardware).

6. Does your facility have a Metadata catalogue? YES *

a. If yes, what kind?

iCat _____ SciCat _____ Other: *

** Elettra doesn't have a catalog like iCat in production (even if it has deployments of it in testing). It has an database in its digital user office (VUO) that contains metadata relevant to the experiment the datasets. These metadata connect the proposal with the actual data but miss machine metadata that are often in the files. The intention is to fully implement and deploy an iCat catalog with data ingested from the VUO.*

b. If not, when is your facility planning to implement one?

Elettra will implement a fully-fledged iCat catalog in the time frame of the ExPaNDs and PANOSC projects (within 2020).

1	2	3	4

7. Have you validated and implemented a data management plan?

NO

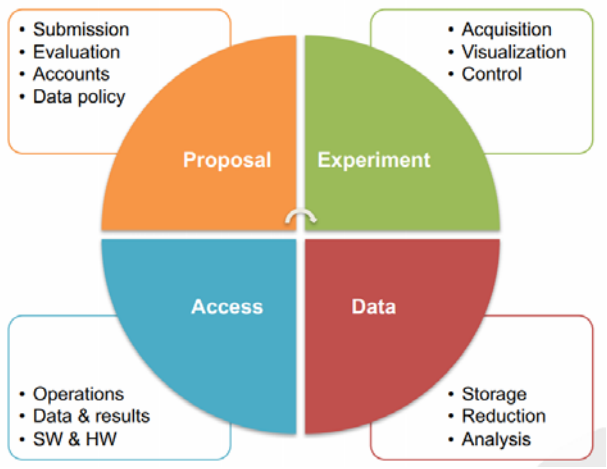
a. If yes, date of implementation: _____

b. If no, when are you planning to fully implement it: *Elettra will implement one during 2020 according to its needs (national) but also based on early feedback from projects like CALIPSOplus, ExPaNDs and PANOSC.*

1	2	3	4

8. Please, describe the data life cycle process inherited from your data policy, based on your Metadata catalogue storage and processes:

(The description of this process can be lengthy and has been the topic of conference papers – for details see [1]) In brief:



[1] Billè, Fulvio & Borghes, R & Brun, Francesco & Chenda, Valentina & Curri, Alessio & Duic, Venicio & Favretto, Daniele & Kourousias, Georgios & Lonza, M & Prica, Milan & Pugliese, Roberto & Scarcia, Martin & Turcinovich, Michele. (2015). DATA LIFECYCLE IN LARGE EXPERIMENTAL PHYSICS FACILITIES: THE APPROACH OF THE SYNCHROTRON ELETTRA AND THE FREE ELECTRON LASER FERMI.
https://www.researchgate.net/publication/302962157_DATA_LIFECYCLE_IN_LARGE_EXPERIMENTAL_PHYSICS_FACILITIES_THE_APPROACH_OF_THE_SYNCHROTRON_ELETTRA_AND_THE_FREE_ELECTRON_LASER_FERMI

9. Please, attach **a copy of your facility's data policy and data management plan**, where applicable.

See: https://vuo.elettra.eu/vuo/cgi-bin/download-tm4.py?frm_user_id=8707&frm_iddocumenttype=14&frm_iddocument=208454&frm_hash=efb046860392bdb106d0e8c91e9cea57e465cf7e

Section related to integration with NA1 WP 2 Task 2.3 - CALIPSOPlus PROJECT

1. Is the facility's user office connected to UMBRELLA? YES

1	2	3	4

2. How is it available?

Umbrella is available through the main page of Elettra's VUO: https://vuo.elettra.eu

Survey

As part of **NA1 WP2 Task 2.3 Deliverables 2.8 and 2.9**

Good Data Management is a prerequisite for Data integration.

The purpose of this survey is to:

- ✓ **learn about the status of your data policy as input for the harmonization of data policies roadmaps (NA1 WP2 Task 2.3)**
- ✓ **learn more about how your facility engages in research data management;**
- ✓ **learn about the degree of implementation of the facility's data management policies;**

Name of Institution: European Synchrotron Radiation Facility (ESRF)

Date of submission 01 / 04 / 2019

Please, note: Deadline for submission is April 1st 2019 h 12:00 PM

In the questions below, please specify your institution's level of engagement

0 – No activity

1 – Information gathering, explorative

2 – Specification and development

3 – Partly implemented

4 – Fully implemented

1	2	3	4

Section related to INTERNAL Data Policy and Data Management Plan of your facility

1. Does your facility have a data policy? YES **X** NO __

1	2	**X**	4

2. If yes, is it compatible with PanData? YES **X** NO __

    a. Date of implementation: 2015 - 2020

1	2	3	4

3. Is your data policy compliant with the GDPR and the implementation of the policy?

    YES in principle (but does not mention GDPR) NO __

1	2	3	5

4. Is your data policy aligned with the F.A.I.R. data principles?

    YES in principle (but does not mention FAIR) NO __

5. What is the facility's strategy on data and/ or Metadata management?

Please, describe
Pre-defined metadata are collected automatically on beamlines during experiments and then stored in a metadata catalogue forever. Raw data are curated and stored in archive for 10 years. Metadata definitions follow the Nexus standard as close as possible.

6. Does your facility have a Metadata catalogue? YES **X** NO __

a. If yes, what kind?

iCat **X** SciCat _____ Other _____

b. If not, when is your facility planning to implement one?

Please, specify

_____

_____

1	2	3	4

7. Have you validated and implemented a data management plan?

YES ___ NO X

a. If yes, date of implementation: _____
b. If no, when are you planning to fully implement it: not planned

1	2	3	4

8. Please, describe the data life cycle process inherited from your data policy, based on your Metadata catalogue storage and processes:

See above

9. Please, attach **a copy of your facility's data policy and data management plan**, where applicable.

http://www.esrf.eu/files/live/sites/www/files/about/organisation/ESRF%20data%20policy-web.pdf

Section related to integration with NA1 WP 2 Task 2.3 - CALIPSOPlus PROJECT

1. Is the facility's user office connected to UMBRELLA? YES X NO __

1	2	3	X

2. How is it available?

   As an alternative authentication mechanism. Not many users use Umbrella.

Survey

As part of **NA1 WP2 Task 2.3 Deliverables 2.8 and 2.9**

Good Data Management is a prerequisite for Data integration.

The purpose of this survey is to:

- ✓ **learn about the status of your data policy as input for the harmonization of data policies roadmaps (NA1 WP2 Task 2.3)**
- ✓ **learn more about how your facility engages in research data management;**
- ✓ **learn about the degree of implementation of the facility's data management policies;**

Name of Institution: European XFEL_____

Date of submission 26 /03 / 2019

Please, note: Deadline for submission is April 1st 2019 h 12:00 PM

In the questions below, please specify your institution's level of engagement

0 – No activity

1 – Information gathering, explorative

2 – Specification and development

3 – Partly implemented

4 – Fully implemented

1	2	3	4

Section related to INTERNAL Data Policy and Data Management Plan of your facility

1. Does your facility have a data policy? YES _X_ NO __

1	2	3	4

2. If yes, is it compatible with PanData? YES _X_ NO __

   a. Date of implementation:
   ___30.06.2017_____

1	2	3	4

3. Is your data policy compliant with the GDPR and the implementation of the policy?

   YES _X_ NO __

1	2	3	4

4. Is your data policy aligned with the F.A.I.R. data principles?

   YES ___ NO X

5. What is the facility's strategy on data and/ or Metadata management?
   Please, describe
   Data strategy is based on Scientific Data Policy document.
   Raw data kept for 5 years, striving for 10, in archive. Access to data and analysis facility provided to experiment team. After 3 years embargo period access to data is open.
   Metadata catalogue maintains information who has access to the data, any modifications are managed there and propagated to low level services.
   Users are encouraged to deposit data analysis routines together with data.
6. Does your facility have a Metadata catalogue? YES _X_ NO __

   a. If yes, what kind?

   iCat _____ SciCat _____ Other ____myMdC_____

   b. If not, when is your facility planning to implement one?

Please, specify

_____
_____
_____

1	2	3	4

7. Have you validated and implemented a data management plan?

   YES ___ NO _X_

a. If yes, date of implementation: _____
b. If no, when are you planning to fully implement it: _no deadline, investigation ongoing

1	2	3	4

8. Please, describe the data life cycle process inherited from your data policy, based on your Metadata catalogue storage and processes:

Raw data is stored in the online system, then after assessment by users if it is of good quality or if the quality cannot be determined it is copied to the offline storage and the data archive. Calibrated data is produced automatically if configured in myMdC. Calibrated data is not archived but can be reproduced on request. Raw data is stored in the offline disk system for 2-3 years and then can be restored from archive in the period of 5-10 years.

9. Please, attach **a copy of your facility's data policy and data management plan**, where applicable.

https://www.xfel.eu/users/experiment_support/policies/scientific_data_policy/index_eng.html

Section related to integration with NA1 WP 2 Task 2.3 - CALIPSOPlus PROJECT

1. Is the facility's user office connected to UMBRELLA? YES ___ NO _X_

1	2	3	4

2. How is it available?

_____
_____
_____
_____
_____

Survey

As part of **NA1 WP2 Task 2.3 Deliverables 2.8 and 2.9**

Good Data Management is a prerequisite for Data integration.

The purpose of this survey is to:

- ✓ **learn about the status of your data policy as input for the harmonization of data policies roadmaps (NA1 WP2 Task 2.3)**
- ✓ **learn more about how your facility engages in research data management;**
- ✓ **learn about the degree of implementation of the facility's data management policies;**

Name of Institution: FELIX Laboratory, Radboud University_____

Date of submission 01/ 04/ 2019

Please, note: Deadline for submission is April 1st 2019 h 12:00 PM

In the questions below, please specify your institution's level of engagement

0 – No activity

1 – Information gathering, explorative

2 – Specification and development

3 – Partly implemented

4 – Fully implemented

1	2	3	4

Section related to INTERNAL Data Policy and Data Management Plan of your facility

1. Does your facility have a data policy? YES 2/3 NO

1	2	3	4

2. If yes, is it compatible with PanData? YES 2 NO __

    a. Date of implementation: the FELIX data policy will be compatible with PanData and will be fully implemented in 2019.

1	2	3	4

3. Is your data policy compliant with the GDPR and the implementation of the policy?

    YES 2 NO __

    The data policy to be fully implemented will be compliant with the GDPR.

1	2	3	5

4. Is your data policy aligned with the F.A.I.R. data principles?

YES 2 NO __

5. What is the facility's strategy on data and/ or Metadata management?

Please, describe
FELIX aims to store data and metadata in a reliable and reusable way. We will develop tools and protocols based on the PanData best practice experience and align to the FAIR principles.

_____

6. Does your facility have a Metadata catalogue? YES ___ NO 2

a. If yes, what kind? To be decided

iCat _____ SciCat _____ Other _____

b. If not, when is your facility planning to implement one?

Please, specify

It is currently explored which format and catalogue will be used for storage of data and metadata including the RU internal option requested by the university in 2020.

_____

1	2	3	4

7. Have you validated and implemented a data management plan?

YES ___ NO 2/3

a. If yes, date of implementation: _____
b. If no, when are you planning to fully implement it: ____in 2019_____

1	2	3	4

8. Please, describe the data life cycle process inherited from your data policy, based on your Metadata catalogue storage and processes:

Currently all data and metadata is stored by the central computer department of Radboud University which also provides the back-up service. The data are presently stored for an infinite period which may become 10 years in the future in compliance with the national legal requirements.

_____

_____

9. Please, attach **a copy of your facility's data policy and data management plan**, where applicable.

Section related to integration with NA1 WP 2 Task 2.3 - CALIPSOPlus PROJECT

1. Is the facility's user office connected to UMBRELLA? YES 2 NO __

1	2	3	4

2. How is it available?

   The FELIX user office has access and connects to Umbrella. Presently FELIX has ist own user portal for proposal submission but the connection to Umbrella is under consideration.

   _____

Survey

As part of **NA1 WP2 Task 2.3 Deliverables 2.8 and 2.9**

Good Data Management is a prerequisite for Data integration.

The purpose of this survey is to:

- ✓ **learn about the status of your data policy as input for the harmonization of data policies roadmaps (NA1 WP2 Task 2.3)**
- ✓ **learn more about how your facility engages in research data management;**
- ✓ **learn about the degree of implementation of the facility's data management policies;**

Name of Institution: _ Helmholtz Zentrum Berlin für Materialien und Energie _____

Date of submission 29 / 03 / 2019

Please, note: Deadline for submission is April 1st 2019 h 12:00 PM

In the questions below, please specify your institution's level of engagement

1 – No activity

2 – Information gathering, explorative

3 – Specification and development

4 – Partly implemented

5 – Fully implemented

1	2	3	4

Section related to INTERNAL Data Policy and Data Management Plan of your facility

1. Does your facility have a data policy? YES _X_ NO __

1	2	3	4

2. If yes, is it compatible with PanData? YES _X_ NO __

    a. Date of implementation: _data policy adopted 2016-06-14, implementation is an ongoing process _____

1	2	3	4

3. Is your data policy compliant with the GDPR and the implementation of the policy?

    YES _X_ NO __

1	2	3	6

4. Is your data policy aligned with the F.A.I.R. data principles?

    YES _X_ NO __

5. What is the facility's strategy on data and/ or Metadata management?

Please, describe
We differentiate between administrative metadata (proposal, title, abstract, users) and physical metadata (experiment parameters). Administrative metadata are imported from the user portal GATE into ICAT before the begin of the experiment. Experimental metadata are collected (automatically, whenever possible) during the experiment and stored directly with the measured raw data. Physical metadata that are considered relevant for the search are additionally stored in ICAT.

6. Does your facility have a Metadata catalogue? YES _X_ NO __

a. If yes, what kind?

ICAT _X___ SciCat _____ Other _____

b. If not, when is your facility planning to implement one?

Please, specify

_____

_____

1	2	3		4
7. Have you validated and implemented a data management plan? YES ___ NO _X_ a. If yes, date of implementation: _____ b. If no, when are you planning to fully implement it: _ Not sure if I understand the question. Individual scientific projects have their respective DMP, but a DMP for the facility as a whole, not sure what this is supposed to mean. _____				
1	2	3		4

8. Please, describe the data life cycle process inherited from your data policy, based on your Metadata catalogue storage and processes:

Raw data and related physical metadata are curated at the experimental station. The preferred file format is NeXus, but this may vary between experimental stations. After curation, the data is transferred to central storage systems managed by ICAT and kept for at least ten years. They become open accessible after the end of the embargo period, which is five years after the end of the experiment by default.

9. Please, attach **a copy of your facility's data policy and data management plan**, where applicable.

Section related to integration with NA1 WP 2 Task 2.3 - CALIPSOPlus PROJECT

1. Is the facility's user office connected to UMBRELLA? YES ___ NO _X_

1	2	3		4

2. How is it available?

_____
_____
_____
_____
_____

Survey

As part of **NA1 WP2 Task 2.3 Deliverables 2.8 and 2.9**

Good Data Management is a prerequisite for Data integration.

The purpose of this survey is to:

- ✓ **learn about the status of your data policy as input for the harmonization of data policies roadmaps (NA1 WP2 Task 2.3)**
- ✓ **learn more about how your facility engages in research data management;**
- ✓ **learn about the degree of implementation of the facility's data management policies;**

Name of Institution: **Helmholtz-Zentrum Dresden-Rossendorf e.V.**

Date of submission 21 / 03 / 2019

Please, note: Deadline for submission is April 1st 2019 h 12:00 PM

In the questions below, please specify your institution's level of engagement

0 – No activity

1 – Information gathering, explorative

2 – Specification and development

3 – Partly implemented

4 – Fully implemented

1	2	3	X

Section related to INTERNAL Data Policy and Data Management Plan of your facility

1. Does your facility have a data policy? YES X_ NO __

1	2	3	X

2. If yes, is it compatible with PanData? YES X_ NO __

    a. Date of implementation: 01.05.2018

1	2	3	X

3. Is your data policy compliant with the GDPR and the implementation of the policy?

    YES X_ NO __

1	2	X	4

4. Is your data policy aligned with the F.A.I.R. data principles?

    YES X_ NO __

5. What is the facility's strategy on data and/ or Metadata management?

Please, describe
Data Management is supported by the HZDR Data repostitory RODARE (in Invenio Instance). Metadata format is recommended to be based on DataCite Scheme 4.1. Since HZDR is a multidisciplinary research center individual metadata catalogues are necessary. Starting in June a data librarian will support the quality control of metadata.

_____

_____

6. Does your facility have a Metadata catalogue? YES X__ NO __

a. If yes, what kind?

iCat _____ SciCat _____ Other RODARE (Invenio)

b. If not, when is your facility planning to implement one?

Please, specify

_____

_____

1	2	X	4

7. Have you validated and implemented a data management plan?

YES X_ NO __

a. If yes, date of implementation: 01.05.2018 _____
b. If no, when are you planning to fully implement it: _____

1	2	X	4

8. Please, describe the data life cycle process inherited from your data policy, based on your Metadata catalogue storage and processes:

The user registration system Gate is linked to the DataPolicy and DMP. The Workflow system ROBIS controls the publication process for texts, data and software. It is closely linked to the system RODARE which is used to store and publish the data. The system for the ingest of data directly from the experiments and as data analysis platform is still in a development and test phase (HZDR has multiple facilities / data sources).

9. Please, attach **a copy of your facility's data policy and data management plan**, where applicable.
-> https://www.hzdr.de/datapolicy

Section related to integration with NA1 WP 2 Task 2.3 - CALIPSOPlus PROJECT

1. Is the facility's user office connected to UMBRELLA? YES X_ NO __

1	2	3	X

2. How is it available?

   It is available via the user registration system GATE. Login to RODARE is currently prepared and tested.

Von:	Jean-François Perrin
An:	van Daalen Mirjam (PSI)
Cc:	Piffer Valentina (PSI)
Betreff:	Re: [jra2] FW: [na1] WP2 NA1 -
Datum:	Dienstag, 26. März 2019 13:57:03

HI Mirjam,

As promised the ILL response to your DP survey.

Name : Institut Laue - Langevin (ILL)

1. Yes 4
2. Yes 4
2.a Date of initial publication: Sept 2011, revised in July 2017. 1st Techinal implementation: Oct 2012
3. Yes 4, on the GDPR side the information are completed by our Personal Data Protection Policy (users have to consent electronically like for the DP if they want to participate in a proposal)
https://www.ill.eu/fileadmin/user_upload/ILL/1_About_ILL/Documentation/Legal_information/GDPR/ILL_Personal_Data_Protection_Policy.pdf
4. Yes 3, the "machine actionable" part is missing.
5. We currently preserve data and metadata for life. We try to automate as far as possible dat/metadata acquisition without the need for user interaction. We mint DOIs for proposal data sets.
6. Yes 4
6.a Home made solution.
7. Our vision is that DMP are more for users than the facility. Our Plan Inside the PaNOSC project is to automate the filling of standard DMPs with the information we have.
8. ACLs and administrative information are derived from our proposal database. During experiment data are transferred automatically to the archive and immediately ingest in the catalogue. A DOI is minted at the end of the experiment cycle.
9. https://www.ill.eu/DataPolicy

Kind regards
--
Jean-François

Survey

As part of **NA1 WP2 Task 2.3 Deliverables 2.8 and 2.9**

Good Data Management is a prerequisite for Data integration.

The purpose of this survey is to:

- ✓ **learn about the status of your data policy as input for the harmonization of data policies roadmaps (NA1 WP2 Task 2.3)**
- ✓ **learn more about how your facility engages in research data management;**
- ✓ **learn about the degree of implementation of the facility's data management policies;**

Name of Institution: INFN- DAFNE_Light Facility

Date of submission 29/ 03 / 2019

Please, note: Deadline for submission is April 1st 2019 h 12:00 PM

In the questions below, please specify your institution's level of engagement

0 – No activity

1 – Information gathering, explorative

2 – Specification and development

3 – Partly implemented

4 – Fully implemented

1	2	3	4

Section related to INTERNAL Data Policy and Data Management Plan of your facility

1. Does your facility have a data policy? YES ___ NO __

1	2 X	3	4

2. If yes, is it compatible with PanData? YES ___ NO __

   a. Date of implementation: _____

1	2	3	4

3. Is your data policy compliant with the GDPR and the implementation of the policy?

   YES X NO __

1	2	3	4

4. Is your data policy aligned with the F.A.I.R. data principles?

   YES X NO __

5. What is the facility's strategy on data and/ or Metadata management?
   Please, describe

   All raw data and the associated metadata obtained by Users as a result of publically funded access to the research facility will become open access after a reasonable embargo period. An on-line catalogue will allow linking experimental data to experimental proposals. Access to raw data and metadata will be provided by the Facility via a F.A.I.R. on-line catalogue. All information used to identify the data sets must be controlled and approved by the USERS.

6. Does your facility have a Metadata catalogue? YES ___ NO X

   a. If yes, what kind?

      iCat _____ SciCat _____ Other _____

   b. If not, when is your facility planning to implement one?

Please, specify

We are defining the information needed to identify data taken at the different beamlines to make them F.A.I.R. We are also trying to understand if we must look for something more general or specifically applied to the INFN DAFNE-Light facility. At the moment data are store at each beamline with the proposal acronym.

1	2 X	3	4

7. Have you validated and implemented a data management plan?

   YES ___ NO X

   a. If yes, date of implementation: _____
   b. If no, when are you planning to fully implement it: In the near future

1	2 X	3	4

8. Please, describe the data life cycle process inherited from your data policy, based on your Metadata catalogue storage and processes:

Access to raw data and the associated metadata will be restricted to the experimental team for a period of 3 years after the end of the experiment. After this period each year the principal investigator will be asked for permission to give open access to the raw data. If not given the access to raw data will remain restricted. A copy of the data will be saved in the storage system of the research facility for period up to 10 years. Each data set will have a unique identifier.Access to raw data and metadata will be provided by the Facility via a F.A.I.R. on-line catalogue

9. Please, attach **a copy of your facility's data policy and data management plan**, where applicable.

Section related to integration with NA1 WP 2 Task 2.3 - CALIPSOPlus PROJECT

1. Is the facility's user office connected to UMBRELLA? YES ___ NO X

1	2	3	4

2. How is it available?

Since the facility is very small, the connection between Users and the Facility is by e-mail, filling in and sending/receiving the requested data/information.

Survey

As part of **NA1 WP2 Task 2.3 Deliverables 2.8 and 2.9**

Good Data Management is a prerequisite for Data integration.

The purpose of this survey is to:

- ✓ **learn about the status of your data policy as input for the harmonization of data policies roadmaps (NA1 WP2 Task 2.3)**
- ✓ **learn more about how your facility engages in research data management;**
- ✓ **learn about the degree of implementation of the facility's data management policies;**

Name of Institution: KIT (IBPT) ANKA

Date of submission _26_ / 03 / 2019

Comments apply to the data management of scientific data of the accelerator. There is no data management or policy for the beamline data yet. There are currently no plans to implement one.

Please, note: Deadline for submission is April 1st 2019 h 12:00 PM

In the questions below, please specify your institution's level of engagement

0 – No activity

1 – Information gathering, explorative

2 – Specification and development

3 – Partly implemented

4 – Fully implemented

1	**X**	3	4

Section related to INTERNAL Data Policy and Data Management Plan of your facility

1. Does your facility have a data policy? YES ___ NO X __

1	X	3	4

2. If yes, is it compatible with PanData? YES ___ NO __X

   a. Date of implementation: _____

1	2	3	4

3. Is your data policy compliant with the GDPR and the implementation of the policy?

   YES _X__ NO __

1	2	3	5

4. Is your data policy aligned with the F.A.I.R. data principles?

   YES _X__ NO __

5. What is the facility's strategy on data and/ or Metadata management?

Please, describe

_____

Strategy is being developed. We are investigating the EUDAT services for data management and are harvesting metadata of the accelerator. Prototype is in operation._____

_____

6. Does your facility have a Metadata catalogue? YES ___ NO _x

a. If yes, what kind?

iCat _____ SciCat _____ Other __EUDAT_____

b. If not, when is your facility planning to implement one?

Please, specify

_____

_____

1	x	3	4

7. Have you validated and implemented a data management plan?

YES ___ NO _x_

a. If yes, date of implementation: _____

b. If no, when are you planning to fully implement it: _____

1	x	3	4

8. Please, describe the data life cycle process inherited from your data policy, based on your Metadata catalogue storage and processes:

_____
_____
_____
_____

9. Please, attach **a copy of your facility's data policy and data management plan**, where applicable.

Section related to integration with NA1 WP 2 Task 2.3 - CALIPSOPlus PROJECT

1. Is the facility's user office connected to UMBRELLA? YES _x__ NO __

1	2	3	4

2. How is it available?

__Login to proposal management system https://proposal.ibpt.kit.edu/anna/
_____
_____
_____

Edited by Dr. Michael Hagelstein and Dr. Wolfgang Mexner
for IBPT at Karlsruhe Institute for Technology.

Survey

As part of **NA1 WP2 Task 2.3 Deliverables 2.8 and 2.9**

Good Data Management is a prerequisite for Data integration.

The purpose of this survey is to:

- ✓ **learn about the status of your data policy as input for the harmonization of data policies roadmaps (NA1 WP2 Task 2.3)**
- ✓ **learn more about how your facility engages in research data management;**
- ✓ **learn about the degree of implementation of the facility's data management policies;**

Name of Institution: _MAX IV laboratory

Date of submission 27 / 03 / 2019

Please, note: Deadline for submission is April 1st 2019 h 12:00 PM

In the questions below, please specify your institution's level of engagement

0 – No activity

1 – Information gathering, explorative

2 – Specification and development

3 – Partly implemented

4 – Fully implemented

1	2	3	4

Section related to INTERNAL Data Policy and Data Management Plan of your facility

1. Does your facility have a data policy? YES 3 NO __

1	2	3	4

2. If yes, is it compatible with PanData? YES ___ NO 2

   a. Date of implementation: 19th January 2017 (https://www.maxiv.lu.se/users/user-policies/) Experimental Data Policy

1	2	3	4

3. Is your data policy compliant with the GDPR and the implementation of the policy?

   YES / NO (unclear, no activity)

1	2	3	4

4. Is your data policy aligned with the F.A.I.R. data principles?

YES ___ NO 3

5. What is the facility's strategy on data and/ or Metadata management?
   Max IV is funded for 5 years to upgrade the Data Storage and Management in a project called (DataStaMP). Data will be curated for many years and in addition extra services will be implemented to increase the value of the stored data and improve access and tools for researchers._

6. Does your facility have a Metadata catalogue? YES (for MX) NO __

   a. If yes, what kind?

   iCat _____ SciCat __X_ Other _ISPyB _____

   b. If not, when is your facility planning to implement one?

Please, specify

_____
_____
_____

1	2	3	4

7. Have you validated and implemented a data management plan?

   YES ___ NO X

   a. If yes, date of implementation: _____
   b. If no, when are you planning to fully implement it: no date set, it is pending to request to user office to start asking users if they have a DMP when they apply for beam time.

1	2	3	4

8. Please, describe the data life cycle process inherited from your data policy, based on your Metadata catalogue storage and processes:

   All MAX IV systems use the same ID management for users so that their data can be automatically handled from generation to central storage. This is a work in progress, including the development of functionality in the meta-data catalogue to provide data management services.

9. Please, attach **a copy of your facility's data policy and data management plan**, where applicable:
https://www.maxiv.lu.se/users/user-policies/) Experimental Data Policy

Section related to integration with NA1 WP 2 Task 2.3 - CALIPSOPlus PROJECT

1. Is the facility's user office connected to UMBRELLA? YES ___ NO 0

1	2	3	4

2. How is it available?

_____
_____
_____
_____
_____

Survey

As part of **NA1 WP2 Task 2.3 Deliverables 2.8 and 2.9**

Good Data Management is a prerequisite for Data integration.

The purpose of this survey is to:

- ✓ **learn about the status of your data policy as input for the harmonization of data policies roadmaps (NA1 WP2 Task 2.3)**
- ✓ **learn more about how your facility engages in research data management;**
- ✓ **learn about the degree of implementation of the facility's data management policies;**

Name of Institution: _____PAUL SCHERRER INSTITUT_____

Date of submission 01 / 04 / 2019

Please, note: Deadline for submission is April 1st 2019 h 12:00 PM

In the questions below, please specify your institution's level of engagement

0 – No activity

1 – Information gathering, explorative

2 – Specification and development

3 – Partly implemented

4 – Fully implemented

1	2	3	4

Section related to INTERNAL Data Policy and Data Management Plan of your facility

1. Does your facility have a data policy? YES X NO __

1	2	3	4

2. If yes, is it compatible with PanData? YES X NO __

    a. Date of implementation: AUGUST 2016

1	2	3	4

3. Is your data policy compliant with the GDPR and the implementation of the policy?

    YES X NO __

1	2	3	5

4. Is your data policy aligned with the F.A.I.R. data principles?

    YES ___ NO X

5. What is the facility's strategy on data and/ or Metadata management?

Please, describe
WE ARE DEVELOPING THE SciCat METADATA CATALOGUE IN COLLABORATION WITH ESS. THIS CATALOGUE IS ROLLED OUT BEAMLINE BYB BEAMLINE AT PSI AND WILL BE FULLY COMPATIBLE WITH iCaT.

6. Does your facility have a Metadata catalogue? YES X NO __

a. If yes, what kind?

iCat _____ SciCat X Other _____

b. If not, when is your facility planning to implement one?

Please, specify

_____

_____

1	2	3	4

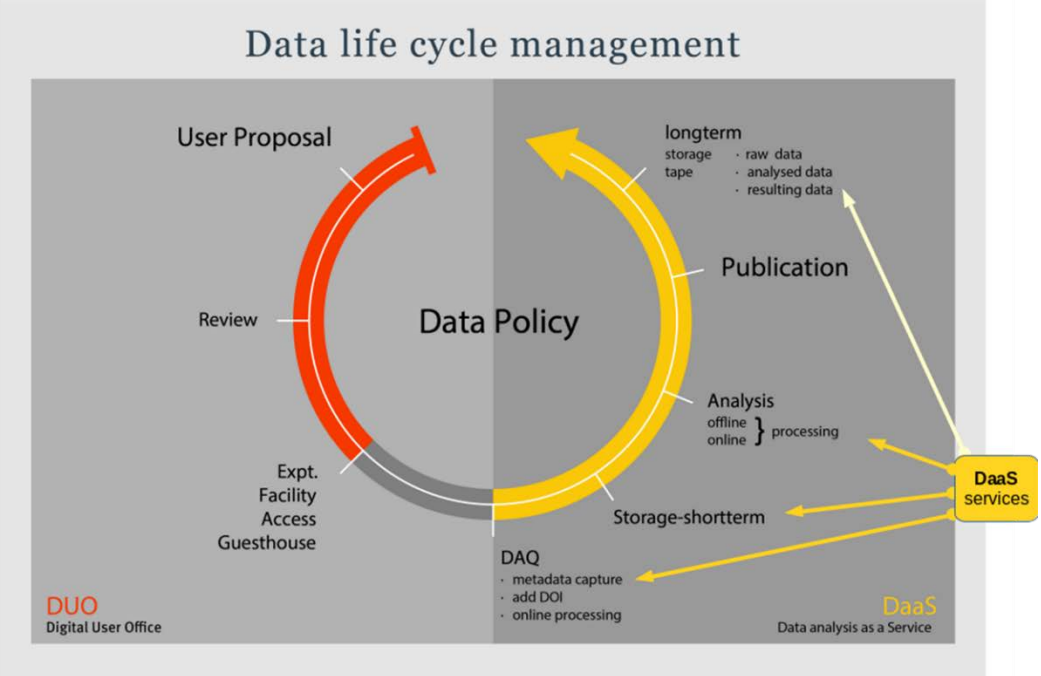
7. Have you validated and implemented a data management plan?

YES X NO __

a. If yes, date of implementation: SEPTEMBER 2017
b. If no, when are you planning to fully implement it: _____

1	2	3	4

8. Please, describe the data life cycle process inherited from your data policy, based on your Metadata catalogue storage and processes:



9. Please, attach **a copy of your facility's data policy and data management plan**, where applicable.

Section related to integration with NA1 WP 2 Task 2.3 - CALIPSOPlus PROJECT

1. Is the facility's user office connected to UMBRELLA? YES X NO __

1	2	3	4

2. How is it available?

IT IS AVAILABLE THROUGH DIGITAL USER OFFICE DUO.
ACCESS FOR EXTERNAL USERS HAPPENS FROM THE PARK INNOVAARE

Survey

As part of **NA1 WP2 Task 2.3 Deliverables 2.8 and 2.9**

Good Data Management is a prerequisite for Data integration.

The purpose of this survey is to:

- ✓ **learn about the status of your data policy as input for the harmonization of data policies roadmaps (NA1 WP2 Task 2.3)**
- ✓ **learn more about how your facility engages in research data management;**
- ✓ **learn about the degree of implementation of the facility's data management policies;**

Name of Institution: SESAME

Date of submission: 10 April 2019

Please, note: Deadline for submission is April 1st 2019 h 12:00 PM

In the questions below, please specify your institution's level of engagement

0 – No activity

1 – Information gathering, explorative

2 – Specification and development

3 – Partly implemented

4 – Fully implemented

1	* 2	3	4

Section related to INTERNAL Data Policy and Data Management Plan of your facility

1. Does your facility have a data policy? YES ___ NO __

NO

1	2	3	4

2. If yes, is it compatible with PanData? YES ___ NO __

   a. Date of implementation: _____

1	2	3	4

3. Is your data policy compliant with the GDPR and the implementation of the policy?

   YES ___ NO __

1	2	3	4

4. Is your data policy aligned with the F.A.I.R. data principles?

   YES ___ NO __

Yes, it will address the relevant aspects of making data FAIR – Findable, Accessible, Interoperable and Re-usable

5. What is the facility's strategy on data and/ or Metadata management?
Please, describe

We are developing our strategy, no clear ideas so far. Most likely, SESAME will keep both of the experimental raw data and metadata for each experiment for a long-term (a minimum of 5 years) on a best effort basis

6. Does your facility have a Metadata catalogue? YES ___ NO NO

a. If yes, what kind?

iCat _____ SciCat _____ Other _____

b. If not, when is your facility planning to implement one?

Please, specify

In the context of BEATS project, metadata catalogue will be ready within the next three years

1	2	3	4

7. Have you validated and implemented a data management plan?

YES ___ NO __

a. If yes, date of implementation: _____

b. If no, when are you planning to fully implement it: Within one year (in the context of BEATS Project)

1	2	3	4

8. Please, describe the data life cycle process inherited from your data policy, based on your Metadata catalogue storage and processes:

Data policy is not ready yet, however we have a proposed data life cycle which can be generally described as "1. Data will be collected from the data acquisition system and temporary stored at the beamline 2. Experimental data and metadata will be stored in hdf5 format for 6 months (short-term storage) 3. Any data older than 6 months will be archived in the long-term storage for 5 years at least"

9. Please, attach **a copy of your facility's data policy and data management plan**, where applicable.

Section related to integration with NA1 WP 2 Task 2.3 - CALIPSOPlus PROJECT

1. Is the facility's user office connected to UMBRELLA? YES ___ NO NO

1	2	3	4

2. How is it available?

_____
_____
_____
_____
_____

Survey

As part of **NA1 WP2 Task 2.3 Deliverables 2.8 and 2.9**

Good Data Management is a prerequisite for Data integration.

The purpose of this survey is to:

- ✓ **learn about the status of your data policy as input for the harmonization of data policies roadmaps (NA1 WP2 Task 2.3)**
- ✓ **learn more about how your facility engages in research data management;**
- ✓ **learn about the degree of implementation of the facility's data management policies;**

Name of Institution: SOLARIS National Synchrotron Radiation Centre

Date of submission 26 / 03 / 2019

Please, note: Deadline for submission is April 1st 2019 h 12:00 PM

In the questions below, please specify your institution's level of engagement

0 – No activity

1 – Information gathering, explorative

2 – Specification and development

3 – Partly implemented

4 – Fully implemented

1	2	3	4

Section related to INTERNAL Data Policy and Data Management Plan of your facility

1. Does your facility have a data policy? YES _x_ NO __

1	2	3	4

2. If yes, is it compatible with PanData? YES __ NO _x_

    a. Date of implementation: _____

1	2	3	4

3. Is your data policy compliant with the GDPR and the implementation of the policy?

YES _x_ NO __

1	2	3	4

4. Is your data policy aligned with the F.A.I.R. data principles?

YES ___ NO _x_

5. What is the facility's strategy on data and/ or Metadata management?

Please, describe:
1. Provide access to Digital User Office via Umbrella or eduGAIN.
2. Implement F.A.I.R. Data Management Principles.
3. Adopt ESRF solutions on metadata tagging based on ICAT.
4. Provide data storage mechanism based on HDF5 or NeXus.

6. Does your facility have a Metadata catalogue? YES ___ NO _x_

a. If yes, what kind?

iCat _____ SciCat _____ Other _____

b. If not, when is your facility planning to implement one?

Please, specify

_____Year 2019_____

_____

1	2	3	4

7. Have you validated and implemented a data management plan?

YES ___ NO _x_

a. If yes, date of implementation: _____
b. If no, when are you planning to fully implement it: __Year 2019_____

1	2	3	4

8. Please, describe the data life cycle process inherited from your data policy, based on your Metadata catalogue storage and processes:

SOLARIS Data Policy is based on our actual experience (our practical procedures to save and use the data from the experiments performed up to 2019). We still do not have an official Data Policy.

_____

9. Please, attach **a copy of your facility's data policy and data management plan**, where applicable.

Section related to integration with NA1 WP 2 Task 2.3 - CALIPSOPlus PROJECT

1. Is the facility's user office connected to UMBRELLA? YES ___ NO _x_

1	2	3	4

2. How is it available?

_____
_____
_____
_____
_____

Survey

As part of **NA1 WP2 Task 2.3 Deliverables 2.8 and 2.9**

Good Data Management is a prerequisite for Data integration.

The purpose of this survey is to:

- ✓ **learn about the status of your data policy as input for the harmonization of data policies roadmaps (NA1 WP2 Task 2.3)**
- ✓ **learn more about how your facility engages in research data management;**
- ✓ **learn about the degree of implementation of the facility's data management policies;**

Name of Institution: Synchrotron SOLEIL

Date of submission 27 / 03 /2019

Please, note: Deadline for submission is April 1st 2019 h 12:00 PM

In the questions below, please specify your institution's level of engagement

0 – No activity

1 – Information gathering, explorative

2 – Specification and development

3 – Partly implemented

4 – Fully implemented

1	2	3	4

Section related to INTERNAL Data Policy and Data Management Plan of your facility

1. Does your facility have a data policy? YES X NO _

1	2	3 - X	4

2. If yes, is it compatible with PanData? YES X NO _

   a. Date of implementation: _starting October 2018___

1	2	3 - X	4

3. Is your data policy compliant with the GDPR and the implementation of the policy?

   YES X NO _

1	2	3	4 - X

4. Is your data policy aligned with the F.A.I.R. data principles?

   YES X NO _

5. What is the facility's strategy on data and/ or Metadata management?
   Please, describe

SOLEIL strategy is to encourage its Users and Scientists to produce FAIR data & to facilitate this with appropriate data management services, in coordination with EXPANDS (EOSC Photon and Neutron Data Services Project) activities and by taking into account results of other collaborative works or projects as JRA2. SOLEIL requires a DMP for each proposal; will provide Users with automatically collected metadata for all experiments carried out on its beamlines; will act as a custodian for raw data and associated metadata - read-only, long-term curation of up to five years (striving for 10 years), DOI, made open access (CC-BY license) after an embargo period, available via a searchable on-line catalog; aims at providing means for reduction and/or processing of raw data.

6. Does your facility have a Metadata catalogue? YES ___ NO X

   a. If yes, what kind?

      iCat _____ SciCat _____ Other _____

   b. If not, when is your facility planning to implement one?

Please, specify

Aiming 2020-2021
iCat previously tested at SOLEIL; other solutions to be compared in coordination with EXPANDS activities

1 - X	2	3	4

7. Have you validated and implemented a data management plan?

   YES ___ NO X

   a. If yes, date of implementation:_____
   b. If no, when are you planning to fully implement it:_____Aiming 2020-2021_____

1 - X	2	3	4

8. Please, describe the data life cycle process inherited from your data policy, based on your Metadata catalogue storage and processes:

Raw data and associated metadata will be read-only for the duration of their life-time, migrated or copied to long-term storage facilities for long-term curation of up to five years (striving for 10 years) starting with the end of the respective beamtime. The experimental team will have exclusive access to data for a three-year embargo period, renewable twice a year (or more if needed). After the embargo, the data will be released under CC-BY license with open access to anyone registered on the SOLEIL dedicated portal. Data generated on the beamlines will automatically have a unique persistent identifier (DOI) making it possible to cite data.

9. Please, attach **a copy of your facility's data policy and data management plan**, where applicable.

see SOLEIL data policy at https://www.synchrotron-soleil.fr/fr/file/11309/download?token=tSyINMmx

Section related to integration with NA1 WP 2 Task 2.3 - CALIPSOPlus PROJECT

1. Is the facility's user office connected to UMBRELLA? YES X NO _

1	2	3	4 - X

2. How is it available?

via the SOLEIL user portal (i.e.: SUN set at https://sun.synchrotron-soleil.fr/sunset)

Survey

As part of **NA1 WP2 Task 2.3 Deliverables 2.8 and 2.9**

Good Data Management is a prerequisite for Data integration.

The purpose of this survey is to:

- ✓ **learn about the status of your data policy as input for the harmonization of data policies roadmaps (NA1 WP2 Task 2.3)**
- ✓ **learn more about how your facility engages in research data management;**
- ✓ **learn about the degree of implementation of the facility's data management policies;**

Name of Institution: Turkish Accelerator and Radiation Laboratory –TARLA / Ankara University Institute of Accelerator Technologies

Date of submission 01 / 04 / 2019

Please, note: Deadline for submission is April 1st 2019 h 12:00 PM

In the questions below, please specify your institution's level of engagement

0 – No activity

1 – Information gathering, explorative

2 – Specification and development

3 – Partly implemented

4 – Fully implemented

1			

Section related to INTERNAL Data Policy and Data Management Plan of your facility

1. Does your facility have a data policy? YES X NO __

1			

2. If yes, is it compatible with PanData? YES X_ NO __

    a. Date of implementation: _____

1			

3. Is your data policy compliant with the GDPR and the implementation of the policy?

YES _X__ NO __

1			

4. Is your data policy aligned with the F.A.I.R. data principles?

TARLA Data Policy is in preparation phase at the moment. The data policy has been documented in accordance with PanData, GDPR and F.A.I.R. data principles.

5. What is the facility's strategy on data and/ or Metadata management?

Please, describe

Decision how data/metadata would be used in the TARLA
Data/metadata Governance and Stewardship approach
Data/metadata usages
Identify sources and where data/metadata will be stored
Determine ownership data/metadata
Maintaining and securing the metadata, etc.

6. Does your facility have a Metadata catalogue? YES ___ NO _X_

We are planning to have an Online-Catalogue. It will prepared in accordance with ICAT.

a. If yes, what kind?

iCat _____ SciCat _____ Other _____

b. If not, when is your facility planning to implement one?

Please, specify

within two years when the TARLA instruments open to serve

1			

7. Have you validated and implemented a data management plan?

YES ___ NO _X_

a. If yes, date of implementation: _____
b. If no, when are you planning to fully implement it: within 2 years

1			

8. Please, describe the data life cycle process inherited from your data policy, based on your Metadata catalogue storage and processes:

In draft TARLA data Policy we are planning to perform the following procedure:

Access to Raw Data and Metadata obtained from an experiment is restricted to the Experimental Team for an embargo period of three (3) years after the end of the experiment. Thereafter, the data will become openly accessible. Any PI that wishes data to retain restricted access for a period longer than three (3) years will have this possibility on a yearly basis on a maximum prolongation of two (2) years. For longer extension a written request shall be submitted, specifying the reasons for the proposed prolongation, to the head of the corresponding TARLA division who decides on the request, all at their sole discretion. Data can always be made openly accessible earlier on simple request of the PI. In exceptional circumstances the head of the corresponding TARLA division can grant access to official committees at any time for the purpose of verifying data integrity.

9. Please, attach **a copy of your facility's data policy and data management plan**, where applicable.

Section related to integration with NA1 WP 2 Task 2.3 - CALIPSOPlus PROJECT

1. Is the facility's user office connected to UMBRELLA? YES ___ NO X

1			

2. How is it available?

Since the TARLA is under construction at the moment, there is no beamtime available. Therefore membership of UMBRELLA already exists but the connection between user office and UMBRELLA is still in progress

ANNEX 2

DATA POLICIES

Generic data management policy at ALBA CELLS.

This document describes the scientific data management policy at the synchrotron ALBA at the CELLS consortium.

1 General principles

1.1 This data management policy covers the access and ownership of the experimental raw data and metadata collected and/or stored at ALBA-CELLS.
1.2 Acceptance of this policy is a condition of the award of public beam time.
1.3 Users must not attempt to access, exploit or distribute data or metadata under the terms of this policy unless they are entitled to do so.
1.4 Deliberate infringements of the policy may lead to denial of access to data and metadata and the denial of future beam time proposals.
1.5 All data and metadata will be managed subject to the Spanish Data Protection legislation (Ley Orgánica de Protección de Datos LOPD).

http://www.cells.es/en/users/applying-for-beam-time

2 Definitions

Experiments carried out at ALBA can be classified as public access (including the regular user calls, friendly users and training experiments, in-house research, access derived from formal collaboration agreements and the use of management contingency beam-time) and proprietary access.
Public access experiments are covered by this policy.
Proprietary access experiments are carried out keeping the proposal, raw data and results confidential and therefore are not covered by this policy.

For the purposes of this policy:

2.1 The term facility refers to the ALBA Synchrotron at CELLS.
2.2 The term ***raw data*** refers to data collected from experiments performed on Beamlines or other instruments at ALBA-CELLS. This definition includes data that are created automatically or manually by facility specific software and/or ALBA-CELLS staff expertise in order to facilitate subsequent analysis of the experimental data, unless otherwise agreed.
2.3 The term ***metadata*** describes information referring to data collected from ALBA-CELLS instruments, including (but not limited to) the context of the experiment, the experimental team, experimental conditions and other logistical information.
2.4 The term ***principal investigator*** (PI) pertains to the PI identified on the ALBA-CELLS academic proposal (for peer-reviewed experiments) or the Safety approval form for any other proposals (like in-house, training and user-friendly experiments, special international access, etc.).
2.5 The term ***experimental team*** includes the PI and any other person to whom the PI designates the right to access resultant raw data and associated metadata.

2.6 The term ***on-line catalogue*** refers to a computer database of metadata containing links to raw data files, that can be accessed by a variety of methods, including (but not limited to) www-based browsers.

2.7 The term ***result*** refers to data, intellectual property, and outcomes arising from the analysis of raw data. This does not include publications.

2.8 The term ***open access*** means belonging to the community at large, unprotected by copyright or patent and subject to appropriation by anyone. It also means unrestricted (but not anonymous) and free-of-charge access. The ALBA-CELLS data archive will be made available under CC-BY

(Creative Commons BY, http://creativecommons.org/licenses/by/4.0/legalcode)

3 Raw data and associated metadata

3.1 Access to raw data and associated metadata

3.1.1 All raw data and the associated metadata obtained as a result of public access experiments, as defined in section 2 above, will be open access after an initial embargo period during which access is restricted to the experimental team, represented by the PI.

3.1.2 ALBA-CELLS is the custodian of the raw data and associated metadata.

3.1.3 All raw data and the associated metadata obtained as a result of proprietary access to ALBA-CELLS, will be owned exclusively by the client who purchased the access and is not covered by this data policy. Permissions and relevant procedures are in place to guarantee the confidentiality. Data from proprietary research will be removed after the experiment from the ALBA-CELLS storage systems, unless otherwise agreed with ALBA-CELLS management before the start of any experiment.

3.2 Curation of raw data and associated metadata

3.2.1 Raw data may or may not be curated at ALBA-CELLS premises. The raw data will be curated in well-defined formats, for which the means of reading the data will be made available by ALBA-CELLS.

3.2.2 Metadata that is automatically captured by instruments will be curated either within the raw data files, within an associated on-line catalogue, or within both.

3.2.3 Only data with metadata generated by ALBA-CELLS will be archived.

3.2.4 Raw data and metadata will be read-only for the duration of their life time.

3.2.5 Raw data and metadata will be migrated or copied to archival facilities for long-term curation.

3.2.6 It is planned that each experiment and data set will have a unique persistent identifier. Anybody publishing results based on open access data must quote the same identifier (and related publications if available and required).

3.2.7 High level metadata such as Title, Authors, Abstract, Beamline will be made public as soon as the experiment has been carried out. This information will be available via the persistent identifier landing page on the web.

3.3 Access to raw data and metadata

3.3.1 Access to raw data and metadata will be via suitable list/search tools.

3.3.2 Access to data will be restricted to those who are registered as users at ALBA-CELLS. Any retrieval of data may require access to permanent store with the consequent latencies and supplementary waiting times.

3.3.3 ALBA–CELLS shall custody the data for a minimum of 5 years. On a best effort basis, ALBA-CELLS will keep the data accessible online for one year, after which it can be archived on a tape library or a warehouse with the consequent latencies and waiting times. Access to raw data and the associated metadata obtained from an experiment is restricted to the experimental team for 3 years. After the embargo period is over the data and metadata will become publicly accessible. Any PI that wishes their data not to become "publicly accessible" for a longer period will be required to make a special case to the Director of ALBA–CELLS.

3.3.4 It is the responsibility of the PI to ensure that the data is stored into the directories enabled to that purpose and that the experiment number is correctly entered into the metadata for each raw data set, in order to correctly associate each data set with the PI. If this is not done, the experimental team will not be able to access the data, or other users may inadvertently be given access rights to the data.

3.3.5 Authorized ALBA-CELLS staff (e.g. instrument scientists, computing group members) have access to any facility curated data or metadata for facility related purposes. ALBA-CELLS undertakes the compromise that they will preserve the confidentiality of such data during the embargo period.

3.3.6 The on-line catalogue will enable linking experimental data to experimental proposals. Access to proposals will only ever be provided to the experimental team and appropriate ALBA-CELLS staff, unless otherwise authorized by the PI during the 3-year embargo period.

3.3.7 The PI has the possibility to transfer parts of the totality of her/his rights during the embargo period to another registered person.

3.3.8 The PI has the possibility to create and distribute copies of the raw data.

4 Results

4.1 Ownership of results

4.1.1 Ownership of all results (intellectual property) derived from the analysis of the raw data is determined by the contractual obligations of the person(s) performing the analysis, as well as by the copyright agreements assumed upon publication.

4.1 Curation of results

4.1.1 The results of the analysis performed at ALBA-CELLS can be stored at ALBA-CELLS premises if the PI decides so. It will not be the responsibility of ALBA-CELLS to fully curate this data e.g. to ensure that software to read / manipulate this data is available.

4.1.2 ALBA-CELLS can not be made liable in case of unavailability or loss of data, results or data analysis software.

4.2 Access to results

4.2.1 Access to the results of analysis performed on raw data and metadata during the embargo period is restricted to the experimental team in charge of the analysis, unless otherwise requested by the members of that team.

4.2.2 Appropriate facility staff (e.g. instrument scientists, computing group members) has access to any facility-curated data or metadata for facility related purposes. Alba-CELLS will undertake that confidentiality of such data is preserved during the embargo period.

5 Good practice for metadata capture and results storage

5.1 The experimental team is encouraged to ensure that experimental 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.

5.2 ALBA-CELLS undertakes to provide means, on a best effort basis, 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.

5.3 Researchers who aim to carry out analyses of raw data and metadata which are openly accessible should, where possible, contact the original PI to inform her/him and suggest a collaboration if required. Researchers must acknowledge the source of the data and cite its unique identifier as well as any publications linked to the same raw data.

5.4 PIs 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 publicly accessible.

6 Publication information

6.1 Publications related to data from experiments carried out at ALBA-CELLS must cite the persistent identifier of the experiment and data in their publication.

6.2 References for publications related to experiments carried out at ALBA-CELLS must be deposited in the ALBA-CELLS publications database within six months of the publication date, or during any new application for beamtime, whichever is the earlier. Non-compliance to this rule may result on not getting further beamtime allocation.

DIAMOND DATA POLICY:

https://www.diamond.ac.uk/Users/Policy-Documents/Policies/Experimental-Data-Management-Pol.html

Elettra
Sincrotrone
Trieste

Scientific Data Policy

Introduction

PANdata (Photon And Neutron data infrastrucure) brings together thirteen major world-class European research infrastructures to create a fully integrated, pan-European, information infrastructure supporting the scientific process (http://pan-data.eu).

PANdata like other similar initiatives in other research areas takes into account what reported by the high level expert group on scientific data (http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf) and what results from years of scientific research in the field of ICT and e-Infrastructures funded by the EC during FP6 and FP7 (http://cordis.europa.eu/fp7/ict/e-infrastructure/publications_en.html). All these efforts are trying to provide solutions to the data deluge and its related issues. Similar initiatives are progressing all around the world.

PaNdata has recently concluded the FP7 supported project named PaNdata Europe. PaNdata Europe, a pure support action, has developed a policy framework and laid the basic foundation for a sustainable data infrastructure, like an agreement on a data format standardization based on HDF5, design of a global authentication system or the layout of a software catalogue as a basis for generating automatic workflows.

PaNdata Open Data Infrastructure (PaNdata ODI), another FP7 supported projected that started November 2011 and finished in November 2014, took these developments to create a federated open data infrastructure, seamlessly integrating the existing user and data management systems of the European photon and neutron facilities. This initiative required each facility to define a scientific data policy. A policy means guidelines to develop the systems, common rules and in the end value for the researchers.
Here are links to data policy documents of some research infrastructures:
http://www.isis.stfc.ac.uk/user-office/data-policy11204.html
http://www.stfc.ac.uk/Resources/pdf/STFC_Scientific_Data_Policy.pdf
http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx
http://www.ill.eu/users/ill-data-policy/

Elettra Sincrotrone Trieste Scientific Data Policy

The scientific data policy issue involves the research facility and its users. **The data policy will be** published by the facility and **signed by the users before the beamtime.**
All raw **data** and the associated metadata obtained as a **result of publically funded access** to the research facility **will become open access after a reasonable embargo period**, with the research facility acting as the custodian.
All raw data and the associated metadata obtained as a result of proprietary research will be owned exclusively by the clients who purchased the access.
Before the start of any experiment, proprietary users must agree with the facility management on how their raw data and metadata will be managed.
All raw data will be curated in well-defined formats, for which the means of reading the data will be made available by the facility.
Metadata that is automatically captured by instruments will be curated either within the raw data files, or within an associated on-line catalogue, or within both.
Data will be read-only for the duration of its lifetime. **A copy of the data will be saved in the storage system of the research facility for period ranging from a minimum of 15 days to 10 years.** Data will be migrated or copied to archival facilities for long-term curation. It is planned that each data set will have a unique identifier.
It is the **responsibility of the principal investigator to ensure that the proposal number is correctly entered** into the metadata for each raw data set. The principal investigator is the proponent of the beamtime.
An on-line catalogue will allow linking experimental data to experimental proposals. Access to raw data and metadata in the facility will be provided via a searchable on-line catalogue.
Access to raw data and the associated metadata is **restricted to the experimental team for a period of 3 years** after the end of the experiment.
After this period each year the principal investigator will be asked for permission to give open access to the raw data.

In case the permission is not given the access to raw data will remain restricted to the experimental team.

In case the permission is given the raw data will become openly accessible.

Data can always be made openly accessible earlier upon request of the principal investigator.

The principal investigator has the right to transfer or grant parts or all of his rights to another registered person and to create and distribute copies of raw data.

Ownership of all results (intellectual property) derived from the analysis of the raw data is determined by the contractual obligations of the person(s) performing the analysis.

The facility will provide a means for users to upload results and associated metadata to the facility and enable them to associate these results with raw data collected at the facility.

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.

Trieste, December 2014

Elettra - Sincrotrone Trieste S.C.p.A.
Prof. Alfonso Franciosi

30 November 2015

The ESRF Data Policy

The ESRF aims to implement a Data Policy starting as soon as possible in 2016. The main elements of this policy comprise:

- **Data ownership**
- **Data curation**
- **Data archiving**
- **Open access to data**

This policy follows largely the recommendations of the PaN-data Europe Strategic Working Group laying out a common framework for scientific data management at photon and neutron facilities (Deliverable D2.1, PaN-data Europe, co-funded by the European Commission under the 7th Framework Programme)

1. General Principles

1.1. The present data management policy pertains to the ownership of, the curation of and access to experimental data and metadata collected and/or stored at the ESRF.

1.2. Acceptance of this policy is a condition for the award of beam time.

1.3. Users must not attempt to access, exploit or distribute raw data or metadata unless they are entitled to do so under the terms of this policy.

1.4. Deliberate infringements of the policy may lead to denial of access to raw data or metadata and/or denial of future beam time requests at the ESRF, as well as actions of the ESRF in the court of law.

1.5. All data and metadata will be subject to the data protection legislation of France.

2. Definitions

For the purposes of this policy:

2.1. The term **raw data** pertains to data collected from peer-reviewed and in-house experiments performed on ESRF's instruments and includes data collected from peer-reviewed experiments performed on CRG beamlines. This definition includes data that are created automatically or manually by facility specific software and/or facility staff expertise in order to facilitate subsequent analysis of the experimental data.

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

2.3. The term **principle investigator** (PI) pertains to the PI identified on the Experiment Proposal (for peer-reviewed experiments) or the Safety Approval Form (for in-house experiments).

2.4. The term **experimental team** includes the PI and any other person to whom the PI designates the right to access resultant raw data and associated metadata.

2.5. The term **public research** refers to research done through peer review or access via in-house research beam time.

2.6. The term **proprietary research** refers to research done through purchased (commercial) access.

2.7. The term **on-line catalogue** pertains to a computer database of metadata containing links to raw data files, that can be accessed by a variety of methods, including (but not limited to) web-based browsers.

2.8. The term **result** pertains to data, intellectual property, and outcomes arising from the analysis of raw data. This does not include publications.

2.9 The term **custodian** refers to the Institute storing, curating and providing access to raw data, metadata and results.

2.10. The term **long-term** means a minimum of 5 years and the ESRF will strive for 10 years. This will depend on the type and volume of data concerned and the economical consequences associated with long-term data storage. Thus the ESRF reserves the right to restrict the storage periods or data sets in consultation with the respective communities of high data rate instruments.

2.11. The term **open access** means belonging to the community at large, unprotected by copyright or patent and subject to appropriation by anyone. The ESRF data archive will be made available under CC-BY

(Creative Commons BY, http://creativecommons.org/licenses/by/4.0/legalcode).

3. Raw data and associated metadata

3.1 Access to raw data and associated metadata

3.1.1. All raw data and the associated metadata obtained as a result of peer reviewed access to the ESRF, in-house research and use of Management Contingency beamtime excluding proprietary research will be open access after an initial embargo period during which access is restricted to the experimental team, represented by the PI.

3.1.2 ESRF is the custodian of the raw data and associated metadata.

3.1.3. All raw data and the associated metadata obtained as a result of proprietary research will be owned exclusively by the client who purchased the access and is not covered by the ESRF data policy. Data from proprietary research will be removed after the experiment from ESRF disk storage, unless otherwise agreed with ESRF management before the start of the experiment.

3.2 Curation of raw data and associated metadata

3.2.1. All raw data and metadata will be curated in well-defined formats, for which the means of reading the data will be made available by the ESRF.

3.2.2. Metadata that are automatically captured by instruments will be curated either within the raw data files, within an associated on-line catalogue, or within both.

3.2.3. Only data with metadata generated by ESRF software will be archived.

3.2.4. Raw data and metadata will be read-only for the duration of their life time.

3.2.5. Raw data and metadata will be migrated or copied to archival facilities for long-term curation.

3.2.6. It is planned that each experiment and data set will have a unique persistent identifier. Anybody publishing results based on open access data must quote the same identifier (and related publications if available & required).

3.2.7 High level metadata such as Title, Authors, Abstract, Beamline will be made public as soon as the experiment has been carried out. This information will be available via the persistent identifier landing page on the web.

3.3 Access to raw data and metadata

3.3.1. Access to raw data and metadata is foreseen to be via a searchable on-line catalogue.

3.3.2. Access to the on-line catalogue of the ESRF will be restricted to registered users of the on-line catalogue. The ESRF sets up the on-line procedure to become a registered user of the on-line catalogue.

3.3.3. Access to raw data and the associated metadata obtained from an experiment is restricted to the experimental team for an embargo period of 3 years after the end of the experiment. Thereafter, the data will become openly accessible. Any PI that wishes data to retain *restricted access* for a period longer than three years will have this possibility by submitting a written request, specifying the reasons for the proposed prolongation, to the ESRF Directors of Research who decide on the request. In exceptional circumstances, data can be made openly accessible earlier than 3 years if the PI or the ESRF Directors of Research inform the ESRF to do so.

3.3.4. It is the responsibility of the PI to ensure that the experiment number is correctly entered into the metadata for each raw data set.

3.3.5. Authorized ESRF staff (e.g. instrument scientists, computing group members) have access to any curated data or metadata for facility related purposes. ESRF will undertake that confidentiality of such data is preserved during the embargo period.

3.3.6 The on-line catalogue will enable linking experimental data to experimental proposals. Access to proposals will only be provided to the experimental team and appropriate facility staff, unless otherwise authorized by the PI.

3.3.7. The PI has the possibility to transfer parts or the totality of her/his rights during the embargo period to another registered person.

3.3.8. The PI has the possibility to create and distribute copies of the raw data.

4. Results

4.1 Ownership of results

4.1.1. Ownership of all results (intellectual property) derived from the analysis of the raw data is determined by the contractual obligations of the person(s) performing the analysis.

4.2 Curation of results

4.2.1 The ESRF will provide curation of results on a best effort basis, and acts as custodian of results in the long term.

4.2.2. The ESRF cannot be made liable in case of unavailability or loss of data or results.

4.2.3. The ESRF cannot be made liable in case of unavailability or loss of data analysis software.

4.3 Access to results

4.3.1. Access to the results of analysis performed on raw data and metadata is restricted to the person or persons performing the analysis, unless otherwise requested by those persons. However, if the raw data being analysed is still restricted, access to the analysis results must be granted by the PI on request.

5. Good practice for metadata capture and results storage

5.1. The experimental team is encouraged to ensure that experiments metadata are as complete as possible, as this will enhance the possibilities for everybody to search for, retrieve and interpret the data in the long term.

5.2. ESRF provides 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.

5.3. Researchers who aim to carry out analyses of raw data and metadata which are openly accessible should, where possible, contact the original PI to inform her/him and suggest a collaboration if required. Researchers must acknowledge the source of the data and cite its unique identifier as well as any publications linked to the same raw data.

5.4. PIs and researchers who carry out analyses of raw data and metadata are encouraged to link the results of these analyses to the raw data / metadata using the facilities provided by the on-line catalogue. Furthermore, they are encouraged to make such results openly accessible.

6. Publication information

6.1. Publications related to data from experiments carried out at ESRF must cite the persistent identifier of the experiment and data in their publication.

6 June 2017

Scientific Data Policy of European X-Ray Free-Electron Laser Facility GmbH

as approved by the Council on 30 June 2017

1 Preface

European XFEL GmbH aims to implement a Scientific Data Policy starting from 1 July 2017 in order to be prepared for the first beam operation of the European XFEL Facility.

The main elements of the Scientific Data Policy comprise:

- Data ownership
- Data curation
- Data archiving
- Open access to data

This Scientific Data Policy closely follows the recommendations of the PaN-data Europe Strategic Working Group laying out a common framework for scientific data management at photon and neutron facilities (Deliverable D2.1, PaN-data Europe, co-funded by the European Commission under the 7th Framework Programme).

2 Definitions

For the purposes of this Scientific Data Policy:

2.1. The term ***European XFEL instruments*** shall mean all instruments designed to perform scientific experiments and located at the European XFEL Facility, including those contributed by third party organizations and user consortia.

2.2. The term ***scientific data*** shall mean data collected from peer-reviewed and in-house experiments performed on European XFEL 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 in order to facilitate subsequent analysis of the experiment data.

2.3. The term ***raw data*** shall mean a category of the scientific data that is recorded during experiments, cannot be derived from other persistent data, and is registered in the metadata catalogue.

2.4. The term ***processed data*** shall mean a category of the scientific data that is derived from raw data.

2.5. The term ***calibrated data*** shall mean a subcategory of the processed data that is obtained from the raw data by applying detector-specific corrections.

2.6. The term ***calibration data*** shall mean a subcategory of the processed data that describes detector correction factors.

2.7. The term ***alignment data*** shall mean a subcategory of the processed data that is obtained from dedicated alignment measurements, providing the position of detectors and their relevant components.

2.8. The term ***results*** pertains to a subset of processed data and other outcomes arising from the analysis of raw data, excluding publications based on such analysis and intellectual property rights out of such analysis or resulting from the registration of a right based on results.

2.9. The term ***metadata*** shall mean information collected in relation to the scientific data, including (but not limited to) the information regarding the context of the experiment, the experiment team, experiment conditions, and other logistical information.

2.10. The term ***principal investigator*** (***PI***) pertains to the person identified as PI on the experiment proposal. For experiments outside of the European XFEL GmbH's proposal system, the PI is the person who initiated the experiment proposal and obtained the management board approval of the European XFEL GmbH.

2.11. The ***proposal team*** includes the PI and any other registered user of the European XFEL Facility who is defined as a member of the proposal team in the European XFEL user portal (UPEX).

2.12 The term ***experiment team*** includes the PI and any other registered user of the European XFEL Facility to whom the PI designates the right

to access resultant raw data and associated metadata in the European XFEL GmbH's proposal system. For experiments outside of the European GmbH's proposal system, the experiment team is defined in the metadata catalogue.

2.13. The term ***beamtime*** shall mean time period allocated by European XFEL GmbH to the experiment team for performing scientific experiments on a European XFEL instrument. Every beamtime has a well-defined start date, end date, and experiment team.

2.14. The term ***public research*** refers to research done through peer review or access via in-house research beamtime.

2.15. The term ***proprietary research*** refers to research done through purchased (commercial) access.

2.16. The term ***metadata catalogue*** pertains to a computer database of metadata maintained and provided by the European XFEL GmbH containing links to data files, which can be accessed by a variety of methods, including but not limited to web-based browsers.

2.17. The term ***custodian*** in the context of this Scientific Data Policy refers to the European XFEL GmbH which will store, curate, and provide access to raw data, metadata, and results.

2.18. The term ***long-term*** means a minimum of five years and the European XFEL GmbH will strive for 10 years (***storage period***), starting with the end of the respective beamtime. The precise period will depend on the type and volume of data concerned and the economic consequences associated to long-term data storage. Thus, the European XFEL GmbH reserves the right to restrict the storage periods of data sets in consultation with the respective communities of high data-rate instruments.

2.19. The term ***open access*** shall mean the generally unrestricted (but not anonymous) and free-of-charge access and use for the academic community. With respect to the European XFEL GmbH, open access shall mean that data will be published under CC-BY

(Creative Commons BY, http://creativecommons.org/licenses/by/4.0/legalcode).

2.20. The term ***European XFEL GmbH*** shall mean European X-Ray Free-Electron Laser Facility GmbH.

2.21. The term ***European XFEL Facility*** shall mean the research facility of European XFEL GmbH located in Hamburg (Bahrenfeld, Osdorfer Born) and in Schenefeld, Schleswig-Holstein, Germany.

2.22. The term ***embargo period*** shall mean a period of three years, starting from the end of the beamtime.

3 General principles

3.1. The present Scientific Data Policy pertains to the ownership, curation, archiving and access to scientific data and metadata collected and/or stored at the European XFEL GmbH. European XFEL GmbH may use subcontractors to perform its obligations under this Scientific Data Policy. These subcontractors act as data processors in case personal data are concerned.

3.2. Acceptance of this Scientific Data Policy is a condition for the award of beamtime.

3.3. Users must not attempt to access, exploit or distribute scientific data or metadata unless they are entitled to do so under the terms of this Scientific Data Policy.

3.4. Culpable and repeated non-culpable infringements of the Scientific Data Policy may lead to denial of access to scientific data or metadata and/or denial of future beamtime at the European XFEL GmbH, as well as actions of the European XFEL GmbH in the court of law.

3.5. If and to the extent scientific data and metadata include personal data, the data protection legislation of the Federal Republic of Germany and the European Union, respectively, will be applicable. This Scientific Data Policy will be governed by and constructed in accordance with the law of

the Federal Republic of Germany. Exclusive place of jurisdiction is Hamburg, Germany.

3.6. Users of the European XFEL Facility must be personally registered at the user portal (UPEX); for details regarding the collection and processing of personal data in connection with UPEX, the UPEX Privacy Policy applies[1]. For details regarding the collection and processing of personal data in connection with the use of the European XFEL Facility the data handling provisions of the Terms and Conditions for the non-proprietary user access to the European XFEL Facility[2] apply.

4 Raw data and associated metadata

European XFEL GmbH is the custodian of the raw data and the associated metadata during the storage period.

4.1 Access to raw data and associated metadata

4.1.1. All raw data and the associated metadata obtained as a result of public research will be made available as open access after an initial embargo period during which access is restricted to the experiment team, represented by the PI.

4.1.2. All raw data and the associated metadata obtained as a result of proprietary research will be owned exclusively by the client who purchased the access and is not covered by this Scientific Data Policy. Data from proprietary research will neither be made available as open access nor curated or stored at European XFEL GmbH. Data from proprietary research will be removed after the experiment from the European XFEL GmbH's storage, unless otherwise agreed with the

[1] The UPEX Privacy Policy is ready and available on the webpage of European XFEL GmbH.

[2] The Terms and Conditions for the non-proprietary user access to the European XFEL Facility are still under negotiations. A final version is not yet available.

management board of the European XFEL GmbH before the start of the experiment.

4.2 Curation of raw data and associated metadata

4.2.1. All raw data and metadata will be curated in well-defined formats, for which the means of reading the data will be made available by the European XFEL GmbH.

4.2.2. Metadata that are automatically captured by instruments will be curated either within the raw data files, within an associated metadata catalogue, or within both.

4.2.3. Only data with metadata generated by European XFEL GmbH's software will be archived.

4.2.4. Raw data will be read-only for the duration of its storage.

4.2.5. Raw data and metadata will be migrated or copied to archival facilities for long-term curation.

4.2.6. It is planned that each data set will have a unique persistent identifier. Anyone providing data with the same identifier must make sure that the copy is identical to the data in the facility repository. Anyone publishing results based on open access data must quote the same identifier (and related publications if available and required).

4.2.7. High-level metadata, such as title, authors, instrument, will be made public as soon as the experiment has been carried out. This information will be available via the persistent identifier landing page on the web.

4.3 Access to raw data and metadata

4.3.1. Access to raw data and metadata is foreseen to be via a searchable metadata catalogue.

4.3.2. Access to the metadata catalogue of the European XFEL GmbH will be restricted to registered users. The European XFEL GmbH sets up the

online procedure to become a registered user of the metadata catalogue.

4.3.3. 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.

4.3.4. It is the responsibility of the PI to ensure that the proposal, experiment numbers, and sample description are correctly entered through the control system or directly into the metadata for each raw data set.

4.3.5. Authorized European XFEL staff (e.g. instrument and detector scientists, computing group members) have access to any curated data or metadata for facility related purposes. European XFEL GmbH will undertake that the confidentiality of such data is preserved during the embargo period.

4.3.6. The metadata catalogue will enable linking scientific data to experiment proposals. Access to proposals will be provided only to the proposal team, the experiment team, the members of the proposal review panels, and appropriate facility staff, unless otherwise authorized by the PI.

4.3.7. The PI has the possibility to transfer the totality of her/his data-related rights and responsibilities during the embargo period to another person registered in the user portal (UPEX) as a user of the European XFEL Facility. The transfer must be documented in writing.

4.3.8. The PI has the possibility to delegate parts or the totality of her/his rights under this Scientific Data Policy during the embargo period to another person registered in the user portal (UPEX) as a user of the European XFEL Facility. The delegation must be documented in writing.

4.3.9. The PI has the possibility to create and distribute copies of the raw data. During the embargo period other experiment team members must obtain agreement from the PI to do so.

4.3.10. As a matter of precaution (and without prejudice to the question of ownership), all members of the experiment team grant the European XFEL GmbH the unlimited and unrestricted right to use the raw data and the metadata to the extent necessary to curate and make available the raw data and the metadata in accordance with this Scientific Data Policy.

5 Processed data and results

5.1 Ownership of results

5.1.1. Ownership, including any IP rights that might arise, of all results derived from the analysis of the raw data is determined by the contractual obligations of the person(s) performing the analysis respectively the applicable law.

5.1.2. In addition the clauses on intellectual property of the Terms and Conditions for the non-proprietary user access to the European XFEL Facility apply.

5.2 Curation of processed data and results

5.2.1. Unless otherwise stated in this Scientific Data Policy, the processed data from the interim analysis steps and the associated metadata hereto will not be curated for long-term by European XFEL GmbH. In particular, calibrated data and the associated metadata hereto will not be curated for long-term, but the tools will be provided to regenerate calibrated data from raw data using facility curated calibration data. A temporary storage will be provided to enable data analysis. The exact amount of available storage space and storage period for beamtime related processed data will be published by European XFEL GmbH.

5.2.2. All calibration and alignment data and the associated metadata obtained in the context of public research through the European XFEL GmbH provided services will be curated for long-term by the European XFEL Facility.

5.2.3. If agreed between PI and European XFEL GmbH and to the extent results are entered into its system, the European XFEL GmbH will provide curation of results on a best effort basis, and act as custodian of results for the storage period.

5.2.4. European XFEL GmbH will not be responsible for the full curation of the results, e.g. to ensure that software to read/manipulate this data is available.

5.2.5. As a matter of precaution (and without prejudice to the question of ownership), all members of the experiment team grant the European XFEL GmbH the unlimited and unrestricted right to use the processed data and the metadata hereto to the extent necessary to curate and make available the processed data and the metadata hereto in accordance with this Scientific Data Policy.

5.3 Access to results

5.3.1. Calibration and alignment data and associated metadata will be made available as open access without applying embargo period, except as otherwise agreed with European XFEL GmbH.

5.3.2. Access to the results of analysis performed on raw data and metadata and curated by European XFEL GmbH is restricted to the experiment team, unless otherwise requested by PI.

6 Warranty and liability regarding scientific data, metadata and results

6.1. European XFEL GmbH will at its own discretion use reasonable efforts to ensure an accurate storing and curating as well as an uninterrupted

access in accordance with the acknowledged IT standard. However, failures caused by technical or human mistakes cannot be ruled out regarding any data processing. European XFEL GmbH cannot warrant an absolutely accurate storing and curating. Also, access might be temporarily limited or impossible, especially due to necessary maintenance or overhaul services or failure of third-party service providers.

6.2. European XFEL GmbH shall not be liable in case of lost, inaccurate, or defective scientific data, metadata, or results as well as for access being limited or unavailable unless European XFEL GmbH, a representative, agent, or employee of European XFEL GmbH acted in a grossly negligent manner or intentionally.

7 Good practice for metadata captures and results storage

7.1. The experiment team is encouraged to ensure that metadata are as complete as possible, as this will enhance the possibilities for everyone to search for, retrieve, and interpret the data during the storage period.

7.2. European XFEL GmbH provides 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.

7.3. PIs and researchers who carry out analyses of scientific data and metadata are encouraged to link the results of these analyses to the raw data/metadata using the metadata catalogue. Furthermore, they are encouraged to make such results available on an open access basis.

8 Publication information

8.1. Publications related to raw data and metadata from experiments carried out at European XFEL Facility must cite the persistent identifier of the beamtime data in their publications.

8.2. References for publications related to experiments carried out at the European XFEL Facility must be deposited in the European XFEL publications database within three months of the publication date, or during any new application for beam time, whichever is earlier.

9 Termination of custodianship or metadata catalogue

If the European XFEL GmbH decides to not continue to act as custodian and/or to maintain and provide the metadata catalogue, the European XFEL GmbH will inform the PIs concerned in a timely manner and provide them with effective means to make a copy of the respective raw data, metadata, calibration data, alignment data, and results, provided European XFEL GmbH is aware of the effective email address of the PI at that time.

Data retention policy for the European XFEL

Storage class	Quota	Safety	Lifetime	Comment
dcache.raw	None	Tape Archive	6 months	Raw data on commodity disks
raw	None	None	2 months	Very fast accessible raw data, lifetime not guaranteed
usr	5TB	Snapshots + Tape Backup	24 months	User data, results
proc	None	None	6 months	Processed data (e.g. calibrated)
scratch	None	None	6 months	Temporary data (lifetime not guaranteed)
dcache.cal	None	Tape Archive	10 years	Calibration constants on commodity disks
cal	None	Tape archive	6 months	Very fast accessible calibration data
user home	20GB	Snapshots + Tape Backup	Lifetime of the account	Home folder for user account
archive.raw	None	-	Long-term	“Long term” means 5 years and XFEL will strive for 10 years
archive.cal	None	-	10 years	

HZB Data Policy

Introduction

The proper management of scientific data is imperative for safeguarding the integrity and reproducibility of scientific findings. The Deutsche Forschungsgemeinschaft (DFG) recommends in the Proposals for Safeguarding Good Scientific Practice [1]: "Primary data as the basis for publications shall be securely stored for ten years in a durable form in the institution of their origin."

Furthermore, the concept of open access to scientific results gains increasingly appreciation. In 2007, the Organisation for Economic Co-operation and Development (OECD) formulated Principles and Guidelines for Access to Research Data from Public Funding [2]. This document emphasizes the importance of openness and the free exchange of ideas, information and knowledge for the advancement of science and postulates that research data from public funding should be openly accessible. The Alliance of German Science Organisations took this idea up in the Principles for the Handling of Research Data [3] where 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 [4] in the context of their funding regulations.

The HZB approves the principle of open access to research data. In doing so, it strives for a careful balance between aspects of competition and collaboration in science. HZB supports its users to fulfill the requirements placed by funders and the scientific community. To this end, it provides the necessary infrastructure for the data management and regulates the access to the data in the present policy.

Last but not least, data management at HZB should be considered as a service to the researchers. It aims to help them to document their results and to ease the burden of the data archiving.

The present policy is based on a model formulated by the PaNdata Europe project [5].

1. General principles

1.1 This policy sets the rules for the management of scientific data collected by public research at HZB's large-scale facilities. This includes the ownership of, the curation of and access to the data.

1.2 Acceptance of this policy is a condition of the award of beamtime.

1.3 For the data from proprietary research, users must make a separate agreement with HZB management how they wish their data to be managed before the start of any experiment.

Version 1.1, 2017-01-19

2. Definitions

For the purposes of this policy:

2.1 The term *raw data* pertains to data collected from experiments performed on HZB instruments. This definition includes data that are created automatically or manually by facility specific software and/or facility staff expertise in order to facilitate subsequent analysis of the experimental data.

2.2 The term *results* pertain to data, intellectual property, and outcomes arising from the analysis of raw data. This does not include publications.

2.3 The term *metadata* describes information pertaining to other data, including (but not limited to) the context of the experiment, the experimental team, experimental conditions and other logistical information.

2.4 The term *public research* refers to research done during public funded beamtime granted by HZB after peer review of a proposal. Furthermore, all in-house research at HZB is public research for the purposes of this policy.

2.5 The term *proprietary research* refers to research done through purchased (commercial) access to HZB's facilities.

2.6 The term *experimental team* includes the proposer that have been awarded beamtime, the co-proposers, and any person involved in the creation of the raw data.

2.7 The term *data access team* includes the proposer, the co-proposers, and any other person designated by him/her to have access to the scientific data. Additional persons may be added on request by any member of the data access team, if no other member objects.

2.8 The term *open access* or *openly accessible* means that data is made freely available to everyone.

3. Raw data and associated metadata

3.1 The experimental team waives all copyright and related or neighboring rights together with all associated claims and causes of action with respect to the raw data and associated metadata according to the Creative Commons CC0 Dedication (see appendix).

3.2 All raw data will be curated in well-defined formats, for which the means of reading the data will be made available by HZB.

3.3 Associated metadata will be curated either within the raw data files, within an associated on-line catalogue, or within both.

3.4 Raw data and associated metadata will be stored by HZB for at least ten years.

3.5 Access to raw data and the associated metadata is restricted to the data access team for a period of five years after the end of the experiment. Thereafter, it will become openly accessible. Any member of the data access team that wishes their data to remain restricted access for a longer period will be required to make a special case to HZB management. Data can always be made openly accessible earlier on simple request of any member of the data access team, if no other member objects.

3.6 In deviation from 3.5, appropriate facility staff (e.g. instrument scientists, computing group members) has access to any HZB curated data or metadata for facility related purposes. HZB will undertake that they will preserve the confidentiality of such data.

3.7 Any member of the data access team has the right to create and distribute copies of his raw data.

4. Results

4.1 Copyright and related or neighboring rights with respect to results are not affected by this policy.

4.2 HZB will provide a means for the data access team to upload results and associated metadata to the facility and enable them to associate these results with raw data. The storage period for results is determined by the storage of associated raw data. The upload of results and associated metadata may be subject to volume restrictions.

4.3 HZB does not take any responsibility to fully curate results e.g. to ensure that software to read / manipulate this data is available.

4.4 Access to results is restricted to the data access team. They may be made openly accessible on request of any member of the data access team, if no other member objects. Paragraph 3.6 applies accordingly.

5. Good practice for metadata capture and results storage

5.1 The experimental team is encouraged to ensure that experimental 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.

5.2 HZB 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.

5.3 Researchers who aim to carry out analyses of raw data and metadata which are openly accessible should, where possible, contact the original experimental team to inform them and suggest a collaboration if appropriate.

5.4 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.

6. Privacy and other legal requirements

6.1 The proposer must ensure in the design and preparation of the experiment that the raw data comply with the legal data protection regulations (Bundesdatenschutzgesetz) and do not contain any personal or otherwise particularly sensitive data. The HZB is not responsible for the compliance of the user data with data protection legislation.

A. Creative Commons CC0 1.0 Universal Public Domain Dedication

A.1. Statement of Purpose

The laws of most jurisdictions throughout the world automatically confer exclusive Copyright and Related Rights (defined below) upon the creator and subsequent owner(s) (each and all, an "owner") of an original work of authorship and/or a database (each, a "Work").

Certain owners wish to permanently relinquish those rights to a Work for the purpose of contributing to a commons of creative, cultural and scientific works ("Commons") that the public can reliably and without fear of later claims of infringement build upon, modify, incorporate in other works, reuse and redistribute as freely as possible in any form whatsoever and for any purposes, including without limitation commercial purposes. These owners may contribute to the Commons to promote the ideal of a free culture and the further production of creative, cultural and scientific works, or to gain reputation or greater distribution for their Work in part through the use and efforts of others.

For these and/or other purposes and motivations, and without any expectation of additional consideration or compensation, the person associating CC0 with a Work (the "Affirmer"), to the extent that he or she is an owner of Copyright and Related Rights in the Work, voluntarily elects to apply CC0 to the Work and publicly distribute the Work under its terms, with knowledge of his or her Copyright and Related Rights in the Work and the meaning and intended legal effect of CC0 on those rights.

A.2. Copyright and Related Rights

A Work made available under CC0 may be protected by copyright and related or neighboring rights ("Copyright and Related Rights"). Copyright and Related Rights include, but are not limited to, the following:

i. the right to reproduce, adapt, distribute, perform, display, communicate, and translate a Work;

ii. moral rights retained by the original author(s) and/or performer(s);

iii. publicity and privacy rights pertaining to a person's image or likeness depicted in a Work;

iv. rights protecting against unfair competition in regards to a Work, subject to the limitations in paragraph A.5(a), below;

v. rights protecting the extraction, dissemination, use and reuse of data in a Work;

vi. database rights (such as those arising under Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, and under any national implementation thereof, including any amended or successor version of such directive); and

vii. other similar, equivalent or corresponding rights throughout the world based on applicable law or treaty, and any national implementations thereof.

A.3. Waiver

To the greatest extent permitted by, but not in contravention of, applicable law, Affirmer hereby overtly, fully, permanently, irrevocably and unconditionally waives, abandons, and surrenders all of Affirmer's Copyright and Related Rights and associated claims and causes of action, whether now known or unknown (including existing as well as future claims and causes of action), in the Work (i) in all territories worldwide, (ii) for the maximum duration provided by applicable law or treaty (including future time extensions), (iii) in any current or future medium and for any number of copies, and (iv) for any purpose whatsoever, including without limitation commercial, advertising or promotional purposes (the "Waiver"). Affirmer makes the Waiver for the benefit of each member of the public at large and to the detriment of Affirmer's heirs and successors, fully intending that such Waiver shall not be subject to revocation, rescission, cancellation, termination, or any other legal or equitable action to disrupt the quiet enjoyment of the Work by the public as contemplated by Affirmer's express Statement of Purpose.

A.4. Public License Fallback

Should any part of the Waiver for any reason be judged legally invalid or ineffective under applicable law, then the Waiver shall be preserved to the maximum extent permitted taking into account Affirmer's express Statement of Purpose. In addition, to the extent the Waiver is so judged Affirmer hereby grants to each affected person a royalty-free, non transferable, non sublicensable, non exclusive, irrevocable and unconditional license to exercise Affirmer's Copyright and Related Rights in the Work (i) in all territories worldwide, (ii) for the maximum duration provided by applicable law or treaty (including future time extensions), (iii) in any current or future medium and for any number of copies, and (iv) for any purpose whatsoever, including without limitation commercial, advertising or promotional purposes (the "License"). The License shall be deemed effective as of the date CC0 was applied by Affirmer to the Work. Should any part of the License for any reason be judged legally invalid or ineffective under applicable law, such partial invalidity or ineffectiveness shall not invalidate the remainder of the License, and in such case Affirmer hereby affirms that he or she will not (i) exercise any of his or her remaining Copyright and Related Rights in the Work or (ii) assert any associated claims and causes of action with respect to the Work, in either case contrary to Affirmer's express Statement of Purpose.

A.5. Limitations and Disclaimers

a. No trademark or patent rights held by Affirmer are waived, abandoned, surrendered, licensed or otherwise affected by this document.

b. Affirmer offers the Work as-is and makes no representations or warranties of any kind concerning the Work, express, implied, statutory or otherwise, including without limitation warranties of title, merchantability, fitness for a particular purpose, non infringement, or the absence of latent or other defects, accuracy, or the present or absence of errors, whether or not discoverable, all to the greatest extent permissible under applicable law.

c. Affirmer disclaims responsibility for clearing rights of other persons that may apply to the Work or any use thereof, including without limitation any person's Copyright and Related Rights in the Work. Further, Affirmer disclaims responsibility for obtaining any necessary consents, permissions or other rights required for any use of the Work.

d. Affirmer understands and acknowledges that Creative Commons is not a party to this document and has no duty or obligation with respect to this CC0 or use of the Work.

B. Version history

0.1	2015-10-20	R. Krahl	First german draft.
0.2	2015-12-10	R. Krahl	Introduction, mention in-house research in 2.4, move discussion in appendix.
0.3	2015-12-18	R. Krahl	Add notes on implementation.
0.4	2016-03-22	R. Krahl	English version (without discussion and implementation).
0.5	2016-04-26	R. Krahl	Extend embargo period to five years. Eliminate the principle investigator in favour of the experimental team.
0.6	2016-06-01	R. Krahl	Modified formulation for 6.1, essentially changing from "should" to "must". Drop 6.2.
1.0	2016-06-14	R. Krahl	Final version, no changes.
1.1	2017-01-19	R. Krahl	Introduce distinction between experimental team and data access team.

References

[1] Deutsche Forschungsgemeinschaft (ed.). Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Wiley-VCH, 2. edition, 2013. URL http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf.

[2] Organisation for Economic Co-operation and Development (OECD). Principles and Guidelines for Access to Research Data from Public Funding, April 2007. URL http://www.oecd.org/sti/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm.

[3] Alliance of German Science Organisations. Principles for the Handling of Research Data, June 2010. URL http://www.allianzinitiative.de/en/core-activities/research-data/principles.html.

[4] Deutsche Forschungsgemeinschaft. Leitlinien zum Umgang mit Forschungsdaten, September 2015. URL http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf.

[5] PaN-data Europe. Common Policy Framework on Scientific Data, December 2010. URL http://wiki.pan-data.eu/imagesGHD/0/08/PaN-data-D2-1.pdf.

Terms and Conditions for the Storage, Access and Curation of Research Data

Table of Contents

List of Annexes

Appendix 1	Checklist for a Data Management Plan
Appendix 2	Data Cite Metadata Schema v4.1

List of Revisions

Page	Rev.-No	Date	Reason for revision
1-9	0	01.05.2018	New Regulation

List of Abbreviations

CC BY	Creative Commons Attributive License
CC0	Creative Commons Universal License
DMP	Data Management Plan
DOI	Digital Object Identifier
FAIR data	Data that is findable, accessible, interoperable and reusable
HZDR	Helmholtz-Zentrum Dresden - Rossendorf e. V.
PI	Principal Investigator
RODARE	Rossendorf research Data Repository

To enhance the readability, no distinction has been made between male and female.

Preamble

The proper management of research data is imperative to ensure that scientific findings are findable, accessible, interoperable and reusable (FAIR). The national and international research organizations support the long-term safeguarding of and the open access to research data from publicly funded research, as laid down in the "Berlin Declaration on Open Access" of 2003 [1] and the "Guidelines on Data Management in Horizon 2020" [2]. The Alliance of German Science Organizations, the Deutsche Forschungsgemeinschaft and the Helmholtz Association took up this idea in their guidelines for the handling of research data.

The HZDR approves the principle of open access to research data. The HZDR supports its guests in the fulfillment of the requirements of funders and of the scientific community. To this end, it provides the necessary infrastructure for data management and regulates the access to research data by these terms and conditions. Open access to research data should be ensured wherever possible taking the pathway of the citable data publication.

1 Definitions

For the purposes of these terms and conditions:

(1) The term Research Data refers to all data exclusively in anonymous form which are involved in the research process related to this research. This includes (but is not limited to) raw data, results, other data (e.g. validated data or simulation data), third party data, automatically or/and manually generated data and associated metadata.

  (1-a) The term Raw Data means originally generated data.

  (1-b) The term Result Data pertains to data that are outcomes arising from the analysis of raw data. This does not include publications.

  (1-c) The term Metadata describes the context in which the data has been generated including information about (but not limited to) the person generating the data, the setting, the surrounding and the facilities which were used when generating the data.

(2) The term Public Research refers to research done with means of basic funding and/or other public funding.

(3) The User Group consists of the natural and legal persons that are involved in the specific research project.

(4) The Principal Investigator (PI) is the natural or legal person who accepts these terms and conditions by signature and acts as leader and speaker of the User Group implementing the individual research project at the HZDR. The PI in particular is responsible for the coordination and definition of data access and usage rights.

(5) The Technician Group consists of persons who support the individual research process. This includes (but is not limited to) archiving and curating of data.

(6) The Right Holder Group consists of natural and legal persons that hold rights on the research data.

(7) Anonymous Data is information which does not relate to an identified or identifiable natural person or personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable.

2 General Principles

(1) These terms and conditions set the rules for the storage, the curation of and the access to research data exclusively in anonymous form collected in the framework of publically funded research at research facilities and research projects of HZDR.

(2) The User Group shall document the quality of the data (such as accuracy, completeness, integrity, confidentiality) and the data management parts (including responsibility, publication) as part of a Data Management Plan (DMP). Templates and support for such plans will be provided by the HZDR.

(3) The principles of Open Access and Technology Transfer shall be balanced.

(4) Mandatory legal provisions such as the Employee Invention Act [Gesetz über Arbeitnehmererfindungen] are not affected by this Terms and Conditions for Storage, Access and Curation of Research Data.

3 Research Data Management

(1) The sustainable utilization of research data requires a quality management. This covers the entire lifecycle of the data ranging from data collection to processing, storage and to a controlled deletion. The documentation of the processes and their application in the context of specific projects are part of a DMP. A checklist for creating a DMP is annexed (Appendix 1).

(2) The PI has to ensure that data generated during a specific project is anonymized as soon as possible to enable access to research data according to these terms and conditions.

(3) During data collection metadata should be recorded which allow conclusions on the context and the quality of the data collected. If possible, open and free data formats should be used. It is recommended to use the “DataCite Metadata Schema for the Publication and Citation of Research Data” (Appendix 2) [3].

(4) Access to research data that forms the basis of citable publications must be ensured by publication in a suitable and trustworthy research data repository with the Digital Object Identifier (DOI). This enables the exchange and reuse of research data with collaboration partners. In addition, verification of the research results is enabled in this way. Appropriate data repositories are provided by the HZDR [4] or by external facilities.

(5) For research data the storage and safeguarding for at least ten (10) years corresponds to "good scientific practice" [5].

4 Raw Data and associated Metadata

(1) Raw Data stored at the HZDR will be curated in well-defined formats, for which the means of reading the data will be made available by HZDR.

(2) Associated Metadata will be curated either within the raw data files, within an associated on-line catalogue, or within both.

(3) Raw data and associated metadata will be stored by HZDR for at least ten (10) years as far as legally permitted.

(4) Access to raw data and the associated metadata may be restricted to the members of the User Group for an embargo period of five (5) years after the end of the experiments of the individual research project. Thereafter, if legally permitted or if not necessarily required for Technology Transfer, it will be made openly accessible with HZDR acting as custodian. Any member of the Right Holder Group that wishes to maintain the restricted access to its data for a longer period will be required to file a corresponding request to the HZDR management.

(5) Members of the technician group have access to data or metadata curated by HZDR for facility related purposes. HZDR will undertake to preserve the confidentiality of such data unless it has been made openly accessible.

(6) In the DMP regulations have to stipulate if and how data shall be deleted at the end of its life cycle and how this deletion process will be documented.

5 Result Data

(1) Rights to ownership and other rights of use with respect to Result Data are not affected by these terms and conditions.

(2) HZDR will provide means for the entitled members of the User Group to upload results and associated Metadata to the facility and enable him/her to associate these results with Raw Data. The storage period for Result Data is determined by the storage of associated Raw Data or the DMP. The upload of results and associated Metadata may be subject to volume restrictions.

(3) The User Group is requested to ensure that associated Metadata is as complete as possible, as this will enhance the chance to search for, retrieve and interpret the data in the future. For the long-term usability of data it is essential to use open or standard formats or to ensure that software to read / manipulate this data is made available.

(4) HZDR is in charge of the curation of Result Data stored in HZDR repositories.

(5) As far as legally permitted, the HZDR undertakes to provide means for the complementation 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.

(6) Access to Result Data is restricted to the User Group. As far as legally permitted or if not necessarily required for Technology Transfer, the Result Data may be made openly accessible upon request of the PI.

(7) Publications related to experiments carried out at HZDR shall acknowledge the support of HZDR, including the facilities used, supporting staff or any other assistance.

6 Legal Requirements

(1) The User Group must ensure by the design and the preparation of the experiment that the Research Data to be stored are Anonymous Data. If Research Data are going to be published, the names of the authors and contributors involved in the specific project have to be specified in the Metadata set (see Appendix 2).

(2) The User Group must ensure that no third party holds rights to the Research Data stored in HZDR repositories by this User Group.

(3) The storage and processing of data which are not anonymous must be explicitly defined within in the DMP. The data protection officer of the HZDR or the authorized entity has to be involved if non anonymous data have to be processed.

(4) It is not permitted to store non anonymous data in the HZDR data repository RODARE.

(5) It is recommended to publish the Research Data associated with a publication in accordance with the Open Access Guidelines [6]. The choice of an open license (actual Creative Commons, [7]) is recommended for easy reuse. This corresponds to the requirements of the research funders and project partners.

(6) The HZDR assumes no warranty or representation as to the accuracy, integrity, timeliness and correctness or to the availability of Research Data and of Software stored in HZDR repositories.

(7) The HZDR shall neither be liable for the usability of Research Data or Software nor for possible damages resulting therefrom nor for consequential loss. HZDR excludes any trustee relationship with respect to the Research Data and any representation of right holders to the Research Data stored in HZDR repositories.

(8) The HZDR shall not be liable for the Research Data to be free of third party rights, viruses, bugs, defects, backdoor functions, malware or other malfunctions.

(9) The HZDR including its senior executives, legal representatives and vicarious agents shall only assume liability in the case of fraud, intent and gross negligence. The liability of the HZDR for slightly negligent violations of duty is excluded unless these violations affect duties essential to these Terms and Conditions, damages resulting in wrongful death, personal injury or health impairment or guaranties or claims in accordance with the German product liability law. It should be noted that the provisions of 6.8 shall also apply to the limitations and exclusions of liability according to 6.5, 6.6 and 6.7.

7 Taking Effect

This directive takes effect when signed by the HZDR´s Board of Directors.

References

[1] Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 2003, http://openaccess.mpg.de/67605/berlin_declaration_engl.pdf (20.02.2018)

[2] H2020 Program Guidelines on FAIR Data Management in Horizon 2020, 2016, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (20.02.2018)

[3] DataCite Metadata Working Group. (2017). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.1., DataCite e.V., http://doi.org/10.5438/0014

[4] Terms and Conditions for User Access to the Experimental Facilities (HZDR Regulation B210)

[5] Safeguarding Good Scientific Practice and Proceeding in Case of Scientific Malpractice (HZDR Regulation B110)

[6] Open Access Policy of the Helmholtz Association, 2016, http://os.helmholtz.de/open-science-in-der-helmholtz-gemeinschaft/open-access-richtlinien/open-access-richtlinie-der-helmholtz-gemeinschaft-2016/open-access-policy-of-the-helmholtz-association-2016/ (12.02.2018)

[7] Creative Commons Licenses, https://creativecommons.org/licenses (20.02.2018)

Appendix 1: Checklist for a Data Management Plan[1]

DMP component	Issues to be addressed
1. General information	☐ Project name ☐ Project ID ☐ Principal investigator (project coordinator) ☐ Contact ☐ Date
2. Data summary	☐ State the purpose of the data collection/generation ☐ Explain the relation to the objectives of the project ☐ Specify the types and formats of data generated/collected ☐ Specify if existing data is being re-used (if any) ☐ Specify the origin of the data ☐ State the expected size of the data (if known) ☐ Outline the data utility: to whom will it be useful
3. FAIR data 3.1. Making data findable, including provisions for metadata	☐ Outline the discoverability of data (metadata provision) ☐ Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? ☐ Outline naming conventions used ☐ Outline the approach towards search keyword ☐ Outline the approach for clear versioning ☐ Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how
3.2 Making data openly accessible	☐ Specify which data will be made openly available? If some data is kept closed provide rationale for doing so ☐ Specify how the data will be made available ☐ Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? ☐ Specify where the data and associated metadata, documentation and code are deposited ☐ Specify how access will be provided in case there are any restrictions

1 H2020 Programme Guidelines on FAIR Data Management in Horizon 2020, Summary Table 1, 2016, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (20.02.2018)

3.3. Making data interoperable	□ Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. □ Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
3.4. Increase data re-use (through clarifying licences)	□ Specify how the data will be licensed to permit the widest reuse possible □ Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed □ Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why □ Describe data quality assurance processes □ Specify the length of time for which the data will remain re-usable
4. Allocation of resources	□ Estimate the costs for making your data FAIR. Describe how you intend to cover these costs □ Clearly identify responsibilities for data management in your project □ Describe costs and potential value of long term preservation
5. Data security and handling of sensitive data	□ Specify where data will be stored and how data recovery as well as secure storage will be ensured □ Describe how data shall be deleted at the end of its life cycle □ Describe how sensitive data (data which is not anonymous) is handled, which law is used, which technical process has been used for processing of them □ Describe which persons have access to the sensitive data
6. Ethical aspects	□ To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
7. Other	□ Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

It is recommended to use the Digital Curation Centre DMPonline tool, which offers DMP templates that match the demands and suggestions of the Guidelines on Data Management in Horizon 2020 (https://dmponline.dcc.ac.uk/, 28.02.2018).

Appendix 2: Data Cite Metadata Schema v4.1[2]

There are three different levels of obligation for the metadata properties:

- **Mandatory (M)** properties *must* be provided,
- **Recommended (R)** properties are optional, but strongly recommended for interoperability and
- **Optional (O)** properties are optional and provide richer description."

Table 1: DataCite Mandatory Properties

ID	Property	Obligation
1	Identifier (with mandatory type sub-property)	M
2	Creator (with optional family name, given name, name identifier and affiliation sub-properties)	M
3	Title (with optional type sub-properties)	M
4	Publisher	M
5	PublicationYear	M
10	ResourceType (with mandatory general type description sub-property)	M

Table 2: DataCite Recommended and Optional Properties

ID	Property	Obligation
6	Subject (with scheme sub-property)	R
7	Contributor (with optional family name, given name, name identifier and affiliation sub-properties)	R
8	Date (with type sub-property)	R
9	Language	O
11	AlternateIdentifier (with type sub-property)	O
12	RelatedIdentifier (with type and relation type sub-properties)	R
13	Size	O
14	Format	O
15	Version	O
16	Rights	O
17	Description (with type sub-property)	R
18	GeoLocation (with point, box and polygon sub-properties)	R
19	FundingReference (with name, identifier, and award related sub-properties)	O

"Those clients who wish to enhance the prospects that their metadata will be found, cited and linked to original research are strongly encouraged to submit the Recommended as well as Mandatory set of properties."

2 DataCite Metadata Working Group. (2017). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.1, S. 8, Table 1 u. 2, DataCite e.V., http://doi.org/10.5438/0014

Cover Page for Internal MAX IV Laboratory Use

MAX IV User Access Policy

DNR: STYR 2016/1056.
Date: 28 March 2018
Revision: 2

Prepared by:

- Marjolein Thunnissen, User Office Coordinator
- Mirja Carlsson Möller, Research Coordinator
- Pernilla Skantze/Henrik Wiebe, Legal Counsels

Reviewed/Approved by:

- Olivier Balmes, Group Manager Hard X-ray group
- Uwe Mueller, Group Manager Diffraction and scattering group
- Katarina Norén, Group Manager Spectroscopy group
- Rajmund Mokso, Group Manager Imaging group
- MAX IV Laboratory Directors (DM-Ex)

Changes made in this revision

Rev	*Change made*	*Date*	*By*	*Approved*
2	4.3.1 removed: *Such results and all related publications, must be submitted to MAX IV, no later than six (6) months after the project/Experiment was carried out at MAX V.* 8 changed *Peer Review* into *peer review.*	28 March 2018	Marjolein Thunnissen	

POLICY
DNR: STYR 2016/1056.
DATE: 28 March 2018
REVISION: 2

MAX IV User Access Policy

1. *Policy Statement*

   This policy presents the overall framework governing user access to MAX IV.

2. *To whom does this policy apply?*

   This policy applies to all MAX IV Users and User Institutions when working at the MAX IV Facilities.

3. *Who is responsible for this policy?*

   The directors of MAX IV have overall responsibility for this policy. The directors have delegated practical management of the policy to the User Office Coordinator. Any queries or suggestions relating to this policy should be sent to policies@maxiv.lu.se.

4. *Policy*

   4.1. Access modes

   MAX IV Facilities are available to the scientific community in the following access modes:

   - Open Access
     - Peer-Reviewed Access
       - Regular Access
       - Privileged Peer-Reviewed Access
         - Investor Access
         - External Access Programmes
       - Access for Education and Training
     - Non Peer-Reviewed Access
       - Access for In-House Research and Partners having Participated in the Beamline Build-Up
       - Directors Discretionary Access
   - Proprietary Access

   Access preparation and registration are managed through DUO. Scientific, technical and safety reviews are carried out at different levels depending on access mode. Remote access is considered equivalent to normal access if not otherwise agreed in writing.

4.2. Responsibilities of Principal Investigator

4.2.1 Each Experiment must be sufficiently staffed to conduct the Experiment safely and the Users must have correct and adequate training and education in order to perform the Experiment. If this cannot be assured it must be clearly stated in the Experimental Proposal

4.2.2 The Principal Investigator (PI) is the point of contact and has the responsibility for the Experimental Team. The PI has the right to transfer parts or all of his/her responsibilities to another User provided that it is clearly communicated to MAX IV in writing. Further, if the PI is not physically present during the Experiment, a contact person within the Experimental Team, who must be physically present, must be assigned.

4.3. Publication of results

4.3.1. Research outcome from all Experiments, except those made under Proprietary Access, are to be published in scientific publications in accordance with good international standards for the publication of research results. Any published result must include appropriate acknowledgment of MAX IV as a facility and beamline personnel as scientists. In accordance with good scientific practice, Users are encouraged to offer co-authorship to those working at MAX IV having made genuine scientific contributions to their work.

4.3.2. Experimental Reports are required for evaluation of follow up proposals and must be submitted to MAX IV no later than three (3) months after the project/Experiment.

4.3.3. References for publications related to Experiments carried out at the Facilities must be deposited in the MAX IV publications' database within three (3) months of the publication date.

4.4. Financial framework

MAX IV is entitled to charge User Institution for services outside MAX IV's ordinary range of services according to MAX IV's at all time applicable prices.

This section 4.4. does not apply to Proprietary Access. The financial framework for Proprietary Access will be regulated in individual agreements.

5. Feedback

Any feedback, suggestions for improvements and/or complaints relating to User Access as well as appeals related to allocation of beamtime, should be addressed to MAX IV User Office (useroffice@maxiv.lu.se).

6. Amendments to this policy

This policy may at any time be revised by MAX IV.

7. *Related resources*

- European Charter for Access to Research Infrastructures
  ec.europa.eu/research/infrastructures/pdf/2016_charterforaccessto-ris.pdf
- MAX IV General Terms and Conditions for Open Access
  www.maxiv.lu.se/users/user-policies/
- Experimental Data Policy (DNR: STYR 2016/1057)
  www.maxiv.lu.se/users/user-policies/
- MAX IV Laboratory Education policy (DNR: STYR 2015/719)
  www.maxiv.lu.se/education-training/projects
- IT Rules & Guidelines
  www.maxiv.lu.se/users/user-policies/
- Guidelines for user access to MAX IV
  www.maxiv.lu.se/users/user-guide/
- Safety for users
  www.maxiv.lu.se/safety/safety-for-users
- Safety for staff, including radiation safety
  www.maxiv.lu.se/safety/safety-for-staff

8. *Definitions*

For the purpose of this policy, the following definitions shall apply:
Access for Education and Training: Peer-Reviewed Access within the frame of academic or external education programmes
Access for In-House Research: Non Peer-Reviewed Privileged Access for MAX IV Staff
Directors Discretionary Access: Non Peer-Reviewed Privileged Access mechanism that allows assignment of access at the discretion of the Director
DUO (Digital User Office): web-based proposal submission and management system used for requesting all types of access.
Experiment: A set of tests or procedures carried out under controlled conditions at the MAX IV Facilities to determine the validity of a hypothesis or make a discovery within a research field.
Experimental Report: document to be provided by the User within the DUO system describing the immediate result of the Experiment.
Experimental Team: individuals, which may include the PI, who carry out an Experiment at the MAX IV Facility.
External Access Programmes: Peer-Reviewed Privileged Access for programmes with external funding, e.g. EU-projects.
Investor Access: Peer-Reviewed Privileged Access for organisations investing in MAX IV, distributed according to special agreement between investor and MAX IV.
Facilities: all research facilities made available by and at MAX IV, which may include equipment, services, information and other material, with or without MAX IV scientist collaboration, for purposes of performing Experiments at MAX IV.
MAX IV: the Swedish national synchrotron light laboratory as set out in regulation SFS 1994:946 (*Förordning (1994:946) om den nationella forskningsanläggningen i elektron-acceleratorlaboratoriet (MAX IV-laboratoriet) i Lund)*, being a part of Lund University.
MAX IV Programme Advisory Committees (PAC): assist in allocation of beamtime by evaluating and ranking applications for Regular Access.

MAX IV Staff: individuals working at MAX IV under a contract of employment or consultancy with Lund University.
Non Peer-Reviewed Access: Participating Research Partner Access, Access for In-House Research and Directors Discretionary Access
Open Access: Peer-Reviewed and Non Peer-Reviewed access to beamlines, accelerators or similar research infrastructure at MAX IV Laboratory, which is free of charge provided that all results are published and registered with the Facility.
Peer-Reviewed Access: access available for allocation to Users through peer review after application in DUO. Research outcome from Experiments must provide an Experimental Report and successful Experiments must be published in scientific literature.
Principal Investigator, PI: the PI identified on the Experiment Proposal. For E xperiments outside of the Experimental Proposal system, the PI is the person initiating the Experiment.
Privileged Peer-Reviewed Access: access for Investors and External Access Programmes.
Proprietary Access: non peer-reviewed, market driven, access.
Regular Access: access available for allocation to general Users through the MAX IV Programme Advisory Committees (PAC) after application in DUO.
User: all individuals making use of the Facilities including scientists, engineers and students from academia, research councils and charitable institutions, researchers from commercial and non-commercial organisations and MAX IV Staff.
User Institutions: all institutions, universities, companies and other organisations, regardless of legal form, making use of the Facilities.

POLICY
DNR: STYR 2016/1057
DATE: 19 January 2017
REVISION: 2

Experimental Data Policy

1. *Policy purpose*

   This policy applies to Experimental Data collected at MAX IV. For the avoidance of doubt, it does not apply to diagnostic data needed by MAX IV Staff.

2. *To whom does this policy apply?*

   This policy applies to all Users and User Institutions.

   Experimental Data from Proprietary Research may be regulated in a separate agreement which will prevail over this policy.

3. *Who is responsible for this policy?*

   The directors of MAX IV have overall responsibility for this policy. The directors have delegated practical management of the policy to the manager for Controls & IT. Any queries or suggestions relating to this policy should be sent to policies@maxiv.lu.se.

4. *Policy*

   4.1. General principles

   4.1.1. Violation of this policy may lead to denial of access to Experimental Data and/or denial of future beamtime requests at MAX IV.
   4.1.2. Users and User Institutions must not attempt to access, exploit or distribute Experimental Data unless they are entitled to do so under the terms of this policy.
   4.1.3. All Experimental Data will be processed in accordance with Swedish data protection legislation. Information about the processing of personal data in MAX IV's digital user systems can be found here: duo.maxiv.lu.se/duo/disclaimer.php.
   4.1.4. MAX IV claims no ownership to Experimental Data. This does not apply to Experimental Data originating from experiments carried out solely by MAX IV Staff.
   4.1.5. Users and User Institutions are encouraged to apply an Open Access policy to Results.
   4.1.6. MAX IV does not guarantee that data is not unintentionally lost or accessed, nor that it will be available at all times

4.2. *Curation of Experimental Data*

4.2.1. All Raw Data will be curated in well-defined formats, for which the means of reading the data will be made available by MAX IV.

4.2.2. Metadata automatically captured by instruments will be curated either within the Raw Data files, within an associated On-line Catalogue, or within both.

4.2.3. Experimental Data will be read-only while stored at MAX IV.

4.2.4. Each data set in the On-line Catalogue will have a persistent identifier (PID). Anybody publishing Results must quote the PID of the data set (and related publications if available and appropriate).

4.3. *Access to Experimental Data*

4.3.1. Experimental Data will be stored in the MAX IV database for a maximum time of three months. MAX IV reserves the right to restrict the storage time, for high data rate instruments. Such restriction of storage time will be decided by MAX IV after consultation with the respective user communities and will be communicated to Users and User Institution in advance of each experiment.

4.3.2. Experimental Data will be made available to the User and/or User Institution. MAX IV assumes no responsibility for long-term storage of Experimental Data.

4.3.3. Experimental Data will be available via a searchable On-line Catalogue. The On-line Catalogue will enable the linking of Experimental Data to Experimental Proposals via the Experimental Proposal ID. Access to Experimental Proposals and Experimental Data will only be provided to the Experimental Team and MAX IV Staff who needs access, unless otherwise authorised by the PI.

4.3.4. The Experimental Team is responsible for ensuring that the Experimental Data collection uses the correct Experimental Proposal ID. Failing to do so may result in data loss, inability for the Experimental Team to access the data via the On-line Catalogue or unwarranted dissemination of data.

4.3.5. User acknowledges that MAX IV Staff (such as instrument scientists and computing group members) has access to any MAX IV-curated Experimental Data for MAX IV related purposes.

4.4. *Good practice*

4.4.1. The Experimental Team is encouraged to ensure that their Metadata is as complete as possible, as this will enhance the possibilities to search for, retrieve and interpret Experimental Data.

4.4.2. MAX IV will provide its best effort to capture Metadata items not automatically captured by an instrument in order to facilitate the recording of the fullest possible description of the Raw Data.

5. *Feedback*

Any feedback, suggestions for improvements and/or complaints regarding the MAX IV Experimental Data handling should be addressed to MAX IV Controls & IT (issues@maxiv.lu.se).

6. *Amendments to this Policy*

   This policy may at any time be revised by MAX IV.

7. *Related resources*

   - Swedish Research Council's guidelines for Open Access: www.vr.se/inenglish/aboutus/activities/analysisevaluationandfollowup/nationalguidelinesforopenaccesstoresearchinformation.4.18f425dd146e9437d292db46.html
   - Swedish Information Security legislation www.informationssakerhet.se/forfattningar/
   - LU Open Access Policy www.staff.lu.se/sites/staff.lu.se/files/policy-on-open-access-publication-at-lund-university.pdf

8. *Definitions*

   For the purpose of this policy, the following definitions shall apply:
   Analysed Data: Raw Data and Metadata that has been manually or automatically evaluated.
   Experimental Data: Raw Data, Analysed Data and associated Metadata originating from experiments at MAX IV Facilities
   Experimental Proposal: electronic document comprising the description of the proposed research and all associated beamtime requests.
   Experimental Team: individuals, which may include the PI, who carry out an experiment physically on site at the MAX IV Facility.
   Facilities: all research facilities made available by and at MAX IV, which may include equipment, services, information and other material, with or without MAX IV scientist collaboration, for purposes of performing experiments at MAX IV.
   MAX IV: the Swedish national synchrotron light laboratory as set out in regulation SFS 1994:946 (*Förordning (1994:946) om den nationella forskningsanläggningen i elektron-acceleratorlaboratoriet (MAX IV-laboratoriet) i Lund)*, being a part of Lund University.
   MAX IV Staff: individuals working at MAX IV under a contract of employment or consultancy with Lund University.
   Metadata: information pertaining to Raw Data, including (but not limited to) the context of the experiment, the participants of the Experimental Team, experimental conditions and other logistical information.
   On-line Catalogue: a database of Metadata containing links to Raw Data files. The On-line Catalogue can be accessed by a variety of methods, including (but not limited to) web-based browsers to registered Users.
   Open Access: free, unrestricted, on-line access to peer reviewed research papers, journal articles or conference papers that have been submitted and accepted for publication.
   Persistent Identifier, PID: a long-lasting reference to a digital resource. Typically it has two components: a unique identifier; and a service that locates the resource over time even when its location changes.
   Principal Investigator, PI: the PI identified on the Experiment Proposal. For experiments outside of the Experimental Proposal system, the PI is the person initiating or performing the experiment.
   Proprietary Research: research done through purchased (commercial) access to MAX IV.

Raw Data: data collected from experiments performed on MAX IV instruments.
Results: any inventions, designs, information, know-how, specifications, formulae, Experimental Data, processes, methods, techniques and other technology arising out of peer reviewed research.
Users: all individuals making use of the Facilities including scientists, engineers and students from academia, research councils and charitable institutions, researchers from commercial and non-commercial organisations and MAX IV Staff.
User Institutions: all institutions, universities, companies and other organisations, regardless of legal form, making use of the Facilities.

Data policy for PSI research data (the “Policy”)

The Policy defines the rules for the following topics:

- **Data ownership**
- **Data curation**
- **Data archiving**
- **Open access to data**

1 General Principles

1.1. The present Policy pertains to the ownership of, the curation of and access to experimental data and Metadata collected and/or stored by PSI Research Infrastructure.

1.2. Acceptance of this Policy is a condition for the award of access to Research Infrastructures and a binding part of each PSI employment agreement.

1.3. Users must not attempt to access, exploit or distribute data or Metadata unless they are entitled to do so under the terms of this Policy.

1.4. Deliberate infringements of the Policy may lead to actions concerning the employment of PSI employees to denial of access to data or Metadata and/or denial of future beam time requests at PSI, as well as to legal actions.

1.5. All data and Metadata will be subject to Swiss law.

2 Definitions

For the purposes of this Policy the term:

2.1 **Custodian** refers to the head of the institute divisions (or a delegated person) storing, curating and providing access to Raw Data, Metadata and Results.

2.2 **Experimental Team** includes the PI and any other person to whom the PI designates the right to access data and associated Metadata related to this proposal.

2.3 **Long-term** means a minimum of five (5) years and PSI will strive for ten (10) years, depending on the type and volume of data concerned and the economic consequences associated with Long-term data storage. Thus, PSI reserves the right to restrict the storage periods or data sets in consultation with the respective communities.

2.4 **Metadata** means information pertaining to data collected from Research Infrastructures at PSI, including (but not limited to) the scientific and administrative context of the experiment, the Experimental Team and the experimental conditions.

2.5 **On-line Catalogue** pertains to a computer database of Metadata containing links to data files, that can be accessed by a variety of methods.

2.6 **Open Access** means belonging to the public at large, unprotected by most copyrights or patents and subject to appropriation by anyone. Those data will be made available under CC-BY-SA (https://creativecommons.org/licenses/by-sa/4.0/).

2.7 **Principle Investigator** (PI) means the main proposer identified on the experiment proposal for peer-reviewed experiments or the leader of the Experimental Team for non-peer-reviewed experiments, both at the PSI large-scale facilities, or at other Research Infrastructures.

2.8 **Proprietary** research refers to research done through commercial access.

2.9 **Public Research** refers to research done through peer reviewed experiments and experiments done during in-house research or management contingency beam time.

2.10 **Research Infrastructure** includes but is not limited to PSI facilities SLS, SINQ, SμS, SwissFEL, and other PSI research infrastructures.

2.11 **Raw Data** means data collected from experiments performed at PSI. This includes data that are created automatically or manually by facility specific software and/or facility staff expertise in order to facilitate subsequent analysis of the experimental data.

2.12 **Result** pertains to data, intellectual property, and outcomes arising from the analysis of Raw Data (i.e. derived data).

Unless there is something inconsistent in the subject or context, words denoting the singular number include the plural and vice versa; words denoting one gender include the other gender and the neuter.

3 Raw Data and Metadata

3.1 <u>Data resulting from Public Research and Proprietary Research</u>

3.1.1. All Raw Data and Metadata obtained as a Result of Public Research will be Open Access after an initial embargo period during which access is restricted to the Experimental Team, represented by the PI.

3.1.2. PSI is the Custodian of the Raw Data and Metadata.

3.1.3. All Raw Data and Metadata obtained as a Result of Proprietary Research will be owned exclusively by the client who purchased the ac-

cess and is not covered by this Policy. Data from Proprietary Research will be removed after the experiment from PSI storage, unless otherwise agreed with the PSI management before the start of the experiment.

3.2 Curation of Raw Data and Metadata

3.2.1. All Raw Data and Metadata will be curated in well-defined formats, for which the means of reading the data will be made available by PSI.

3.2.2. Metadata that are automatically captured by instruments will be curated either within the raw data files, within an associated On-line Catalogue, or within both.

3.2.3. Raw Data and Metadata will be read-only for the duration of their life time.

3.2.4. Raw Data and Metadata will be migrated or copied to archival facilities for Long-term curation.

3.2.5. Each experiment and data set will have a unique persistent identifier. Anybody publishing Results based on open access data must quote the same identifier (and related publications if available & required).

3.2.6. High level Metadata such as title, authors, abstract, specific Research Infrastructure will be made public as soon as the experiment has been carried out. This information will be available via the persistent identifier landing page on the web.

3.3 Access to Raw Data and Metadata

3.3.1. Access to Raw Data and Metadata is foreseen to be via a searchable On-line Catalogue in addition to a direct access protected file based access within the PSI network.

3.3.2. Access to the On-line Catalogue of PSI will be restricted to registered users of the On-line Catalogue. PSI sets up the on-line procedure to become a registered user of the On-line Catalogue.

3.3.3. Access to Raw Data and Metadata obtained from an experiment is restricted to the Experimental Team for an embargo period of three (3) years after the end of the experiment. Thereafter, the data will become openly accessible. Any PI that wishes data to retain *restricted access* for a period longer than three (3) years will have this possibility on a yearly basis on a maximum prolongation of two (2) years. For longer extension a written request shall be submitted, specifying the reasons for the proposed prolongation, to the head of the corresponding PSI division who decides on the request, all at their sole discretion. Data can always be made openly accessible earlier on simple request of the PI. In exceptional circumstances the head of the corresponding PSI division can grant access to official committees at any time for the purpose of verifying data integrity.

3.3.4. Raw Data and Metadata explicitly used for peer-reviewed publication will become Open Access at the time of such publication.

3.3.5. It is the responsibility of the PI to ensure that the experiment number and requirements concerning the confidentiality, integrity and availability of the data is correctly entered into the Metadata for each Raw Data set.

3.3.6. Authorized PSI staff (including but not limited to facility management, instrument scientists, computing group members) have access to any curated data or Metadata for facility related purposes. PSI will undertake that confidentiality of such data is preserved during the embargo period.

3.3.7. The On-line Catalogue will enable linking experimental data to experimental proposals. Access to the full proposal text will only be provided to the Experimental Team and appropriate facility staff, unless otherwise authorized by the PI.

3.3.8. The PI has the possibility to transfer any or all of his rights during the embargo period to another person. This transfer has to be documented.

3.3.9. The PI may to create and distribute copies of the Raw Data, without violating against the rules that apply during the embargo period.

3.3.10. To the extent permitted by law, PSI cannot be held liable in the case of unavailability or loss of Raw Data.

4 Results

4.1 Ownership of Results

Ownership of all Results (intellectual property) derived from the analysis of the Raw Data is determined by the contractual obligations of the person(s) performing the analysis.

4.2 Curation of Results

4.2.1. PSI will provide curation of Results on a best effort basis, and acts as Custodian of Results in the long term.

4.2.2. PSI cannot be held liable in case of unavailability or loss of data or Results.

4.2.3. PSI cannot be held liable in case of unavailability or loss of data analysis software.

4.3 Access to Results

Access to the Results of analysis performed on Raw Data and Metadata, is restricted to the person performing the analysis, unless otherwise requested by that person. However, if the Raw Data being analysed is still restricted, access to the analysis results must be granted by the PI on request.

5 Leading practice for Metadata capture and Results storage

5.1. The Experimental Team shall ensure that experiments Metadata are as complete as possible, as this will enhance the possibilities for everybody to search for, retrieve and interpret the data in the long term.

5.2. PSI provides 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.

5.3. Researchers who carry out analyses of Raw Data and Metadata which are openly accessible shall, to the extent practicable, contact the original PI to inform him and suggest a collaboration if required. Researchers must acknowledge the source of the data and cite its unique identifier as well as the original publication linked to the same Raw Data.

5.4. PI and researchers who carry out analyses of Raw Data and Metadata shall link the Results of these analyses to the Raw Data / Metadata using the facilities provided by the On-line Catalogue. Furthermore, PI and researchers shall make such Results openly accessible.

6 Publication information

Publications related to data from experiments carried out at PSI must cite the persistent identifier of the experiment and data in their publication.

Implementation

The data policy for PSI research data will be implemented over time, depending foremost on our capacity to tag raw data with meaningful Metadata. A Metadata catalogue will be installed. A process document describes the details.

[1] http://pan-data.eu/PaNdataEurope

Reference: DSI-DU-P-0056

P

SOLEIL DATA MANAGEMENT POLICY

Date of circulation	Editor	Auditors	Approver	Changes	Index
02/10/2018	B.Gagey	SOLEIL Management SOLEIL Legalist Users' Representatives	J.Daillant		0

PUBLIC

Circulation: For action / For information

Recipients: SOLEIL Users, SOLEIL Staff

Synchrotron SOLEIL – Société Civile au capital de 12.000 €
439 684 903 R.C.S. EVRY – NAF 7219Z – SIRET 439 684 903 00016
L'Orme des Merisiers – Saint-Aubin – BP 48 – 91192 GIF S/YVETTE Cedex
Tél. : 01 69 35 90 xx – Fax : 01 69 35 94 xx - www.synchrotron-soleil.fr

SOLEIL DATA MANAGEMENT POLICY

TABLE OF CONTENTS

SOLEIL Data Management Policy

2 October 2018

As « Modern Science builds on extensive scientific dialogue and advances by improving earlier work »[1], SOLEIL is encouraging its Users and Scientists to produce FAIR (Findable, Accessible, Interoperable, Reusable) data. SOLEIL aims to facilitate this with data management services for storage and retrieval of experimental data and associated metadata collected and/or stored at SOLEIL. This also implies the definition of a framework to access these data and metadata.

This is the purpose of the present data management policy. It is based on a model for scientific data management at photon and neutron facilities formulated by the PaNdata project[2], co-funded by the European Commission under the 7th Framework Program, and on its application at other facilities. Adapted to SOLEIL with the help of science data experts from CNRS, it will be implemented starting 2018.

1. GENERAL PRINCIPLES

1.1. The present data management policy pertains to the ownership of, the curation of, and access to experimental data and associated metadata collected, reduced, processed and/or stored at SOLEIL.

1.2. SOLEIL may use subcontractors to perform its obligations under this data management policy. These subcontractors act as data processors in case personal data are concerned.

1.3. Acceptance of this policy is a condition for beamtime allocation.

1.4. Users must not attempt to access, exploit or distribute data or metadata unless they are entitled to do so under the terms of this policy.

1.5. Deliberate infringements of the policy may lead to denial of access to data or metadata and/or denial of future beamtime requests at SOLEIL.

1.6. If, and to the extent that, data and metadata include personal data, the data protection legislation of France and the European Union, respectively, will be applicable. This data management policy will be governed by and constructed in accordance with the French law.

1.7. Users of the SOLEIL facility must be personally registered via the SOLEIL user portal (SUN set). Each User performing an experiment at SOLEIL must agree and follow the SOLEIL User Charter Rules.

[1] Cf. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, European Commission, March 2017

[2] Cf. PaNdata Europe. Common Data Policy Framework on Scientific Data, December 2010

2. DEFINITIONS

For the purposes of the present data management policy:

2.1. The term **experimental data**, see Figure 1, pertains to data collected from experiments performed on SOLEIL 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 SOLEIL instruments, including (but not limited to) the context of the experiment, the experimental team, experimental conditions and other logistical information.

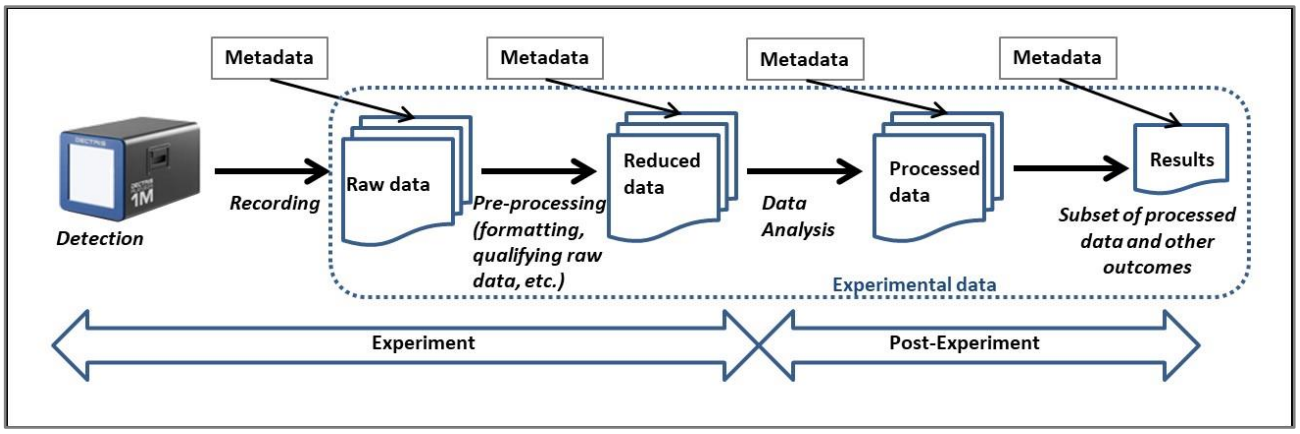


Figure 1 : Simplified illustration of Data Classes

2.7. The term **beamtime** means the time period, allocated by SOLEIL to the experimental team, for performing scientific experiments on a SOLEIL instrument. Every beamtime has a well-defined start date, end date, and experimental team.

2.8. The term **cycle of beamtime allocation** pertains to the time period starting from one call of proposals to the next one, typically six months.

2.9. The term **standard proposal** pertains to a proposal for an experiment running on one cycle of beamtime allocation.

2.10. The term **long-term proposal** pertains to a proposal for an experiment running over several cycles of beamtime allocation.

2.11. The term **block allocation group proposal** (BAG) pertains to a proposal for multiple experiments submitted by local, national or international consortia and running over two or

more cycles of beamtime allocation. Each BAG is composed of a BAG coordinator (drawn from the team members of the BAG) and several teams, each led by a **principal investigator (**or PI).

2.12. The term **main proposer** (MP) pertains to the main proposer identified on a standard or long-term proposal or to the BAG coordinator identified on a BAG proposal. The MP is the person who submits the proposal and who is the manager of the proposal for any matter. He/she is also the contact person to SOLEIL for any administrative issue related to the proposal.

2.13. The term **experimental team** includes the MP and all co-proposers of a given proposal, all participants in the experiment(s) performed for the proposal (including the SOLEIL local contact(s)), and any other person to whom the MP designates the right to access resultant experimental data and associated metadata.
In the case of a BAG, each experimental team includes the PI and all co-proposers of the experiment, all participants in this experiment (including the SOLEIL local contact(s)), and any other person to whom the PI designates the right to access resultant experimental data and associated metadata.

2.14. The term **proprietary research** refers to research done through purchased (commercial) access.

2.15. The term **public research** refers to research done through peer reviewed access (e.g. the SOLEIL proposal system), or via in-house research beamtime, or use of Rapid Access beamtime but excluding proprietary research and some specific agreements.

2.16. The term **on-line catalog** pertains to a computer database of metadata, containing links to raw data files, which can be accessed by a variety of methods including (but not limited to) web-based browsers.

2.17. A **registered user** refers to any person who has been granted a personal SUN set account, which can be requested via the login web page of the SOLEIL User portal.

2.18. The term **custodian** refers to the Institute storing, curating and providing access to raw data, metadata and results.

2.19. The term **long-term** means up to five years and SOLEIL will strive for 10 years starting with the end of the respective beamtime. The precise duration will depend on the beamline used, the type and volume of data concerned, and the economic consequences associated with long-term data storage. Thus, SOLEIL reserves the right to restrict the storage periods or data sets in consultation with the respective communities for high data rate instruments. Long-term durations applicable at each beamline or end station will be communicated on the SOLEIL user portal (SUN set), and reminded to the MP on acceptance of the proposal and before end of long-term duration.

2.20. The term **open access** means belonging to the community at large, unprotected by copyright or patent and subject to appropriation by anyone. Data held in long-term storage will be made available under the CC-BY license [3].

2.21. The term **embargo period** means the period, starting from the end of the respective beamtime, during which access to the collected raw data and associated metadata is restricted to the experimental team. This period will be 3 (three) years by default with the possibility of up to 5 (five) years, as detailed in 6.5.

2.22. The term **Data Management Plan**, or DMP, refers to the document describing the data management life cycle for the data to be collected, processed and/or generated by a research project: “the handling of research data during and after the end of the project, what data will be collected, processed and/or generated, which methodology and standards will be applied,

[3] Cf. (https://creativecommons.org/licenses/by/4.0/).

whether data will be shared/made open access, and how data will be curated and preserved (including after the end of the project)"[4]. Such a document may be required by external funders, research institutions and other organizations.

3. OWNERSHIP

3.1. All raw data and the associated metadata obtained via public research conducted at SOLEIL will be open access after an embargo period.

3.2. All raw data and the associated metadata obtained via proprietary research conducted at SOLEIL will be owned exclusively by the client who purchased the access and is not covered by the present data management policy. Data from proprietary research will be removed after the experiments from SOLEIL storage, unless otherwise agreed with SOLEIL management before the start of the experiment.

3.3. In some specific cases (as a national security related experiment, a classified experiment carried out by a ZRR[5] laboratory, etc.), SOLEIL and the organization carrying out the experiment may agree on a specific data management policy. The MP is invited to mention it when submitting the proposal, and to contact SOLEIL Management at the acceptance of the proposal.

3.4. Ownership, including any IP rights that might arise, of all results derived from the analysis of the experimental data is determined by the contractual obligations of the person(s) performing the data analysis and/or data interpretation respectively the applicable law.

4. ROLES AND RESPONSIBILITIES

4.1. Unless otherwise specified, SOLEIL will act as a custodian for raw data and associated metadata.

4.2. SOLEIL aims at providing means for reduction and/or processing of raw data.

4.3. The MP is required to establish a Data Management Plan. SOLEIL can bring help by providing a template or by sending out the available information to the MP if the funder or research institution or other organization has particular data management requirements.

4.4. It is the responsibility of the MP (and/or each PI in a BAG) to ensure that the experiment number is correctly entered into the metadata for each raw data set, in order to correctly associate each data set with the MP and, if relevant, with the PI. If this is not done, the experimental team will not be able to access the data via the on-line catalog (6.1) or other users may inadvertently be given access rights to the data. Normally this will be done by simply entering the experiment number via the data acquisition software.

4.5. The MP (or each PI in a BAG) has to ensure sample descriptions are included in the metadata.

4.6. The experimental team is encouraged to ensure that metadata are as complete as possible, as this will enhance the possibilities for everybody to search for, retrieve and interpret the data in the long term.

4.7. SOLEIL undertakes to provide means for the capture of such metadata items that are not automatically captured by an instrument, including the Data Management Plan, to facilitate recording the fullest possible description of the raw data.

[4] Cf. Guidelines on FAIR Data Management in Horizon 2020, European Commission, July 2016, p. 4

[5] ZRR = « Zone à Régime restrictif », cf. http://www.sgdsn.gouv.fr/missions/protection-du-potentiel-scientifique-et-technique-de-la-nation/

4.8. It is the responsibility of the MP (or each PI in a BAG) to ensure that any information, including personal data, added by the experimental team into the metadata, comply with the French and European data protection regulations. SOLEIL is not responsible for their compliance with data protection legislation. SOLEIL can bring help by providing available information.

4.9. It is the responsibility of the MP to ensure that the experimental team agrees on any change of the embargo period (6.5).

4.10. It is the responsibility of the MP (and/or each PI in a BAG), to manage possible authorizations of access to the data, granted to persons from outside the designated experimental team, during this embargo period (6.4) and in conformity with the legislation.

4.11. The MP (and/or each PI in a BAG) has the possibility to transfer or grant parts of all his/her rights during the embargo period to another registered person.

4.12. The MP (or each PI in a BAG) has the obligation to recover and to save all the data issued from the experiment within the duration specified at the acceptance of the proposal. SOLEIL can advise him/her to find an appropriate data repository for very long-term preservation.

4.13. Should a MP (or a PI) no longer be able to ensure his/her role, he/she may inform SOLEIL management and the experimental team, who will together organize a replacement, if possible.

4.14. SOLEIL will, at its own discretion, make its best to ensure an accurate storage and curation as well as an uninterrupted access to experimental data and metadata, during the duration specified at the acceptance of the proposal. However, failures caused by technical or human mistakes cannot be ruled out regarding any data processing. SOLEIL cannot guarantee an absolutely accurate storing and curating. Also, access might be temporarily limited or impossible, especially due to necessary maintenance or overhaul of services or failure of third-party service providers.

4.15. SOLEIL and MP (or each PI in a BAG) cannot be made liable in case of loss or defect of data, metadata or results, as well as for access being limited or unavailable.

4.16. SOLEIL and MP (or each PI in a BAG) cannot be made liable in case of unavailability or loss of data analysis software.

4.17. SOLEIL cannot be made liable for the consequences of any interpretation of the data.

4.18. As a matter of precaution (and without prejudice to the question of ownership) all members of the experimental team grant SOLEIL the unlimited and unrestricted right to use the experimental data and metadata to the extent necessary to curate them and make them available in accordance with this data management policy.

5. DATA STORAGE AND CURATION

5.1. All raw data will be curated in well-defined formats, for which the means of reading the data will be made available by SOLEIL.

5.2. Metadata, whether automatically captured by instruments or manually recorded with SOLEIL tools, will be curated either within the raw data files, within an associated on-line catalog, or within both.

5.3. Raw data and metadata will be read-only for the duration of their life-time.

5.4. Raw data and metadata will be migrated or copied to long-term storage facilities that could be sub-contracted, for long-term curation.

5.5. It is planned that each experiment and data set will have a unique persistent identifier, so that they will be uniquely attributable.

5.6. Unless otherwise specified, reduced data and the associated metadata will not be curated for long-term by SOLEIL. The tools for regenerating them from raw data will be made available by SOLEIL on a best effort basis (via recent versions of supported tools running on recent operating systems).

5.7. Unless otherwise specified, the processed data from the interim analysis steps and the associated metadata will not be curated long-term by SOLEIL.

5.8. Results issued from analyses performed on raw data and metadata using SOLEIL means will be stored long-term by SOLEIL in case of in-house research, and only on a best effort basis in other cases. It will not be the responsibility of SOLEIL to fully curate this data e.g. to ensure the semantic of this data or that software to read / manipulate this data is available.

6. DATA ACCESS AND REUSE

6.1. Access to raw data, metadata and results stored by SOLEIL will be available via a searchable on-line catalog, under the conditions described below.

6.2. Access to the on-line catalog will be restricted to registered users.

6.3. Download of data may be subject to restrictions of the volume which SOLEIL can provide.
Download of data will be logged and the information made available to the MP (or PI in a BAG), within the limits of the legislation.

6.4. During the embargo period, access to raw data and the associated metadata obtained from an experiment is restricted to the experimental team, except specific transfer or grant of rights is granted by the MP (or PI in a BAG).
If people from outside the designated experimental team request access to the data, SOLEIL will transfer the request to the MP (or PI in a BAG), who will examine the request in coordination with the experimental team.

6.5. After the embargo period, all raw data and the associated metadata will become openly accessible.
At least two months before end of embargo period, SOLEIL will remind the deadline to the MP.
Any MP that wishes data to retain restricted access for a period longer than three (3) years (for example to align with the practices of the scientific disciplines to which the proposal relates) will have this possibility on a yearly basis, with justification (via the SUN set tool), up to a maximum prolongation of two (2) years. For longer extensions, a written request shall be submitted to SOLEIL Scientific Directors, specifying the reasons for the proposed prolongation, who decide on the request. In exceptional circumstances, data can be made openly accessible earlier than the initial embargo period if the MP requests SOLEIL to do so.

6.6. Access to results issued from analyses performed on raw data and metadata is restricted to the person or persons performing the analysis, unless otherwise requested by those persons. However, if the raw data being analyzed is still restricted, access to results must be granted by the MP (or PI in a BAG) on request.

6.7. Authorized SOLEIL staff (e.g. instrument scientists, IT staff) have access to any curated data or metadata for facility related purposes. SOLEIL will undertake that confidentiality of data with restricted access is preserved.

6.8. The on-line catalog will enable linking experimental data to experimental proposals. Access to proposals will only be provided to the experimental team and appropriate facility staff, unless otherwise requested by the MP (or PI in the case of a BAG).

6.9. A limited subset of metadata will be made public immediately on completion of experiments. This information will be available via the unique persistent identifier landing page on the web.

6.10. SOLEIL strongly recommends that researchers, who aim to carry out analyses of raw data and metadata which are openly accessible, should, when possible, contact the original MP (or PI in a BAG), to inform him/her/them, suggest collaboration and propose to be co-author in any future publication arising from the data, if appropriate.

6.11. Anybody publishing results based on open access data must acknowledge the source of the data, cite its unique persistent identifier and any publications linked to the same raw data or the origin of the raw data. Furthermore, SOLEIL encourages making such results openly accessible.

6.12. Allocation of SOLEIL beamtime commits the MP (and/or each PI in a BAG) to respond to requests about data obtained from the experiment once openly accessible.

7. PUBLICATION INFORMATION

7.1. Publications related to data from experiments carried out at SOLEIL must cite the unique persistent identifier of the experiment and data in their publication.

7.2. References for publications related to experiments carried out at the facilities must be deposited in the publications database (function “Register Publications” of the User menu in the SUN set) as soon as the article is published. Failure in providing publication reference can lead to a refusal of new beamtime allocation.

8. TERMINATION OF CUSTODIANSHIP OR ON-LINE CATALOGUE

8.1. If SOLEIL decides to not continue to act as custodian and/or to maintain and provide the on-line catalogue, SOLEIL will inform the MPs (and PIs) concerned in a timely manner and provide them with effective means to make a copy of the respective raw data, metadata, and results, provided SOLEIL is aware of the correct email address of the MP (or PI in a BAG) at that time.

SESAME Scientific Data Policy

Central facilities for neutron scattering and synchrotron X-rays in Europe are working together increasingly to develop and share infrastructure for the data collected there. Such co-operation should make it easier and more efficient for users to access and process their data, and provide more secure means of storage and retrieval. It should also increase the scientific value of the data by opening it up to a wider community for further analysis and fostering new collaborations between scientific groups. Ultimately this should improve the quality and quantity of publications from such data. However, with these developments comes a need to define how such data and any associated metadata are stored and made accessible, and for this a common data policy has been established to provide a suitable working framework.

1. General principles

1.1 This data management policy pertains to the ownership of, the curation of and access to experimental raw data and metadata collected and/or stored at SESAME.

1.2 Acceptance of this policy is a condition of the award of beamtime.

1.3 Users must not attempt to access, exploit or distribute raw data or metadata unless they are entitled to do so under the terms of this policy.

1.4 Deliberate infringements of the policy may lead to denial of access to raw data or metadata and/or denial of future beamtime requests at SESAME.

1.5 All data and metadata will be subject to the data protection legislation of Jordan, in which the data and metadata are stored.

2. Definitions

For the purposes of this policy:

2.1 The term raw data pertains to data collected from experiments performed on SESAME instruments. This definition includes data that are created automatically or manually by SESAME specific software and/or SESAME staff expertise in order to facilitate subsequent analysis of the experimental data.

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

2.3 The term principle investigator (PI) pertains to the Main Proposer identified on the experiment proposal. For experiments outside of SESAME proposal system, the PI is the person initiating or performing the experiment.

2.4 The term experimental team includes the PI and any other person to whom the PI designates the right to access resultant raw data and associated metadata.

2.5 The term public research refers to research done through peer review and leading to publication(s).

2.6 The term proprietary research refers to research done through purchased (commercial) access to the research facility.

2.7 The term on-line catalogue pertains to a computer database of metadata containing links to raw data files, that can be accessed by a variety of methods, including (but not limited to) web browsers.

2.8 The term results pertains to data, intellectual property, and outcomes arising from the analysis of raw data. This does not include publications.

2.9 The term long-term means a minimum of 5 years and SESAME will thrive for 6 years. This may obviously depend on the type and volume of data concerned and the economical consequences associated to long-term data storage. Thus the facility reserves the right to restrict the storage periods in consultation with the respective communities for high data rate instruments.

2.10 The term open access means belonging to the community at large, unprotected by copyright or patent and subject to appropriation by anyone. Open access to research data from public funding should be easy, timely, user-friendly and preferably Internet-based.

3. Raw data and associated metadata

3.1 Access to raw data and associated metadata

3.1.1 All raw data and the associated metadata obtained as a result of publically funded access to the research facilities are open access, with SESAME acting as the custodian.

3.1.2 All raw data and the associated metadata obtained as a result of proprietary research will be owned exclusively by the client who purchased the access. Proprietary users must agree with the facility management how they wish their raw data and metadata to be managed before the start of any experiment.

3.1.3 Data from PRT instruments measured during SESAME beam time will be subject to the same rules as data from SESAME instruments.

3.2 Curation of raw data and associated metadata

3.2.1 All raw data will be curated in well-defined formats, for which the means of reading the data will be made available by SESAME.

3.2.2 Metadata that is automatically captured by instruments will be curated either within the raw data files, within an associated on-line catalogue, or within both.

3.2.3 Data will be read-only for the duration of its life-time.

3.2.4 Data will be migrated or copied to archival facilities for long-term curation.

3.2.5 It is planned that each data set will have a unique identifier. Anybody providing data with the same identifier must make sure that the copy is identical to the data in the facility database. Anybody publishing results based on open access data must quote the same identifier (and related publications if available & appropriate).

3.3 Access to raw data and metadata

3.3.1 Access to raw data and metadata in the facility is foreseen to be available via a searchable on-line catalogue.

3.3.2 Access to the on-line catalogue of the facility will be restricted to those who are registered users of the on-line catalogue. Downloading of data will be logged and the information made available to the PI.

3.3.3 Access to raw data and the associated metadata obtained from an experiment is restricted to the experimental team for a period of 3 years after the end of the experiment. From 3 years to 5 years after the end of the experiment people from outside the designated experimental team may request to access the data, in such situation, the facility management in coordination with PI will study the release of the access. Any PI that wishes his data to remain restricted access for a longer period will be required to make a special case to the respective facility management. Data can always be made openly accessible earlier on simple request of the PI. A limited subset of metadata will be available immediately on completion of experiments. If data can only be stored at the facility for less than three years, then access is exclusive to the PI up to the end of the storage period.

3.3.4 It is the responsibility of the PI to ensure that the experiment number is correctly entered into the metadata for each raw data set, in order to correctly associate each data set with the PI. If this is not done, the experimental team will not be able to access the data via the on-line catalogue or other users may inadvertently be given access rights to the data.

3.3.5 Appropriate SESAME staff (e.g. instrument scientists, computing group members) has access to any SESAME curated data or metadata for SESAME related purposes. SESAME will undertake the preservation of the confidentiality of such data.

3.3.6 Data and metadata may be confidentially reviewed in strict cases based on court order or SESAME Council request.

3.3.7 The on-line catalogue will enable the linking of experimental data to experimental proposals. Access to proposals will only ever be provided to the experimental team and appropriate facility staff, unless otherwise authorized by the PI.

3.3.8 The PI has the right to transfer or grant parts or all of his rights to another registered person.

3.3.9 The PI has the right to create and distribute copies of his raw data.

4. Results

4.1 Ownership of results

4.1.1 Ownership of all results (intellectual property) derived from the analysis of the raw data is determined by the contractual obligations of the person(s) performing the analysis.

4.2 Curation of results

4.2.1 SESAME will provide a means for users to upload results and associated metadata to the facility and enable them to associate these results with raw data collected from the facility.

4.2.2 The upload of results and associated metadata may be subject to volume restrictions.

4.2.3 These results will be stored long-term by SESAME. It will not be the responsibility of SESAME to fully curate this data e.g. to ensure that software to read/manipulate this data is available.

4.2.4 SESAME cannot be made liable in case of unavailability or loss of data.

4.2.5 SESAME cannot be made liable in case of unavailability or loss of data analysis software.

4.3 Access to results

4.3.1 Access to the results of analyses performed on raw data and metadata is restricted to the person or persons performing the analyses, unless otherwise requested by those persons. However, if the raw data being analysed is still restricted, access to the analysis results must be granted to the PI on request.

5. Good practice for metadata capture and results storage

5.1 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.

5.2 SESAME 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.

5.3 Researchers who aim to carry out analyses of raw data and metadata which are openly accessible must, when possible, contact the original PI or their designate (3.3.7) 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.

5.4 PIs 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 SESAME provided by the on-line catalogue. Furthermore, they are encouraged to make such results openly accessible.

6. Publication information

6.1 References for publications related to experiments carried out at the facilities must be deposited in a publications database within 3 months of the publication date, or during any new application for beamtime, whichever is the earlier