

General Analysis

	PaN-Data [Comparing Reference Data Policy]	ESRF	ILL	XFEL	CERIC	ESS
1. General principles	1.1. Subjects of data and metadata management policy 1.2. Condition of the award of beamtime 1.3. Users data managing limitations 1.4. Policy infringements 1.5. Data protection legislation	<u>Differences:</u> None <u>Additions:</u> None <u>Deletions:</u> None	<u>Differences:</u> None <u>Additions:</u> <i>Point 1.6.</i> Unique identifier defined as Digital Object Identifier (DOI) to identify a data set < proposal number [+ reference to the ILL experiment] > <u>Deletions:</u> None	<u>Differences:</u> <i>Point 3.1.(Pt 1.1. in PaN-Data)</i> + Management policy pertains also archiving data and metadata + Possibility to use subcontractors to perform Scientific Data Policy obligations. <i>Point 3.4.(Pt 1.4. in PaN-Data)</i> + Punishment in case of repeated non-culpable infringements <i>Point 3.5.(Pt 1.5. in PaN-Data)</i> + Application of data protection legislation only if contains personal data <u>Additions:</u> <i>Point 3.6.</i> Registration at the user portal (UPEX) is needed for users of XFEL. Application of UPEX Privacy Policy or Terms and Conditions for the non-proprietary user access to the European XFEL Facility or user access to UPEX <u>Deletions:</u> None	<u>Differences:</u> None <u>Additions:</u> None <u>Deletions:</u> None	<u>Differences:</u> <i>Point 1.1.</i> - Ownership and curation are not mentioned - Proprietary access is not considered + The usage of the data pertains to this policy <u>Additions:</u> <i>Point 1.2.</i> Revision to this policy may occurred for any reason at any time and with immediate effect. It must be approved by " Council " even retroactively in case it was previously signed by General Director <i>Point 1.2.</i> The revision will be notified and the affected parties will be allowed to make adjustment <u>Deletions:</u> None
	2.1. facility 2.2. raw data 2.3. metadata 2.4. principle investigator 2.5. experimental team 2.6. public research 2.7. proprietary research 2.8. on-line catalogue 2.9. results 2.10. long-term 2.11. open access	<u>Differences:</u> <i>Point 2.1. (Pt 2. in PaN-Data)</i> ≠ Differentiation between peer-reviewed experiments from in-house experiments <i>Point 2.3. (Pt 4. in PaN-Data)</i> + Introduction of Safety Approval Form that defines a PI for in-house experiments <i>Point 2.11.</i> + Availability of archive under CC-BY (Creative Commons Attribution License creativecommons)	<u>Differences:</u> <i>Point 2.1. (Pt 2.2. in PaN-Data)</i> + specified that data have not yet been reduced or processed by any data-treatment software + specified the subdirectory where raw data are curated <i>Point 2.2. (Pt 2.3. in PaN-Data)</i> + Definition of what metadata include (but are not limited to) with some examples as scientific abstract, logfiles, instrument configuration, state points ecc. + specified the subdirectories where	<u>Differences:</u> <i>Point 2.8. (Pt 2.9. in PaN-Data)</i> ≠ 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 <i>Point 2.10. (Pt 2.4. in PaN-Data)</i> + For experiment outside of proposal system the PI must obtain a management board approval of the Institute	<u>Differences:</u> None <u>Additions:</u> None <u>Deletions:</u> <i>Point 2.10. in PaN-Data</i> [Shortly] long-term means a minimum of 5 years and facilities will thrive for 10 years, depending on the	<u>Differences:</u> <i>Point 2.2.(Pt 2.4 in PaN-Data)</i> ≠ PI outside the proposal system is not considered <i>Point 2.3.(Pt 2.5 in PaN-Data)</i> ≠ Proposal team is defined instead of experimental team, it has the same role <u>Additions:</u> <i>Point 2.1.</i>

General Analysis

PaN-Data [Comparing Reference Data Policy]		ESRF	ILL	XFEL	CERIC	ESS
2. Definitions		<p>Additions:</p> <p><i>Point 2.9.</i> Custodian refers to the Institute storing, curating and providing access to raw data, metadata and results</p> <p>Deletions:</p> <p>None</p>	<p>Specifies the scientific conclusions from metadata are curated.</p> <p><i>Point 2.5. (Pt 2.9. in PaN-Data)</i> + Specified that result refers to treated/analysed data and/or the associated scientific conclusions.</p> <p><i>Point 2.10. (Pt 2.11. in PaN-Data)</i> + Download access only and subject to certain conditions and obligations.</p> <p><i>Point 2.17. (Pt 2.5. in PaN-Data)</i> # Experimental team includes all participants, including the ILL local contact(s), in the experiment(s) performed for a given proposal</p> <p><i>Point 2.19. (Pt 2.8. in PaN-Data)</i> + on-line catalogue refers also software tools allowing access to ILL-curated data (raw, meta- and processed).</p> <p>Additions:</p> <p><i>Point 2.3.</i> processed data refers to raw data, metadata processed or reduced by data-treatment software. Are curated in the /processed/ subdirectory. Processed data can be also considered to be results.</p> <p><i>Point 2.4.</i> Data refers to ensemble of raw, meta- and processed data</p> <p><i>Point 2.6.</i> Curation denotes the long-term storage, backup and protection of data and results.</p> <p><i>Point 2.8.</i> Access refers to read-access only in the case of raw data, to read-access and possibly also write-access for metadata, and to read+write-access for processed data and results.</p> <p><i>Point 2.9.</i> Access Control List (ACL) software to manage users access to data files.</p> <p><i>Point 2.11.</i> Open data refer data enjoying open access are called (released under CC-BY)</p>	<p><i>Point 2.12. (Pt 2.5. in PaN-Data)</i> + Person to whom the PI designates must be registered in the user portal + For experiment outside of proposal system the experiment team is defined in the metadata catalogue</p> <p><i>Point 16. (Pt 2.8. in PaN-Data)</i> # On-line catalogue is called metadata catalogue</p> <p><i>Point 2.19. (Pt 2.11. in PaN-Data)</i> + Data will be published under CC-BY (Creative Commons Attribution License creativecommons)</p> <p>Additions:</p> <p><i>Point 2.1.</i> Instrument refers to all instruments designed to perform scientific experiments and located at the Facility, including those contributed by third party organizations and user consortia</p> <p><i>Point 2.2.</i> Scientific data refers to data collected from peer-reviewed and in-house experiments performed on Institute 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</p> <p><i>Point 2.4.</i> Processed data that refers to scientific data that is derived from raw data</p> <p><i>Points 2.5.-2.6.-2.7.</i> Three types pf data that refer to subcategory of processed data: calibrated data, calibration data and alignment data</p> <p><i>Point 2.11.</i> Proposal team refers to PI and any other registered user of the Institute who is defined as a member of the proposal</p> <p><i>Point 2.13.</i></p>	<p>thrive for 10-years, depending on the type, volume of data, economical consequences. Facility can restrict the storage periods in previous communities consultation.</p>	<p>Instrument(s) refers to all instruments used at ESS</p> <p><i>Point 2.4.</i> Proposal number refers to unique identifiers that represent a specific experiment</p> <p><i>Point 2.7.</i> Scientific data research refers to a collation of raw data and metadata</p> <p><i>Point 2.10.</i> Custodian refers to Data Management and Software Center (DMSC) storing, curation and providing access to the raw data, metadata and results</p> <p>Deletions:</p> <p><i>Point 2.6. in PaN-Data</i> Public research user that pay for their experiments</p> <p><i>Point 2.7. in PaN-Data</i> Proprietary research experiments that have public funding</p> <p><i>Point 2.9. in PaN-Data</i> Results analyzed data</p> <p><i>Point 2.10. in PaN-Data</i> Long-term refers to the storage periods from a minimum of 5 to 10 years</p> <p><i>Point 2.11. in PaN-Data</i> Open access data can be access by the scientific community</p>

General Analysis

PaN-Data [Comparing Reference Data Policy]	ESRF	ILL	XFEL	CERIC	ESS
		<p>CC-BY)</p> <p><u>Point 2.15.</u> Proposal managers are PI and the ILL local contact(s) for a given proposal, they can modify the rights of others to access data pertaining to that proposal.</p> <p><u>Point 2.16.</u> Proposal team includes the main proposer and all co-proposers of the ILL beamtime proposal.</p> <p><u>Point 2.18.</u> Authorized member refers to a person having full access to the data for a specific proposal. Are included proposal team, experimental team(s) end proposal managers.</p> <p><u>Point 2.20.</u> Registered user refers to any individual who has been granted an ILL User Club account</p> <p><u>Deletions:</u></p> <p>None</p>	<p><u>Point 2.13.</u> Beamtime refers to the period allocated to the experiment team for performing scientific Experiments (every beamtime has a well-defined start date, end date, and experiment team)</p> <p><u>Point 2.17.</u> Custodians refers to the Institute storing, curating and providing access to raw data, metadata and results</p> <p><u>Point 2.20.</u> European XFEL GmbH shall mean European X-Ray FreeElectron Laser Facility GmbH.</p> <p><u>Point 2.21.</u> European XFEL Facility shall mean the two research facilities</p> <p><u>Point 2.22.</u> Embargo period refers to the period of three year, starting from the end of the beamtime</p> <p><u>Deletions:</u></p> <p>None</p>		
<p><u>3.1. Access to raw data and associated metadata:</u></p> <p>Raw data and metadata derived from research publically founded are open access and facility is the custodian.</p> <p>Raw data and metadata derived from proprietary research will be</p>	<p><u>Differences:</u></p> <p><u>Point 3.1.</u> + Underlining that all experiments type peer-reviewed, in-house and Management Contingency will be open after embargo period (at least 3 years) # Specified the data and metadata will be removed all after experiment in case of proprietary researches</p>	<p><u>Differences:</u></p> <p><u>Point 3.1.</u> (Called ownership data) + All data pertains to processed data # The embargo period from 3 years to 5 years + Data from CRG instruments obtained during ILL beamtime are subject to the same rules as data from ILL instruments + Data from CRG instruments</p>	<p><u>Differences:</u></p> <p><u>Point 4.1.</u> (Pt 3.1. in PaN-Data) + Specified that there will be an embargo period during which access is restricted # proprietary research is not covered by this Scientific Data Policy furthermore proprietary research derived from data and metadata will be removed all after experiment in</p>	<p><u>Differences:</u></p> <p><u>Point 6.1.</u> (Pt 3.3.3. in PaN-Data) # Open access after a reasonable embargo period (not specified the duration)</p> <p><u>Additions:</u></p>	<p><u>Differences:</u></p> <p><u>Point 3.1.</u> + ESS provided a persistent identifier for each container of data and metadata</p> <p><u>Additions:</u></p>

General Analysis

	PaN-Data [Comparing Reference Data Policy]	ESRF	ILL	XFEL	CERIC	ESS
metadata	<p>owned exclusively by the users who must agree with the facility management how they wish their data to be managed before the start of any experiment.</p> <p>3.2. Curation of raw data and associated metadata:</p> <p>Raw data are curated in well-defined formats, facility provide means of reading them.</p> <p>Metadata that is automatically captured by instruments are curated within the raw data files or/and within on-line catalogue.</p> <p>Data will be read-only for the duration of its life-time.</p> <p>Data will be migrated or copied to archival facilities for long-term curation.</p> <p>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. Anybody publishing results based on open access data must quote the same identifier (and related publications if available & appropriate).</p> <p>3.3. Access to raw data and metadata:</p> <p>On-line catalogue allow to access to raw data and metadata</p> <p>On-line catalogue of the facility will be either open access or restricted to registered users of the on-line catalogue.</p> <p>Raw data and metadata is restricted to the experimental team for a period of 3 years after the end of the experiment, thereafter it will become openly accessible. PI can request to extend that period or to make data open accessible earlier</p> <p>PI ensure that the experiment</p>	<p>in case of proprietary researches unless previous agreement</p> <p>Point 3.2. ≠ Only metadata generated by software will be archived + High level metadata such as Title, Authors, Abstract, Beamline will be made public as soon as the experiment has been carried out</p> <p>Point 3.3. ≠ Open access to the on-line catalogue is restricted to the registered user + Specified that ESRF will set up an on-line procedure to become a registered user of the on-line catalogue.</p> <p>Additions:</p> <p>None</p> <p>Deletions:</p> <p>None</p>	<p>+ Data from CRG instruments obtained during CRG beamtime, but using ILL's software tools and IT infrastructure for data curation are subject to the same rules as data from ILL instruments or ILL beamtime</p> <p>Point 3.2. + Specified the directory in which raw data will stored /rawdata/ ≠ Metadata automatically captured are curated either within raw data files (in /rawdata/ directory) either in auxiliary files logfiles or parameters survey (in /logfiles/ /histo/ dir.) and either in processed files (in /processed/ dir.) + Proposal-related metadata are stored in a ILL proposal database + Process data are curated as files of various formats in /processed/ dir. + After the end of experiment a DOI is generated and it is linked to all data of related proposal + Accepted beamtime proposal have empty data directories accessible via ILL Data Portal using beamtime proposal number</p> <p>Point 3.3. ≠ Access to all data is restricted to staff and registered user + The ILL provides registered user with a personal computer account in order to use ILL computing facilities, disk space, and access to data from workstations via ACLs. MyData is symbolically linked to data file of accepted beamtime proposals. User is responsible for the secure computer account. + ILL Data Portal allows to access to data (raw, meta- and processed) from outside firewall, you can access also via remote login to workstations. User's local contact assumes responsibility for the remote access. ≠ From 3 years to 5 years after the end of the experiment(s) non-authorized members of the proposal who are registered may request access to the data, ILL and PI will come to an agreement about granting access ≠ limited subset of metadata like sample composition or proposal abstract will be release when user will release Experimental Report + All metadata and processed data</p>	<p>be removed all after experiment in case of proprietary researches unless previous agreement</p> <p>Point 4.2. (Pt 3.2. in PaN-Data). ≠ Only metadata generated by software will be archived + High level metadata such as Title, Authors, Abstract, Beamline will be made public as soon as the experiment has been carried out</p> <p>Point 4.3. (Pt 3.3. in PaN-Data). ≠ Open access to the metadata (on-line) catalogue is restricted to the registered user + Specified that XFEL will set up an on-line procedure to become a registered user of the on-line catalogue + PI ensures that proposal and sample description (in addition of experiment number) are correctly entered in metadata (via control system) ≠ Metadata catalogue will enable linking scientific data to experiment proposal + The access is restricted to proposal team, and member of proposal review + Specified that PI can transfer or delegate his right only during the embargo period + Experiment team members can create or distribute copy of raw data if they obtain agreement with PI</p> <p>Additions:</p> <p>None</p> <p>Deletions:</p> <p>None</p>	<p>None</p> <p>Deletions:</p> <p>None</p>	<p>...</p> <p>Deletions:</p> <p>...</p>

Raw data and associated m	PaN-Data [Comparing Reference Data Policy]	ESRF	ILL	XFEL	CERIC	ESS
	<p>number is correctly entered into the metadata for each raw data set. (otherwise it could involve in access problems).</p> <p>Appropriate facility staff has access to any curated data or metadata. They will preserve the confidentiality of such data.</p> <p>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.</p> <p>The PI can transfer or grant parts or all of his rights to another registered person.</p> <p>The PI can create and distribute copies of his raw data.</p>		<p>+ All metadata and processed data remain modifiable only by authorized members of the proposal.</p> <p>+ Downloading of open data will be logged (identity of the downloader too) and PI will be inform.</p> <p>+ Open data are released under CC-BY</p> <p>+ Non-electronic, eletronic that are not fully curated, become open data "indirectly". Downloader can request PI to access to them</p> <p>+ Specified that authorized member during the embargo have full access to data, in particular, read-only permission for raw data and auxiliary files and read+write permission for process data</p> <p>+ Curated metadata have generally read+write permission (by authorized members) and read-only for that within raw data files (beamtime proposal, certain logfiles)</p> <p>+ Processed data are necessarily read+write-able by authorized members of the proposal</p> <p>+ Authorized members can always upload additional metadata or processed data, sftp could be use to upload that files. Change metadata or processed data must not caused the degradation of usability</p> <p>+ Specified that PI ensures that the proposal number is correctly entered into an instrument control computer at the start of the experiment</p> <p>≠ Access to the electronic files of submitted beamtime proposals will however never become open, (scientific description, probably not up-to-date, could be less useful than the content of publications resulting from the proposal)</p> <p>Additions:</p> <p>None</p> <p>Deletions:</p> <p><i>Point 3.3.3. in PaN-Data</i> Data can be made openly accessible before the embargo on simple request of the PI</p>			

General Analysis

PaN-Data [Comparing Reference Data Policy]		ESRF	ILL	XFEL	CERIC	ESS
Results	<p><u>4.1. Ownership of results:</u></p> <p>Intellectual property is determined by the contractual obligations of the persons performing the analysis on raw data.</p> <p><u>4.2. Curation of results:</u></p> <p>Provide means for users to upload results and metadata and allow to associate results with raw data.</p> <p>The upload of results and metadata may be subject to volume restrictions.</p> <p>Results will be stored long-term by the originating facility. It will not be the responsibility of each facility to fully curate this data (e.g. to ensure that software to read / manipulate this data is available).</p> <p>Not liability in case of unavailability or loss of data and/or data analysis software.</p> <p><u>4.3. Access to results:</u></p> <p>Access to results is restricted to the persons performing the analyses, unless otherwise requested by</p>	<p><u>Differences:</u></p> <p>None.</p> <p><u>Additions:</u></p> <p>None.</p> <p><u>Deletions:</u></p> <p><i>Point 4.2.1. in PaN-Data</i> Provide means for users to upload results and metadata and allow to associate them with raw data</p>	<p><u>Differences:</u></p> <p><i>Point 4.1.</i> + Results can be produced either members of the proposal or other registered users (after embargo) + Results stored in the /processed/ subdirectory are treated as processed data destined to become open data</p> <p><i>Point 4.2.</i> + You shouldn't store results in the limited amount of diskspace provided that is not accessible via Data Portal</p> <p><i>Point 4.3.</i> # The curated results will be openly accessible after embargo (if they are stored in /processed/) + Registered user can continue to upload results after embargo.</p> <p><u>Additions:</u></p> <p>None.</p> <p><u>Deletions:</u></p> <p><i>Point 4.2.4. in PaN-Data</i> Facility can't be made liable in case of unavailability or loss of data</p>	<p><u>Differences:</u></p> <p><i>Point 5.1. (Pt 4.1. in PaN-Data)</i> + IP rights that might arise from a results is determined by the contractual obligation + Clauses on intellectual property of Terms and Conditions are applied.</p> <p><i>Point 5.2. and 6 and 9 (Pt 4.2. in PaN-Data)</i> # There must be an agreement between PI and XFEL for curation of the results. + Specified that curation is provided on best effort basis, and XFEL act as custodian + XFEL can decides to not continue to act as custodian, maintain and/or provide the metadata catalogue, PIs will be informed timely, XFEL will provide them means to make a copy all data</p> <p><i>Point 5.3. (Pt 4.3. in PaN-Data)</i> + Calibration, alignment data and metadata will be made available as open access after the experiment, except as otherwise agreed # Access to the results is restricted to the experiment team, unless otherwise requested by PI</p> <p><u>Additions:</u></p>	<p><u>Differences:</u></p> <p>None.</p> <p><u>Additions:</u></p> <p>None.</p> <p><u>Deletions:</u></p> <p>None</p>	<p><u>Differences:</u></p> <p>...</p> <p><u>Additions:</u></p> <p>...</p> <p><u>Deletions:</u></p> <p>...</p>

General Analysis

	PaN-Data [Comparing Reference Data Policy]	ESRF	ILL	XFEL	CERIC	ESS
	those persons. If the raw data still restricted , access to the analysis results must be granted to the PI on request.		<p><u>Point 4.2.5, in PaN-Data</u> Facility can't be made liable in case of unavailability or loss of data software analysis.</p> <p><u>Point 4.3.1, in PaN-Data</u> Registered user can request access to the analysis results before the end of embargo</p>	<p><u>Point 5.2.1</u> Processed data (e.g. calibrated data) from the interim analysis steps and the associated metadata hereto will not be curated for long-term but the tools will be provided to regenerate processed data. A temporary storage will be provided to enable data analysis.</p> <p><u>Point 5.2.2</u> All calibration and alignment data and the associated metadata will be curated for long-term.</p> <p><u>Deletions:</u> None</p>		
Practice for metadata capture and results storage	<p>5.1. Metadata as complete as possible for ease of searching, retrieving and interpretation.</p> <p>5.2. Facility provide means to capture metadata that are not automatically captured by instruments.</p> <p>5.3. Researchers who use open access data should contact PI and acknowledge the source of data and cite the unique identifier and linked any related publication.</p> <p>5.4. PI and researchers who analyzed raw data and metadata are encouraged to link the results via on-line catalogue and make them openly accessible.</p>	<p><u>Differences:</u> None.</p> <p><u>Additions:</u> None.</p> <p><u>Deletions:</u> None</p>	<p><u>Differences:</u> None.</p> <p><u>Additions:</u> <u>Point 5.3.</u> Managers control treatment or processing of raw data and metadata that involves write-access to the /processed/ before and after embargo.</p> <p><u>Point 5.5.</u> ILL cannot be held responsible for the misuse, misinterpretation, or misrepresentation of open data.</p> <p><u>Deletions:</u> None</p>	<p><u>Differences:</u> None.</p> <p><u>Additions:</u> None.</p> <p><u>Deletions:</u> None</p>	<p><u>Differences:</u> None.</p> <p><u>Additions:</u> None.</p> <p><u>Deletions:</u> None</p>	

General Analysis

Good pr	PaN-Data [Comparing Reference Data Policy]	ESRF	ILL	XFEL	CERIC	ESS
Publication information	<p>6.1. Reference for publications must be deposited in database within 3 months the publications date (or during any new application for beamtime).</p>	<p><u>Differences:</u></p> <p>None.</p> <p><u>Additions:</u></p> <p><i>Point 6.1</i> Publications related to data from experiments must cite the persistent identifier.</p> <p><u>Deletions:</u></p> <p><i>Point 6.1 in PaN-Data</i> [Shortly] References for publications must be deposited in a database within 3 months of the publication date</p>	<p><u>Differences:</u></p> <p>None</p> <p><u>Additions:</u></p> <p><i>Point 6.1.</i> Anyone publishing results must cite the DOI corresponding to the ILL beamtime proposal in the references section of the publication.</p> <p><i>Point 6.3.</i> It is encouraged to submitted preprints, postprints and reprints of publications to the library staff.</p> <p><u>Deletions:</u></p> <p>None</p>	<p><u>Differences:</u></p> <p>None.</p> <p><u>Additions:</u></p> <p><i>Point 8.1</i> Publications related to data from experiments must cite the persistent identifier.</p> <p><u>Deletions:</u></p> <p>None</p>	<p><u>Differences:</u></p> <p>None.</p> <p><u>Additions:</u></p> <p>None.</p> <p><u>Deletions:</u></p> <p>None</p>	<p><u>Differences:</u></p> <p>...</p> <p><u>Additions:</u></p> <p>...</p> <p><u>Deletions:</u></p> <p>...</p>

	PaN-Data [Comparing Reference Data Policy]	NOTES / QUESTIONS / ISSUES
1. General principles	<p>1.1. Subjects of data and metadata management policy</p> <p>1.2. Condition of the award of beamtime</p> <p>1.3. Users data managing limitations</p> <p>1.4. Policy infringements</p> <p>1.5. Data protection legislation</p>	<p><u>Issues:</u></p> <p><u>ILL:</u></p> <p>1) DOI could be defined in the next data policy section (<i>Definitions</i>).</p> <p><u>XFEL:</u></p> <p>1) Maybe define the archiving of data is not needed, curation definition should already include archiving.</p> <p>2) Is really needed to add repeated non-culpable infringements? Are there any actual examples?</p> <p>3) Registered user could be define in the next data policy section (<i>Definitions</i>)</p> <p>4) Why are metadata subjected to the country law just in case that contains personal data?</p> <p><u>Maybe useful additions:</u></p> <p>1) Specified the possibility to use subcontractor to provide the data curation.</p>
	<p>2.1. facility</p> <p>2.2. raw data</p> <p>2.3. metadata</p> <p>2.4. principle investigator</p> <p>2.5. experimental team</p> <p>2.6. public research</p> <p>2.7. proprietary research</p> <p>2.8. on-line catalogue</p> <p>2.9. results</p> <p>2.10. long-term</p> <p>2.11. open access</p>	<p><u>Issues:</u></p> <p><u>ESS:</u></p> <p>1) Why proprietary and in-house experiments are not considered ?</p> <p><u>ESRF:</u></p> <p>1) Is really needed to differentiate the peer-reviewed and in-house experiments ?</p> <p><u>ILL:</u></p> <p>1) Is really needed to regroup experiment participants in principal investigator, proposal manager (PI and ILL contact, they decide the access right) , proposal team (PI and co-proposal), experiment team (All</p>

	PaN-Data <i>[Comparing Reference Data Policy]</i>	NOTES / QUESTIONS / ISSUES
2. Definitions		<p>participants)</p> <p>2) Why did you call on-line catalogue metadata catalogue? Can you access only to the metadata? And the raw data?</p> <p><u>XFEL:</u></p> <p>1) Should be the registration mandatory to be an authorized member?</p> <p>2) Why call metadata catalogue instead of on-line catalogue?</p> <p>3) Is really necessary to differ among many types of data and metadata? Could cibrated, calibration and alignment be treated as metadata or processed data?</p> <p><u>ALL:</u></p> <p>1) Is it necessary to define "process data" ? Maybe they could be consider just as results? Otherwise it is needed to well-define the concept to avoid misunderstanding</p> <p>2) Could make sense to introduce concept of "experiment outside the proposal system" ?</p> <p>.</p> <p><u>Maybe useful additions:</u></p> <p>1) "Embargo period" definition that all RIs use in their own policy even if only XFEL define it;</p> <p>2) "Registered user" definition that must be well-defined, it must to be clear whom the definition pertains (i.e. to the on-line catalogue users or the facility users);</p> <p>3) "Unique identifier" definition that is a pillar definition in the Data Policy;</p> <p>4) "Custodian" definition that almost all RIs use in their own data policy but not all define it (ESRF, XFEL);</p> <p>5) "Open data", if it can be well-define</p> <p>6) Specify that open data will be published under CC-BY</p> <p>7) "Curation" definition that must be well-defined, and shouldn't lead to misunderstanding</p> <p>8) "beamtime" definition could be helpful.</p> <p>.</p> <p><u>Maybe NOT useful:</u></p> <p>1) Expose the implementation of the data policy, e.g. the directories used to stored the data, ACL system and so on (ILL)</p> <p>2) "Data" definition could lead to misunderstanding</p> <p>3) "Access" definition could lead to misunderstanding</p> <p>.</p> <p>.</p> <p>.</p>

PaN-Data <i>[Comparing Reference Data Policy]</i>	NOTES / QUESTIONS / ISSUES
<p><u>3.1. Access to raw data and associated metadata:</u></p> <p>Raw data and metadata derived from research publically founded are open access and facility is the custodian.</p> <p>Raw data and metadata derived from proprietary research will be</p>	<p><u>Issues:</u></p> <p>ESRE:</p> <p>1) Is really needed to differentiate the peer-reviewed and in-house experiments, Management Contingency ?</p> <p>2) Why high level metadata are directly open data?</p> <p>ILL:</p> <p>1) Why don't you specified that PI can request to make</p>

	PaN-Data <i>[Comparing Reference Data Policy]</i>	NOTES / QUESTIONS / ISSUES
metadata	<p>owned exclusively by the users who must agree with the facility management how they wish their data to be managed before the start of any experiment.</p> <p>3.2. Curation of raw data and associated metadata:</p> <p>Raw data are curated in well-defined formats, facility provide means of reading them.</p> <p>Metadata that is automatically captured by instruments are curated within the raw data files or/and within on-line catalogue.</p> <p>Data will be read-only for the duration of its life-time.</p> <p>Data will be migrated or copied to archival facilities for long-term curation.</p> <p>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. Anybody publishing results based on open access data must quote the same identifier (and related publications if available & appropriate).</p>	<p>data open before embargo period ?</p> <p>2) Why do you store proposal-related metadata in a separated database?</p> <p>3) Why release subset of metadata before embargo (sample composition, abstract) ?</p> <p><u>XFEL:</u></p> <p>1) Why proposal and sample description must be inserted metadata?</p> <p><u>ALL:</u></p> <p>1) Could be useful to specify a default behavior for the data derived from property research? (as deletion for ESRF)</p> <p>2) Should the open data be accessed just by registered user of the catalogue ?</p> <p>3) Should the embargo period be well-define, an exact number of years?</p> <p>4) Should data (raw and meta) be accessible only for staff and registered users after embargo?</p> <p>5) Should we release some metadata before embargo period ? (like XFEL(Author, Title) and ILL(sample, proposal))</p> <p>6) Should we log downloading operations from on-line catalogue ?</p> <p>7) We should decide and well-define what type of metadata will be curated and which one won't.</p>
	<p>3.3. Access to raw data and metadata:</p> <p>On-line catalogue allow to access to raw data and metadata</p> <p>On-line catalogue of the facility will be either open access or restricted to registered users of the on-line catalogue.</p> <p>Raw data and metadata is restricted to the experimental team for a period of 3 years after the end of the experiment, thereafter it will become openly accessible. PI can request to extend that period or to make data open accessible earlier</p> <p>PI ensure that the experiment</p>	<p><u>Maybe useful additions:</u></p> <p>1) Specified that only metadata generated by software can be curated</p> <p>2) Clarify what kind of metadata will be curated and what type that won't be.</p>

Raw data and associated metadata	PaN-Data <i>[Comparing Reference Data Policy]</i>	NOTES / QUESTIONS / ISSUES
	<p>number is correctly entered into the metadata for each raw data set. (otherwise it could involve in access problems).</p> <p>Appropriate facility staff has access to any curated data or metadata. They will preserve the confidentiality of such data.</p> <p>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.</p> <p>The PI can transfer or grant parts or all of his rights to another registered person.</p> <p>The PI can create and distribute copies of his raw data.</p>	

	PaN-Data <i>[Comparing Reference Data Policy]</i>	NOTES / QUESTIONS / ISSUES
Results	<p>4.1. Ownership of results:</p> <p>Intellectual property is determined by the contractual obligations of the persons performing the analysis on raw data.</p> <p>4.2. Curation of results:</p> <p>Provide means for users to upload results and metadata and allow to associate results with raw data.</p> <p>The upload of results and metadata may be subject to volume restrictions.</p> <p>Results will be stored long-term by the originating facility. It will not be the responsibility of each facility to fully curate this data (e.g. to ensure that software to read / manipulate this data is available).</p> <p>Not liability in case of unavailability or loss of data and/or data analysis software.</p> <p>4.3. Access to results:</p> <p>Access to results is restricted to the persons performing the analyses, unless otherwise requested by these persons. If the raw data will</p>	<p>Issues:</p> <p><u>ILL:</u></p> <p>1) Where should a user store results ?</p> <p><u>ALL:</u></p> <p>1) Should results become open data by default after embargo ?</p> <p>2) Is an agreement necessary for the curation of results ?</p> <p>3) Should a facility be able to decide to stop the data curation unilaterally ?</p> <p><u>Maybe useful additions:</u></p> <p>1) Underling that results can be produced by member of the proposal or by other registered users</p>

	PaN-Data <i>[Comparing Reference Data Policy]</i>	NOTES / QUESTIONS / ISSUES
	<p>those persons. If the raw data still restricted, access to the analysis results must be granted to the PI on request.</p>	
Practice for metadata capture and results storage	<p>5.1. Metadata as complete as possible for ease of searching, retrieving and interpretation.</p> <p>5.2. Facility provide means to capture metadata that are not automatically captured by instruments.</p> <p>5.3. Researchers who use open access data should contact PI and acknowledge the source of data and cite the unique identifier and linked any related publication.</p> <p>5.4. PI and researchers who analyzed raw data and metadata are encouraged to link the results via on-line catalogue and make them openly accessible.</p>	<p><u>Maybe useful additions:</u></p> <p>1) No responsibility in case of misuse, misinterpretation, or misrepresentation of open data.</p>

Good pr	PaN-Data [Comparing Reference Data Policy]	NOTES / QUESTIONS / ISSUES
Publication information	6.1. Reference for publications must be deposited in database within 3 months the publications date (or during any new application for beamtime).	<u>Maybe useful additions:</u> 1) Citation of the unique identifier in publications