Ref. Ares(2022)8327460 - 01/12/2022

This partnership has received funding from the European Union's "EURATOM" research and innovation program under the 101061037 grant agreement

PIANOFORTE Partnership

European Partnership for Radiation Protection Research

Horizon-Euratom – 101061037

D 5.2: Data Management Plan

Lead Authors: Omid Azimzadeh (BfS), Paul Schofield (UCAM), Liz Ainsbury (UKHSA) and Simon Bouffler (UKHSA)

With contributions from WP5 leads/contributors: Laure Sabatier (CEA), Jean-Michel Dolo (CEA), Evaristo Cisbani (ISS), De Angelis Cinzia (ISS), Alan Tkaczyk (UTARTU) and Spyros Andronopoulos (NCSRD)

Reviewer(s): [PIANOFORTE Coordination Team]

Work package / Task	**WP5**	**5.5**
Deliverable nature:	**Data Management Plan**	
Dissemination level: (Confidentiality)	**Public**	
Contractual delivery date:	**30 November 2022 - M6**	
Actual delivery date:	**30 November 2022 - M6**	
Version:	**1.0**	
Total number of pages:	**19**	
Keywords:	**Data Management Plan (DMP), FAIR, Data sharing, STORE**	
Approved by the coordinator:	**30.11.2022**	
Submitted to EC by the coordinator:	**30.11.2022**	

Disclaimer:

The information and views set out in this report are those of the author(s). The European Commission may not be held responsible for the use that may be made of the information contained therein.

Abstract

PIANOFORTE Data Management Plan (DMP) describes the strategies, measures and tools for data sharing according to the FAIR (Findable, Accessible, Interoperable and Re-Usable) principles in the radiation research community. The DMP includes the security, regulatory and ethical aspects to ensure safe, transparent and efficient data storage, maintenance, sharing and use.

The DMP will address the key issues of the data structure, standardised strategies for data collection and archiving, and secure and transparent platforms for data sharing and exploitation in proposed and funded PIANOFORTE projects.

The integration of a project DMP into project planning and evaluation is a crucial issue for PIANOFORTE. Each Project will be required to produce its own DMP based on the overall PIANOFORTE DMP that will be taken into account in the assessment for funding of the proposed projects.

Content

1. Aim of the document

1.1. Introduction

The PIANOFORTE Partnership aims to improve the radiological protection of members of the public, patients, workers and the environment in all exposure scenarios and provide solutions and recommendations for optimised protection following the Basic Safety Standards. The focus will be on research and innovation efforts, and original Research and Innovation projects (RI) will be supported through a competitive open-call process. These projects will generate the original data for PIANOFORTE.

Research projects focusing on identified research and innovation priorities will be selected through a series of three competitive open calls (months 10, 22 and 34) based on the Joint Road Map (JRM) developed by the H2020 CONCERT EJP and on the priorities defined by the European Radiation Protection Research Platforms including the MEENAS platforms (MELODI, ALLIANCE, EURADOS, SHARE, EURAMED, NERIS) as well as national Programme Owners / Managers (POMs) and the Stakeholders of PIANOFORTE.

This PIANOFORTE Data Management Plan (DMP) is a living and descriptive document that defines the strategies and tools for managing the data generated during and after the end of the project by the funded RI Projects and thus represents a statement of the adopted policy of PIANOFORTE with which each project will be expected to comply. The integration of a project DMP into project planning and evaluation is an emerging issue for PIANOFORTE. Each Project will be required to produce its own DMP based on the overall PIANOFORTE plan that will be taken into account in the assessment for funding of the proposed projects.

The DMP clarifies the technical and organizational aspects of managing project data and deliverables as early and as easily as possible. The PIANOFORTE DMP addresses security, legal, and ethical issues to ensure secure, transparent, and efficient storage, maintenance, sharing, and use of data. The template for a DMP for research projects introduced in Horizon 2020 program is used for PIANOFORTE projects. Although the DMP is foreseen as a guide for handling data during and after the completion of the project, it should be well aligned with the project objectives and data structure. Therefore, it is advisable to include and consider the DMP early enough in the project planning. The DMP also needs to be updated during the project if significant changes occur.

This DMP provides a platform for optimal data management to address the issues of making PIANOFORTE data findable, accessible, interoperable, and reusable (FAIR). PIANOFORTE is committed to sharing as much primary and derivative data and research outputs as possible and also to making use of existing routes for sharing such as STORE database as a dedicated sharing platform, along with the development of a new Open Biological and Biomedical Ontology to describe PIANOFORTE data.

This DMP fulfils the requirement of the Open Research Data Pilot (ORD pilot) and was created using the Digital Curation Centre's DMP online tool, based on the "Horizon 2020 DMP" template provided by European Commission. The PIANOFORTE-funded projects will have broad-ranging aspects, resulting in a very diverse range of data types and content including omics data and clinical data, through environmental monitoring, to behavioural and attitudinal surveys.

This deliverable represents the Detailed DMP which will be revisited annually and revised as needed, in response to developments within the project and in the light of the RI Projects funded in the Open Calls, following the topics of each RI project as they are funded and as data generation and outputs

evolve. It will act as a set of criteria against which data sharing by RI Projects will be checked and validated.

1.2. Objectives

The PIANOFORTE DMP is an essential and descriptive document that defines the strategies and tools for managing the data generated in a project proposed and funded by PIANOFORTE. The DMP clarifies the technical and organizational aspects of managing project deliverables as early and as simply as possible. PIANOFORTE's DMP addresses security, legal, and ethical issues to ensure safe, transparent, and efficient storage, maintenance, sharing, and use of data.

1.3. Purpose and scope of this deliverable

This document provides a detailed template for a DMP for research projects applied for and funded under the PIANOFORTE partnership. The DMP includes a guide to simplify the content and assist applicants. Each project must prepare its own DMP based on the overall PIANOFORTE plan, which will be considered during the evaluation for funding of the proposed projects (and this will be a clear criterion outlined in the call text).

2. Data Summary

This DMP describes the overarching principles for the management of data within PIANOFORTE. Following this, each Research and Innovation (RI) project funded by the Open Calls will be expected to produce a project-specific DMP including basic information for the project, clarify the project goal and purpose of data generation, collection and use, the quality (type and format) and quantity (volume and duration) of available and generated data, administrative and legal aspects, storage and accessibility, rights and responsibilities, and finally costs and resources.

2.1. State the purpose of the data collection/generation

The purpose of the PIANOFORTE Partnership is to improve the radiological protection of members of the public, patients, workers and the environment in all exposure scenarios and provide solutions and recommendations for optimised protection following the European radiation protection Basic Safety Standards. PIANOFORTE is designed to initiate research and technical development in support of European Union Member States, Associated Countries and the European Commission. Research projects focusing on research and innovation priorities identified by the consortium with input from the stakeholders as described below will be selected through a series of competitive open calls. The data collected in these research projects, funded and executed in fulfilment of the aims of PIANOFORTE will permit dissemination, reuse and accountability of the research carried out in support of the objectives of the project. Further project documents are/will be available on PIANOFORTE homepage.

2.2. Explain the relation to the objectives of the project

The proposed multidisciplinary and inclusive research projects initiated by and funded through PIANOFORTE to realise its objectives will be based on the priorities defined by the Joint Road Map (JRM) developed in the H2020 CONCERT EJP and the European Radiation Protection Research Platforms including the MEENAS platforms (MELODI, ALLIANCE, EURADOS, SHARE, EURAMED, NERIS) as well as national Programme Owners / Managers (POMs) and the Stakeholders of PIANOFORTE. At

the outset of the partnership, high priority was expected to be dedicated to medical applications considering that 1) medical exposures are, by far, the largest artificial source of exposure of the European population and 2) the fight against cancer is a top priority of the present European Commission. To ensure an appropriate continuity in the research goals and methodologies, in line with the contents of the CONCERT JRM, two other priorities have been identified to further understand and reduce uncertainties associated with health risk estimates for exposure at low doses to consolidate regulations and improve practices and to further enhance a science-based European methodology for emergency management and long-term recovery. These high-priority areas are likely to evolve over the course of PIANOFORTE taking into account further stakeholder input and when research projects are selected. The RI Project calls will be made at approximately 10, 22, and 34 months following the launch of PIANOFORTE, and requirements for data management and FAIR data included in the individual project contracts with the successful applicants.

2.3. Specify the types and formats of data generated/collected

As a very wide range of studies in different disciplines is anticipated to be conducted by PIANOFORTE partners, a variety of data types and formats are expected. Proposals submitted to the calls will be required to provide clear details of the types of data they will produce – expected chiefly to be experimental, observational and numerical in nature, in the form of epidemiological, radiobiological, and clinical data sets as well as new models.

The range of data types is expected to be extremely broad that presents an unusual, if not unique, challenge to data management through the 58 partner institutions and potential additional beneficiaries. There will be much data collected as spreadsheets, either excel files or other formats (e.g. TSV), word documents, PDFs, images (e.g. JPG, pyramidal TIFF, as well as zoomable files), DNA and RNA sequence files, mass spectrometry, gas and liquid chromatography raw files, dosimetry data, and data in a range of modelling languages, and GPS data.

2.4. Specify if existing data is being re-used (if any)

Each project is encouraged to re-use existing data to the extent possible. The project DMPs will require individual specifications of data resources and datasets to be re-used. In particular, on a large scale, we anticipate the re-use of human clinical data and dosimetric time series data.

2.5. Specify the origin of the data

The DMPs of the funded projects must specify the source of the data newly generated. The funded projects are expected to generate new data from sources including (but not necessarily limited to):

- Radiation measurements in dwellings and workplaces; measurement in laboratories; environmental and sampling analysis, and environmental monitoring.

- Experimental and physical studies in laboratories involving abiotic materials, animals and plants

- Human clinical and occupational health records, registry and disease data from hospitals and national health data collection services, data collected by other already ongoing epidemiological studies and those from commercial entities.

- Clinical radiotherapy data from electronic health records and related datasets
- Data on disaster preparedness, plans and models, assessments of public perception of risk
- Socioeconomic data gathered from national governments, the European Commission and specific studies.

2.6. State the expected size of the data (if known)

It is envisaged that the size of the data resources generated will vary by several orders of magnitude according to the funded projects, from the very small in pilot style or proof-of-concept studies in the more novel disciplines to big data in its truest form for example for multi-omic radiobiological studies.

2.7. Outline the data utility: to whom will it be useful

The data will be useful to radiation safety regulators and investigators concerned with the biological and health effects of radiation in environmental, medical and industrial contexts. This will improve the radiological protection of members of the public, patients, workers and the environment in all exposure scenarios and provide solutions and recommendations for optimised protection. The data will contribute to the aims of several EU platforms including the MEENAS platforms (MELODI, ALLIANCE, EURADOS, SHARE, EURAMED, NERIS), International Organisations, such as the International Atomic Energy Agency and the International Commission on Radiological Protection, as well as national Regulators and agencies.

3. FAIR Data

The PIANOFORTE DMP aims to establish a culture of effective data handling following "FAIR" principles so that data are findable, accessible, interoperable, and reusable to facilitate effective open science and sharing among researchers, stakeholders, and policymakers.

Full, open-access publications which include the raw data, either within the publications or with a link to a permanent record, e.g. a Digital Object Identifier (DOI), will be required. Early open sharing of research data, models, and other research outputs will be encouraged (including through preregistration, presentation at international conferences and preprints), as well sharing of new data in the context of existing knowledge. At the very least, research data will be required to be made available to the community when the associated manuscripts are published. Only open-call proposals which clearly describe how they will meet these criteria will be considered for funding.

It is expected that the PIANOFORTE DMP will build on the use of existing data-sharing resources including the STORE database which already contains a set of relevant data and is well used by the community (see section 3.2.4).

3.1. Making data findable, including provisions for metadata

To make data findable, the PIANOFORTE DMP recommends well-structured data information. Data produced and/or used in the project should be identifiable and findable through metadata, using a standard identification mechanism (e.g. persistent and unique identifiers such as DOI). The naming conventions, search keywords and version numbering should be clear and well-documented. The type and standards of metadata created and the process of data generation should be well explained.

3.1.1. Outline the discoverability of data (metadata provision)

The funded projects are encouraged to deposit generated data in public repositories that meet FAIR criteria. PIANOFORTE will strongly recommend the use of existing and well-established resources such as the STORE database which already contains a range of relevant data and is well-used by the community. The STORE hosted by the German Federal Office for Radiation Protection (BfS) is a central access portal to information from radiobiology research distributed across scientific institutions (see section 3.2.4).

There are efforts to use public repositories that make use of Open Biological and Biomedical Ontology (OBO) for metadata annotation whenever possible. Many of these provide programmatic access and data discovery is possible through manual searches. Development and implementation of a specific Radiation Biology Ontology (RBO) were planned and progressed during the RadoNorm project. This will improve radiation biology metadata uniformity and transparency. A critical part of the work of WP 5 will be the standards and definitions relating to the vocabulary of data to facilitate interoperability.

Individuals depositing data will be identified by Open Researcher and Contributor ID (ORCID). Data referenced in publications and reports will be linked via database accession identifiers or DOI numbers minted by the repository into which they are placed. Documentation on the source of the data, tools needed to read or use the data, its purpose and sources of legacy data that may have been used in its analysis or combined with new data are available as a free text narrative in the database which can be searched lexically.

3.1.2. Outline the identifiability of data and refer to the standard identification mechanism

Data deposited in public databases, as well as some institutional databases, will attract a persistent digital object identifier (DOI) and database-specific identifier.

3.1.3. Outline naming conventions used

It is suggested that the Technical Information Library Services (TILS) naming convention to be used in the formal deposition of records in the repository databank. This will include the number of the work packages from which they were generated and versioning using the ISO8601data format standard.

3.1.4. Outline the approach towards search keywords

Data need to be annotated with standard ontologies and datatypes as described above. Other high-quality public repository databases use almost all standard OBOs and string searches in a similar way.

3.1.5. Outline the approach for clear versioning

Versions need to be specified in the naming conventions described above and will be attached as metadata.

3.1.6. Specify standards for metadata creation (if any)

Metadata standards are often developed by a specified user community as schemas to provide the most appropriate description of a given type of information and specific purposes.
The Development and implementation of a specific ontology for radiation biology (RBO) were planned and progressed as part of the RadoNorm project. The RBO is under cooperative and continuous development with the Ames laboratory at NASA for additional use with NASA space radiation databases and with the NURA archive at Northwestern University. Because ERA and NURA both used formal ontology for their metadata over the last decade semantic interoperability between all of these archives is proposed to be established in PIANOFORTE using its implementation into STORE (see section 3.2.4). RBO is constructed according to the principles of the OBO Foundry and has been accepted as an OBO foundry ontology. RBO contains new classes, and classes imported from other OBO ontologies or directly cross-referenced to them. For example, all species metadata in RBO is derived from National Center for Biotechnology Information (NCBI) taxon ontology, disease from Human Disease Ontology and environmental features from the Environment (ENVO) ontology.
It is very important that if there are no existing standards in a specific discipline chosen to receive funding through the Open Calls, the recipients must describe what metadata will be created and how this will be standardized.

3.2. Making data openly accessible

Open accessibility of data and results is a fundamental aim for activities under PIANOFORTE. It is foreseen that most data that may not be freely available in the public domain can be shared through a request to the principal investigator or the local data access committee to ascertain that the legal and ethical conditions for data sharing are fulfilled.

3.2.1. Specify which data will be made openly available

Open accessibility of data and results is a fundamental aim for activities under PIANOFORTE. Most data is expected to be licensed under the Creative Commons Attribution license (CC-BY). The acceptable reasons for not publicly sharing data are conflicting intellectual property and privacy issues. In some cases data or part of datasets may be owned by commercial entities; in some cases, personal or sensitive health-related data is involved and consenting or local legal restrictions do not allow open sharing. The software will be shared when there are no IP conflicts under standard licenses.

3.2.2. Specify how the data will be made available

Early open sharing of research data, models, and other research outputs will be encouraged (including through preregistration, presentation at international conferences and preprints), as well sharing of new data in the context of existing knowledge. Full, open-access publications which include the raw data, either within the publications or with a link to a permanent record, e.g. a DOI, will be required. At the very least, research data will be required to be made available to the community when the associated manuscripts are published, which must be in an open-access format.

Data need to be made available as soon as it is curated, subject to publication or IP-related embargo, in the following ways:

- Deposition in public repository databases
- Application to public web services of resources to allow programmatic access in addition to HTML GUIs.
- Application to the data holder

3.2.3. Specify what methods or software tools are needed to access the data?

If the software is required to access the data, the associated documentation for the relevant software (e.g. in the form of open-source or CC-BY licensed code) must be included. Most data only require commonly-used software for reading, for example, Word and generic equivalents, text readers, Excel, and Adobe Acrobat. Excel files may be read using a standard text editor or imported as TSV files. Specific machine readouts and models need dedicated software to use easily. In some cases, there are only obtainable commercially on licence from the machine manufacturers, and in other cases by request to the data originators when the software cannot be freely shared due to IP constraints. Where possible software will be made publicly available through the database, GitHub or the institutional platform. The use of standard, open file forms is encouraged. Specification of the software needed to read the files will be included in the descriptive free text associated with each dataset together with links to where the software can be obtained. Users will be recommended to use tools such as the open OMERO platform to access any image formats, as this includes alternative access to many otherwise proprietary formats.

3.2.4. Specify where the data and associated metadata, documentation and code are deposited

Data deposited in public databases as well as some institutional databases will attract a persistent DOI and database-specific identifier.

Public data-type and theme-specific databases

Domain-specific data types will mostly be deposited in appropriate databases such as Array express, the European Genome Archive, Genbank etc. Each major database has associated with metadata and most have a brief documentation facility, usually a free text element to describe the data and its associated project.

The STORE database

The database was developed with funding from the Euratom Research and Training Programme and is intended to provide a repository of primary data to support publications, protect data at risk of being lost to the community, and maintain legacy data and links to archives, as well as links to biological resource collections for radiobiology projects to facilitate systematic data sharing and archiving. Furthermore, the database provides Standard Operating Procedures (SOPs) for the storage and use of biological samples.

The use of the existing and well-established STORE database is strongly recommended in PIANOFORTE. The database contains a wide range of data and is used by the radiation community and is an infrastructure for effective resource sharing and a central access portal to radiation research data and

records distributed across scientific institutions. STORE is administered by the German Federal Office for Radiation Protection (BfS).

The project data can be deposited in STORE as data files or as references to accession numbers/DOIs for the data in other databases. The organisation of STORE allows for files of multiple types to be grouped as a coherent reflection of data generation in a study irrespective of data type, providing a relationship between the data and an overall context which we believe aids data discovery. Currently, STORE uses structured metadata to describe file contents in considerable granularity using controlled vocabularies. As part of the STORE and RADONORM projects, a formal RBO has been developed.

As RBO is cross-referenced or imports classes from other OBO ontologies STORE will be semantically compatible with databases in the ELIXIR platforms, for example, Proteomics Identifications database (PRIDE), Array Express, the European Nucleotide Archive etc. Data deposited in STORE is in principle semantically interoperable and discoverable within the ELIXIR context. Data deposited in the ELIXIR family of deposition databases will attract the same metadata standards. The adoption of RBO as a standard will assist data integration and discovery globally. STORE also provides the facility for a descriptive narrative, (available for string searches) PUBMED referencing and version identification.

In STORE each dataset and data item is assigned a persistent STORE ID and a DOI which can be used for reference. STORE is registered with re3Data, FAIR-sharing and the ELIXIR Indentifiers.org register of persistent identifiers and is compliant with the FAIR data principles.

The STORE database is free to access and open and permits users to upload and share environmental, experimental, observational or epidemiological data from legacy, recently completed or ongoing studies. Users can archive primary or derived data, and maintain control over dissemination through Creative Commons licensing and user-defined security. STORE also contains links and pointers to large datasets with extensive formalised metadata and contact details and facilitates the discovery of biomaterials and samples of interest.

The STORE database is underwritten by the BfS indefinitely, but in the event of an unforeseen contingency, data will be migrated to the ELIXIR Biosamples database where shared semantic standards will facilitate integration. Data on STORE will be kept live for 15 years with a review of use beginning after ten years. If the data has had low access (<5 requests) after ten years then a further five-year review will determine if it remains live or is archived in "cold storage". Most Institutional databases have a policy of maintaining data for ten years at least. In cases where this lifespan is less than ten years and in the absence of a deposit elsewhere, data can be moved onto STORE and cold archived at the host institution.

STORE database is held on the German Federal government platform of the Federal Office of Radiation Protection and is protected by processes and procedures required by the German Federal Government. Data and platforms are dynamically mirrored at other geographically distant locations and are frequently backed up. The infrastructure is subject to disaster preparedness provisions of the Federal government in case of civil disaster or nuclear accident. Data is transferred using secure protocols.

Institutional databases

Some data will in addition be stored in institutional databases to which the public does not have access and in the cases where users wish to obtain data they will do so by application to the data originator.

Software

Software to be openly shared will be deposited in hosting portals (e.g. GitHub) or made available from institutional websites.

Personal data

Personal and genomic data sharing may be restricted due to GDPR, privacy legislation and ethical considerations. We expect personal data to be made available in negotiation with the data owner, where possible, taking into account privacy and intellectual property issues.

3.2.5. Specify how access will be provided in case there are any restrictions

If the restrictions are foreseen, the DMP should state why and how long they will apply. The restrictions need to be considered, as they will be imposed for:

- Embargo until publication
- Reasonable embargo until IP issues are determined and resolved as specified in the PIANOFORTE guidelines and the individual contracts issues under the Open Calls.
- Proprietary data may be accessed by negotiation with the data holders and owners and by the establishment of licensing agreements
- Personal and genomic data will be restricted due to General Data Protection Regulation (GDPR), privacy legislation and ethical considerations. Data may be accessible by negotiation with the data holder and relevant Institutional data access and ethics committees.

3.3. Making data interoperable

With the wide-ranging research programme envisaged within PIANOFORTE, the development of shared standards for data and research output interoperability will be essential. To make data interoperable, the DMP requires documenting permissible data exchange and reuse between researchers, institutions, organisations and countries (i.e. following standards for formats, compatibility with available (open) software applications where possible, and in particular facilitating recombination with different datasets from different origins). The DMP should describe the mechanism of data exchange, and explain ontologies applied in data and metadata, as well as standards and methods. It should promote the use of universal standard vocabularies for all data types to enable interdisciplinary interoperability, or provide mappings and specify specific ontologies or vocabularies if this is unavoidable.

3.3.1. Assess the interoperability of your data

The DMP requires specifying what data and metadata vocabularies, standards or methodologies are needed to facilitate interoperability. Data formats follow community norms and are compliant with generally available software, although this in some cases might be proprietary. We anticipate no problems in combining PIANOFORTE project datasets with others and indeed the extensive reuse we intend to make of legacy data requires formats that allow data integration. The adoption of OBO ontologies for metadata allows the identification of measurement types and units.

3.3.2. Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability?

The DMP recommends using standard vocabulary from OBO ontologies for all data types and where new ontology classes are coined for the RBO, equivalence mappings through database cross-references (DBxrefs) where available and synonyms are included for each class. If not, the project DMP should provide mapping to more commonly used ontologies.

3.4. Making data re-useable (through clarifying licenses)

The projects proposed and funded by PIANOFORTE are anticipated to produce a wide range of data types that should be reusable by other researchers.

3.4.1. Specify how the data will be licenced to permit the widest reuse possible

PIANOFORTE projects are expected to generate a broad range of data types, ranging from exome sequences to behavioural questionnaires. Most are expected to be released under the CC-BY licence with the remainder not publicly released or released under specific licence conditions negotiated between the originator and user. Software generation during the projects is not expected to be extensive, but existing software will be modified making licensing complex and involving multiple IP issues. As far as possible, the software will be licensed under General Public License (GNU GPL) or under a permissive license such as the Massachusetts Institute of Technology (MIT) license, as appropriate and agreed upon between developers.

3.4.2. Specify when the data will be made available for re-use

DMP of proposed and funded projects requires specifying when data are available for re-use. Some data may require an embargo period, for example until publication, or made available to specific collaborators subject to individual agreements or the permission of Data Access Committees, specifically for human health-related or personal data. The DMP requires to state why and for what period a data embargo is needed.

3.4.3. Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project?

Most of the data generated by the project will be reusable by third parties subject to the restrictions described above. The main reasons for not sharing publicly (disregarding publication embargo) are the ethical and legal constraints on sharing human health-related and personal data. It is expected that most of the data generated will be available through some mechanism after its generation. We expect data to be available on a public repository database (such as STORE) for more than ten years and will

remain available after the end of the project. If the re-use of some data is restricted, DMP needs to explain the reasons.

3.4.4. Describe data quality assurance processes

Responsibility for data quality rests with the coordinator and management of the projects resulting from the open calls with oversight from the PIANOFORTE. Each project will be required to include the role of a data manager (data wrangler) who will oversee and check files uploaded to chosen platforms when they are archived for sharing. Before each periodic report, a census will be taken of data shared during the reporting period and sampled for file availability and integrity. PIANOFORTE partners are expected to act according to the European Code of Conduct for Research Integrity (ALLEA, 2017; https://allea.org/code-of-conduct/). This code applies to research in all scientific and scholarly disciplines and presents a set of fundamental principles of research integrity. These principles guide PIANOFORTE researchers in their work as well as in their engagement with the practical, ethical and intellectual challenges inherent in research.

3.4.5. Specify the length of time for which the data will remain re-usable

Data held on large public databases such as the ELIXIR deposition databases will be retained according to individual database policy but is currently at least ten years and expected to be much longer in most cases.

4. Allocation of resources

The DMP should address which person or organisation is responsible for FAIR data management in each project and describe the resources, costs, time and conditions required for the long-term preservation of generated data. DMP recommends appointing a well-trained data wrangler in each project who monitors the quality and integrity of data (files and documents) uploaded to the archiving platforms for sharing.

4.1. Estimate the costs for making data FAIR and describe the potential source to cover the cost

Costs related to data management are expected to be limited and will be covered by the beneficiaries' budget in the Open Call. In most cases, institutional services and support for the implementation of local data management policies are part of the overheads claimed by beneficiaries. The proposed projects are recommended to use the STORE repository for publishing research data, at no cost for service or data storage. The costs for data curation, validation and deposition are included in the costs of the tasks. Individual investigators within the projects have personal responsibility for the curation of their shared data.

4.2. Identify responsibilities for data management in your project

The responsibility for data management rests with the Project Coordinators with oversight of PIANOFORTE. They will be required to appoint individual data wranglers for their WP tasks in a pattern dependent on the type of task and its location. Wranglers are scientists working with the data

themselves and providing support to other scientists involved in these tasks. They have individual training in data submission support and principles of FAIR data.

4.3. Describe the costs and potential value of long-term preservation

The actual cost of long-term preservation depends on the platform where data is deposited. For public repositories data is maintained without charge to the user, but there is obviously a contribution in kind to the project from this. The main value of long-term preservation is the contribution to longitudinal studies of clinical and epidemiological cohorts, changes and impacts of scientific education, evolution and changes in environmental exposure. This value is greatest for primary data collected in the project, which can also be re-analysed in the future in the light of technical and conceptual advances. Rather specific to this project is the opportunity to capture environmental or personal data at a given point of time, or from experiments which cannot be repeated. Derivative data is also useful for the future application of new models or hypotheses. Long-term preservation of data can exert a sparing effect on the use of model animals and avoid duplication of studies.

5. Data security

The DMP requires policies and tools for data security and data recovery, as well as storage and transmission of sensitive data. Data security and recovery fall into two categories; local and institutional data management and management of the public repository such as STORE.

International repository databases

Most repository databases used are within the ELIXIR platform or those supervised by the US NCBI, such as CLINVAR or GEO. These very large platforms have state-of-the-art data backup and recovery systems. Data and platforms stored in STORE are dynamically mirrored at other geographically distant locations and are frequently backed up (see section 3.2.4).

Local and Institutional data management

Institutions and laboratories will have their own policies for data management and data security before data is placed into public repositories. These will be submitted to PIANOFORTE for approval during the assessment of the Open Calls. Individual laboratories use secure local servers for routine data backup regularly and then data is moved either directly to public repositories or institutional platforms for medium to long-term archiving where industry standard backup systems will be used.

Sensitive health and personal data will need to be maintained on servers complying with ISO: 27001 certified Safe Haven requirements and approved by the local Information Governance Office or its equivalent. If a transfer is necessary, data will be moved with encryption using Transport Layer Security (TLS) over the deprecated Single Sockets Layer (SSL) to maintain transmission security for data in motion. Both HTTPs and SFTP may be used. Alongside the use of TLS encryption protocols, an Advanced Encryption Standard (AES) of at least 128-bit will be used, though some will use 256-bit encryption for added protection. Personal data that is assessed as level 4 or higher, will not be kept on unencrypted workstations, laptops or other media.

6. Ethical and legal issues

The DMP recommends including ethical and legal issues that impact data sharing. These should be covered in the context of the ethics review, the ethics section of the description of the action (DoA) and ethics deliverables. DMP recommend including references and related technical aspects if not covered by the former Ethical issues in the project related to:

- Protection of individual data subject to the General Data Protection Regulation (GDPR)
- Sociological and behavioural survey data
- Use of experimental animals
- Human health-related and genetic data

Ethical issues impacting data sharing relate to compliance with European ethical and legal guidelines and legislation for experiments involving animals, and personal data protection. Individual projects will be expected to submit ethical approval for each proposal. The full details of the ethical aspects should also be described in the project proposal. Each proposal/project will be reviewed by WP8 of PIANOFORTE to ensure that the funded research meets the high ethical and scientific standards expected by society. It ensures that research involving human subjects is carried out according to the ethical standards and scientific merit of research; that the rights of research participants and researchers are protected; and that society, which provides the resources for research will ultimately benefit from it.

Personal data protection

Individual institutions undertaking projects will abide by local guidelines and be subject to oversight by an Institutional Data Protection Officer (DPO) and nominated Data Controller. The primary role of the DPO is to ensure that the processing of the personal data of any stakeholders or any other individuals (also referred to as data subjects) comply with the applicable data protection rules and EU General Data Protection Regulation (Regulation [EU] 2016/679). The DPO will have expert knowledge of data protection, personal and professional qualities as well as a good understanding of the project.

The DPO will:

a) ensure that controllers and data subjects are informed about their data protection rights, obligations and responsibilities and raise awareness about them

b) give advice and recommendations to the project partners about the interpretation or application of the data protection rules

c) create a register of processing operations within the project and notify the European Data Protection Supervisor (EDPS) of those that present specific risks (so-called prior checks)

d) ensure data protection compliance within the project and help the latter to be accountable in this respect

e) handle queries or complaints on request by the project, the controller, other person(s), or on his own initiative

f) cooperate with the EDPS (responding to his requests about investigations, complaint handling, inspections conducted by the EDPS, etc.)

g) draw the project's attention to any failure to comply with the applicable data protection rules.

Article 89 of Regulation [EU] 2016/679 provides a mechanism for national derogations from GDPR together with derogations for research data providing flexibility for limitations on the purpose of data collection and storage. Long-term storage of personal, anonymised and pseudoanonymised data will be supported by derogation under article 89 according to principles of proportionality and necessity. There may be differences in the interpretation of regulation and the application of article 89, competences by different national legislations which will be taken into account when considering the long-term storage of such data.

Experimental animals

The use of experimental animals is governed by Directive 2010/63/EU on the protection of animals used for scientific purposes, and local national animal experimentation legislation with specific ethical oversight operated by the institutional animal use and ethics committees. Compliance with legislation and ethical guidelines is a condition of publication of non-human animal research almost without exception and will be a condition of success in the open calls.

Human health-related and genetic data

Research concerning epidemiological data is governed by national and international research ethical and legal requirements. The projects requiring such permissions have data access committees and local ethics committees that oversee the collection and distribution of human health-related and genetic data together with the European Directive on processing and free movement of personal data (Directive 95/46/EC). A detailed description of the provisions for the ethical conduct of consenting, data gathering, transfer, use and storage is contained within the DoA Section 5.1. All data will be anonymised or pseudoanonymised. Beneficiaries will abide by the following legislation:

In addition to national regulations, the relevant international regulations and guidelines will be observed:

- The Charter of Fundamental Rights of the European Union of 12 December 2000.
- The Convention on Human Rights and Biomedicine of the Council of Europe (Oviedo, 04.IV, 1997).
- Additional Protocol to the Convention on Human Rights and Biomedicine, concerning Biomedical Research (Strasbourg, 25.I.2005).
- The Revised World Medical Association Helsinki Declaration (Fortaleza, 2013).
- Recommendations of the Council of Europe.
- Recommendation No. R(64)1 on human tissue banks of 14 March 1994.
- Recommendation No. R (97) 5 on the protection of medical data of 13 February 1997.
- Recommendation No. R (97) 18 concerning the protection of personal data collected and

processed for statistical purposes of 30 September 1997.

- Recommendation Rec (2006) 4 on research on biological materials of human origin of 15 March 2006.
- EU General Data Protection Regulation (Regulation [EU] 2016/679) effective from May 2018
- ICH-GCP Guidelines; Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95), Sept. 1997.
- International Ethical Guidelines for Biomedical Research involving Human Subjects, Council for International Organizations of Medical Sciences (CIOMS), Geneva 1993.
- WHO: Operating Guidelines for Ethics Committee that Review Biomedical Research, Geneva

7. Others

The issues related to local, national, institutional and individual rules and exceptions need to be discussed transparently and documented in the DMP.