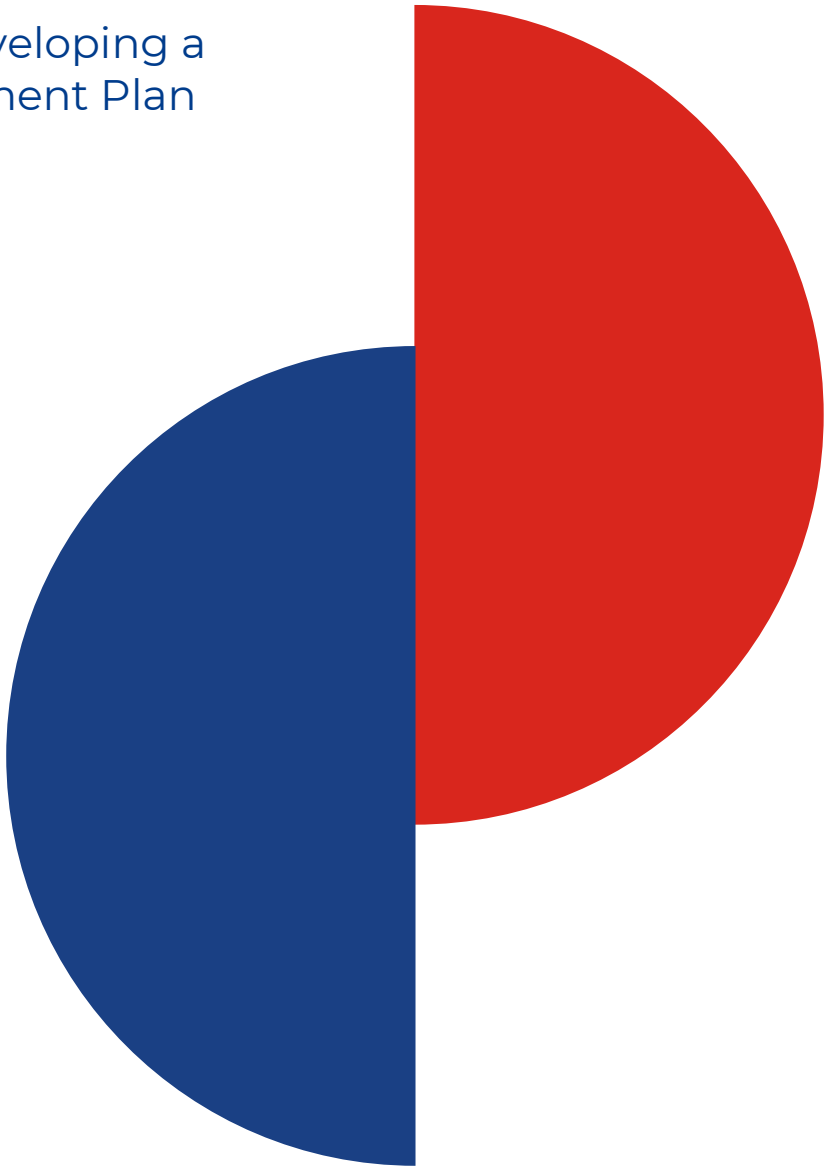
Instituto de Investigación Biomédica de Salamanca

DATA MANAGEMENT PLAN.

A guide for developing a Data Management Plan



Guide to the development of a data management plan

Version 5, March 2024

The Data Management Plan is a formal document that describes the management and processing procedures for research data collected or generated throughout a research project. This document is regularly updated and refined as the project progresses, becoming increasingly detailed and precise. The aim of this document is to serve as a comprehensive guide for researchers required to develop Data Management Plans for various funding opportunities. These plans should specify the measures taken by researchers to ensure their data adheres to the FAIR prociples: findable, accessible, interoperable and reusable.

This document includes:

- A description of the required contents for National Calls (ISCIII, AEI) along with a template for their development.
- A description of the required contents for Horizon 2020 Data Management Plans along with a corresponding template.

Description of contents to be included

The call does not currently require the development of a full Data Management Plan, but rather an executive summary that addresses the following elements:

1. Data typology and format

- Description of the type of data to be obtained:
  - Numerical (e.g., Databases, spreadsheets)
  - Text (e.g., documents)
  - Multimedia (e.g., images, audio, video or a combination thereof)
- File formats (extension) for the datasets or files must be specified.
- Specify which personal data will be collected and used in the research process and clarify whether this will be pseudonymised or anonymised. Recommended tools for anonymization include Manesia Tool by OpenAIRE (http://amnesia.openaire.eu), endorsed by ISCIII, or the Data Protection Agency Anonymisation Tool (https://www.aepd.es/descargas/herramienta-anonimizacion-pdpc).
- Specify methodology for data collection, including software used (e.g. RedCap) and the quality control measures to be applied must be described.
- Indicate whether public or private databases will be used, and confirme the necessary permissions for their use.

2. Procedure for access

- List the individuals in the research tema who will have access to the data.
- Use of database platforms (e.g. RedCap) and whether they are stored on secure servers, on the cloud or on personal computers. Include information about data backups during the project and the periodicity of these backups.
- Specify certain security measures if data is to be sent:
  - Refrain from using services like Gmail, Hotmail or similar services; instead, use always corporate email.
  - Ensure that data is encrypted and transmitted with strong passwords.
  - Data should be sent pseudo-anonymised, after removing any unnecessary indirect or direct identifiers (e.g. using the year of birth instead of the full date birth).

3. Data ownership

- Ownership of the data will depend on its origin:
  - Patient ownership : Data obtained from patient tests or records (e.g., analytical, histopathological tests), as well as any data extracted with informed consent.
  - Institution ownership: Data derived from raw data or animal studies. If patient samples are involved, ownership will be shared between the patient and the institution.

- When data is extracted from external sources (papers or repositories) or from private or public databases (genomic sequences, proteomic sequences, etc.), the ownership resides with the originating entity.

4. Repository for Data Preservation

- All data underlying the research, i.e. raw data generated or produced during the research, must be deposited in the repositories.
- The final data necessary to ensure the verification and reproducibility of the results presented should be published together with the scientific articles.
- The protection of personal data and protection against use by unauthorized third parties, as well as data privacy and confidentiality, must be considered in the storage and publication.
- The IBSAL recommends registering it in Zenodo (https://zenodo.org/), a repository supported by the European Union, as there is an IBSAL community for this purpose. They can be registered in institutional repositories (e.g. www.gredos.usal.es) or in the repository of the Instituto de Salud Carlos III (www.repisalud.isciii.es), when available. Specify in this section the name of the selected repository.
- Determine who will be responsible for developing, archiving in the repository and updating the Data Management Plan (Data Manager).
- Include the type of license to be used for the data and files (Creative Commons license). A helping tool to help in the selection is https://chooser-beta.creativecommons.org/

5. Ethical requirements procedure

- If personal data are collected, describe the measures in place to ensure compliance with personal data protection regulations.
- Specify any ethical or legal conditions that govern the availability and/or its re-use of the data, particularly when the data are protected or protectable by intellectual or industrial property rights.

Working Template

Data typology and format

Clinical and demographic data will be obtained from hospital electronic medical records, structured interview, and other relevant sources. The required variables for the study are described in the methodology section. Data will be collected and managed using the system's electronic data capture tools such as REDCap (Research Electronic Data Capture) or other appropriate systems. This platform ensures traceability, exports in formats such as CSV, R and XML, and facilitates integration with common statistical software packages (SPSS, SAS, STATA, R). The platform will ensure pseudo-anonymisation of data, in compliance with FAIR principles. A detailed description of the data and the methodology will be provided in a readme.txt file accompanying the dataset, along with standardised metadata.

Access

Only authorised research staff will have access to the documents containing personal/relevant data through a personal account. Specific security measures will be implemented to prevent unauthorised identification and access by third parties. Data downloading will be restricted to the IP and user roles. Back-up copies of the database shall be made monthly, semi-annually, or annually and stored on secured devices. Quality controls procedures such as calibration, repeat measurements and validation of data entry will be carried out. Statistical analyses will be performed on fully anonymised data.

Data ownership

According to the proposed regulation for the European Parliament and the Council on the European Health Data Space, the holder of the health data information is the study participant. The data controller is the body where the study is conducted (SACYL, USAL, CSIC, etc.) and the sponsor/principal investigator is the data processor within the scope of the research. The data to be processed will be stored in devices under the control of the responsible body, complying with the relevant requirements in the dimensions of data integrity, authenticity, confidentiality, traceability and conservation. The release of data outside the technical control of the responsible body shall be exceptional and shall require prior justification by the principal investigator, authorised by the responsible body.

Repositories for Data preservation

The data included in the repositories will be anonymised, by using the amnesia tool if necessary, and will be deposited in the open digital repository Zenodo, within the IBSAL Community created for this purpose. The designated researcher (specify their name if it is known) responsible for developing and maintaining the Data

Management Plan will ensure that the plan is delivered within the first six months of the project. The data will be stored in an exportable format (e.g., .csv) and will be published under a **Creative Commons CC-BY** license. A DOI and handle will be assigned to enhance the reusability of the data.

Ethical and legal procedures

The project will recieve prior approval from the Ethics Committee for Research with Medicines (CEIm) of the Salamanca Health Area. The project will be conducted out in strict adherence with ethical standards, the Declaration of Helsinki and relevant data protection legislation including Law 3/2018 and European Regulation 2016/679. The confidentiality and security of the data will be guaranteed, non-anonymised data will be removed and access will be restricted for proprietary data or patentable information.

EUROPEAN CALLS FOR PROPOSALS

Researchers submitting applications for projects under the Horizon Europe program are required to meet a series of requirements regarding research data, which must be included in the Data Management Plan. This document must be submitted within the first 6 months of the project as a deliverable.

Participating projects must complete 2 steps (https://bibliotecas.csic.es/node/195):

- Deposit the data, including its metadata, in a designated repository
- Adopt measures to ensure the accessibility, exploitation, reproduction and dissemination of the data. This may involve using licenses such as Creative Commons (CC-BY or CC-0). The repository should also provide information on the tools required to validate the results.

The structure and information to be included in a Data Management Plan, as outlined by the Horizon Europe Data Management Plan Template and the DMP Checklist is as follows:

1.1 Project Information

This section should include the name and identification of the project, description of the project, institution, the funding agency, the principal investigator and their identifier (ID), contact details, and the version history of the Data Management Plan.

1.2 Summary of data

This section should describe the data to be used in the project, including the type and format, purpose, size, origin and utility of the data.

- Re-use of data: specify whether the data will be re-used and the purpose of such reuse, as well as the original source of the data. Also address the copyright considerations for re-used data.
- Data type and format: describe the type of data the project will generate, (observational, experimental, simulation, derived or compiled, reference, other) and the format of the data (jpeg, pdf, docs, xlsx...).
- Justification for format choice: Provide reasoning for the chosen data formats. This may be based on the expertise of the organization's staff, a preference for open formats, or adherence to recognized data repository standards.
- Purpose of data and relation to project objective: specify the purpose of the data collection and its use, as well as the relation of the data to the project objectives.
- Data size: estimate the expected size of the data to be generated. Provide an estimate of the dataset's volume in MB/ GB/ TB and detail its expected growth overtime.

- Data origin: Indicate the source of the data. For generated data, provide the methodology used, and for reused data, cite the source and the original authors.
- Utility of the data: Describe the potential value of the data for future research.

Additionally, the plan should outline how the data will be organized during the project, including details on version control and folder/file structure.

1.3 FAIR Data

The DMP must ensure that the data adeheres to the FAIR (Findable, Accessible, Interoperable, Re-usable) principles, ensuring the data meets the necessary requirements to be Findable, Accessible, Interoperable and Re-usable.

1.3.1 Making data findable

Metadata should be provided to facilitate the discoverability of the data. This section should specify:

- Persistent identifiers (such as DOI). Indicate whether persistent identifiers they are provided for datasets, individual datasets, data collections or subsets, and clarify which ones. If they are not to be used, explain why.
- Keywords that optimize data reuse.
- Describe data using comprehensive metadata
- Nomenclatures used: Specify any naming conventions applied.
- Versioning: Indicate the number of versions for each dataset.
- Sepcify the metadata standards used (e.g. DDI, TEI, EML, MARC, CMDI). If no established standards exist for the project area, describe the type metadata that will be created and the method of creation.

1.3.2 Making data accessible

This section must specify the repository where the data will be deposited, as well as details on the accessibility of the data and its accompanying metadata.

- Specify the repository where the data will be deposited (e.g., IBSAL community in Zenodo for IBSAL) (LINK).
- Datasets must be asigned DOI and URLs.
- Indicate which data produced and/or used in the project will be open access. If some datasets cannot be shared or require restricted access, explain the legal or contractual reasons for these restrictions and detail the access procedures.
- Clarify wether metadata will be open access.
- Describe the methods or software tools required to access the data (e.g. specific scripts, codes, algorithms, or specifi software versions developed during the project).

1.3.3 Making data interoperable

Data produced in the project must be interoperable, enabling exchange and reuse among researchers, institutions, organizations and countries.

- Indicate the vocabulary and standards to be used.
- Indicate the standards, formats or methodologies that will be used to ensure data exchange and interoperability.

1.3.4 Increasing the re-use of data

This section should document the provenance of the data and provide the necessary information to validate the data and facilitate its reuse, such as README files. Additionally, data usage licenses should be provided.

- Specify the type of license used to maximize the potential for data reuse.
- Indicate when the data will be available for reuse. If there is a delay in availability, provide the reasons for this delay.
- Indicate whether third parties will mbe permitted to reuse the data. If certain data cannot provide the reasons for the restriction.
- Indicate how long the data will remain available for reuse.
- Describe the processes for verifying data quality, such as calibration, repeat sampling or measurement, standardized data capture, data entry validation, peer review, or using controlled vocabularies.

1.4 Other research findings

This section addresses how FAIR principles can be applied to the management of other research outputs generated or reused in the project, such as software, workflows, protocols, new materials, samples, etc. as well as how they will be managed, shared, or made available for reuse, in accordance with the FAIR principles.

1.5 Allocation of resources

Details of both human and financial resources allocated for data management should be provided.

- Describe the roles and responsibilities related to data management and preservation (e.g. data capture, metadata production, data quality, storage and backups, data archiving and data sharing). Indicate the responsible individuals where possible.
- Indicate who is responsible for implementating, reviewing and modifying the Data Management Plan.
- Outline the costs associated with making the data FAIR (e.g. storage, archiving, reuse, security, etc.) and identify the individuals or entities responsible for managing the data.

1.6 Data security

It is essential to ensure that data is securely stored in trusted repositories to facilitate for long-term preservation and retention.

Indicate where the data will be stored and backed up during the project, including the frequency of back up. It is recommended to store data in at least two separate locations and to use robust, managed storage solutions with automatic backup functionality. It is not recommended to store data on laptops, standalone hard disks or external storage devices such as USB sticks.

- Explain the procedures for data recovery in the event of an incident.
- Specify who will have access to the data during the research phase and describe the mechanisms for controlling access, especially in collaborative projects involving multiple research groups.
- Address data protection measures, particularly if the data includes sensitive information (e.g. personal data or trade secrets). Indentify key risks and explain how they will be mitigated.
- Describe any institutional policies in place regarding data protection.

1.7 Ethics

This section should outline any ethical or legal issues that may have an impact on data sharing. These issues may also be addressed within the framework of an ethical review.

- Describe potential ethical issues related to data collection, storage, processing and archiving, along with the procedures for obtaining ethical approval for the project.
- Include procedures for obtaining informed consent for data sharing and the long-term retention of questionnaires containing personal data.
- Consider the anonymization of personal data before retention and/or sharing. If pseudonymization is used instead, specify the methods applied, noting that pseudonymization is a reversible process.

1.8 Other aspects

Provide details on any additional considerations related to data management, including national, funding agency, sectoral, or departmental requirements.

Support documents

There are several tools that can facilitate the development of the PGD:

- Data Stewardship Wizard (https://ds-wizard.org/)
- Argos (https://argos.openaire.eu/splash/index.html)
- DMP Online (https://dmponline.dcc.ac.uk/)
- DMP Tool (https://dmptool.org/)
- PGD Online (https://pgd.consorciomadrono.es/home)

Data Management Plan - Template

Action Number: [insert project reference number]

Action Acronym: [insert acronym]

Action Acronym: [insert acronym]

Action Acronym: [insert acronym

Action title: [insert project title]

Date: [insert date]

Action title: [insert project title]

Date: [insert date]

DMP version: [insert DMP version]

1. Data summary

1.A What is the purpose of the data generation or re-use and its relation to the objectives of the project?

1.A a) Description

Provide a brief explanation of the purpose of the data generated and/or collected and how it relates to the project objectives.

Drafting examples:

- Facilitating data exchange within the project discipline.
- Collecting experimental data to fulfil funding agency commitments.
- Supporting project objectives and in peer-reviewed conferences and publications.
- Enhancing the credibility and the quality of scientific publications based on research data.
- Providing a foundation for future research projects ensuring the long-term impact of this EC-funded project beyond its period.

1.A b) Real example

Ex. 1 The data will be derived from measurements, calibrations, comparisons and validations. It will support project objectives and will be used in conference and peer-reviewed publications.

Experimental data will be collected by the consortium to meet objectives 1

- Measurement and calibration data will result from objectives 1 and 3 and comparison and validation data from objectives 2 and 4. Data from questionnaires and market surveys will be used to support end-user adoption (objective 5).

Ex. 2 Data will be collected and made available to analyze superconducting materials, enhancing the credibility and quality of related scientific publications. This will also facilitate data exchange within the Consortium, enabling distributed sample characterization using different methods. Follow-up projects and future students can build upon existing datasets to validate results and document material and production technique improvements.

The project aims to advance superconducting wire and thin film performance by understanding material behavior, manufacturing influences, and performance limitations (e.g., quality factor for superconducting thin films, current limits under high magnetic fields). A managed collection and publication strategy will establish a long-term library of results for tracking performance evolution and supporting independent verification.

1.B What types and formats of data will the project generate or re- use?

1.B a) Description

Describe the scope and content of the data, which may include:

- Observational data: in real time captured data (e.g., neuroimages, sample data, sensor data, survey data.).
- Experimental data: data captured by laboratory equipment (gene sequences, chromatograms, magnetic field data, etc.).
- Simulation data: test model-generated data sets (e.g., climate, mathematical, economic).
- Derived or compiled data: reproducible but complex data sets, such as text and data mining, 3D models, compiled databases, etc.
- Reference: aggregated datasets (e.g., sequences databases, chemical structure repositories, spatial data portals).
- Other data Formats (e.g., text, numeric, image) must also be indicated.

1.B b) Real example

Ex. 1 The majority of the data will be stored in ASCII (American Standard Code for Information Interchange) format, eg comma separated variable (CSV) format, for

easy import into rich-text files for word-processing or into spreadsheets. If specialized software is used, information about free readers will be provided. Data will be generated in the following formats:

- Graphics: jpeg, odg, pdf, png, pttx
- Tables: odsu, opj, xlsx
- Text: docx, pdf, txt
- Other: nb, cpp

Ex. 2 Openly accessible datasets will include fully characterized sample data, allowing researchers to compare results and build on historical data.

- Data will be stored as Open Document Spreadsheet files (.ODS) for small datasets.
- Large numerical datasets will be formatted in UTF-8 encoded CSV files with detailed column descriptions.
- High-resolution images and raw measurement data will be maintained on an internal data storage platform.
- Proprietary raw instrument data will not be published.
- Metadata (e.g., author, version history, experiment details) will be recorded in a separate METADATA.ODS file for each dataset.

1.C Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded

1.C a) Description

If existing data is reused, specify its source (e.g., research repositories, databases). If purchasing or reusing existing data sources, address copyright and IPR issues.

When creating new data sources, explain why existing data could not be used.

1.C b) Real example

Ex. 1 Some of the project tasks will use existing data in hdf, txt and xlsx formats to validate project results.

Ex. 2 Data from ongoing R&D projects in superconducting wires and thin films will serve as a basis for the data files.

Ex. 3 Selected images and data from partner museums (<Partner1>, <Partner2>, <Partner3>...) will be used in specific tests, such as the storage tests within WP6. Project deliverables containing information will be preserved and shared with future researchers working on <AcronymProject>.

1.D What is the origin/provenance of the data, either generated or re-used?

1.D a) Description

If the data is generated within the project, specify its origin. If it is collected, state the external source from which it was extracted.

1.D b) Real example

Ex. 1 The data will originate from several sources, including pre-existing partner datasets, scientific literature, real-world measurement and simulation experiments. Data from domestic properties will remain confidential and will not be included in the repository.

Ex. 2 The data stem from experiments and measurement campaigns performed by ESRs and colleagues at the beneficiary institutes: 1. Phase A: Superconducting wires and tapes: <Partner1>, <Partner2>, <Partner3>. 2. Phase B: Superconducting thin films: <Partner1>, <Partner2>, <Partner3>.....

Ex. 3 These datasets have been digitized in previous research projects.

1.E What is the expected size of the data that you intend to generate or reuse?

1.E a) Description

Provide an estimate of the dataset's volume (MB/GB/TB) and expected growth over time. Consider storage, backup, cost and access requirements.

1.E b) Real example

Ex. 1 The expected data size is not currently known, but it is likely to be <10 GB with individual files ≤1 MB.

Ex. 2 Data size is initially unknown but will be revised based on management storage needs. The main data volume will come from high-resolution images of microscopic sample characteristics, stored in bitmap format. The total dataset size for a single sample characterization is expected to be in the order of tens of MB only.

Ex. 3 The size of the data handled by <AcronymProject> is quite small, less than 10 GB, except in the tests of the data infrastructure in WP6, where the project needs experience of managing large volumes of data, as explained above.

1.F To whom might your data be useful ('data utility'), outside your project?

1.F a) Description

State the group/s who may be interested in the data.

1.F b) Real example

Ex. 1 The data will be suitable for use by other research groups working on the following topics: biogas, biomethane, energy gases. It will also be useful for standards committees including ISO/TC193/SC1/WG25 Biomethane Working Group, ISO/TC 158 Analysis of Gases and regulators.

Ex. 2 Within the Consortium:

The data sets will be shared within the consortium as the working baseline to produce the scientific publications, to verify and validate the results through repeated experiments at different locations and as a baseline for a comprehensive documentation of the superconducting material performance evaluation in the scope of the world-wide Future Circular Collider Technology R&D program.

Beyond the Consortium:

The data can be used by independent researchers to understand better the contents and conclusions of the scientific publications, which base their findings on the data. Furthermore, independent researchers can use the files to produce figures and publications, showing comparisons of their own results and the

<AcronymProject> results. Scientists can also use the data files to repeat the experiments and measurements to verify and validate the <AcronymProject> research. Finally, the data sets may also be used by scientific writers and the press to produce high-quality infographics, demonstrating the impact potentials of the technology.

Ex. 3 The data from these limited pilots will be useful for users and institutions who may be considering similar technologies in their digitisation and data management work. This applies in particular to the experiments carried out by WP6, but also the others. In particular, the digitised data from the experiments in WP3 will make apparent the quality of the digitisation results achieved with the new technologies. The data in the experiments of WP5 will be useful for the museums.

2. FAIR Data

2.1 Making data findable, including provisions of metadata

2.1.A Will data be identified by a persistent identifier?

2.1.A a) Description

Explain how the data and metadata are assigned to a globally unique and eternally persistent identifier (DOI, Handle...).

2.1.A b) Real example

Ex. 1 The CORA - Research Data Repository (RdR) (https://dataverse.csuc.cat) assigns DOIs for persistent identification and identifiability of the dataset.

Ex. 2 The repository assigns Handle/DOIs for persistent identification and citability of the dataset.

2.1.B Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

2.1.B a) Description

Rich metadata provided: The metadata should document how the data were generated, under what license and how they can be re-used. Also, metadata helps to discover the data and provide the context for proper interpretation by other researchers.

Metadata created and standards: State the metadata standards that will be used. We recommend using metadata standards that are specific to the discipline. Consult metadata standards.

If metadata standards are not used, state what metadata will be generated (manually or automatically) and how.

2.1.B b) Real example

Ex. 1 The metadata standard used to describe the dataset will be the Dublin Core Schema, as it is a flexible and common used standard and is also the one adopted by the European OpenAIRE repository.

Ex. 2 Metadata are created manually by depositors in the deposit form at the repository.

Ex. 3 (1) The data are expected to be provided in ANSI SQL, XML or text (ASCII) format. For this dataset, data citation and metadata practices derived from the community will be considered.

(2) There are no standards for these logs. A possible solution is project servers such as AAA servers. In this case, the logs would include the attributes defined by "project".

Ex. 4 Each file associated with data will be accompanied with unique specified metadata to allow ease of access and re-usability. Below, the form to be followed is presented.

Ex. 5 Standards such as the Dublin Core and ISO/IEC 11179 Metadata Registry (MDR), which addresses issues in the metadata and data modelling space, will be considered.

Ex. 6 There are many different metadata standards for many different types of data and it may not be possible to find one that fits all purposes. Therefore, a pragmatic and feasible approach is to agree on a common and minimal catalogue metadata schema for those datasets that are published in public catalogues and data repositories and to use data-type specific schema extensions, if necessary. In general, the Zenodo deposition metadata domain model which is based on DataCite's metadata schema minimum and recommended terms will be used for open data generated by the project and deposited in an appropriate repository.

2.1.C Will search keywords be provided in the metadata to optimise the

possibility for discovery and then potential re-use?

2.1.C a) Description

State how content search keywords will be created to optimise retrieval and reuse.

2.1.C b) Real example

Ex.1 Data has to be findable easily, rapidly and identically. Therefore, exact and standard measures have to be used to identify the data sets. This can include the definition and use of naming conventions, search keywords, version numbers, metadata standards and standard data identifiers.

Ex. 2 All open project results deposited in a repository will provide search keywords together with their metadata. Keywords for open data will be selected from controlled vocabularies that are suitable for the specific type of data.

2.1.D Will metadata be offered in such a way that it can be harvested and indexed?

2.1.D a) Description

Metadata should be provided structured using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) in such a way as to allow exchange with other repositories. At the same time, the metadata provided should be as detailed as possible to allow it to be indexed so that the data is searchable and retrievable.

2.2. Making data accessible

2.2.A Which data produced and/or used in the project will be made openly available as the default?

2.2.A a) Description

Description of whether and how data will be shared, including access procedures, embargo periods (if any), and definition of whether access will be wide open or restricted to specific groups. If some cannot be made openly available, you must justify why.

2.2.A b) Real example

Ex. 1 All data produced by the experiments of WP3, WP4, WP5, and WP6, which has been described above, will be made openly available. This is any imagery and results of automatic or computer-assisted human interpretation of the data, which can be seen in the imagery. This does not mean that also the details of the

equipment used, and algorithms used in the interpretation will be made openly

available, as these may contain proprietary information. In Zenodo, the option exists to provide open access, embargoed access, closed access.

Ex. 2 All of the data associated with scientific publications will be made openly available as the default unless there is a specific reason not to publish the data. Datasets which cannot be shared - voluntary restrictions Other data may be made available on a case-by-case basis if it is relevant for third parties.

The following data will not be made publicly available:

- Data obtained with the permission of third parties, but the third parties have not agreed to make the data publicly available.
- Data that discloses the identity of a manufacturer.
- Data that compromises the protection of a partner(s) intellectual property. The level of data made available will also be considered, for example, pre-processed data will not be provided unless there is a clear reason for doing so.

Datasets which cannot be shared - legal and contractual reasons All of the data from the project will be made available, with the exception of market or customer survey data, which are commercially sensitive and cannot be shared.

2.2.B How will the data be made available (e.g. by deposition in a repository)

2.2.B a) Description

Describe how the data will be shared, i.e., who will have access to the dataset. You can create a procedure to temporarily make the data accessible to other group members, project partners, and the general public. You should state whether the data will be open access and in what reasonable period. One possibility is to offer them together with the publications. If embargo periods are required, this is where you need to specify them.

2.2.B b) Real example

Ex. 1 The data will be deposited in the storage systems which will be tested by WP6, as appropriate (<Repository1>, <Repository2>, <Repository3>). Links from <AcronymProject> website will be provided to these storage systems. By their service definition, the data stored at <Repository1> remains permanently available. Permanent access to the data on national <Repository2> and <Repository3> tests is not foreseen. Data from the digitisation pilots may remain permanently available if published on <Repository4>. These arrangements will be revisited after the data from the pilots has been created.

Ex. 2 Once processing, quality control, organisation, analysis and publication are complete, the data will be made accessible by deposition in open access repositories (eg Zenodo).

2.2.C What methods or software tools are needed to access the data

2.2.C a) Description

Include any technical requirements for access to and reuse of data.

2.2.C b) Real example

Ex. 1 All the data generated will be accessible via commonly used software: MS Office, Open Office, Adobe Reader, Image Viewer, etc. If any specific software is used,the necessary documentation relating to that software will be provided to facilitate access to the technical data. For software developed under the project, the source code will be deposited in the repository.

Ex. 2 Where possible, standard publicly available software will be used, but if specialised software tools are developed a short text file (e.g. ASCII) will be provided with the data file to explain the required software.

2.2.D Is documentation about the software needed to access the data included?

2.2.D a) Description

You must also include the documentation on the software that is needed to access the data.

2.2.D b) Real example

Ex. 1 If accessed through the API, documentation will be needed.

Ex. 2 Standard publicly available software will be used where possible, but if specialist software tools are developed, i.e. created within Matlab, a short text file (e.g. ASCII) will be provided with the data file to explain the software required.

2.2.E Is it possible to include the relevant software (e.g. in open source code)?

2.2.E a) Description

In case it's specific software, it includes software if possible. For example, in the open source code.

2.2.E b) Real example

Ex. 1 Any such software has already been released by the providers of these storage systems.

Ex. 2 The majority of the software programmes are available as commercial products or as freeware. For the software developed in the project, the source code will be deposited in the repository (eg Zenodo).

2.2.F Where will the data and associated metadata, documentation and code be deposited?

2.2.F a) Description

State the repository in which the data and associated metadata, documents and code will be stored. It can be the same repository or different repositories for the different types of content, for instance, code could be deposit in a specific repository for code.

It is important to use a repository that provides permanent links (DOI, handle) to data in order to facilitate findability and citation.

2.2.F b) Real example

Ex. 1 The data will be deposited in the storage systems which will be tested by WP6, as appropriate (national OSC, EUDAT, Zenodo). Links from <AcronymProject> website will be provided to these storage systems.

Ex. 2 The data and associated metadata, documentation and code will either be deposited in the open access repository called Zenodo or in Open Access Repository (<ExampleURL>).

2.2.G Have you explored appropriate arrangements with the identified repository?

2.2.G a) Description

State if you have explored appropriately which are the requirements of the identified repository.

2.2.G b) Real example

Ex. 1 We have already explored the appropriate arrangements with the national cloud services in Finland (CSC), EUDAT through the work of <AcronymProject> pilot, and Zenodo through the work of the <Disciplinary> Literature Repository community.

Ex. 2 Yes, Open Access Repository is functional and it correctly labels datasets with a metadata scheme that is compatible with DataCite).

2.2.H Specify how access will be provided if there are restrictions on use

2.2.H a) Description

In case public access to data is restricted for any justified reason please specify if data would be accessible to an individual partner, to all partners or under request. Specify procedures of how to request access to restricted data and under which conditions it would be granted. Moreover, specify if restrictions will be lifted after a period of time.

2.2.H b) Real example

Ex. 1 There are no restrictions on use, except when CC BY-NC license has been chosen. <AcronymProject> should address question of sensitive data (e.g. location of protected plants), but <AcronymProject> will avoid working with any sensitive data. If personal data is received in questionnaires, which <AcronymProject> will receive, such data shall be anonymised before making available outside the project.

Ex. 2 There are no restrictions on the use of the published data, but users will be required to acknowledge the consortium and the source of the data in any resulting publications.

2.2.I Is there a need for a data access committee?

2.2.I a) Description

Specify why or why not it is necessary for a data access committee.

2.2.I b) Real example

Ex. 1 Because of the small scale of these experiments, there is no need for a data access committee.

Ex. 2 This consortium will have a data access committee. Their remit will be to select the data that will be openly accessible on a case by case basis. Ethical aspects and data security, including intellectual property requirements, will be considered. If necessary, some or all of a potential publication's data will be withheld. This will be decided in consultation with the relevant partner(s).

2.2.J Are there well described conditions for access (i.e. a machine readable license)?

2.2.J a) Description

Describe which are the conditions for access defined by the repository you have chosen (you can also indicate the URL where the information comes from). For example: a machine-readable license.

2.2.J b) Real example

Ex. 1 The Creative Commons licenses supported by the GBIF will be used. These include CC0, CC-BY, and CC BY-NC (see <ExampleURL>). Zenodo supports a large array of widely used as well as domain specific, machine-readable licences. The owner of the data will determine which of these licenses will be used when data is posted on <AcronymProject> repositories. However, it is the project's recommendation to choose CC0 for data and CC-BY for media and avoid CC-BY-NC which has issues in some national jurisdictions.

Ex. 2 Yes, Zenodo provides well-described conditions for access (see http://about.zenodo.org/policies/).

2.2.K How will the identity of the person accessing the data be ascertained?

2.2.K a) Description

Describe the procedure established by the repository to determine the identity of the person accessing the data.

2.2.K b) Real example

Ex. 1 Identity of the person accessing the data will not be directly ascertained. However, we expect users to follow the standard norms of scientific citation and use of the data in this context will be tracked through scientific citation.

Ex. 2 Users are required to register to use the repository.

2.3 Making data interoperable

2.3.1 Specify what data and metadata vocabularies, standards or methodologies you will follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

A(a) Description

Explain what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. The data interoperability of the project allows the exchange and reuse of data between researchers, institutions, organisations, countries, etc. Adhere to the standards of formats that are, as far as possible, compatible with open programs and applications.

A(b) Real example

Ex. 1 The data produced in the project will be interoperable as the datasets will adhere to standardised formats: ASCII, txt, csv, xml, tiff. If MS Office, pdf viewer or image viewer cannot be used, a text (ASCII) file will be provided with the dataset that explains where a free reader can be obtained.

Ex. 2 Other types of data have been registered following internal codifications, clearly specified within the file.

2.4 Increase data reuse (through clarifying licenses)

2.4.A How the documentation needed to validate data analysis and facilitate data re-use will be provided?

2.4.A a) Description

You must also include the documentation that is needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.).

2.4.A b) Real example

Ex. 1 The data will be accompanied by a readme file for deposit with complementary information to aid the interpretation and re-use of the data.

2.4.B How the data will be licenced to permit the widest reuse possible, in line with the obligations set out in the Grant Agreement?

2.4.B a) Description

If the data are made available to other researchers and the general public, you need to specify what degree of reuse is allowed. This level of reuse will be marked by the establishment of licences. The EC proposes the use of Creative Commons CC-BY or CC0 licences, but there are others.

2.4.B b) Real example

Ex. 1 The data will be licensed under Creative Commons CC-BY licences and will continue to be re-usable after the end of the project by anyone interested in the project, with no access or time restrictions.

2.4.C Are the data produced and/or used in the project useable by third parties, after the end of the project? If the re-use of some data is restricted, explain why.

2.4.C a) Description

In principle, the data should be made available to other researchers and the general public with the fewest possible restrictions. However, there may be several reasons for not sharing them: ethical reasons, protection of personal data, the involvement of intellectual and/or industrial property rights, commercial interests, etc. You must specify the reasons why a dataset will not be shared.

2.4.C b) Real example

Ex. 1 The full dataset will be confidential and only the members of the consortium will have access to it. Furthermore, if it is decided to make specific portions of it (e.g. metadata, statistics, etc.) widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section. Of course, these data will be anonymized so as not to have any potential correlation and identification of the ethical issues with their publication and dissemination.

Ex. 2 Each archived data set will have its own permanent repository ID and will be easily accessible. We expect most of the data generated to be made available without restrictions and only data sets subject to IPR and confidentiality issues will be restricted. Where this is going to be the case, agreements will be made based on the individual data sets. Requests for the use of the data by externals will be approved by the project consortium.

2.4.D How the provenance of the data will be documented using the appropriate standards?

2.4.D a) Description

It is necessary to include information about entities, activities and people involved in producing data.

2.4.E Are data quality assurance processes described?

2.4.E a) Description

Describe what are your data quality assurance processes. How/when internal data quality assessments will be implemented?

The data quality can be ensured by different measures. These include validation of the sample, replication and comparison with results of similar studies and control of systematic distortion.

2.4.E b) Real example

Ex. 1 The quality of the dataset is guaranteed by the platform functioning.

Ex. 2 The data quality is ensured by different measures. These include validation of the sample, replication and comparison with results of similar studies and control of systematic distortion.

Ex. 3 Data quality assurance and control is central and the raison d'étre of this project. About 80% of the efforts spent in our Thematic Centres is directed at data quality assurance.

3. Other research outputs

3.A Will there be other research outputs that may be generated or re-used throughout their project?

3.A a) Description

Explain what other research outputs have been generated in the execution of the project. They can be digital outputs such as software, workflows, protocols, models, etc. or physical outputs such as new materials, antibodies, reagents, samples, etc.

3.A a) Real example

Ex. 1 Where other research outputs are generated (only digital outputs are foreseen), the management of those products is shared with their creators

3.B Will there be other research outputs that may be generated or re-used throughout their project?

3.B a) Description

If you get other research outputs, explain how to make them findable, accessible, interoperable, and reusable. For more information, see 2. FAIR Data:

- Explain if other research outputs will be identified by a persistent identifier, will rich metadata provided to allow discovery, harvested, and indexed, what the metadata will be created and what disciplinary or general standards will be followed, will search keywords be provided in the metadata.
- Explain how these research output will be accessible from the repository, how open or restricted they will be, which license to use.
- Explain what data and metadata vocabularies, standards, formats, or methodologies will you follow to make your research outputs interoperable to allow these outputs exchange and re-use within and across disciplines.
- Explain how you will make these search outputs re-use. What documentation will you provide, what re-use license will you apply, etc.

4. Allocation of resources

4.A What will the costs be for making data or other research outputs FAIR in your project?

4.A a) Description

Estimate the costs for making your data FAIR.

4.A b) Real example

Ex. 1 There are no costs associated to the described mechanisms to make the database FAIR and long term preserved.

Ex. 2 The costs for depositing the dataset with the project, and subsequent resources required to make the dataset publicly available have been included within specific Work Packages within the project.

4.B How will these be covered?

4.B a) Description

State how you plan to cover the cost for making your data FAIR, including additional costs of archiving and preservation: direct and indirect costs related to storage, archiving, re-use, security, etc.

Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions).

4.B b) Real example

Ex. 1 <AcronymProject> is managed and supported by a team of experts and is free of charge.

Ex. 2 The cost of preserving the database will be assumed by the <Partner1>.

Ex. 3 (1) A dedicated hard disk drive will probably be allocated for the dataset. No costs are currently foreseen regarding its preservation.

(2) The cost will be covered at the local hosting institute in the context of the project.

(3) The cost will be covered at the local hosting institute as a part of the standard network system maintenance.

4.C Who will be responsible for data management in your project?

4.C a) Description

Explain the responsibilities for data management in your project.

4.C (b) Real example

Ex. 1 The project coordinator has the ultimate responsibility for the data management in the project.

Ex. 2 Each partner has to respect the policies set out in this DMP. Datasets have to be created, managed and stored appropriately and in line with applicable legislation.

Ex. 3 The Project Coordinator has a particular responsibility to ensure that data shared through the website are easily available, but also that backups are performed and that proprietary data are secured.

Ex. 4 Quality control of these data is the responsibility of the relevant WP leader, supported by the Project Coordinator.

4.D How will long term preservation be ensured?

4.D a) Description

State how you plan regarding long-term preservation and who decides on what data will be kept and for how long.

4.D b) Real example

Ex. 1 Regarding the question of long-term data preservation, no specific arrangements has been done in the consortium yet. However, with a great degree of confidence, it can be confirmed that it is the project coordinator with the help of local <AcronymProject> resources who will play the major role in this task.

5. Data security

5.A What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

5.A a) Description

Briefly describe the technical measures that will be implemented in the short to medium term to ensure data integrity (data backup), recoverability (prevention of data loss) and security (to prevent unauthorised access).

5.A b) Real example

Ex. 1 Data collected from the research group for the Project will be digitised and stored on the University's <StorageService> which is subject to regular back-up that is controlled by the University's IT personnel. The IT department performs operations by type: mission-critical (user data, virtual machines, scientific results, etc.) and static (scientific data sets, intermediate files, etc.). Content will be checked regularly to preserve its integrity, security, and durability. These procedures are designed, set and applied in order to fully comply with personal data as ruled by current national legislation and institutional regulations. Research team members will have an appropriate access level according to their role in the project.

5.B Will the data be safely stored in trusted repositories for long term preservation and curation?

5.B a) Description

Describe where the data will be safely stored in a trusted repository for long term preservation and curation. Also briefly describe the security setting of the chosen repository.

5.B b) Real example

Ex. 1 At the end of the project, all the final data will be deposited at the CORA.RDR data repository. The following list describes their security settings:

- Versions: Data files are versioned. Records are not versioned. The uploaded data is archived as a Submission Information Package. Derivatives of data files are generated, but original content is never modified. Records can be retracted from public view; however, the data files and records are preserved.
- Replicas: All data files are stored Consorcio Madroño in Madrid. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.
- Retention period: Items will be retained for the lifetime of the repository. The RDR has defined a lifetime for the repository of the next 10 years minimum.
- Functional preservation: The RDR makes no promises of usability and understandability of deposited objects over time.
- File preservation: Data files and metadata are backed up nightly and replicated into multiple copies in the online system.
- Fixity and authenticity: All data files are stored along with an MD5 checksum of the file content and the tabular file is stored with Universal Numerical Fingerprint (UNF).
- Files are regularly checked against their checksums to assure that file content remains constant.
- Succession plans: In case of closure of the repository, a guarantee has been made from RDR to migrate all content to suitable alternative institutional and/or subject-based repositories.

6. Ethics

6.A Are there, or could there be, any ethics or legal issues that can have an impact on data sharing?

6.A (a) Description

Description of there is, or could there be, any ethics or legal issues that can have an impact on data sharing. These can be discussed in the context of the ethics review. If relevant include references to ethics deliverables and ethics chapter in the Description of the Action (DoA)

Specify if the informed consent for data sharing a long-term preservation is will be included in questionnaires dealing with personal data.

If your research activities involve children, patients, vulnerable populations, the use of human embryonic stem cells (hESCs) and human embryos (hEs), humans, Human cells or tissues, personal data, animals, Non-EU Countries, Environment, Health and safety, Artificial intelligence, Other ethics issues (man-machine interaction, develops in nanotechnology…), and Crosscutting issue: potential misuse of results (Activities that involve or generate materials, methods, technologies or knowledge that could be misused for unethical purposes) you must comply with ethical principles and relevant national, EU and international legislation.

6.A(b) Real example

Ex. 1 All the activities carried out under the <AcronymProject> project comply with ethical principles and relevant national, EU and international legislation, for example the Charter of Fundamental Rights of the European Union and the European Convention on Human Rights. The tasks for <AcronymProject> only concern basic research activities and the project does not involve humans, animals or cells. Due to the fact that the main domain of the <AcronymProject> project activity is related to materials science with the focus on refractory materials, the risk of having ethics issues during the project is extremely limited. Either way, within the <AcronymProject> DoA Part A, the workpackage 8 is devoted to the ethics issues which sets out the 'ethics requirements' that the <AcronymProject> project must comply with. One deliverable will be provided: D8.1 NEC -Requirement No. 1. In the framework of D8.1, all beneficiaries and partner organisations must confirm that the ethical standards and guidelines of Horizon2020 will be rigorously applied, regardless of the country in which the research is carried out.

Ex 2. The transfer of data on human subjects to the <AcronymProject> repository is only considered when: informed consents, ethics approval and - when applicable

- approval by local data protection authorities cover the purpose that the data are envisaged to be used within <AcronymProject> and allow transfer of individual or aggregated data to the <AcronymProject> repository. All data that are transferred to the <AcronymProject> repository shall be either pseudonymised or completely anonymized. The Data Owner/Data Provider is responsible for the anonymization or prseudonymation process and for ensuring that identifiable variables are not transferred to the <AcronymProject> repository. Directly identifiable variables include - but are not limited to - national ID number, name, phone number, ZIP-code, e-mail address, address, geographical coordinates (at a resolution that risks identification). One shall also be aware that a combination of just of few indirect identifying variables (such as birth data, gender, and zip-code) can be used to identify a large portion of individuals on any dataset. In this context, the Data Owner/Data Provider shall only provide such variables at the lowest possible resolution that is necessary for analysis, e.g. district instead of zip-code; year of

birth or age instead of birth date.

7. Other issues

7.A Do you, or will you, make use of other national/funder/sectorial/ departmental procedures for data management?

7.A(a) Description

Explain the national/funder/sectorial/departmental procedures for data management that you are using.

7.A(b) Real example

Ex. 1 As part of <University>'s commitment to ensuring FAIR and Open data, all research active staff (Postdoctoral fellows, PhD students) are expected to prepare DMPs for their own data, as per the University's Research Data Management Policy. The <University> data management policy defines research data as "the evidence that underpins the answer to the research question and can be used to validate findings regardless of its form". Thus, data covers quantitative and qualitative statements, raw data from measurements and derived data-either cleaned or extracted from a researcher's primary dataset or derived from an existing source.

Ex. 2 As well as European Commission policies on open data management, Project Partners must also adhere to their own institutional policies and procedures for data management:

Ex. 3 Each of the partners will follow their national and institutional procedures for data management, in addition to this <AcronymProject> DMP.