

Data Management Plan for Horizon Europe

(Version 2, December 2023)

(Doc. CO24/02) (4 RDM\DMP\Eina DMP\Plantilles\HorizonEurope\DataManagementPlan_HorizonEurope-EN_v2.docx, 17.05.24)

This document is intended to support researchers in creating their Data Management Plan (DMP). It is specifically aimed at projects financed under the EU's Horizon Europe programme to create a FAIR data management plan.

Key to the numbering:

- A number indicates the fields that are required in Horizon Europe.
- A capital letter indicates the elements that should be considered when filling in each field.
- A lowercase letter indicates the descriptions of each element and a sample of real examples.

This document was prepared by the Research Support Working Group of CSUC, which is composed of representatives from the following universities: University of Barcelona, Universitat Autònoma de Barcelona, Universitat Politècnica de Catalunya, Pompeu Fabra University, University of Girona, University of Lleida, Universitat Rovira i Virgili, Open University of Catalonia, University of Vic-Central University of Catalonia, Ramon Llull University and University of the Balearic Islands.

Examples¹ of Data Management Plans are available online.

This document is licensed under the Creative Commons Attribution (<http://creativecommons.org/licenses/by/4.0/>).

Digital version: <http://hdl.handle.net/2072/537188>

¹ Actris (Grant 654109), Citilab (Grant 635898), ConnectingGEO (Grant 641538), DIMENSION (Grant 688003), DR-BOB (Grant 696114), EGI-Engage (Grant 654142), FREME (Grant 644771), iCirrus (Grant 644526), MAGIC (Grant 689669), MAMI (Grant 688421), MMT (Grant 645487), RAMCIP (Grant 643433), SatisFactory (Grant 636302), Solidus (Grant 649489), Step (Grant 649493), Tandem (Grant 654206), UMobile (Grant 645124), U-Turn (Grant 635773), WaterInnEU (Grant 641821)

Preliminary information

The DMP deliverable must include other preliminary information: the project's logo, the dissemination level, the review history, a table of contents and a list of acronyms.

Consult the “Periodic report template” (or the Web forms under the Participant Portal) or contact your institution.

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

A short introduction text explaining the purpose of the data collection/generation or re-use and the relation of the data to the objectives of the project.

1.A b) Example

Ex. 1 The data will originate from measurements, calibrations, comparisons and validations. It will be used in meeting the project's objectives and in conference and peer-reviewed publications.

Experimental data will be collected by the [institution name] in order to meet objectives 1 - 4. 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 uptake.

Ex. 2 Collecting and making available the data of the analysis of superconducting materials to support the credibility and raise the quality of the scientific publications based on those data. Ease the exchange of data within the [institution name] and promote the distributed characterization of samples with different methods. Permit follow-up projects and further generations of students continuing the work to build upon existing datasets, to validate the results and to document the improvement of materials and production techniques in a verifiable manner. This approach will ensure a durable impact of this EC funded project beyond the project period.

The objective of the project is to advance the performance of superconducting wires and at a later stage thin films by gaining a better understanding of the material behavior, the influence on the production techniques on the performance and to elucidate performance limitations (e.g. quality factor for superconducting thin films on substrate, current limits in wires under high-magnetic field conditions). Managed collection and publication of the data shall help establishing a durable library of results that can help documenting the performance evolution across several years and to permit other researchers validating the results independently.

Ex. 3 The purpose of data generation is related to the achievement of the main objective of the research project, which is to determine the leaching mechanism of REEs in NdFeB magnet scrap for a selective recovery of REEs by means of electrochemical leaching in organic acids.

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

1.B a) Description

Description of the content and scope of the data. Research data are generated for various reasons and through various processes, and may be of the following types:

- Observational: data captured in real time (neuroimages, sample data, sensor data, survey data, etc.).
- Experimental: data captured by laboratory equipment (gene sequences, chromatograms, magnetic field data, etc.).
- Simulation: data generated from test models (climate, mathematical, economic, etc.).
- Derived or compiled: data that are reproducible but difficult to reproduce (text and data mining, 3D models, compiled databases, etc.).
- Reference: conglomerated datasets (databases of gene sequences, chemical structures, spatial data portals, etc.).
- Others

Format of the data (text, numeric, image, etc.) must also be indicated.

1.B b) Example

Ex. 1 Some of the project's tasks will use existing data in [hdf, txt and xlsx] formats. These data will be used in the validation of the project's results.

Ex. 2 Existing data from ongoing R&D projects in the scope of the [project name] study on superconducting wires and thin films will serve as a basis for the data files.

Ex. 3 Selected, existing images and data from the databases of the partner museums ([Partner1], [Partner2], [Partner3]...) will be used in specific tests, such as the storage tests in [WP6]. The final kind of data that will be created is that which is information in project deliverables, which must be preserved, made accessible and passed on to subsequent persons working in [AcronymProject].

Ex. 4 I won't reuse any research data because there is no data available that answers our research questions.

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 you reuse a dataset, specify the source from which it was extracted for example from a relevant repository. If purchasing or reusing existing data sources, explain how issues such as copyright and IPR have been addressed.

When creating new data sources, explain why existing data sources cannot be reused.

1.C b) Example

Ex. 1 The majority of the data will be in [ASCII (American Standard Code for Information Interchange)] data files, eg [comma separated variable (CSV)] format, which can be imported into rich-text files for word-processing or into spreadsheets. If specialised software is used,

then information about free readers will be provided. Data will be generated in the following formats:

- Graphics: [jpeg, odg, pdf, png, pptx]
- Tables: [ods, opj, xlsx]
- Text: [docx, pdf, txt]
- Other: [nb, cpp]

Ex. 2 The openly accessible data will be the comprehensive result datasets of characterized samples that are used to create the figures and plots in scientific publications, such that other researchers can compare their results easier and such that further results including historic data can be produced quicker. The data are value tables in [Open Document Spreadsheet format (.ODS)] for limited amounts of data with typed columns. For larger quantities of numeric data, [UTF-8 encoded, comma separated value in textual format files (.CSV) with column value and data format description (FORMAT.TXT)] will be used. In addition, images and raw measurement data files as provided by the measurement instruments will be stored on a project-internal data storage platform. Data files and images will be included in the open datasets. Proprietary raw data delivered by the measurement instruments will not be published. For all published files, a document record and change track will be included (author contact information, status, version, change reason and date, description of contents, title, origin of the data including a brief description of the measurement and/or experiment setup) in a separate metadata file for each characterization action called [METADATA.ODS.].gu

Ex. 3 Whenever possible, we will use file formats suitable for long-term preservation and re-use of research data. From electrochemical measurements I expect to obtain data as [*.xlsx and *.cvs format; from SEM-EDS and EBSD I will obtain images as *.jpg].

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

1.D a) Description

If the data are generated within the project, state the source of the data.

If the data are collected, state the source from which they were extracted.

If the data are re-used, state the source from which they were extracted.

1.D b) Example

Ex. 1 The existing data will originate from several sources, which will include: partner's pre-existing data, data from the scientific literature, real-world measurement data and data from simulation experiments. The data collected 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 the ESRs and their 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 data have been digitised in diverse earlier projects.

Ex. 4 The data will be collected/generated via [surveys/ interviews/ workshop] by [name] for the purpose of [data analysis for my PhD research]. Data analysis will be done by [name]. The steps taken for data collection, analysis and visualization will be documented in [Word, Excel, Miro, Zoom].

Ex. 5 The data will be collected/generated via experiments by [name] for the purpose of determining the leaching mechanism of REEs in NdFeB magnet scrap for a selective recovery of REEs by means of electrochemical leaching in organic acids. Data analysis will be done by [researcher]. The steps taken for data collection, analysis and visualization will be documented in [Word, Excel, PowerPoint, Notepad]. Additionally, all data files will be named using the following elements in the file name:

- Date or date range of experiment: [YYYYMMDD]
- Descriptive file name
- Initials of the person who last modified the file
- Version number of file

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

1.E a) Description

State the approximate volume of the datasets. Consider the implications of data volumes in terms of storage, backup, cost, and access. Estimate the volume of data in MB/GB/TB and how this will grow to make sure any additional storage and technical support required can be provided.

1.E b) Example

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

Ex. 2 The size of the data is today not known. Initial experience with storing results from different kind of measurements will permit revising this initial data management plan. The main relevant data sizes will stem from images such as microscopic sample characteristic that are stored in high-resolution bitmap format. However, 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, such as [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.

Ex. 4 The expected size of the data will be [less than 250 GB].

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) 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 institution:

The datasets will be shared within the institution 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 institution:

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 datasets 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.

Ex. 4 The data underlying the figures and conclusions in academic papers could be suitable for reuse for researchers in our field of [Adaptive reuse].

2. FAIR data

2.1 Making data findable, including provisions for 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) Example

Ex. CORA.RDR All data will be made openly available through CORA.RDR, a trusted institutional data repository. Every dataset will be assigned a Digital Object Identifier (DOI), to make them citable and persistently available. In the case of tabular data, a Universal Numerical Fingerprint (UNF) is also generated. The repository provides a unique URL to access the document with the format <https://dataverse.csuc.cat/dataset.xhtml?persistentId=doi:10.34810/dataXXX>.

Ex. 1 The repository assigns [Handle/DOI] 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) Example

Ex. CORA.RDR All datasets will be accompanied by rich metadata to ensure that they are findable. The metadata standard used to describe the dataset in CORA.RDR is the DDI's metadata schema compatible with the Dublin Core, flexible and commonly used standard that is also adopted by the European OpenAIRE repository.

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 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 [repository name] 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 optimize the possibility for discovery and then potential re-use?

2.1.C a) Description

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

2.1.C b) Example

Ex. CORA.RDR All datasets deposited in CORA.RDR will provide search keywords together with their metadata. Keywords for open data can be selected from controlled vocabularies that are suitable for the specific type of data.

Ex. 1 Data must 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.

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.

2.1.D b) Example

Ex. CORA.RDR Datasets published in CORA.RDR will be harvested and indexed using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). They will be also indexed in EOSC, OpenAIRE, RECOLECTA, Google Dataset Search, B2FIND and Mendeley Data.

Ex. 1 All data (underlying the published papers) will be made openly available through [repository name] which uses [schema.org] metadata, meaning that all datasets are indexed in [Google Dataset Search].

2.2 Making data accessible

2.2.A Will the data be deposited in a trusted repository?

2.2.A a) Description

Specify in which repository the data will be deposited and whether the repository is trustworthy. A trustworthy repository is one that meets the established requirements for the reliable, secure, and long-term preservation and access of data. The data should be deposited in a repository that provides these guarantees.

2.2.A b) 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, except for market or customer survey data, which are commercially sensitive and cannot be shared.

2.2.B Have you explored appropriate arrangements with the identified repository where your data will be deposited?

2.2.B a) Description

Verify if there are specific agreements with the repository where the data will be deposited. This may include retention policies, access conditions, and other terms that may affect the availability and use of the data.

2.2.B b) Example

Ex. CORA.RDR This Data Management Plan has been prepared with the support of our institution research support staff who works closely with the CORA.RDR managers (CSUC) and has informed us about the policies and procedures of the CORA.RDR repository.

Ex. 1 Different storage solutions were evaluated, and others are still under evaluation. The current solution was chosen as the best solution in terms of cost/benefits.

2.2.C Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

2.2.C a) Description

Explain whether the repository assigns persistent identifiers permanently (such as DOI, handle, etc.) and ensures their durability to facilitate referencing and access to the data.

2.2.C b) Example

Ex. CORA.RDR All data will be made openly available through CORA.RDR, which is a trusted institutional data repository and assigns a Digital Object Identifier (DOI) to datasets and code to make them citable and persistently available.

Ex 1. [Repository name] applies digital object identifiers to datasets.

2.2.D Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions.

2.2.D a) Description

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. Specify whether all data will be accessible through open access or if there will be access restrictions. If there are restrictions, differentiate between legal reasons, such as copyright or patent protection, and intentional restrictions based on legitimate interests or other contractual limitations.

2.2.D b) Example

Ex. CORA.RDR All of the data will be made openly available in CORA.RDR. In case there is data that must have a control of access, these files will be restricted under the proper use conditions.

Ex. CORA.RDR. All of the data associated with scientific publications will be made openly available as default unless there is a specific reason not to publish the data. 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.

All of the data from the project will be made available, except for market or customer survey data, which are commercially sensitive and cannot be shared.

Ex. CORA.RDR. All data will be deposited in CORA.RDR, but restricted; therefore, users won't be able to access them. To gain access to these data, it will be necessary to request it through the repository.

Ex. CORA.RDR. Personal and sensitive data won't be deposited in the repository.

Ex. 1 All raw data (underlying the published papers) will be retained for [name] on [institution name] servers for the purposes of validation, with consent from the participants. All anonymised or aggregated data, and/or all other non-personal data (underlying the published papers) will be uploaded to [repository name] with public access, with consent from the participants. All pseudonymised data (underlying the published papers) will be uploaded to [repository name] with restricted access, with consent from the participants.

Ex. 2 Not all data will be made available. Some data will be business confidential, and thus cannot always be shared outwardly. Studies will be altered to use certain variations of a material so that publications can be made, and data can be made as available as possible.

2.2.E If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

2.2.D a) Description

Detail whether an embargo period (access restriction time) will be required for the data and explain the reasons behind this embargo, as well as the expected duration. This may be necessary to allow time for publication or to protect intellectual property for the shortest possible period.

2.2.D b) Example

Ex. CORA.RDR Data won't need an embargo period. All data (underlying the published papers) will be made openly available through CORA.RDR.

Ex. CORA.RDR It is possible that data may need an embargo period, depending on the publisher conditions. CORA.RDR allows to define an embargo period for a maximum of two years.

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. There is available a document on recommendations to select research data repositories from CSUC (in Catalan).

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) Example

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

2.2.G If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

2.2.G a) Description

In the event that public access to the data is restricted for any justified reason, specify whether the data would be accessible, during the project, to an individual partner, to all partners, or upon request. Outline the procedures for requesting access to restricted data and under what conditions access would be granted. Additionally, specify if restrictions will be lifted after a certain period. Also, indicate how access restrictions to the data will be managed once the project is concluded.

2.2.G b) Example

Ex. CORA.RDR It won't be any restricted files in the dataset. All data (underlying the published papers) will be made openly available through CORA.RDR.

Ex. CORA.RDR. Some files will be restricted, so third parties and users of CORA.RDR could ask for access through the "Contact" button in the dataset page.

1. There should not be any long-term restrictions to the data and will all be made available. It will be possible to access data independently, without any requests etc needing to be made.

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

2.2.H a) Description

Describe the procedure established by the repository to determine the identity of the person accessing the data, if applicable, especially in the presence of access restrictions. This may include identifiers, access control lists, or other methods used to ensure security and authentication.

2.2.H b) Example

Ex. CORA.RDR Identity of individuals accessing the data on CORA.RDR can only be established when restricted access is in place. Guestbooks also can be used if the depositor wants more information from the users accessing data.

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

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

2.2.I a) Description

Indicate whether a data access committee will be required to assess or approve access requests, especially if the data contains personal or sensitive information. Provide details on the functions and responsibilities of this committee. If not required, state the reason.

2.2.I b) Example

Ex.1 It is difficult to speculate as to whether a data access committee will be required. However, no personal/sensitive data will be generated and so at this stage, it seems unlikely that a committee will be required.

Ex 2. Because of the small scale of these experiments, there is no need for a data access committee.

Ex. 3 This [institution name] 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 Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

2.2.J a) Description

Confirm whether metadata will be available openly and with a license that allows unrestricted use, such as CC0 (Creative Commons Zero). Indicate whether the metadata will contain information to enable users to find, access, and reuse the data.

2.2.J b) Example

Ex. CORA.RDR In CORA.RDR the metadata will be made openly available and licenced under a public domain dedication CC0. Information about the accessibility of the dataset is also included in the metadata. Metadata will be made openly available and licensed.

2.2.K How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

2.2.K a) Description

Detail the expected duration for data availability and how it will be ensured that metadata remains accessible after the data is no longer public. Include retention policies and long-term archiving policies.

2.2.K b) Example

Ex. CORA.RDR Research data will be deposited in CORA.RDR, which will archive and preserve the data for at least 10 years. When data is no longer available, the metadata will remain accessible.

1. It is difficult to speculate at this stage, however I see no reason currently as to why data shouldn't be available indefinitely. Should the data be made unavailable, then metadata should still be available.
2. Data will remain available as long as the repository remains active.

2.2.L Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

2.2.L a) Description

Indicate whether specific software will be required to access or read the data and whether documentation or open-source code will be provided to facilitate the use of the necessary software. This ensures that the data is accessible and interpretable.

2.2.L b) Example

Ex. 1 If software is needed, documentatiton will be included in the dataset, as well as the open source code software.

Ex. 2 No additional software is needed to access or read the data.

2.3 Making data interoperable

2.3.A 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?

2.3.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, organizations, countries, etc. Adhere to the standards of formats that are, as far as possible, compatible with open programs and applications.

2.3.A b) 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 Final data will be published in CORA.RDR, the repository of Consorci de Serveis Universitaris de Catalunya (CSUC). This repository follows the Open Archives Initiative model, which allows interoperability with the OAI-PMH metadata transmission protocol (Open Archive Initiative - Protocol for Metadata Harvesting). This protocol allows visibility of the documents from different platforms and collectors: Google Scholar, BASE, CORE, etc. This data repository is OpenAIRE compliant and meets all the requirements of metadata required by the European Commission.

Ex. 3 Persistent IDs are provided for each document (DOI) and author identifiers (ORCID) are included in the metadata. The metadata standard used to describe the dataset is the DDI's metadata schema compatible with the Dublin Core, a flexible and commonly used standard that is also adopted by the european OpenAIRE repository. We use the vocabulary [to be completed] being a standard vocabulary in this field.

2.3.B In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

2.3.B a) Description

Explain which ontologies, vocabularies, or specific languages will be used. Also, detail whether this information will be openly published and if equivalences to standardized ontologies, vocabularies, or languages will be provided to encourage their reuse, improvement, optimization, or expansion.

Ontologies encapsulate scientific knowledge in a specific scientific domain. Therefore, ontologies are crucial for achieving FAIR data: data can only be reused if well-described, classified, and of high quality [<https://www.excelra.com/our-thinking/blogs/ontologies-and-the-fair-data-principles>].

2.3.B b) Example

Ex. 1 There is no standard vocabulary for this type of data. However, the most common definition of the relevant scientific community is used as much as possible.

Ex. 2 We will use common terminology used in the field to describe the data..

Ex. 3 We will generate a specific ontology named [ontology name]. We will publish this ontology and mapping in the dataset to allow its reuse.

2.3.C Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?

2.3.C a) Description

Regarding qualified references, the objective is to create as many meaningful links as possible between (meta)data resources to enrich contextual knowledge about the data. (Source: <https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/>).

2.3.C b) Example

Ex. CORA.RDR CORA.RDR has different metadata that allows to link the dataset to other research outputs, through persistent identifiers, such as DOIs or handles, and citation to related publications.

Ex. 1 To ensure interoperability the [DOIs] of research outputs are linked whenever possible.

Ex. 2 All published datasets must include qualified references to the broadest level possible.

2.4 Increase data re-use (through clarifying licences)

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) Example

Ex. CORA.RDR All documentation needed to validate data analysis and facilitate data re-use will accompany the data via a README file created in accordance with the CORA.RDR template and guidelines.

Ex. 1 Metadata records will accompany the data files in order to describe, contextualise and facilitate external users to understand and reuse the data.

Ex. 2 All datasets published in the [association] installation must have comprehensive documentation addressing the data structure, the definition of variables, and the units of measurement. Ideally, the documentation is published in peer-review journals.

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 licenses. The EC proposes the use of Creative Commons CC-BY or CC0 licences, but there are others.

2.4.B b) Example

Ex. CORA.RDR All data and accompanying documentation needed to validate the results of the research will be made available under a CC-BY license using CORA.RDR, in line with the Grant Agreement.

Ex. 1 Wherever possible the data will be shared right after production following the [Creative Commons 4.0 International License with Attribution (CC BY)]. Experimental data and test data will in some cases only become available after the end of the project or publication of the results, whatever comes first, and will be shared used the same [CC BY] license.

Ex. 2 The deliverables associated to the dataset are licensed through an [All rights reserved] license as they are working papers not intended to be re-used. Nevertheless, the database should be shared as a possible reusable dataset. For this reason, when deposited to the repository, an [Attribution-NonCommercial license (by-nc)] will be requested. The data is currently available for re-use from the project website and will also be findable and reusable through the final depositing repository and from [OpenAire], the latest by the end of the project.

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) Example

Ex. CORA.RDR The data underlying the publications will be available at CORA.RDR under a CC-BY license, allowing third parties to use the data.

Ex. 1 Data access and sharing activities will be rigorously implemented in compliance with the privacy and data collection rules and regulations, as they are applied nationally and in the EU, as well as with the Horizon Europe rules. Raw data collected through the interviews from externals may be available to the whole institution or specific partners

upon authorization of the owners. This kind of data will not be available to the public. The results of the project will become publicly available.

Ex. 2 The full dataset will be confidential and only the members of the institution 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. 3 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 datasets 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 datasets. Requests for the use of the data by externals will be approved by the project institution.

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.D b) Example

Ex. CORA.RDR The documentation and metadata of each dataset recognize the data provenance through proper citation of the source of information and entities using the formats usually accepted by the relevant scientific community.

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) 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'être* of this project. About 80% of the efforts spent in our Thematic Centres is directed at data quality assurance.

Ex. 4 For our research data collection, the quality control of the data can happen at various stages during the quality assurance process. Initial quality control is needed at the local level and early in the collection process. Additional controls will take place at a later stage of the data lifecycle. Final quality control of metadata takes place during its input into IMIS. The initial quality control of the data, during data collection, is the primary responsibility of the project data creator/owner, who must ensure that the recorded data reflect the actual facts, responses, observations and events. The quality of the data collection methods used strongly influences data quality, and documenting in detail how data are collected provides evidence of such quality. Errors can also occur during data entry. Data are digitised, transcribed, entered in a database or spreadsheet, or coded. Here, quality is ensured by standardised and consistent procedures for data entry with clear instructions.

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 b) Example

Ex. 1 (Yes) It will be a series of new materials and samples derived from this research.

Ex 2. (No) It won't be any other research output.

Ex 3. All other research outputs such as the software related to the decision model will be carefully managed and planned

3.B Specify which of the questions pertaining to FAIR data, can apply to the management of other research outputs

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 interoperables 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.

3.B b) Example

See 2. FAIR Data real examples.

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

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

4.A b) 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.

Ex 3 Personnel costs

- Data publication in data repositories other than [repository name] (where [institution name] researchers can publish free of charge for up to 1TB of data per researcher per year).
- Publication of papers about datasets or software in dedicated journals.

Ex 4 The maintenance of the [association] installation is presently estimated to be in the order of [money amount]/year.

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.

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) Example

Ex. 1 These costs will be covered through the funds from the Horizon Europe grant.

Ex. 2 <AcronymProject> is managed and supported by a team of experts and is free of charge.

Ex. 3 Storage, archiving, re-use, and security costs are partly covered by [Partner1] and [Partner2] institutional funding. When additional resources are necessary, they are sought through project funding.

Ex. 4 (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) Example

Ex. 1 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.

- 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.
- WP1 leader, will ensure dataset integrity and compatibility for its use during the project lifetime by different partners.
- Validation and registration of datasets and metadata is the responsibility of the partner that generates the data in the WP.
- Backing up data for sharing through open access repositories is the responsibility of the partner possessing the data.
- Quality control of these data is the responsibility of the relevant WP leader, supported by the Project Coordinator.

Ex. 2 The project coordinator has the ultimate responsibility for the data management in the project and so, for the Marketplace platform management.

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) Example

Ex. CORA.RDR CORA.RDR will take care of data archiving and preservation for at least 10 years.

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.

Ex. 2 The data will be retained indefinitely. Long-term preservation is ensured by storing data on the [institution name] IT infrastructure.

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 unauthorized access).

5.A b) 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 Directive 95/46/EC ([General Data Protection Regulation](#)) and other current national legislation and institutional regulations. Research team members will have an appropriate access level according to their role in the project.

Ex. 2 During the course of the research project, all data will be stored on local servers maintained and automatically backed up by [institution name]. Every night the data will be automatically backed up. The data will be replicated over multiple sites/data centers. Data can be recovered with the help of [institution name] services in the event of an incident. Only team members have access to the designated server, limited to the principal investigator of the project. The storage security is ensured by [institution name] services. The [position name] will provide additional advice, as needed, on data storage during the research project. This data storage solution offers secure storage and transfer.

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

5.A 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.A b) Example

Ex. CORA.RDR 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 in the CSUC Centre, primarily in Barcelona, with replicas in 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.

It is important to remark here any point that was mentioned in Article 34 of the grant Agreement “[Article 34 — Ethics and research integrity](#)”.

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) 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 to for analysis, e.g. district instead of zip-code; year of birth or age instead of birth date.

6.B Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

6.B a) Description

If the dataset or datasets to be published collect personal data, prior consent of the individuals involved will be required for that purpose, whether implicit or explicit. In the case of special categories of data, consent must always be explicit, not implicit, following the instructions of the data protection officer of the institution and the ethics committee if applicable.

6.A b) Example

Ex CORA.RDR No personal data is collected or distributed by CORA.RDR, so informed consent is not needed.

Ex. 1 Research uses questionnaires dealing with personal data, so informed consent will be included.

Ex. 2 No questionnaires dealing with personal data will be created.

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) Example

Ex. 1 As part of [institution name]’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 [institution name]’s Research Data Management Policy. The [institution name] 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 [name policies] on open data management, Project Partners must also adhere to their own institutional policies and procedures for data management:

[institution name]:

Recommended file storage options

Encrypt sensitive information.

[institution name]:

Regolamento per la gestione dell’innovazione e della proprietà intellettuale e industriale.
Rectoral Decree n.1597, 19/10/2015

Codice di comportamento dell’[institution name]. Rectoral Decree n. 2653, 23/12/2014

Information Security

Research Code of Practice

IT policies for the company are set out in written policies which are subject to periodic review.

[institution name] has its own set of internal policies and procedures on data management.

Ex. 3 We are going to be in compliance with the [institution name] Policy stating that research data, code and any other materials needed to reproduce research findings are appropriately documented and shared in a research data repository in accordance with the FAIR principles (Findable, Accessible, Interoperable and Reusable) for at least [number of years] from the end of the research project, unless there are valid reasons not to do so. The project will be conducted in line with the [policy name] of the [country name] which contains a framework for good research practice.