Consorci de Serveis Universitaris de Catalunya

Data Management Plans

(Version 3, June 2019)

(Doc.CO19/04) (4 RDM\Plans Gestio Dades\PlansdeGestiodeDades_v3Publica_juny19-EN.docx, 17.06.19)

This document is intended to support researchers in creating their Data Management Plans (DMP). It is specifically aimed at projects financed under the EU's Horizon 2020 programme to create a FAIR data management plan (version 3.0, June 2019).

Key to the numbering:

- A number indicates the fields that are required in Horizon 2020.
- A capital letter indicates the elements that should be taken into account 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 Universitat Jaume I.

Examples[1] of Data Management Plans are available online.

Digital version: http://hdl.handle.net/2072/357988

[1] AMECRYS (Grant 712965), ASSEMBLE (727610), ATHOR (764987), DIALLS (770045), EASITRAIN (764879), EMPIR (14IND99), HBM4EU (733032), ICEDIG (777483), NANOCOMMONS (731032), RINGO (730944)

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 State the purpose of the data collection/generation 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 and the relation of the data collected or generated to the objectives of the project.

1.A b) Real 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 consortium 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 (objective 5).

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 Consortium 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 data sets, 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.

1.B Specify the types and formats of data will generate/collect

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) Real 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, pttx
• Tables: odsu, opj, xlsx
• Text: docx, pdf, txt
• Other: nb, cpp

Ex. 2 The openly accessible data will be the comprehensive result data sets 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 data sets. 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.

1.C Specify if existing data is being re-used and how

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) Real 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 <AcronymProject> 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>.

1.D Specify the origin of the data

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.

1.D b) Real 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.

1.E State the expected size of the data

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) Real 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 data set 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.

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

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 for metadata

2.1.A Outline the discoverability and identifiability of data produced/used

2.1.A a) Description

Discoverability of data (metadata provision): The metadata should document how the data were generated, under what license and how they can be re-used. Also, metadata help to discover the data and provide the context for proper interpretation by other researchers. See also 2.1.E

Identifiability of data and refer to the standard identification mechanism: 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 institutional repository provides a unique URL to access the document with the format https://repository/record/1234.

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

2.1.B Outline naming conventions do you follow

2.1.B a) Description

Describe how the data will be organized: the structure and name of the files.

2.1.B b) Real example

Ex. 1 The Project dataset identification follows the naming: Data_<WPno>_<SerialNumberDataset>_<DatasetTitle>. Example: Data_WP2_1_01.

Ex. 2 Files will be structured in terms of project and lead partner and publication id and figure and filenames.

2.1.C Outline the search keywords are provided that optimize possibilities for re-use

2.1.C a) Description

State how content search keywords will be created to optimize 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 Outline the approach for clear version numbers

2.1.D a) Description

Describe how version control will be organized.

2.1.D b) Real example

Ex. 1 Version control mechanisms should be established and documented before any data are collected or generated.

Ex. 2 Open source software will follow the semantic versioning schema (<ExampleURL>). The same can also be applied to datasets. Additionally, all open data, publications and open source software deposited in the Zenodo repository will use DOI versioning. DOI versioning allows for updating a dataset after it has been published and to cite either a specific version of a dataset or all versions of a dataset.

2.1.E Specify what metadata will be created

2.1.E a) Description

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.E 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 in order 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 taken into account.

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.

2.2 Making data openly accessible

2.2.A Specify 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 Specify how the data will 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 Specify 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 Web browser and/or application programming interfaces (API) offered by these storage systems, complemented by customized tools developed by users in specific domains. Zenodo provides basic robust, fast services. Anything on top of it is envisioned to be layered, and not necessarily part of the Zenodo infrastructure. For example, viewing and searching multiple images has to be handled outside Zenodo, e.g., by using <ExampleURL> that is currently being developed by <Partner1> for the domain-specific Biodiversity Literature Repository.

Ex. 2 The data will be accessible using the following software: MS Office, Matlab, Mathematica, Origin, Open Office, Adobe Reader, Image Viewer.

2.2.D Specify the documentation on 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 Specify if it is possible to include the relevant software

2.2.D a) Description

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

2.2.D 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 Specify where 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) 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 Specify whether you have explored appropriate arrangements with the 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 Specify 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

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 Determine the identity of the person accessing the data

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.A Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability

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) 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.

2.3.B Specify whether you will be using standards vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability

2.3.B a) Description

Specify whether you will be using standards vocabularies for all types of data. If this is not your case, indicate their correspondence with the most common classifications of your speciality.

2.3.B b) Real example

Ex. 1 Other types of data have been registered following internal codifications, clearly specified within the file.

2.4 Increase data re-use (through clarifying licences)

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

2.4.A 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.A b) Real example

Ex. 1 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 (the institutional one or Zenodo) and from OpenAire, the latest by the end of the project.

Ex. 2 Wherever possible the data will be shared right after production following the Creative Commons 4.0 International License with Attribution (CC4BY). Experimental data 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 CC4BY license.

2.4.B Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

2.4.B a) Description

Describe when will you make the data available for re-use. You can create a procedure to temporarily make the data accessible to other group members, project partners, and then to 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.

2.4.B b) Real example

Ex. 1 The data will remain re-usable after the end of the project by anyone interested in it, with no access or time restrictions.

Ex. 2 Data collected under this project will be made available for reuse upon completion of the experiment. As pointed out above, for reasons of competitive advantages a data embargo may apply, including the completion of a PhD thesis, which case an embargo of three years will be upheld.

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

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 IPRs and Privacy Issues. 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 H2020 rules. Raw data collected through the interviews from externals the consortium sources may be available to the whole consortium 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 based on the IPRs, as described in the Consortium Agreement.

Ex. 2 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. 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 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 Describe data quality assurance processes

2.4.D 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.D 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.

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.

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

2.4.E a) Description

How long do you expect your data to remain re-usable?

2.4.E b) Real example

Ex. 1 For re-usability the data will be stored on the webpage or on a repository system when implemented for at least ten years.

Ex. 2 The data will remain reusable until Zenodo discontinues the dataset(s) (i.e. warrantied for a minimum of 20 years).

Ex. 3 The infrastructure has a time horizon of at least <Number> years, the data will remain useful and usable beyond that period. For example, now the time-series generated since <Year> of CO2 concentrations at <Place> are still being used.

3. Allocation of resources

3.A Estimate the costs for making your data FAIR in your project

3.A a) Description

State the approximate cost for making your data FAIR and how you plan to cover them.

3.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.

3.B Specify how will be covered

3.B a) Description

State how you plan to cover the cost for making your data FAIR, including additional costs of archiving and preservation.

3.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.

3.C Specify who will be responsible for data management in your project

3.C a) Description

Explain the responsibilities for data management in your project.

3.C b) Real example

Ex. 1 The project coordinator has the ultimate responsibility for the data management in the project and so, for the Marketplace platform management.

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.

- 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.

3.D Describe the resources for long-term preservation

3.D a) Description

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

3.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.

4. Data security

4.A Address data recovery as well as secure storage and transfer of sensitive data

4.A a) Description

State what provisions are in place for data security. Includes storage & backup, long term preservation and how to keep research data safe and secure.

4.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>. The <StorageService> Agreement includes Terms and Conditions that are compliant with EU Data Protection Law and the National Bioethics Committee rules and regulations. <StorageService> provide EU Model clauses in agreements, hold <ISONumber> and <ISONumber> certifications and operate their data centres within the European Economic Area. Only researchers working for the Project will have an access to the data, using their username and passwords to access the files. The research group will also keep a backup copy on a hard-disk that will be stored in the School Building in <Room>. Only Principal Investigator nominated DPO for the project will have access to the hard-disk. Safety is ensured for the building and the room with 24 hour.

Ex. 2 The data confidentiality and integrity are implemented at various levels:

- Data at rest-stored at the JRC Data Repository- is protected against unauthorised access by means of standard EU Login (former ECAS authentication). Appropriate access levels will be granted by the creation of groups
- Data in transit is secured by means of secure data transfer mechanisms, such as TLS 1.2.2 (Transport Layer Security)
- Data access is logged by a tamper-proof logging mechanism built into <NameSoftware>, the log files are stored within an encrypted file system, and configured in append-only
- Consortium partners will impose a strict policy on all employees, co-workers, subcontractors … having access to the data. This policy will include, but is not limited to,
    - allowing copies on local devices only during processing of the data with guaranteed
    - erasure after being processed
    - extending the access control policies to the local copies
    - contractual clauses
    - agreement to terms and conditions before access is granted
- Data will be pseudonymised up to the level as to not interfere with the quality of the research.
- Lastly, awareness on data privacy and security will be enhanced (a.o. by attending a webinar on this matter prior to be granted access to the repository; attending this webinar shall be mandatory at least yearly during the course of the project.).

5. Ethical aspects

5.A Ethical or legal issues that may affect the collection and sharing of data

5.A a) Description

Description of ethical issues of collecting, archiving, processing, sharing and archiving data. Regarding personal data from questionnaires requires informed consent in order to be able to share and preserve long term. 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, privacy and data protection issues, or research on animals and non-human primates, you must comply with ethical principles and relevant national, EU and international legislation.

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

6. Other issues

6.A Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

6.A a) Description

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

6.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:

Imperial College London:

- Recommended file storage options
- Encrypt sensitive information

UCAL

- Regolamento per la gestione dell'innovazione e della proprietà intellettuale e industriale. Rectoral Decree n.1597, 19/10/2015
- Codice di comportamento dell'Università della Calabria. Rectoral Decree n. 2653, 23/12/2014

University of Strathclyde Glasgow

- Information Security
- Research Code of Practice
- Research Data Policy

CPI

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

FDB

- FDB has its own set of internal policies and procedures on data management.

7. Further support in developing your DMP

7.A Explain what resources have supported you to develop your DMP

7.A a) Description

List procedures, policies or references you have used in developing your DMP.

7.A b) Real example

Ex. 1 This DMP has been created with the tool "EinaDMP" (https://dmp.csuc.cat).

Ex. 2 Each of the partners will follow their national and institutional procedures for data management, in addition to this <AcronymProject> DMP.

Ex. 3 References:

- Data Management in the context of Horizon 2020: http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm
- Guidelines on Data Management & FAIR data principles under H2020 OA policy: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi -oa-data-mgt_en.pdf
- The Open Research Data (ORD) Pilot in H2020: https://www.openaire.eu/what-is-the-open-research-data-pilot
- OpenAIRE Guidelines for Literature Repositories, Data Archives, and CRIS Managers based on CERIF-XML: https://guidelines.openaire.eu/en/latest.