Planes de Gestión de Datos
(Versión 1, Enero 2020)

(Doc.CO20/02) (4 RDM\Plans Gestio Dades\PlanesdeGestiondeDatos_v1_enero20-ES.docx, 03.02.20)

Este documento pretende servir de apoyo a los investigadores en el momento de crear sus Planes de Gestión de Datos (*Data Management Plan, DMP*); concretamente, para los proyectos financiados en el marco del Plan Estatal de Investigación Científica y Técnica y de Innovación.

A continuación, se muestra:

- Con un número, los diferentes apartados de la plantilla.
- Con una letra mayúscula, los elementos que se deberían tener en cuenta en el momento de rellenar el apartado.
- Con una letra minúscula, las descripciones y una muestra de ejemplos reales.

Este documento ha sido elaborado por el Grupo de Trabajo de Apoyo a la Investigación del CSUC que está formado por representantes de las siguientes universidades: Universitat de Barcelona, Universitat Autònoma de Barcelona, Universitat Politècnica de Catalunya, Universitat Pompeu Fabra, Universitat de Girona, Universitat de Lleida, Universitat Rovira i Virgili, Universitat Oberta de Catalunya, Universitat de Vic-Universitat Central de Catalunya, Universitat Ramon Llull, Universitat de les Illes Balears y Universitat Jaume I.

Los ejemplos son una muestra[1] de diferentes DMPs disponibles en la red.

Este documento está sujeto a la licencia de Reconocimiento de Creative Commons (http://creativecommons.org/licenses/by/4.0/).

Versión digital: http://hdl.handle.net/2072/370510

---

[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)

Información preliminar

El entregable también suele incluir otra información preliminar como: el logotipo del proyecto, el nivel de diseminación, las revisiones históricas, la tabla de contenido y una lista de acrónimos utilizados.

1. Resumen de los datos

1.A Describir los objetivos de los datos generados y/o recolectados y su relación con los objetivos del proyecto

1.A a) Descripción

Explicar de manera breve el objetivo de los datos generados y/o recolectados y su relación con los objetivos del proyecto.

1.A b) Ejemplo real

Ej. 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).

Ej. 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 Especificar el tipo y el formato de los datos que se generarán/recolectarán

1.B a) Descripción

Describir el contenido y alcance de los datos. Los datos de investigación se generan por diferentes motivos a través de diferentes procesos y pueden ser de los siguientes tipos:

- Observacionales: datos capturados en tiempo real (neuroimágenes, datos de muestras, datos de sensores, datos de encuestas, etc.).
- Experimentales: datos capturados en equipos de laboratorio (secuencias de genes, cromatogramas, datos de campos magnéticos, etc.).
- Simulación: datos generados a partir de modelos de prueba (climatológicas, matemáticas o modelos económicos, etc.).
- Derivadas o compiladas: datos reproducibles pero de difícil reproducción (texto y minería de datos, modelos 3D, base de datos compilada, etc.).
- De referencia: conglomerado o conjunto de datos (bancos de datos de secuencias de genes, estructuras químicas, portales de datos espaciales, etc.).
- Otros.

El formato de los datos (texto, numérico, imagen, etc.) también debería indicarse.

1.B b) Ejemplo real

Ej. 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

Ej. 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 Especificar si se reutilizan los datos existentes y como

1.C a) Descripción

Si se reutiliza un conjunto de datos (dataset), especificar la fuente de donde se ha extraído, por ejemplo, de un repositorio relevante. Si se compran o reutilizan fuentes de datos existentes, explica cómo se han tratado temas como los derechos de autor y la ley de propiedad intelectual.

Cuando se crean nuevas fuentes de datos, explicar por qué no se reutilizan las fuentes ya existentes.

1.C b) Ejemplo real

Ej. 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.

Ej. 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.

Ej. 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 Especificar el origen de los datos

1.D a) Descripción

Indicar el origen de los datos, si éstos se generan dentro del proyecto.

Indicar la fuente de donde se han obtenido, si éstos son recolectados.

1.D b) Ejemplo real

Ej. 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.

Ej. 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>....

Ej. 3 These data have been digitised in diverse earlier projects.

1.E Indicar el tamaño esperado de los datos

1.E a) Descripción

Indicar, de manera aproximada, el volumen del conjunto de datos (datasets). Hay que considerar el volumen de los datos en términos de almacenamiento, copias de seguridad, costes y acceso. Estimar el volumen de los datos en MB / GB / TB y el crecimiento gradual para asegurar que el almacenamiento y el soporte técnico adicional puedan ser proporcionados.

1.E b) Ejemplo real

Ej. 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.

Ej. 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.

Ej. 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 Describir la utilidad de los datos: para quienes pueden ser útiles

1.F a) Descripción

Indicar para qué colectivo pueden ser de interés los datos.

1.F b) Ejemplo real

Ej. 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.

Ej. 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.

Ej. 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. Datos FAIR

2.1 Datos encontrables (incluyendo los metadatos)

2.1.A Describir la descubierta y la identificación de los datos producidos/usados

2.1.A a) Descripción

<u>Descubierta de los datos (incluyendo los metadatos)</u>: Los metadatos deben documentar como se generaron los datos, bajo qué licencia y como se reutilizarán. Además, los metadatos ayudan a descubrir nuevos datos y proporcionan el contexto para otras interpretaciones adecuadas. Ver también el 2.1E.

<u>Identificar los datos y consultar el mecanismo de identificación estándar:</u> Explicar cómo los datos y los metadatos asignan un identificador globalmente eterno y permanente (DOI, Handle…).

2.1.A b) Ejemplo real

Ej. 1 The institutional repository provides a unique URL to access the document with the format https://repository/record/1234.

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

2.1.B Describir cómo se nombrarán y se estructurarán los archivos

2.1.B a) Descripción

Describir la organización de los datos: estructura y nombre de los ficheros.

2.1.B b) Ejemplo real

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

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

2.1.C Proporcionar palabras claves que optimicen las posibilidades de reutilización

2.1.C a) Descripción

Indicar como se crearán las palabras clave para optimizar la recuperación y la reutilización de los datos.

2.1.C b) Ejemplo real

Ej. 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.

Ej. 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 Describir el planteamiento a seguir para un buen control de versiones

2.1.D a) Descripción

Describir cómo se organizará el control de versiones de los ficheros.

2.1.D b) Ejemplo real

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

Ej. 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 Especificar que metadatos se crearan

2.1.E a) Descripción

Describir los estándares de metadatos y como se utilizarán. Se recomienda la utilización de estándares específicos para cada disciplina. En el caso que no se utilice ningún estándar, describir los metadatos que se generaran (manual o automáticamente) y cómo.

2.1.E b) Ejemplo real

Ej. 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.

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

Ej. 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".

Ej. 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.

Ej. 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.

Ej. 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 Datos accesibles

2.2.A Especificar qué datos producidos o utilizados en el proyecto estarán disponibles de manera abierta por defecto

2.2.A a) Descripción

Describir cómo se compartirán los datos, incluyendo el procedimiento para acceder a los mismos, los periodos de embargo (si los hay) y definir si el acceso será totalmente abierto o restringido a un grupo específico. Si algunos datos no pueden ser abiertos totalmente debe justificarse el porqué.

2.2.A b) Ejemplo real

Ej. 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.

Ej. 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 Especificar cómo se accederá a los datos (ej. depositando en un repositorio)

2.2.B a) Descripción

Describir cómo se compartirán los datos, es decir, quién tendrá acceso al conjunto de datos. Se puede crear un procedimiento para que temporalmente los datos se hagan accesibles al resto de miembros del grupo, los socios del proyecto, y al público en general. Hay que indicar si los datos se pondrán en acceso abierto y en qué plazo razonable. Una posibilidad es ofrecer juntamente con las publicaciones. Si hay que establecer un periodo de embargo, es aquí donde hay que especificarlo.

2.2.B b) Ejemplo real

Ej. 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.

Ej. 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 Especificar qué métodos o softwares son necesarios para acceder a los datos

2.2.C a) Descripción

Incluir cualquier requerimiento técnico para acceder a los datos y reutilizarlos.

2.2.C b) Ejemplo real

Ej. 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.

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

2.2.D Especificar en la documentación qué software es necesario para acceder a los datos

2.2.D a) Descripción

Incluir la documentación del software que se necesita para acceder a los datos.

2.2.D b) Ejemplo real

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

Ej. 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 Especificar, si es posible, el software relevante

2.2.D a) Descripción

En caso de que sea un software específico, incluirlo si es posible. Por ejemplo, el código.

2.2.D b) Ejemplo real

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

Ej. 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 Especificar donde se depositan los datos, metadatos, documentación y el código

2.2.F a) Descripción

Indicar el repositorio en el que los datos, los metadatos, la documentación y el código se almacenarán. Puede ser en el mismo repositorio o en diferentes según el tipo de contenido. Por ejemplo, el código debería depositarse en un repositorio específico. Se puede consultar el documento de recomendaciones para seleccionar un repositorio de datos del CSUC.

Es importante usar repositorios que proporcionen a los datos enlaces permanentes (DOI, Handle) para facilitar el descubrimiento y la citación.

2.2.F b) Ejemplo real

Ej. 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.

Ej. 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 Especificar si se han examinado los requisitos del repositorio

2.2.G a) Descripción

Indicar si se han explorado adecuadamente cuáles son los requisitos del repositorio identificado.

2.2.G b) Ejemplo real

Ej. 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.

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

2.2.H Especificar cómo se da acceso si hay restricciones de uso

2.2.H a) Descripción

En caso de que el acceso a los datos sea restringido a todo el público por un motivo justificado, especificar qué datos serán accesibles a socios individuales, a todos los socios o bajo solicitud. Especificar los procedimientos para solicitar el acceso a los datos restringidos y en qué condiciones se concederán. Además, especificar también si habrá más restricciones en un período de tiempo.

2.2.H b) Ejemplo real

Ej. 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.

Ej. 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 Especificar si es necesario un comité de acceso a los datos

2.2.I a) Descripción

Especificar por qué es o no necesario un comité de acceso a los datos.

2.2.I b) Ejemplo real

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

Ej. 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 Describir correctamente las condiciones de acceso

2.2.J a) Descripción

Describir cuáles son las condiciones de acceso definidas por el repositorio de vuestra elección (indicando el URL de donde proviene la información). Por ejemplo: con una licencia legible por máquina.

2.2.J b) Ejemplo real

Ej. 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.

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

2.2.K Determinar la identidad de las personas que acceden a los datos

2.2.K a) Descripción

Describir el procedimiento establecido por el repositorio para determinar la identidad de la persona que accede a los datos.

2.2.K b) Ejemplo real

Ej. 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.

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

2.3 Datos interoperables

2.3.A Evaluar la interoperabilidad de los datos. Especificar qué datos y qué vocabularios de metadatos, estándares o metodologías se seguirán para facilitar la interoperabilidad

2.3.A a) Descripción

Explicar qué datos y qué vocabularios de metadatos, estándares o metodologías se siguen para facilitar la interoperabilidad.

2.3.A b) Ejemplo real

Ej. 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 Especificar si se usarán vocabularios estandarizados para todos los tipos de datos del conjunto de datos (dataset), para facilitar la interoperabilidad entre disciplinas

2.3.B a) Descripción

Especificar si se usan vocabularios estandarizados para los distintos tipos de datos. Si no es el caso, indicar la correspondencia entre las clasificaciones más usadas en tu especialidad.

2.3.B b) Ejemplo real

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

2.4 Datos reutilizables

2.4.A Especificar cómo se asignan las licencias para permitir la máxima la reutilización de los datos

2.4.A a) Descripción

Si los datos se ponen a disposición de otros investigadores y del público en general, especificar el grado de reutilización permitido. Se establecerá el grado de reutilización a través de licencias. La CE propone el uso de licencias Creative Commons CC-BY o CC0, pero existen otras.

2.4.A b) Ejemplo real

Ej. 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.

Ej. 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 Especificar cuando los datos estarán disponibles para su reutilización. Si procede, especificar porqué y durante qué período será necesario embargar los datos

2.4.B a) Descripción

Describir cuando los datos estarán disponibles para su reutilización. Se podrá elaborar un procedimiento para que los datos puedan ser accesibles a otros miembros del grupo, a los socios de proyecto y, posteriormente al público en general. Se debe indicar si los datos serán de acceso abierto y cuando. Una posibilidad es hacerlos públicos junto con las publicaciones. Si se requieren periodos de embargo, se debe indicar en este apartado.

2.4.B b) Ejemplo real

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

Ej. 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 Especificar si los datos producidos y/o utilizados en el proyecto podrán ser utilizados por terceros, en particular, una vez finalizado el proyecto. Si se restringe la reutilización de algunos datos se debe concretar los motivos

2.4.C a) Descripción

En principio, los datos deben estar a disposición de otros investigadores y del público en general con las menos restricciones posibles. Sin embargo, existen motivos para no compartir los datos: razones éticas, protección de datos personales, implicación de derechos de propiedad intelectual y/o industrial, intereses comerciales, etc. Si procede, se debe especificar los motivos por los cuales no se compartirán algunos conjuntos de datos.

2.4.C b) Ejemplo real

Ej. 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.

Ej. 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.

Ej. 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 Describir el proceso que garantiza la calidad de los datos

2.4.D a) Descripción

Describir el proceso que garantiza la calidad de los datos. Especificar cómo i cuando se llevarán a cabo las evaluaciones internas.

La calidad de los datos se puede asegurar a través de diferentes medidas. Estas incluyen la validación de la muestra, las réplicas y la comparación con resultados de estudios similares y el control de la distorsión sistemática.

2.4.D b) Ejemplo real

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

Ej. 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.

Ej. 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.

Ej. 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 Especificar los plazos en los que los datos podrán ser reutilizados

2.4.E a) Descripción

Indicar el tiempo de espera para que los datos puedan ser reutilizados.

CSUC

2.4.E b) Ejemplo real

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

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

Ej. 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 CO_2 concentrations at <Place> are still being used.

3. Asignación de recursos

3.A Estimar el coste para hacer que los datos sean FAIR en el proyecto

3.A a) Descripción

Indicar el coste aproximado de hacer que los datos sean FAIR.

3.A b) Ejemplo real

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

Ej. 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 Especificar como se prevé cubrir el coste

3.B a) Descripción

Indicar como se quiere cubrir el coste para hacer que los datos sean FAIR, incluyendo los costes adicionales de archivo y conservación.

3.B b) Ejemplo real

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

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

Ej. 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 Especificar quién es el responsable de la gestión en cada proyecto

3.C a) Descripción

Explicar las responsabilidades de la gestión de los datos en cada proyecto.

3.C b) Ejemplo real

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

Ej. 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 Describir los recursos para la preservación a largo plazo

3.D a) Descripción

Indicar como se prevé preservar a largo plazo y quién decide qué datos se conservarán y durante cuánto tiempo.

3.D b) Ejemplo real

Ej. 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. Seguridad de los datos

4.A Tratamiento de recuperación de datos, así como el almacenaje y la transferencia de datos confidenciales

4.A a) Descripción

Describir las medidas adoptadas respecto la seguridad de los datos. Incluye el almacenaje y las copias de seguridad, la preservación a largo plazo y cómo se mantendrán los datos seguros y protegidos.

4.A b) Ejemplo real

Ej. 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.

Ej. 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. Aspectos éticos

5.A Problemas éticos o legales que pueden afectar la recogida e intercambio de datos

5.A a) Descripción

Describir los posibles problemas éticos durante la recogida, el almacenaje, el procesamiento y el archivo de los datos, junto a los procedimientos de aprobación ética relacionados con el proyecto. Es importante destacar en este punto cualquier aspecto mencionado en el artículo 34 de los acuerdos de financiamiento "Article 34 — Ethics and research integrity".

Si las actividades de investigación involucran a niños, pacientes, población vulnerable, uso de células madre embrionarias, cuestiones de privacidad y protección de datos o investigación en animales y primates, se deben cumplir los principios éticos y la legislación nacional pertinente de la Unión Europea e internacional, y en este punto es donde debe tratarse.

5.A b) Ejemplo real

Ej. 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.

Ej 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. Otros temas

6.A Referenciar otros procedimientos nacionales/sectoriales/institucionales/de financiación usados en la gestión de datos de investigación

6.A a) Descripción

Explicar los procedimientos nacionales/sectoriales/institucionales/de financiaciones usados en la gestión de datos de investigación.

6.A b) Ejemplo real

Ej. 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.

Ej. 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. Otro soporte en el desarrollo del plan

7.A Explicar los recursos utilizados para el desarrollo del plan

7.A a) Descripción

Lista los procedimientos, las políticas o las referencias que se han usado en el desarrollo del plan.

7.A b) Ejemplo real

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

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

Ej. 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.