Data Management Plan

Template & Guideline for Horizon Europe Projects

Version control

Version	Date	Author/reviewer	Comments
1.0	17/01/2024	Jordi Moretón Lorena López	First version

SJD Sant Joan de Déu Institut de Recerca

This document is intended to support Institut de Recerca Sant Joan de Déu (IRSJD) researchers in creating their Data Management Plan (DMP). It is specifically aimed at projects financed under the EU's Horizon Europe programme to create a FAIR data management plan, although it can be used as a general template & guideline for any DMP.

Key to the numbering:

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

Institutional information

IRSJD institutional answers and/or key information is provided for several fields of this document. Those are provided to guide and help answer each item according to institutional stablished policies, protocols and systems, but might need modifications to better adapt each projects' specifics. Please contact **frecerca.suport-dmp@sjd.es** if you need support in the elaboration of your DMP.

This document was adapted by IRSJD from the version 1 DMP from October 2022 for Horizon Europe prepared by the Research Support Working Group of CSUC, which is composed of representatives from the following universities: University of Barcelona, Universitat Autònoma de Barcelona, Universitat Politècnica de Catalunya, Pompeu Fabra University, University of Girona, University of Lleida, Universitat Rovira i Virgili, Open University of Catalonia, University of Vic-Central University of Catalonia, Ramon Llull University and University of the Balearic Islands.

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).

Template / Table of contents

1. Data summary

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

1.A. a) Description

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

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

1.B a) Description

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

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

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

Institutional information

You can find information of formats (accepted and recommended for long term preservation) in the following link:

https://confluence.csuc.cat/display/RDM/Recomanacions+de+formats

File type	Recommended formats	Not recommended but commonly accepted formats
Text documents	PDF/A (.pdf) ODT (.odt)	Microsoft Word (.doc) Office Open XML (.docx) Rich Text File (.rtf) PDF different to a PDF/A (.pdf)
Plain text	Unicode text (.txt)	Non-Unicode text (.txt)
Markup languages	XML (.xml) HTML (.html) YAML (.yaml) JSON (.json) ReStructuredText (.rst) Related files: .css, .xslt, .js, .es	SGML (.sgml) Markdown (.md)
Programming languages	NetCDF TextFabric R (.r)	MATLAB (.mat)

	Octave (.mat)	
Spreadsheets	ODS (.ods) CSV (.csv)	Microsoft Excel (.xls) Office Open XML Workbook (.xlsx) PDF/A (.pdf)
Database	SQL (.sql) SIARD (.siard) CSV (.csv) FITS (.fits, .fit, .fts) (Apache) Parquet (.parquet)	Microsoft Access (.mdb, .accdb) dBase (.dbf) HDF5 (.hdf5, .he5, .h5)
Statistical data	SPSS (.dat/.sps) STATA (.dat/.DO) R (.rdat/.rdara)	SPSS Portable (.por) SPSS (.sav) STATA (.dta) SAS (.7dat; .sd2; .tpt)
Images (bitmap)	JPEG (.jpg, .jpeg) TIFF (.tif, .tiff) PNG (.png) JPEG 2000 (.jp2) DICOM (.dcm)	
Vector images	SVG (.svg)	Adobe Illustrator (.ai) EPS (.eps) WMF/EMF (.wmf, .emf) CDR (.cdr)
Audio (Container and Codec)	BWF (.bwf)	
Audio (container)	MXF (.mxf) Matroska (.mka)	Audio (container)
Audio (Codec)	FLAC (.flac) OPUS	WAVE (.wav) MP3 (.mp3) AAC (.aac, .m4a) AIFF (.aif, .aiff) OGG (.ogg)
Video (Container)	MXF (.mxf) Matroska (.mkv)	MPEG-4 (.mp4, .m4a, .m4v) MPEG-2 (.mpg, .mpeg, .m2v, mpg2) AVI (.avi) QuickTime (.mov, .qt)
Video (Codec)	Theora (.ogv)	
Computer Aided Design (CAD)	AutoCAD DXF version R12 (ASCII) (.dxf) SVG (.svg)	AutoCAD DXF diferent a R12 (ASCII) (.dwg, .dxf) DWG (.dwg) DGN (.dgn)
Geographic Information Systems (GIS)	GML (.gml) MIF/MID (.mif/.mid)	Esri Shapefiles (.shp & related files) MapInfo (.tab & related files) KML (.kml)

		Esri Geodatabase (.gdb) Project files/Workspaces (.mxd, .wor, .qgs)
Georeferenced images	GeoTIFF (.tif, .tiff)	TIFF World File (.tfw & .tif, possibly with additional files) JPEG World File (.jgw & .jpg, possibly with additional files) ERDAS IMAGINE File Format (.img)
Raster GIS	ASCII GRID (.asc, .txt)	Esri GRID (.grd & related files) Surfer Grid (.grd; .srf) ERDAS IMAGINE File Format (.img)
3D	WaveFront Object (.obj) Polygon file format (.ply) X3D (.x3d) COLLADA (.dae)	Autodesk FBX (.fbx) Blender (.blend) 3D PDF (.pdf)
RDF	RDF/XML (.rdf) Trig (.trig) Turtle (.ttl) NTriples (.nt) JSON-LD ATLAS.TI Copy bundle NVivo Project file	
Computer Assisted Qualitative Data Analysis (CAQDAS)	REFI-QDA (Qualitative Data Analysis)	

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

1.C a) Description

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

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

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

1.D a) Description

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

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

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

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

1.E a) Description

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

Institutional information

Make an educated guess of the order of magnitude (1MB, 10MB, 100MB, 1GB, 10GB, 100GB, 1TB).

In regards to the IRSJD FAIR data institutional repository, the CORA – Research Data Repository (RdR) (https://dataverse.csuc.cat), it allows for up 100GB dataset, with individual files up to 10GB.

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

1.F a) Description

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

1.F b) Real example

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

Ex. 2: Within the Consortium:

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

Beyond the Consortium:

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

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

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

2. FAIR data

2.1 Making data findable, including provisions for metadata

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

2.1. a) Description

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

Institutional information

IRSJD FAIR data institutional repository, the CORA – Research Data Repository (RdR) (https://dataverse.csuc.cat), assigns DOIs for persistent identification and citability of the dataset.

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

2.1.B a) Description

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

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

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

Institutional information

IRSJD FAIR data institutional repository, the CORA – Research Data Repository (RdR) (https://dataverse.csuc.cat) uses the Data Documentation Initiative (DDI) standard to describe the dataset and is compliant with the Dublin Core Metadata Terms and DataCite. More information on the metadata: https://guides.dataverse.org/en/5.5/user/appendix.html

2.1.B b) Real example

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

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

Ex. 3: (1) The data are expected to be provided in ANSI SQL, XML or text (ASCII) format. For this dataset, data citation and metadata practices derived from the community will be considered. (2) There are no standards for these logs. A possible solution is project servers such as AAA servers. In this case, the logs would include the attributes defined by "project".

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

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

Ex. 6: There are many different metadata standards for many different types of data and it may not be possible to find one that fits all purposes. Therefore, a pragmatic and feasible approach is to agree on a common and minimal catalogue metadata schema for those datasets that are published in public catalogues and data repositories and to use data-type specific schema extensions, if necessary.

In general, the Zenodo deposition metadata domain model which is based on DataCite's metadata schema minimum and recommended terms will be used for open data generated by the project and deposited in an appropriate repository.

2.1.C Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

2.1.C a) Description

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

Institutional information

IRSJD FAIR data institutional repository, the CORA – Research Data Repository (RdR) (https://dataverse.csuc.cat) provides specific metadata keywords fields to optimize possibilities of discovery and reuse. It's possible to specify if any controlled vocabulary is used.

2.1.C b) Real example

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

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

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

2.1.D a) Description

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

Institutional information

Datasets published in IRSJD FAIR data institutional repository, the CORA – Research Data Repository (RdR) (https://dataverse.csuc.cat), are harvested and indexed using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). They will be also indexed in EOSC, OpenAIRE, RECOLECTA, Google Dataset Search, B2Find and Mendeley Data.

2.2 Making data accessible

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

2.2.A a) Description

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

Institutional information

FAIR principles allow for open, restricted or even -justified- closed data, and the CORA – Research Data Repository (RdR) (https://dataverse.csuc.cat) allows for setting up restricted access and/or embargo periods for datasets. The degree of "openess" will depend on specifics of the data being published (ie. the degree of anonymization and the risks of re-identification (GDRP), confidentiality agreements, patent Applications, etc).

Contact frecerca.suport-dmp@sjd.es to answer this item.

In case of many datasets, you might create a table:

Dataset	Open/Restricted/Closed	Justification for restrictions
...	...	...

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

Ex. 3: All research data underpinning publications will be made available for verification and re-use unless there are justified reasons for keeping specific datasets confidential. The main elements when considering confidentiality of datasets are:

- Protection of intellectual property regarding new processes, products and technologies where the data could be used to derive sensitive information that would impact the competitive advantage of the consortium or its members,
- Commercial agreements as part of the procurements of components or materials that might foresee the confidentiality of data.
- Personal data that might have been collected in the project where sharing them is not allowed by the national and European legislation.

2.2.B How will the data be made accessible (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.

Institutional information
The final data can be deposited in open/restricted/closed access in the IRSJD Fair Data repository, the CORA – Research Data Repository (RdR) (https://dataverse.csuc.cat). This data will be preserved for the lifetime of the repository. If needed, the access to the datasets can be embargoed for a period of time.

2.2.B b) Real example

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

<Repository4>. These arrangements will be revisited after the data from the pilots has been created.

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

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

2.2.C a) Description

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

2.2.C b) Real example

Ex. 1: 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 Is documentation about the software needed to access the data included?

2.2.D a) Description

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

2.2.D b) Real example

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

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

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

2.2.E a) Description

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

Institutional information

The IRSJD Fair Data repository, the CORA – Research Data Repository (RdR) (https://dataverse.csuc.cat) can be used to deposit software code along with the dataset and the relevant documentation. Also, the repository can link the dataset to code deposited in Github.

2.2.E b) Real example

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

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

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

2.2.F a) Description

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

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

Institutional information

The data generated under a project and information supporting preservation and reuse can be deposited in CORA – Research Data Repository (RdR) (https://dataverse.csuc.cat/). It's a trusted multidisciplinary data repository that allows Catalan universities and research centres to publish research datasets in FAIR mode and following the EOSC guidelines. This repository is designed to collect, disseminate, and provide persistent and reliable, long-term open access to the data. The CORA – Research Data Repository (RdR) assigns DOIs for persistent identification and citability of the dataset.

2.2.F b) Real example

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

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

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

2.2.G a) Description

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

Institutional information

The IRSJD has signed an agreement with Consorci de Serveis Universitaris de Catalunya (CSUC), the provider of CORA – Research Data Repository (RdR) (https://dataverse.csuc.cat/), to use it as the institutional repository.

As previously stated, it's a trusted multidisciplinary data repository that allows Catalan universities and research centres to publish research datasets in FAIR mode and following the EOSC guidelines.

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 If there are restrictions on use, how will access be provided?

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.

Institutional information

If, according to the answer given in 2.2B, the final data is to be deposited in restricted or closed access in IRSJD Fair Data repository, the CORA – Research Data Repository (RdR) (https://dataverse.csuc.cat), or it is going to be embargoed for a period of time, explain the reasons.

This item can also be answered in accordance with 2.2J and 2.4B, where the license for the reuse is stated. If the license it's CC-BY-NC or equivalent, no reuse with commercial purposes is allowed.

2.2.H b) Real example

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

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

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

2.2.I a) Description

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

2.2.I b) Real example

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

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

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

2.2.J a) Description

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

Institutional information

The CORA – Research Data Repository (RdR) (https://dataverse.csuc.cat) assigns a CC0 license by default, but other custom licenses are allowed.

IRSJD suggests CC-BY license, or CC-BY-NC license for the data whenever the Innovation unit needs to limit the reusability because of issues with transferability. Consult with the Innovation unit about potential issues before assigning a license to the dataset.

Metadata will always have a CC0 license assigned.

2.2.J b) Real example

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

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

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

2.2.K a) Description

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

Institutional information

The CORA – Research Data Repository (RdR) (https://dataverse.csuc.cat) has an optional guestbook functionality, which can be toggled on a dataset by dataset basis. This allows us to know who accessed the dataset. Also, the dataset access can be restricted by a prior contact form.

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.

2.3. Making data interoperable

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

2.3.A a) Description

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

Institutional information

Besides explaining the data formats and standards used, explain interoperability of the dataset as follows:

The CORA – Research Data Repository (RdR) (https://dataverse.csuc.cat) follows the Open Archives Initiative model, which allows interoperability with the OAI-PMH metadata transmission protocol (Open Archive Initiative - Protocol for Metadata Harvesting). This protocol allows visibility of the documents from different platforms and collectors: Google Scholar, BASE, CORE, etc. This data repository is OpenAIRE compliant and meets all the requirements of metadata required by the European Commission.

Persistent IDs are provided for each document (DOI) and author identifiers (ORCID) are included in the metadata. The metadata standard used to describe the dataset is the DDI's metadata schema compatible with the Dublin Core, a flexible and commonly used standard that is also adopted by the european OpenAIRE repository.

2.3.A b) Real example

Ex. : 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.

2.4. Increase data re-use (through clarifying licences)

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

2.4.A a) Description

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

Institutional information

A README file is needed for every dataset uploaded to the repository. It allows to communicate important information about the dataset, clarify possible questions about the use, creation and/or updating of the data. It is essential to write a good REAME file so that all the research is presented in a compact form. At a minimum, the file must contain the following:

- Dataset title, DOI, contact information
- Methods
- Summary of data and files
- Specific data information
- Reuse conditions

Download a README template in the following link.

Also, if needed, additional supporting documentation can be provided alongside the dataset and README file.

2.4.A b) Real example

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

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

2.4.B a) Description

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

Institutional information

The CORA – Research Data Repository (RdR) (https://dataverse.csuc.cat) assigns a CC0 license by default, but other custom licenses are allowed.

IRSJD suggests CC-BY license, or CC-BY-NC license for the data whenever the Innovation unit needs to limit the reusability because of issues with transferability. Consult with the Innovation unit about potential issues before assigning a license to the dataset.

Metadata will always have a CC0 license assigned.

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

2.4.C a) Description

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

2.4.C b) Real example

Ex. 1: 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 Horizon Europe 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 How the provenance of the data will be documented using the appropriate standards?

2.4.D a) Description

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

2.4.D b) Real example

Ex. 1: The documentation and metadata of each dataset recognize the data provenance through proper citation of the source of information and entities using the formats usually accepted by the relevant scientific community.

2.4.E Are data quality assurance processes described?

2.4.E a) Description

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

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

2.4.E b) 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.

3. Other research outputs

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

3.A a) Description

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

3.A b) Real example

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

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

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

3.B a) Description

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

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

3.B b) Real example

Ex. 1: See 2. FAIR Data Real example**.**

4. Allocation of resources

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

4.A a) Description

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

Institutional information

There are no costs for your project budget on using an deposit in the institutional repository, as it is managed by the Institute's Knowledge Management Unit. This includes all the costs related to long term storage & archival, reuse, security, curation, etc.

*If you don't use the institutional repository, there might be costs in the Project Budget.
*If you have a dedicated budget for making data FAIR during the Project (i.e. a dedicated Data Steward, storage systems, etc), explain it in this section.

4.A b) Real example

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

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

4.B How will these be covered?

4.B a) Description

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

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

4.B b) Real example

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

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

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

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

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

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

4.C a) Description

Explain the responsibilities for data management in your project.

Institutional information
[If CORA RDR institutional is used] The Institute's Knowledge Management Unit manages the FAIR data repository, the CORA – Research Data Repository (RdR) (https://dataverse.csuc.cat). Explain all other actors in the management of data during the Project.

4.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.

4.D How will long term preservation be ensured?

4.D a) Description

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

Institutional information

Items will be retained for the lifetime of the institutional FAIR data repository, the CORA – Research Data Repository (RdR) (https://dataverse.csuc.cat).

4.D b) Real example

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

5. Data security

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

5.A a) Description

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

Institutional Storage Services: Esplugues de Llobregat

Data collected will be stored on the<Storage Service> which is subject to regular back-up that is controlled by the specialized personnel.

The IT department performs operations by type: mission-critical (user data, virtual machines, scientific results, etc.) and static (scientific data sets, intermediate files, etc.).

Content will be checked regularly to preserve its integrity, security, and durability.

Storage Services:

- Redcap
  - Type of server: Virtual
  - Location: Hospital Sant Joan de Déu Data Center (Pg. de Sant Joan de Déu, 2, 08950 Esplugues de Llobregat, Barcelona, Spain)
  - Backups: Daily, Weekly, Montly & Annually. Hospital Sant Joan de Déu Data Center 2 (Pg. de Sant Joan de Déu, 2, 08950 Esplugues de Llobregat, Barcelona, Spain)
  - Recovery: any deleted file can be recovered from the backup server
  - Access: nominal access for each user upon request, if it is external it must be requested by someone from the hospital.

- Filesystem ("Recursos" folders)
  - Type of server: Virtual
  - Location: Hospital Sant Joan de Déu Data Center (Pg. de Sant Joan de Déu, 2, 08950 Esplugues de Llobregat, Barcelona, Spain)
  - Backups: Daily, Weekly, Montly & Annually. Hospital Sant Joan de Déu Data Center 2 (Pg. de Sant Joan de Déu, 2, 08950 Esplugues de Llobregat, Barcelona, Spain)
  - Recovery: any deleted file can be recovered by using Dataprotector software.
  - Access: nominal access

- Microsoft Office365/Onedrive
  - Type of server: Cloud
  - Location: Azure Microsoft (Europe)
  - Backups: Daily, Weekly, Montly & Annually. Hospital Sant Joan de Déu Data Center 2 (Pg. de Sant Joan de Déu, 2, 08950 Esplugues de Llobregat, Barcelona, Spain)
  - Recovery: We have a Sinology Onpremise cabin with backups/restores everything
  - Access: nominal access with MFA (multi factor authentication)

- Genetics
  - Type of server: Physical

- Location: Hospital Sant Joan de Déu Data Center (Pg. de Sant Joan de Déu, 2, 08950 Esplugues de Llobregat, Barcelona, Spain)
- Backups: Upon request. Tape drives library at Hospital Sant Joan de Déu Data Center 1 (Pg. de Sant Joan de Déu, 2, 08950 Esplugues de Llobregat, Barcelona, Spain)
- Recovery: Upon request, restored from the tape drives.
- Access: only Bioinformatics department have access

- uPad
  - Type of server: Cloud
  - Location: Amazon Web Services (AWS)
  - Backups: AWS Backup (pending implementation)
  - Recovery: AWS Backup (pending implementation)
  - Access: a través de front-end: desde el dominio de Active Directory / a través de VPN si se accede desde fuera del dominio

These procedures are designed, set and applied in order to fully comply with personal data as ruled by Directive 95/46/EC (General Data Protection Regulation) and other current national legislation and institutional regulations. Research team members will have an appropriate access level according to their role in the project.

Institutional Storage Services: Parc Sanitari Sant Boi de Llobregat

Data collected from the research group for the Project will be stored on the <Storage Service>, which is subject to regular back-up that is controlled by the IT personnel.

The IT department performs operations by type: mission-critical (user data, virtual machines, scientific results, etc.) and static (scientific data sets, intermediate files, etc.).

Content will be checked regularly to preserve its integrity, security, and durability.

Storage Service:

- Redcap
  - Type of server: Virtual
  - Location: Parc Sanitari Sant Joan de Déu - Data Center 1 (Carrer Doctor Antoni Pujades 42, Sant Boi de Llobregat, Barcelona, Spain)
  - Backups: Daily, Weekly, Montly & Annually from Parc Sanitari Sant Joan de Déu - Data Center 2 (Camí Vell de la Colònia 25, Sant Boi de Llobregat, Barcelona, Spain)
  - Recovery: We hav a Dell Data Domain On-Premise backup disk platform with backups/restores through Dell Avamar software.
  - Access: Nominal access for each user upon request, if it is external it must be requested by someone from the hospital.
- Filesystem
  - Type of server: Physical Huawei storage disk cabin with HA
  - Location: Parc Sanitari Sant Joan de Déu - Data Center 1 (Carrer Doctor Antoni Pujades 42, Sant Boi de Llobregat, Barcelona, Spain) and Data Center 2 (Camí Vell de la Colònia 25, Sant Boi de Llobregat, Barcelona, Spain)
  - Backups: Daily, Weekly, Montly & Annually from Parc Sanitari Sant Joan de Déu - Data Center 2 (Camí Vell de la Colònia 25, Sant Boi de Llobregat, Barcelona, Spain)
  - Recovery: Any deleted file can be recovered from Dell Data Domain On-Premise backup disk platform using Veeam Backup software.
  - Access: Nominal access
- Microsoft Office365/OneDrive
  - Type of server: Cloud
  - Location: Azure Microsoft in Europe

- Backups: Daily, Weekly, Montly & Annually from Parc Sanitari Sant Joan de Déu - Data Center 1 (Carrer Doctor Antoni Pujades 42, Sant Boi de Llobregat, Barcelona, Spain)
- Recovery: We have a Sinology On-Premise storage disk cabin with backups/restores for O365 accounts including OneDrive
- Access: Nominal access with MFA (Multi Factor Authentication)

These procedures are designed, set and applied in order to fully comply with personal data as ruled by Directive 95/46/EC (General Data Protection Regulation) and other current national legislation and institutional regulations. Research team members will have an appropriate access level according to their role in the project.

5.A b) Real example

Ex. 1: Data collected from the research group for the Project will be digitised and stored on the University's <StorageService> which is subject to regular back-up that is controlled by the University's IT personnel. The IT department performs operations by type: mission-critical (user data, virtual machines, scientific results, etc.) and static (scientific data sets, intermediate files, etc.). Content will be checked regularly to preserve its integrity, security, and durability. These procedures are designed, set and applied in order to fully comply with personal data as ruled by Directive 95/46/EC (General Data Protection Regulation) and other current national legislation and institutional regulations. Research team members will have an appropriate access level according to their role in the project.

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

- Data at rest-stored at the <StorageService> 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, o 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.B Will the data be safely stored in trusted repositories for long term preservation and curation?

5.B a) Description

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

Institutional information

The following list describes the security settings of all the final data deposited at the institutional CORA RDR FAIR data repository:

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

6. Ethics

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

6.A a) Description

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

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

It is important to remark here any point that was mentioned in Article 34 of the grant Agreement "Article 34 — Ethics and research integrity".

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

Institutional information

General Data Protection Regulation (GDPR)

The IRSJD is under the obligation to preserve the confidentiality of the personal and clinical data of the study subjects based on the GDPR (EU) 2016/679 of the European Parliament and of the Council of April 27, 2016 on the protection of natural persons with regard to the processing of personal data and the free movement of such data, and repealing Directive 95/46/EC and any other applicable provision regarding data protection. Any processing will be duly analysed and reviewed by each relevant Data Protection Officer. Data generated will be unlinked to an identifiable person and the necessary precautions will ensure their security and the security of the associated data.

Spanish national law (LOPD-GDD)

The IRSJD is under the Spanish legislation on data protection, and the applicable law is the Organic Law 3/2018, of December 5, of protection of personal data (LOPD-GDD); the norm in force in the precepts is not repealed by the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 of April of 2016 relative to the protection of physical persons with regard to the

treatment of personal data and to the free circulation of said data, and repealing Directive 95/46 / EC (GPDR).

Data Protection Officer (DPO)

The IRSJD counts with a Data Protection Officer that will be involved in the project, in particular regarding the GDPR issues. The Data Protection Officer will contribute to the regular and systematic monitoring of data collection, processing, analysis and storage. <If applicable> At the other participating sites involved in data collection and processing, a Data Protection Officer will be appointed. The contacts of the Data Protection Officer at each participating site will be available to all subjects involved in the project and will be added to this document when available.

Consent form

The *Comité de Ética de la Investigación con medicamentos* (CEIm) is an affiliated Ethical Committee for Drug Research that evaluates research projects considering, among others, the following generic criteria established by the standards of good clinical practice and from current Spanish and EU legislation:

1. The relevance of the research project.
2. The suitability of the protocol, its scientific efficiency and the possibility of drawing valid conclusions.
3. The justification of the foreseeable risks and inconveniences in relation to the expected benefits.
4. The justification of the control group.
5. The study follow-up plan.
6. The selection and withdrawal criteria of the study subjects.
7. The suitability of the research teams.
8. The suitability of written information on the characteristics of the study that will be given to potential subjects of the investigation.
9. Verification of the compensation and treatment forecast that will be offered to the participating subjects in case of injury or death attributable to the study, and of the insurance or compensation to cover the responsibilities specified by the corresponding legislation (when necessary).

For study participants who are not in a legal position to consent to participate in a scientific study due to their age (minors), informed consent must be provided to parents/guardians. In addition to consent from parents/guardians, those over 12 years of age must provide informed assent to participate in clinical studies.

In the case of study participants with obstacles for expressing their own will, they will be offered the relevant support measures to grant their consent to participate in a scientific study. If this is not possible, informed consent must be provided to parents/guardians.

Methods of un-identify personal data

Explain anonimization, pseudo-anonimization, etc. techniques you are going to use in the Project in the case there is personal data.

Pseudo-anonymization

One of the ways to mitigate the legal concerns arising from the use of personal data is pseudo-anonymization. Pseudoaonymization substitutes the identity of the data subject in such a manner that the personal data can no longer be attributed to a specific data subject without the use of

additional information. Such additional information must be kept separately and is subject to technical and organisational measures to ensure that data are not traced back to an identifiable natural person. But because of the use of 'additional information' can lead to the identification of the individuals, pseudoanonymous data is still personal data.

In a proces of pseudoanonimization, patients will be indicated and tracked by a univocal code on all documents. Only this unique subject identification number will link the data or samples to the patient. A confidential correspondence list, containing the code associated with each patient name and identity, will be securely stored by the study doctor. Only the study doctor will be able to connect the subject identification number to the patient's personal data. He/she will not share this information except as explained in the Patient Information and Informed Consent Form.

Anonimization

Anonymization is the process of turning data into a form that does not identify individuals and where identification is not likely to take place. Once data is truly anonymous and individuals are no longer identifiable, the data does not fall within the scope of the GDPR, thus not being personal data and allowing for a much ease to openly share and publish.

The proces of anonimization uses methods against identify, membership and attribute disclosure to minimize the risk of reidentification.

Data Protection Impact Assessment (DPIA)

[If a DPIA is required] Based on the nature of the data the present project intends to collect, a Data Protection Impact Assessment (DPIA) might be required under the GDPR (Art. 35). The DPIA is mandatory for projects that are likely to involve "a high risk" to other people's personal information. More specifically, a DPIA is required when[1]:

- Using new technologies
- Tracking people's location or behaviour
- Systematically monitoring a publicly accessible place on a large scale
- Processing personal data related to "racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation"
- Processing is used to make automated decisions about people that could have legal (or similarly significant) effects
- Processing children's data

Additionally, following the EC Guidelines for DPIA[2], it is also required when:

- Evaluation or scoring, including profiles and predictions
- Automated decision making with legal effects or that has a similar and significant effect on the natural person
- Systematic monitoring to observe, monitor or control data subjects of a public access area or networks
- Sensitive data or data of a highly personal nature
- Large-scale data processing

[1] https://gdpr.eu/data-protection-impact-assessment-template/
[2] https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

- Data sets that are matched or merged, obtained for different purposes and / or different sources
- Data concerning vulnerable data subjects
- Innovative use or applying new technological or organisational solutions.
- Processing that in itself prevents the exercise of a right or the use of a service or contract

Although not all of the above apply to the present project, our data collection will partly consist of <insert requirements>.

Moreover, given the age of the participants, data controllers must conduct data protection and privacy impact assessments, considering the specific impact on children's rights and ensuring algorithmic outcomes are in their best interests and do not unduly affect their development.

As outlined in Article 35, the GDPR requires DPIAs to contain the following elements:

- A systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller
- An assessment of the necessity and proportionality of the processing operations in relation to the purposes
- An assessment of the risks to the rights and freedoms of data subjects
- The measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with the GDPR, taking into account the rights and legitimate interests of data subjects and other persons concerned.

6.A b) Real example

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

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

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

7. Other issues

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

7.A a) Description

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

Institutional information

Each of the partners, including IRSJD, will follow their own institutional policies and procedures for research data management, in addition to European policies regarding research data management and open science.

IRSJD:

- The Open Science Policy
- The Institute / Fundació Recerca Sant Joan de Déu Privacy Policy (link)
- The Hospital Sant Joan de Déu Privacy Policy (link)
- Parc Sanitari Sant Joan de Déu Privacy Policy (link)
- "EinaDMP" (https://dmp.csuc.cat)
- The CSUC Horizon Europe Data Management Plan "template" & guidelines with customizations by IRSJD.
- The DPIA Template

7.A b) Real example

Ex. 1: As part of <Institute>'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 <Institute> 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 University of Strathclyde Glasgow
- Information Security
- Research Code of Practice
- Research Data Policy

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

SJD Sant Joan de Déu
Institut de Recerca

Edifici Docent Sant Joan de Déu
C/ Santa Rosa, 39-57, 4a planta
08950 Esplugues de Llobregat
Tel. 93 600 97 51
info@fsjd.org · www.irsjd.org