Plan Estatal
(Version 2, May 2025)
(Doc.CO25/05) (4 RDM\Plans Gestio Dades, 26.05.25)

This document aims to serve as a guide for researchers in drafting their Data Management Plans (*Data Management Plan,* DMP), especially in the context of projects funded under the Plan Estatal de Investigación Científica, Técnica y de Innovación.

The following is shown:

- With a number: the different sections of the template.
- With a capital letter: the elements to be considered when completing the section.
- With a lowercase letter: descriptions and examples from real cases.

This document was prepared by the CSUC Research Support Working Group, composed of representatives from the following universities: 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 y Universitat de les Illes Balears.

The examples are a sample[1] of various DMPs available online.

This document is subject to the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

Digital version: http://hdl.handle.net/2072/484382

Preliminary information

The deliverable also typically includes other preliminary information such as the project logo, level of dissemination, historical revisions, table of contents, and a list of acronyms used.

1. Data summary

1.A Specify the source of the data, its type, format, and size. Describe their usefulness from the point of view of their reuse, and whether they can be shared

1.A a) Description

Briefly specify whether the data will be generated within the project or collected from other sources, as well as the standard, methodology or software used for its generation or collection. If existing data is reused (whether open, purchased, etc.) specify the data source from which it was extracted (DOI, handle, URL, etc.) and explain how issues such as copyright and LPI have been handled.

Describe the data types, formats (text, numeric, image, etc.) and sizes (or approximate estimate) of the data used in the project. Preferably use open standards so that data can be read by multiple programs, making it easier to retain and share with other users. If you don't use them, explain why and strategies for converting them.

Also specify for each dataset the possibilities of sharing in open access (total, restricted or embargoed open access) in repositories, open licenses.

Indicate for which group and for what purpose the project data may be of interest.

1.A b) Real example

E.g. Existing data will come from a variety of sources including pre-existing data from *[http://doi.org/xxxx],* data from the scientific literature, data from real-world measurements, and data from simulation experiments. Data collected by the project itself will also be created through surveys *[XXXXX explain].*

Tabular data will be stored in open formats such as *[.csv (comma-separated variables)],* which can be imported into rich text files for word processing or spreadsheets. If specialized software is used, information about free readers will be provided. Text files will be saved in *[.txt and .pdf]* format. The expected size of the data is not currently known, but it is likely to be *[<10 GB]* and the individual files to be *[≤1 MB].*

E.g. RDR. The data generated by the project will be published in the consortium repository *Repositori de Dades de Recerca (RDR)* Once the project is finished, preferably under the https://creativecommons.org/publicdomain/zero/1.0/ license. The data will be useful for use by other research groups working on the following topics: *[indicate topics].*

2. FAIR Data

2.1 Findable data (including metadata)

2.1.A Specifying the metadata standard used

2.1.A a) Description

Metadata schema that uses the open data repository where the dataset was deposited.

2.1.A b) Real example

E.g. RDR. The data have been deposited in the _Repositori de Dades de Recerca (RDR)_, which is based on _Dataverse_ and uses the _DataCite metadata standard_.

2.1.B Identify the type of persistent identifier assigned

2.1.B a) Description

What type of unique and permanent identifier the data repository grants.

2.1.B b) Real example

E.g. RDR. This repository uses the DOI as a unique and permanent identifier, in the following format: _[doi.org/10.34810/dataXXX]_

2.1.C Indicate if version control is applied

2.1.C a) Description

Specify whether the repository in which the data was deposited performs versioning.

2.1.C b) Real example

E.g. RDR. RDR allows control of 'minors' (V1.1 -> V1.2) and 'majors' versions (V.1.0 -> V2.0).

2.1.D Indicate the institutional or thematic location in the repository

2.1.D a) Description

Mention if a specific Dataverse/Instance has been created to host the data of the institution or research project, the address of the Dataverse/Instance.

2.1.D b) Real example

E.g. All research datasets linked to this project can be consulted from the link _[specify the specific link for the institution or project]_

2.2 Accessible data

2.2.A During the period of execution of the research

2.2.A a) Description

Where the data has been stored. Who has been able to access, with what permissions.

2.2.A b) Real example

E.g. During the period of execution of this project, the data has been stored in the ECM (Enterprise Content Management) of the institution, *[Microsoft SharePoint]*, which allows sharing access to the dataset with different permissions depending on the user.

For this research, the PI of the project has been the administrator of all the datasets generated; On the other hand, each researcher has created their own structure of folders and files, within the space defined for the project. Permissions to other spaces in which you are not an owner have only been for consultation.

2.2.B At the end of the research (in which several cases may occur)

2.2.B.1 a) Description

Where the data has been stored. Who has been able to access, with what permissions?

2.2.B.1 b) Real example

E.g. 1 The data is stored in open access without any restriction on access.

E.g. RDR A. The data are deposited in the RDR, with the following DOI: *[doi.org/10.34810/dataXXX]*. The metadata is available in open access, the files have an embargo period that ends *[dd/mm/yyyy]*, and from that moment on they will be accessible without any restriction.

E.g. RDR B. The research data are deposited in the RDR with the following DOI: *[doi.org/10.34810/dataXXX]*. The metadata is available in open access, the files have restricted access since to be shared, a collaboration agreement must be generated for their reuse.

E.g. RDR C. Research data contains personal data and has not been deposited in compliance with *[....]* / Given that the RDR regulation specifies that this type of data cannot be deposited.

2.2.B.2 a) Description

The methods and programs necessary to access the data should be mentioned. The recommendation is to make the deposit in open formats. In cases where the discipline uses proprietary formats as scope standards, the program and the specific version that was used to create the research dataset should be indicated.

2.2.B.2 b) Real example

E.g. 1 Data have been deposited in *[(.csv, .tiff, .pdf, .tab....)]* that it is open and no specific program is required to access them.

E.g. 2 The data has been generated in *[.raw format]*, which is the standard and proprietary format of *[XXX software version XXX]* used in mass spectrometry. This format requires the use of the program for the opening, visualization and analysis of the data deposited in *[.raw]*. *[Software XXX]* does not allow you to export the data in its entirety to an open format without losing some of the structural information of the experiment (acquisition metadata, instrument parameters, etc.). Therefore, full access to the original data is conditional on having such software.

2.3 Interoperable data

2.3.A Specify data interoperability. Explain what data and what metadata vocabularies, standards, or methodologies will be followed to facilitate interoperability

2.2.A a) Description

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

2.2.A b) Real example

E.g. 1 The data generated in the project will be interoperable, as the datasets will conform to standardized formats: *[ASCII, TXT, CSV, XML, and TIFF]*. If *[MS Office],* a PDF viewer, or an image viewer cannot be used, a text file (ASCII) will be provided along with the dataset explaining where a free reader can be obtained.

E.g. 2 There are no community-agreed metadata standards for this discipline. To ensure interoperability, data will be properly documented (README files, etc.) to allow for reuse and operability across disciplines. In addition, the data underlying the publications will be available through *[repository name]* (which uses *[DataCite and Dublin Core metadata schema]*).

E.g.3 To ensure interoperability between the datasets available through the *[partnership]* portal and other spatial data, the standard protocols *[OGC WMS and WFS]* are adopted. Metadata is published using the standard protocol *[OGC CSW]*. In addition, the availability of [portal] datasets as downloadable files in common formats *[(GeoJSON files, ESRI shapefiles, MapInfo tables, KML)]* makes it easy for users who need it to combine [portal] datasets *with* other geographically referenced data in *[desktop GIS]*.

2.4 Reusable data

2.4.A Specify when the data will be available for reuse and under what conditions (licenses). If applicable, specify why and for what period the data will need to be attached

2.4.A a) Description

If data is made available to other researchers and the general public, specify the degree of reuse allowed through the licenses. The EC proposes *Creative Commons CC-BY* or *CC0* licenses. Software/code-specific licenses are available. For more information, see the Rights and Licenses for Research Data guide.

If the dataset is published to the open, indicate when the data will be available for consultation and reuse (during the project, at the end of the project, etc.). If the dataset has any access restrictions, specify the reason and for what period it will be necessary to seize or restrict.

Indicate the documentation that will accompany the published data (readme file, data dictionaries, code books, etc.) necessary to understand and reuse them. This documentation should include information on how the data was collected, analytical and procedural information, definitions of variables, units of measurement, etc.

Specify how long the data will be available for reuse (different international standards recommend a minimum of 10 years).

2.4.A b) Real example

E.g. RDR. The project data will be published after the project is completed in the *Repositori de Dades de Recerca (RDR)* under the [*CC-BY*] license. The published data will not have access restrictions or embargo. All documentation needed to validate the data analysis and facilitate its reuse will accompany the data through a README file created in accordance with the RDR template and guidelines. The RDR repository ensures preservation and facilitates access to deposited datasets for a minimum of 10 years.

3. Responsibilities

3.A Identify the persons responsible for data management activities and specify issues related to the ownership of intellectual property rights

3.A a) Description

Indicate the persons responsible for the creation, monitoring and updating of the PGD. Specify who is responsible for the management of the data and who determines access to it, as well as its publication, preservation and conservation. If there are intellectual property rights, determine who the owner is. In the case of collaborative projects, determine the

responsibilities of each team and the responsibility for coordination, as well as manage the intellectual property between the different teams. Indicate any applicable institutional regulations.

3.A b) Real example

E.g. 1 The PI of the project will be responsible for data management throughout the period of execution of the research. Their responsibilities have been to create and monitor the shared space that will contain the entire structure of files and folders generated as a result of the investigation process. Before publishing the datasets in the repository, the datasets will be curated by the Library's research support service.

E.g. 2 The project PI will monitor the implementation and progress of the Data Management Plan, ensuring that the goals are met. The PI will also oversee data management efforts, with the involvement of all team members, in the manipulation, preservation, and preservation of data. Researchers will be responsible for ensuring, throughout the project, the integrity of the data, its correct labelling and archiving. Access to the data will be monitored by the IP and the data manager.

4. Data security

4.A Specify security measures for data management and retention

4.A a) Description

It must specify how the data recovery is processed, as well as the secure storage and transfer of sensitive data and the measures taken for data security. This includes storage and backups, long-term preservation, and procedures to keep data safe and secure.

4.A b) Real example

E.g. RDR. During the collection or processing of research data, the guidelines of *[institution X] shall be followed*. And at the end of the project, they will be deposited in the RDR. The following list describes your security settings:

- **Versions:** Data files are versioned. The files are not versioned. The uploaded data is archived as a shipment information package. They are generated by-products from the data files, but the original content is never modified. Records may be removed from public view; however, data files and logs are retained.
- **Replicas:** All data files are stored at the CSUC Centre, mainly in Barcelona, with replicas at the Madroño Consortium in Madrid. Data files are maintained in multiple replicas on a distributed file system, which are backed up to tape nightly.
- **Retention period:** Items will be retained for the life of the repository. The RDR has defined a useful life for the repository of at least the next 10 years.
- **Functional preservation:** The RDR makes no promises about the usability and understandability of deposited objects over time.

- **File retention:** Data and metadata files are backed up nightly and replicated to multiple copies on the online system.
- **Fixity and authenticity:** All data files are stored together with an MD5 checksum of the file contents, and the tabular file is stored with the Universal Numerical Fingerprint (UNF). Files are periodically checked against their checksums to ensure that their contents remain constant.
- **Succession plans:** In the event of a repository closure, RDR has guaranteed the migration. The RDR's contingency plan and technical specifications can be consulted at the following link: Policies/Plans.

4.B Describe ethical coverage in data management

4.B a) Description

Ethical coverage, it must describe, if applicable, the ethical problems in the collection, storage, processing and archiving of the data, as well as in the ethical approval procedures related to the project.

4.B b) Real example

E.g. 1 This investigation has the approval of the ethics committee of the *[institution],* with a favourable report issued by *[competent body in each institution]* on the day *[dd/mm/yyyy]*, as attached in the justification of this project.

E.g. 2 This research does not need the approval of the ethics committee, since no data of a personal, sensitive nature or related to the assumptions that the law establishes for registration in the ethics committee are handled.