UCA Universidad de Cádiz

BibliotecaUCA

Research Data Management Plan

TEMPLATE

Sección de Apoyo a la Investigación y Servicios
Área de Biblioteca, Archivo y Publicaciones

With the aim of aiding researchers at the University of Cádiz with the correct management of research data and the preparation of a Data Management Plan (DMP), the Library of the University of Cádiz has published a *Guide for the Preparation of a Research Data Management Plan*, together with this document which is intended as a template for the preparation of a Data Management Plan.

Both documents are based on the *Horizon Europe Data Management Plan Template* but they are intended to address issues that arise in a DMP even though they are not necessarily carried out under this programme. It can be used with tools such as DMPonline, PGDonline, Argos, etc.

We would like to thank the Research Support Service of the Library of the University of Seville for allowing us to use the template developed by them, which we have adapted to the needs of the Institutional Repository of the UCA (RODIN) as a complementary tool to facilitate the deposit of research data.

In addition, the Library of the University of Cádiz has published a Research Data Guide with various sections on aspects to be taken into account when managing research data. A series of general and UCA-level guidelines and recommendations are included, as well as links to resources for further information, tools, examples of other DMPs, etc.

Research Support and Services Section

March 2023

Horizon Europe

Data Management Plan Template

Version 1. 0

05 May 2021

HISTORY OF CHANGES		
Version	**Publication date**	**Changes**
1.0	05. 05.2021	• Initial version
		•

Action Number: [insert project reference number]

Action Acronym: [insert acronym]

Action title: [insert project title]

Date: [insert date]

DMP version: [insert DMP version]

1. Data Summary

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.

If a dataset is re-used, indicate where it has been extracted from, e.g. a relevant repository. If existing data sources are purchased or re-used, explain how issues such as copyright and intellectual property law have been addressed.

When creating new data sources, explain why existing sources are not re-used.

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

Describe the content and scope of the data. Research data is generated for different reasons through different processes and may be of different types.

Give examples of the types of data that could be referred to and their content. Choose the type that suits your needs, specifying what you consider necessary:

- Observational: data captured in real time (e.g., neuroimages, sample data, sensor data, survey data, etc.).
- Experimental: data captured from laboratory equipment (e.g., gene sequences, chromatograms, magnetic field data, etc.).
- Simulation: data generated from test models (climate, mathematical or economic models, etc.).
- Derived or compiled: data resulting from processing or combining 'raw' data which is very difficult to reproduce (e.g., text and data mining, 3D models, compiled databases, etc.).
- Reference: a conglomeration or collection of datasets (gene sequence databanks, chemical structures, spatial data portals, etc.).
- Other.

Data formats:

- Graphics: jpeg, odg, pdf, png, pttx
- Tables: odsu, opj, xlsx
- Text: docx, pdf, txt

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

Briefly explain the purpose of the data generated and/or collected and its relation to the project objectives.

Drafting examples:

- To facilitate data exchange within the project discipline.
- Experimental data will be collected in order to fulfil the commitments established with the funding agency.
- To meet the project objectives and in peer-reviewed conferences and publications.
- To collect research data under the project and make it available to support the credibility and enhance the quality of scientific publications based on such data.
- To provide data for future research projects to continue with the work. This will ensure the lasting impact of this EC-funded project beyond the project period.

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

Indicate the approximate volume of the dataset. The volume of the data needs to be considered in terms of storage, backups, cost and access. Estimate the volume of data in MB/ GB/ TB and the incremental growth to ensure that additional storage and technical support can be provided.

Drafting examples:

- The expected size of the data is currently unknown but is likely to be <10 GB with individual files of ≤1 MB.
- The size of the data handled by <ProjectAcronym> is quite small (less than 10 GB) except for the testing of the data infrastructure in WP6, for which the project needs to handle large volumes of data as explained above.
- The size of the data is currently unknown. This aspect will be reviewed following this initial data management plan based on initial experiences storing the results of the different types of measurements. The main relevant data sizes will stem from images such as microscopic sample characteristics stored in high-resolution bitmap format. However, the total dataset size for a single sample characterisation is expected to be in the order of tens of MB only.

What is the origin/provenance of the data, either generated or reused?

Specify whether the data has been retrieved from another repository and whether it consists of shared results from other research or another project. If so, indicate the

origin/provenance of the data, identifying the source, research project and authors of the data, and if possible add a brief description of the data and how it has been processed in the current project. If existing data sources are purchased or re-used, explain how issues such as copyright and intellectual property law have been addressed. When creating new data sources, explain why existing sources are not re-used.

If the data does not come from any previous data source, indicate and describe the methodology used for its generation.

Drafting examples:

- The existing data will originate from a variety of sources, including: pre-existing data, data from the scientific literature, real-world measurement data and data from simulation experiments. Data collected from domestic properties will remain confidential and will not be included in the repository.
- The data originates from experiments and measurement campaigns carried out by...
- Some of the project tasks will make use of existing data in formats such as.... This data will be used in the validation of the project results.
- Existing images and data selected from the (SOURCE) databases will be used in specific tests, such as storage tests in (DESTINATION). The last type of data to be created is that which must be preserved, made available and transmitted to subsequent parties working on (PROJECT NAME).

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

Indicate to whom the data may be of interest.

Drafting examples:

- The data may be used by other research groups working on the following topics: ……. The data will also be useful for standards committees, including the Working Group .....
- The data may be used by independent researchers to better understand the content and conclusions of scientific publications which base their findings on the data. In addition, independent researchers can use the files to produce figures and publications and make comparisons of their own results with the results in the data. Scientists can also use the data files to repeat the experiments and measurements to verify and validate the research. Finally, science writers and the press can also use the datasets to produce high-quality infographics that demonstrate the potential impact of the technology.

2. FAIR Data

The content of these sections refers to the characteristics of the Institutional Repository of the University of Cádiz (RODIN).

2.1. Making data findable, including provisions for metadata

Will data be identified by a persistent identifier?

Drafting example for RODIN:

The data generated under the research project will be stored in the Institutional Repository of the University of Cádiz (RODIN), which assigns each record a persistent identifier based on the handle and citability of the dataset.

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.

Drafting example for RODIN:

The research data will be described using the Dublin Core extended metadata model, in line with the descriptive specifications of the OpenAIRE repository. These specifications include identification of the different types of author (creator, collector, manager...) and their affiliation, title and variants thereof, creation and update dates, versions, description of the content, keywords, research projects to which they are linked, usage licences and access details. To the extent possible, this data will be provided in Spanish and English. Compliance with these requirements allows collection and inclusion of the data on different platforms (Recolecta, Google Scholar, OpenAIRE...) and links to the record in the RODIN repository.

In addition, a Readme.txt file will be added to the data files with additional information to facilitate interpretation and re-use of the data.

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

Drafting example for RODIN:

The RODIN repository uses the OAI-PMH-compatible DSpace application for open metadata harvesting, applying the Dublin Core Extended metadata model and the OpenAIRE specifications.

Will search keywords be provided in the metadata to optimise the possibility for discovery and then potential re-use?

Drafting example for RODIN:

The data entry form in the repository has specific metadata for keywords indicating the subject matter of the project and the attachments, as well as geographic and chronological identifiers. In the case of RODIN, these terms are assigned to the dc.subject metadata.

2.2. Making data accessible

REPOSITORY:

Will the data be deposited in a trusted repository?

Drafting example for RODIN:

The research data will be deposited in a repository that complies with the data description recommendations and guidelines, has defined access conditions and possible restrictions and complies with OAI-PMH harvesting protocols. In the case of the RODIN repository, apart from meeting these requirements it is integrated in most of the open science harvesting platforms (Recolecta, OpenAIRE...).

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

Drafting example for RODIN:

The Institutional Repository of the University of Cádiz (RODIN) is fully functional and correctly tags datasets with the Dublin Core metadata schema.

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

Drafting example for RODIN:

The process of integrating research data into the repository will involve automatic assignment of handles consisting of a persistent identifier providing a unique access URL.

DATA:

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

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

Indicate how the data will be shared, including the procedure for accessing the data, embargo periods (if any) and whether access will be fully open or restricted to a specific group. If certain data cannot be fully opened, justification must be given as to why.

Drafting examples:

- All data produced by the experiments described above will be openly available, meaning that all results and images will have open access. There may be exceptions relating to details of the equipment used and algorithms applied in the interpretation, as these may contain other proprietary information. In RODIN, the option exists to provide open access, embargoed access and closed access.
- All data associated with scientific publications will be openly available by default, unless there is a specific reason for not publishing.

- The following data will not be made publicly available:
  - Data obtained with the permission of third parties but who have not given permission to make the data publicly available.
  - Data revealing the identity of a manufacturer.
  - Data compromising the protection of a third party's intellectual property rights.
  - All project data will be made available with the exception of market or customer survey data, which is commercially sensitive and cannot be shared. There are certain datasets that cannot be shared for legal and contractual reasons.
  - Creative Commons licences will be used to determine the use of the data. These include CC0 and CC-BY licences. The owner of the data will determine which of these licences will be used when the data is published in the repositories. However, the project recommends using CC0 for data and CC-BY for media and avoiding use of CC-BY-NC, which causes problems in some national jurisdictions.

- Once processing, quality control, organisation, analysis and publication are complete, the data will be made accessible via deposit in open access repositories (e.g. Zenodo).

Will the data be accessible through a free and standardised open access protocol?

Drafting example for RODIN:

The research data will be stored in a repository compliant with the OAI-PMH open access standard and the Dublin Core description standard to ensure open access and harvesting.

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

If access to the data is restricted to the general public for a justified reason, specify which data will be accessible to individual partners, to all partners or on request. Specify the procedures for requesting access to restricted data and the conditions for granting access. Also specify whether there will be further restrictions over a certain time period. It is also necessary to specify the Creative Commons licences that may be used to identify the possible uses of the data.

Drafting example:

- The working data, raw survey data and personal data of survey participants will be restricted to members of the research project and will only be accessible according to established levels of accessibility.
- There will be no restrictions on the use of published data, but users will be required to cite the consortium and the data source in any resulting publication.
- The data will be licensed under a Creative Commons Attribution 4.0 (CC BY 4.0) licence.

How will the identity of the person accessing the data be ascertained?

Describe the access conditions defined by the repository of your choice (indicating the URL where the information originates from), e.g. with a machine-readable licence.

Drafting examples:

- In the case of open access data no access control will be established, with users being expected to follow the scientific citation guidelines and the terms of the corresponding licences. If there is any access restriction or embargo, the request will be handled via the repository management team.

Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

Drafting examples:

- Due to the small scale of these experiments, there is no need for a data access committee.
- The research team will establish the data access criteria and manage their application. The data with open access will be selected on a case-by-case basis. Ethical aspects and data security, including intellectual property requirements, will be considered. If necessary, some or all of the data may be withheld from a potential publication. This will be decided in consultation with the other members of the project.

METADATA:

Will metadata be made openly available and licensed under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why.

Will metadata contain information to enable the user to access the data?

Drafting example for RODIN:

The metadata standard used to describe the dataset will be the Dublin Core Schema, as it is a flexible and commonly used standard and it has also been adopted by the European OpenAIRE repository. They will be licensed under a CC-BY Licence.

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

Indicate the time limits of the repository for preservation of the data.

Drafting example for RODIN:

The Institutional Repository of the University of Cádiz (RODIN) guarantees preservation of the data without any time limit.

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

Include any technical requirements for accessing and re-using the data, as well as the documentation of the software needed to access the data.

Drafting example:

- All the data generated will be accessible via commonly used software: MS Office, Open Office, Adobe Reader, Image Viewer, etc. If any specific software is used,

the necessary documentation relating to that software will be provided to facilitate access to the technical data. For software developed under the project, the source code will be deposited in the repository.

- Where possible, standard publicly available software will be used, but if specialised software tools are developed a short text file (e.g. ASCII) will be provided with the data file to explain the required software.
- Most of the software programmes are available as commercial products or as freeware. For software developed under the project, the source code will be deposited in the repository (e.g. Zenodo).

2.3. Making data interoperable

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

Drafting example:

The data generated under the project will be interoperable as the datasets will conform to standardised formats: ASCII, txt, csv, xml, tiff.

If you must inevitably use uncommon or generate project-specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the ontologies or vocabularies generated to allow re-use, refinement or extension?

Drafting example:

There is no standard vocabulary for this type of data. However, the most common definition of the relevant scientific community is used to the extent possible.

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

2.4. Increase data re-use

How will you provide documentation 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.)?

Drafting example:

The data will be accompanied by a readme file for deposit with complementary information to aid the interpretation and re-use of the data.

Will your data be made freely available in the public domain to permit the widest possible re-use? Will your data be licensed using standard re-use licences, in line with the obligations set out in the Grant Agreement?

Drafting example:

The data will be licensed under Creative Commons CC-BY licences and will continue to be re-usable after the end of the project by anyone interested in the project, with no access or time restrictions.

Will the data produced under the project be usable by third parties, in particular after the end of the project?

Indicate to whom the data may be of interest.

Will the provenance of the data be thoroughly documented using the appropriate standards?

Drafting example:

The documentation and metadata for each dataset acknowledge the provenance of the data by citing the information source, with appropriate use of formats commonly accepted by the relevant scientific community.

Describe all relevant data quality assurance processes.

Drafting example:

The project team carries out quality control of the data distributed according to a workflow described in the data quality assurance document available in the documentation section of the project portal.

In addition to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

3. Other research outputs

In addition to data management, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs may be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

Drafting example:

Where other research outputs are generated (only digital outputs are foreseen), the management of those products is shared with their creators.

Beneficiaries should consider which of the questions pertaining to FAIR data above may apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed, shared or made available for re-use in line with the FAIR principles.

Drafting example:

Where other research outputs are generated, compliance with the FAIR principles will be managed by their creators.

4. Allocation of resources

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, reuse, security, etc.)?

Indicate the approximate cost of making the data FAIR.

Drafting examples:

- The estimated costs for making data findable, accessible, interoperable and re-usable (FAIR) are €... (staff costs). These costs have been minimised by using an institutional repository (RODIN) and making only relevant data FAIR.
- The data will be deposited in the institutional repository and the University assumes the costs of storage, security and preservation.

How will these be covered? 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)

Drafting example:

The costs will be covered by the financial envelope included in the project grant earmarked for this purpose.

Who will be responsible for data management under your project?

Explain the responsibilities for data management under each project.

Drafting example:

- The project coordinator is ultimately responsible for data management under the project.
- Each partner must respect the policies set out in this DMP. Datasets must be created, managed and stored appropriately and in accordance with the applicable legislation.
- The project coordinator has a particular responsibility to ensure that data shared via the website is readily available and also that backups are made and proprietary data is protected.
- The relevant WP leader is responsible for quality control of the data with the support of the Project Coordinator.

How will long-term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

Drafting example:

The data will be deposited in an institutional repository so there are no costs associated with the long-term preservation of the data.

5. Data security

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)?

Indicate how data security has been ensured in terms of the storage media used during the research, file transfer, data access, backups, etc., especially if it is personal or sensitive data.

Will the data be securely stored in trusted repositories for long-term preservation and curation?

Indicate where the data will be stored to ensure future preservation and curation.

Drafting example for RODIN:

The data will be deposited in the institutional repository (RODIN), which ensures long-term preservation and curation and complies with the necessary protocols to guarantee data security and protection.

6. Ethics

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

Describe any potential ethical issues during data collection, storage, processing and archiving, together with the ethical approval procedures related to the project.

If the research activities involve children, patients, members of vulnerable populations, use of embryonic stem cells, privacy and data protection issues or research on animals and primates, the ethical principles and relevant national, EU and international legislation must be complied with.

Drafting examples:

- The data collection, storage, use, generation and dissemination under this project do not pose any ethical issues.
- The main ethical issues under the project will be related to preservation of the identity of the survey participants and the image rights of minors. Market survey data will be anonymised to comply with the General Data Protection Regulation (GDPR).
- All activities carried out in the framework of the <AcronymProject> comply with ethical principles and relevant national, EU and international legislation, e.g. the Charter of Fundamental Rights of the European Union and the European Convention on Human Rights. The tasks of <AcronymProject> only concern basic research activities and the project does not involve humans, animals or cells. Due to the fact that the main field of the <AcronymProject> is related to..., the risk of ethical issues during the project is extremely limited.

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

Drafting example:

- Informed consent has been requested from all persons participating in the study, along with authorisation to publish the data, even if it is anonymised.
- The project does not collect or distribute personal data.

7. Other issues

Do you or will you make use of other national/ funder/sectorial/departmental procedures for data management? If so, which ones? (please list and briefly describe them)

Drafting example:

No other data management procedure is contemplated. Additional DMPs using the DCC template will be adopted for specific datasets.