



Important information for DMPs¹

Preamble:

The project members are committed to the following underlying principles to facilitate the discoverability, accessibility, dissemination and reusability of the project data:

- Persistent identifiers will be assigned to the shared data and included in its metadata, so to make data *findable*;
- Detailed documentation on the datasets will be provided, with metadata being based on standard vocabulary(ies), whenever possible, making data *accessible*;
- Standardized methodologies and metadata content will be aimed, in order to make data *interoperable*;
- The relevant data collected in the course of the project, except sensitive and personal data, will be made openly available, with respect to possible embargo periods, through data repositories and with suitable licenses, thus making data *re-usable*.
- For data management plans templates available online, please see:
 - ✓ <https://argos.openaire.eu/splash/>
 - ✓ <https://dmponline.dcc.ac.uk/>
 - ✓ <https://dmptool.org/>
 - ✓ [CORDIS – EU Research Results](#)

¹ This annotated DMP has been drafted based on the annotated template by the University of Cádiz and on the Horizon Europe DMP template.

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.

2. 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. The data can be both quantitative and qualitative in nature and will be analyzed for project development and scientific purposes. Metadata and results from statistical analyses should be available in accessible standard formats (PDF, XLSX, CSV, DOC, JPEG, PNG, GIF, TIFF).

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.).

Data formats:

- Graphics: jpeg, odg, pdf, png, pptx
 - Tables: odsu, opj, xlsx
 - Text: docx, pdf, txt
- Type of data

3. 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.

4. 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.

5. 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).

6. 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.

Fair Data

Findable and Interoperable

Data will be documented using metadata standards defined in the DMP, to ensure that the datasets can be understood, shared, and re-used. Appropriate metadata standards should be followed, as the Dublin Core, a highly adopted, domain-agnostic metadata standard, published as ISO Standard 15836 :2017 (<https://www.iso.org/standard/71339.html>).

Metadata records based on the Dublin Core standard ensure that project data is more easily findable and also interoperable. Metadata standards for describing specific project data should be evaluated on a case-by-case basis and reflected in upcoming versions of the DMP.

Drafting examples:

- For the deposit of a dataset in the i3S data repository (NAS), a metadata template will be made available by the DMP manager. This template is to be filled in by the creator of the specific dataset and validated by the DMP manager. The DMP manager will maintain an inventory of all the data deposited, and essential metadata, in the project drive.
- The following file name structure will be adopted: yyyy.mm.dd_(research output) context_creator initials_version.extension

When applicable, links to digital object identifiers (DOI) from open access databases where the corresponding data can be found and downloaded will be provided. The use of persistent unique identifiers namely DOI and open access databases ensures the availability and identifiability of each data file.

- Metadata will be carefully stored - including collection date and, methodology used - and follow the EU Open Data Portal Metadata definitions.
- Research data will be deposited in open data European repositories as soon as possible, unless a decision has been taken to protect results. (in order to choose the best data repository for your project, please see <https://www.re3data.org/>)

Data availability

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.

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:

- o Data obtained with the permission of third parties but who have not given permission to make the data publicly available.
- o Data revealing the identity of a manufacturer.
- o Data compromising the protection of a third party's intellectual property rights.
- o 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.
- o [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 CCBY for media and avoiding use of CC-BY-NC, which causes problems in some national jurisdictions.
- o 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).

Another drafting example:

- o The experimental data generated during the development of the project will be maintained until the results are patented and/or published. Processed data will be disseminated in peer-reviewed scientific articles, published in specialized open access journals and in repositories associated with the host institutions. Scientific articles will be identified by a DOI. Published data will follow the standards to identify the authors, locations and affiliations, and may be reused citing the authorships and origins through the application of a license. Data requested to be publically available by the publishing journal will be made publicly available, and therefore accessible by the article DOI. Metadata in .csv format not included in the publication but relevant will be assigned a unique and persistent identifier D.O.I. through the Zenodo data repository.

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 (see [Creative Commons](#)) 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.

Publications

Outcomes must be published or otherwise disseminated in an appropriate form, although publication or release of findings may be delayed for a reasonable period to allow for protection and commercialization of intellectual property. In this case, publications should be postponed until IP protection has been granted. Once published, data will be made available upon request and under licenses mentioned below as appendixes or supplementary data. Publications will also be deposited on the selected repository. Research papers will be published in open-access platforms, becoming available free of cost. According to the funding agency recommendations, a CC-BY Public Copyright Licence shall be granted to all future Author Accepted Manuscripts (AAMs).

Note: Before publication, see [Think, Check, Submit](#) (to identify trusted publishers for your research) and the [Directory of Open Access Journals](#) (to search for open access journals, including those free of fees)

Licensing

Drafting examples:

- Access to the generated data in the project will be initially given to team members only, and after publication or upon request to other researchers (stating the intended use).
- Creative Commons (CC) licences will be implemented to guide subsequent use of data and other intellectual products, providing robust legal code, while being also free and easily readable by end-users. With this license, the user can distribute, remix and build upon the original, as long as the original creator is given credit, or attribution. However,

a CC-BY-NC or Creative Commons Attribution NonCommercial may be also used, preventing the work from being commercially used, generating profit.

Data preservation

The data generated by the project, provided it does not contain sensitive information, can be deposited, and preserved in *the institutional repositories* for at least 10 years (depending on the project, it can vary).

Personal storage devices, such as PCs and Laptops are convenient for short-term storage and data processing yet, these fall outside the scope of the DMP. However, project partners must ensure that the backup of data is regularly made through suitable networked drives.

Project outputs can be deposited into Zenodo (<https://zenodo.org/signup/>), an open-access catch-all repository to support the dissemination of project achievements, but mostly at a later stage of data life-cycle. Datasets can be deposited with private visibility, in order to respect eventual embargo periods and other access restrictions. Zenodo and the offline hard drive are complementary services for preservation purposes of selected data.

Zenodo also enables the definition of communities responsible for the curation of their resources. Therefore, the setting up of a project community in Zenodo can be considered in order to manage project outputs. These outputs can take the form of processed data, research protocols, reports, the different version of the DMP itself, among other artefacts.

The deposit of data in disciplinary data repositories is also open to accommodate possible data policy requirements while publishing project results. In this case, the project partners can deposit the data in the repositories recommended by the publishers, if they recommend any, or the selection of a disciplinary repository can be made via the data repositories directory [re3data.org](https://www.re3data.org/) -- <https://www.re3data.org/>.

Please note: The person/institution responsible for data curation should be identified.

IP management

Projects must follow reasonable procedures for the identification, protection, management and commercialization of IP, in line with the funding agency Intellectual Property policy. This is particularly relevant when the Grant Activities involve collaboration with, or the contribution(s) of, third parties.

Data security

Data should be stored in servers protected by an access list and password, with regular backup copies to avoid data loss. Backups are responsibility of host institutions.

Drafting examples:

To prevent research loss (either due to loss of bits or loss of rendering capability), “our team/this project members” will:

- Prefer data storage in open data formats;
- Maintain on-site and off-site backups, at least two;
- Routinely backup data (at least monthly).

Specifically:

- Weekly backup of data on NAS/drive of each individual person
- Monthly backup of data on common drive (all information generated during that month; new versions of the documents should indicate so, so there is no need to delete previous files on the drive) – all team members have access to these files
- Monthly backup of data on i3S data repository (NAS) by the project manager (the PM, after everyone has made the backup of their data to the drive, will backup everything to the i3S NAS) – second online site with all data
- Monthly backup of the final versions of the documents on Basecamp
- Every three months backup of data on an offline hard drive, for long-term off-site data preservation
- Every three months backup of relevant data on Zenodo

Drafting example:

As a rule of thumb, the data generated by the project, provided it does not contain sensitive information, will be deposited, and preserved in the i3S data repository for at least 10 years.

Important note: to manage and share data securely, certain data storage platforms are not recommended for long-term storage, such as:

- External portable storage devices, such as external hard drives and USB drives, given their longevity uncertainty and how easily these can be damaged or lost;
- Popular cloud storage platforms, aimed for the general public.

Personal storage devices, such as PCs and Laptops are convenient for short-term storage and data processing yet, these fall outside the scope of the DMP. However, project partners must ensure that the backup of data is regularly made through suitable networked drives.