



# CEA Commissariat à l'énergie atomique et aux énergies alternatives: Scientific Data Management Plan - CEA template (EN)

## History of the document

Versioning

Version number	Date	Description of the modification

*Recommendations:*

OPIDoR doesn't manage versioning of data management plans. It is up to you to back up the DMP, as necessary, before any significant changes are made.

You can create a copy of the plan and add a version number to the new document.

## Information about the project

**Project name\***

**Project acronym**

*Project objectives*

*Call number\**

**Funder (Europe, ANR, organism, Industrial, ...)**

**Grant agreement number**

*Project coordinator and partners*

**Contact**

*Recommendations:*

Name, First name and email

**Employer and affiliation of the contact**

**Project start date**

**Project duration**

*Additional Information*

## Data sets

**Data sets naming rules**

*Recommendations:*

Define the rules to name the dataset during the project. They should be clear and coherent. Moreover the names should NOT contain :

- special or accented characters,
- spaces,
- empty spaces

It is advisable to include dates

*List of data sets*

*Recommendations:*

Give an identifier/reference to each dataset. These identifiers can be used in the "Produits de recherche" section.

**Classification tree**

*Recommendations:*

Provide a tree structure for data classification as soon as possible. It will facilitate data access and storage during the project and also at the end for archiving.

## Data description

**Data set reference and name\***

### *Purpose and relation to the objectives of the project \**

#### *Recommendations:*

Describe the aim of the data collection and the project objectives

### **Data types\***

#### *Recommendations:*

Images, code, text,...

### **File formats\***

#### *Recommendations:*

Use rather open or widely used formats to facilitate sharing, but also stable formats for better preservation and easier archiving.  
Describe the relation between the original format and the one resulting from a conversion.

### **Re-use of existing data\***

#### *Recommendations:*

Specify the origin of the re-used data if it is a critical point for the project.  
Link between the re-used data ONLY if they are stored in an open repository.

### **Data production methods\***

#### *Recommendations:*

How were the data generated?  
Experiments, measuring instrument, observation, compilation, simulation, ...  
Specify the experimental protocols, devices, measurement conditions; The source of the re-used data; The software used (version); Formulas; Algorithms; ...

### *Storage medium*

#### *Recommendations:*

The medium should be chosen according to the desired use of the data

### *Hosting*

#### *Recommendations:*

Material and physical storage site.  
Use of a service provider  
Specific software

### *Expected size of the data\**

#### *Recommendations:*

Estimated Data volume  
To be re-evaluated during the project

### **General data policy**

#### *Recommendations:*

Policy of funders, institutions, ...  
Ex: Open Data Policy, ...

### **Data property**

#### *Recommendations:*

The contract or consortium agreement, governing the project that has generated the data, may contain measures related to the regime of such data.  
Who owns the data?  
Can they, and under what conditions, be re-used?  
What is the legal regime applicable to databases and their content, if any?

### *Data utility\**

#### *Recommendations:*

Intended target audience (general public, scientific community, private actor, internal use of the organization ...)

### *Potential for re-use\**

#### *Recommendations:*

Outlook for application or development of these data

### **Diffusion principles\***

*Recommendations:*

Rules that apply to the majority of the data produced

### **Scientific publications**

*Recommendations:*

Scientific publications, data papers, scientific reports

### **Relevant documentation**

*Recommendations:*

Any information necessary for long-term data intelligibility: codes, abbreviations, version of the reading software, ...

## **FAIR Data - Making data findable**

### **Standards and Metadata format\***

*Recommendations:*

<http://www.dcc.ac.uk/resources/metadata-standards>

<http://rd-alliance.github.io/metadata-directory/>

If no appropriate standards are available in your discipline, explain how and which metadata have been created.

### **Persistent and unique identifier\***

*Exemple de réponse:*

Ex : DOI (Digital Object Identifier)

*Recommendations:*

Durability of the link

### **Naming conventions\***

*Recommendations:*

Convention of data set naming.

### **Search keywords\***

*Recommendations:*

Will keywords be provided to optimize the re-use possibilities?

### **Version numbers\***

*Recommendations:*

Approach foreseen for clear versioning (version numbers)

## **FAIR Data - Making data accessible**

### **Data openly available\***

*Recommendations:*

Specify which data will be available for external third parties.

Specify the external level of access:

- to members of a consortium for a project, for the sole purpose of carrying out the R & D project;
- to members of an international scientific community, for research purposes,
- all people

If data cannot be made available, provide a reason.

### **Tools to read or re-use data\***

*Recommendations:*

Specify whether there is a need for a specific software or tool to read, re-use, reprocess the data and whether it will be included  
Software documentation if necessary.

### **Ways to make data available\***

*Recommendations:*

Specify how the data will be accessible: data repositories,...

### **Data repository\***

*Exemple de réponse:*

<https://zenodo.org/>

*Recommendations:*

Submission in an organization's repository, in an external repository is also possible provided that this is allowed for the data of

the project: in a disciplinary basis, etc...

BUT check carefully that this external repository is permanent and that the entity in charge of this repository does not take the ownership of the data.

Link to the repository ONLY if it is available to everyone on the internet.

#### *Access procedures\**

##### *Recommendations:*

Specify how authorized persons will have access to data in the case of access restrictions

## **FAIR Data - Making data interoperable**

### ***Standards, vocabularies, or methodologies for data and metadata\****

##### *Recommendations:*

Indicate whether data and associated metadata follow standards, vocabulary and methodologies for data interoperability

#### *Inter-disciplinary interoperability\**

##### *Recommendations:*

Specify whether the vocabulary used for all data types in your dataset allows interoperability between disciplines

If not, specify whether a match will be provided

## **FAIR Data - Increase data re-use**

#### *Data licensing\**

##### *Recommendations:*

Specify the chosen licenses to clarify the conditions for the sharing and re-use of the disseminated data and the possible economic impact of this re-use.

- Specific licenses determining the conditions, in particular the financial terms, for access to data
- Creative Commons : <https://creativecommons.org/choose/>

Open Data Commons : <http://opendatacommons.org/>

#### *Date of data release\**

##### *Recommendations:*

Or when known, specify the availability end date.

Specify whether certain data is subject to an embargo, why and for how long (e.g. publication, patents, etc.)

#### *Access to third parties\**

##### *Recommendations:*

Specify whether the data produced and / or used in the project can be used by a third party, especially after the project end.

### ***Restricted re-use : exception to the general diffusion principles\****

##### *Recommendations:*

Specify BRIEFLY the nature of the risk or the constraint justifying access restrictions (ethics, personal data, intellectual property, commercial, private data, security ...).

#### *Data quality assurance processes\**

##### *Recommendations:*

Specify the quality assurance processes used to manage the data

#### *Length of time for re-use\**

##### *Recommendations:*

Length of time that the data will be re-usable

## **Allocation of resources**

### ***Costs for making data FAIR and how to cover these costs\****

##### *Recommendations:*

Specify the costs associated with the implementation of the data management plan

Indicate how these costs will be covered

### ***Data manager responsible during the project\****

##### *Recommendations:*

Name, First name and Email address

## **Responsibilities of partners**

### *Recommendations:*

Describe, where appropriate, the responsibilities of other project partners in managing the data

### *Potential value of long term preservation\**

### *Costs of long term preservation\**

### *Recommendations:*

Be careful to properly assess the costs of long-term conservation. These may be disproportionate to the potential value of preservation. It must be demonstrated that the benefit expected from conservation is greater than the cost of preservation

## **Archiving and preservation**

### *Data at the end of the project*

### *Recommendations:*

Indicate what is planned at the end of the active phase of the project, in particular concerning the dissemination and exploitation of the results.

### *Data selection\**

### *Recommendations:*

After a risk analysis to determine the relevance of data retention (scientific, strategic, legal, heritage value), indicate, if any, the data to be kept.

Main criteria for selecting the data to be archived:

- Legal or contractual obligation to retain certain data for a specified period by making them available on request in certain cases.
- Interest in keeping all the data produced and the capacity of the computer resources to accommodate them. Data sampling to reduce the archive volume can be considered if a selection of data is sufficiently representative of the set.
- Selection of data whose characteristics (standardized format, widespread standard) will facilitate further operation without using any specific software tools
- Transparency of research and traceability. A publisher may request to link the data to the publication. We advise you to avoid depositing them on the publisher's website.
- Data which are non-reproducible or too expensive to reproduce.
- Political and heritage issues: capitalization of data.

The type of data to be archived will also be assessed according to the project and the archiving objectives.

The retention periods will be defined jointly by the contracting authority, the legal department and the referring archivist. They will be reported in a management table.

Pay attention to the cost of data archiving!

Put the data in a format recommended by the archiving platform (e.g. at CINES <http://facile.cines.fr/>) if the conservation time is longer than 5 years

### *Estimated final volume*

### *Recommendations:*

Estimation

### *Recommended preservation duration\**

### *Recommendations:*

Does not only depend on the legal and regulatory requirements but also on the recommendations made during the risk analysis

### *Long term preservation storage\**

### *Relevant paper documentation*

### *Recommendations:*

Reference documentation

## **Data security**

### *Risk analysis*

### *Recommendations:*

Specify the type of risk analysis performed if necessary (Consult your ASSI if necessary)

### *Provisions for data security\**

### *Recommendations:*

Arrangements for the security of information (including data recovery, secure storage and transfer of data, guarantee of confidentiality, integrity and traceability)

For non-publication research data, the data owner who will make them free and available to anyone, gives no guarantee on the use of the data. In particular, in the case of commercial exploitation, they shall not be liable for any accident, incident of any kind, or loss of commercial exploitation that may result from the use of the data.

***Identification of sensitive data sets Beware : this part is for internal use only***

*Recommendations:*

Protected data, data at risk, nominative data, strategic data, ...

*Data management plans (DMP) produced using the OPIDoR tool are stored on **servers outside the CEA**. This tool may not be suitable for DMPs containing confidential data.*

*Security of long term preservation\**

*Recommendations:*

Check that data are securely stored in a repository certified for both the organization and long-term storage of the data.

**Ethical aspects**

*Impact of ethical or legal issues\**

*Recommendations:*

Impact on data sharing and re-use

Is the informed consent of the data owner for personal data sharing, re-use, and long-term archiving, taken into account in questionnaires dealing with personal data?

**Other issues**

*Other data management procedures\**

*Recommendations:*

Indicate if other data management procedures are used (national / funding / sector / ...)