



D6.2 – Data management Plan

Date: 01/03/2023

Authors: Charles TOULEMONDE (EDF), Candice BOUDET (LGI), Gilles QUENEHERVE (LGI)



Table of contents

1. Introduction.....	3
1.1. Definition	3
1.2. Versions	4
2. Scope and lifecycle of the OFFERR data	5
2.1. General framework for data collection	5
2.1.1. WP1 – User Facility Network	5
2.1.2. WP2 – Design and launch call for infrastructure access.....	5
2.1.3. WP3 – Evaluation process monitoring	5
2.1.4. WP4 – European and international interactions and training.....	5
2.1.5. WP5 – Scientific monitoring, dissemination and exploitation	5
2.1.6. WP6 – Project coordination, management and communication.....	6
2.1.7. WP7 – Facilitating access to infrastructure	6
2.2. Use and Re-Use of the data	6
3. FAIR Data Management in OFFERR	6
3.1. Making data findable.....	6
3.2. Making data openly accessible.....	8
3.2.1. Data licensing.....	8
3.2.2. Datasets that could be made openly accessible in OFFERR	9
3.2.3. Datasets to remain confidential	9
3.2.4. Data storage in OFFERR.....	10
3.3. Making data interoperable & increasing re-use.....	10
3.3.1. Making data interoperable	10
3.3.2. Restrictions for re-use	10
3.3.3. Archiving and preservation.....	11
4. Human resources.....	12
5. Ethical aspects	12
5.1. GDPR.....	12



Document information

Grant Agreement	n°101060008
Project Title	eurOpean platForm For accEssing nucleaR R&d facilities
Project Acronym	OFFERR
Project Coordinator	Charles TOULEMONDE (EDF)
Project Duration	June 2022 – June 2026
Related Work Package	WP6
Related Task(s)	Task 6.2 (D6.2 / WP6)
Lead Organisation	EDF
Contributing Partner(s)	LGI, EDF
Due Date	28 February 2023
Submission Date	02/03/2023
Dissemination level	PU

History

Submission Date	Version	Submitted by	Reviewed by	Review Date
2023-01-09	V0	Gilles QUENEHERVE (LGI)	Charles TOULEMONDE (EDF)	2023-02-01
2023-03-02	V1	Gilles QUENEHERVE (LGI)	Charles TOULEMONDE (EDF)	2023-02-01

Summary

This is deliverable D6.2 Data Management Plan of the OFFERR project.

This document is structured in three distinct parts. The first part provides some basic contextualisation on data management and reminds some key definitions necessary to understand the deliverable correctly. The second part aims at explaining the scope of OFFERR Data Management Plan, including the purpose of data collection in the frame of the project, the type, format and the origin of the data to be collected and the use and re-use of the data. The third part details the various actions that OFFERR will implement to make its data findable, accessible, interoperable, and available for re-use.



According to the EU's guidelines regarding the DMP (European Commission, 2016), the document may be updated - if appropriate - during the project lifetime (in the form of deliverables).

1. Introduction

This document is the OFFERR Data Management Plan (DMP), a deliverable required by the European Commission for every project participating in the Open Research Data (ORD) Pilot. According to the European Commission, Open access (OA) refers to the practice of providing online access to scientific information that is free of charge to the end-user and reusable. "Scientific" refers to all academic disciplines. In the context of research and innovation, "scientific information" can mean:

- peer-reviewed scientific research articles (published in scholarly journals);
- research data (data underlying publications, curated data and/or raw data).

The rationale behind the Pilot project is that data management is not a goal in itself but a key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse. It follows from this that the ORD pilot aims to improve and to maximise access to and re-use of research data generated by projects while balancing openness and protection of scientific information, commercialisation and Intellectual Property Rights, privacy concerns, etc .

Data Management Plans (DMPs) are a key element of good data management. As part of making research data Findable, Accessible, Interoperable and Re-usable (FAIR), a DMP should include information on the data life cycle:

- the handling of research and organization data during and after the project,
- what data will be collected, processed, or generated,
- what methodology and standards will be applied,
- whether data will be shared/made open and how,
- how data will be curated and preserved.

The OFFERR DMP was written in reference to the task 6.2 of the Grant Agreement on data management activities.

1.1. Definition

Before explaining the consortium strategy in terms of Data Management, several terms must be defined:

- **Data:** Data refers to unstructured facts and figures, which are not organised in anyway and which provide no further information regarding patterns, context, etc. For instance, data on production, demand, results from technical tests and so on, is unstructured data.
- **Information:** For data to become information, it must be contextualized, categorized, calculated and condensed. Information thus paints a bigger picture; it is data with relevance and purpose. It may convey a trend in the environment, or perhaps indicate a pattern of sales for a given period of time.
- **Knowledge:** Knowledge is closely linked to doing and implies know-how and understanding. The knowledge possessed by every individual is a product of his/her experience and encompasses the norms by which s/he evaluates new inputs from his/her surroundings. For instance, knowledge is related to the know-how acquired in R&D projects, commercial activities or the expertise that is inherent to each partner.

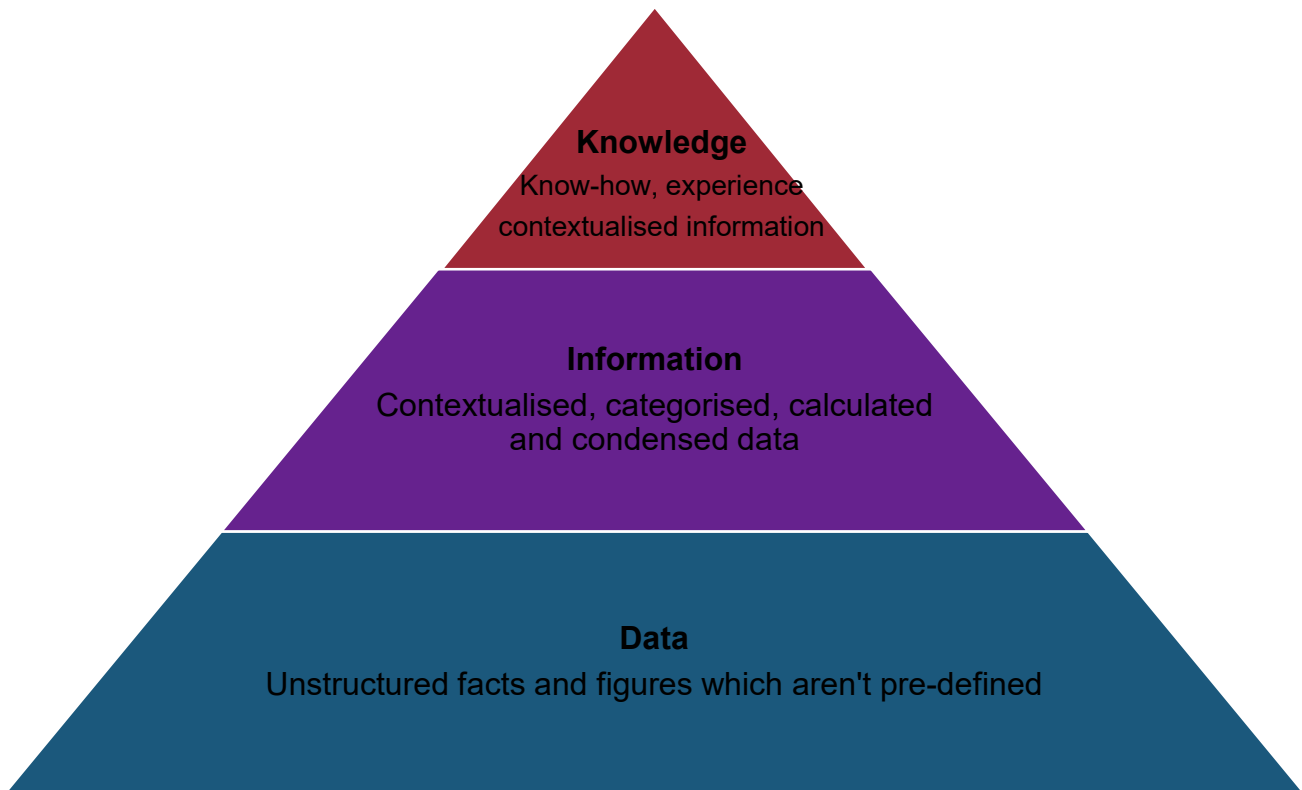


Figure 1: Knowledge Management – Definitions and hierarchy

This present DMP will mainly deal with the data will be managed in order to create information then knowledge. For that reason, we will focus our efforts on the following aspects:

- Naturally, capitalise data, the more we can, in structured database.
- Always keep track of data model to be able to understand or read data produced by the project.
- Always make a link between data produced (experimental data in the end) and metadata (context of the study, definition of the project). For that reason, we propose to save the results (data, reports and publications) of the applications with the platform used to manage the application call.

1.2. Versions

According to the EU's guidelines regarding the DMP (European Commission, 2016), the document may be updated - if appropriate - during the project lifetime (in the form of deliverables). The DMP should be updated as a minimum in time with the periodic evaluation of the project. If there are none, such an update needs to be made in time for the final review at the latest.

The DMP is intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses.



2. Scope and lifecycle of the OFFERR data

2.1. General framework for data collection

In this section, the data to be collected in each WP of the OFFERR project will be presented and described in order to define the purpose of the collection as well as to previously define the type, format and origin

2.1.1. WP1 – User Facility Network

This work package is dedicated to the setup of a network of facilities. Two different kinds of data were identified:

- The data model that describes any facility embedded in the network
 - Factsheet of these facilities are consolidated in deliverable D1.1
- The list of the facilities, following the description given by the above model
 - The list of these facilities is published in deliverable D1.2 and implemented in the EUFN catalogue and OFFERR website.
 - The lists and the catalogue will be periodically updated.

2.1.2. WP2 – Design and launch call for infrastructure access

This work package is dedicated to the drafting of the call text and to the establishment of the call management platform. The different kinds of data produced by the work-package are as following:

- The data model of the different objects in the platform: users and roles, application projects.
- The content of the platform in itself: list of users, list of applications

2.1.3. WP3 – Evaluation process monitoring

This work package is dedicated to the evaluation and to the monitoring of the applications. The different kinds of data produced by the work-package are as following :

- The list of the reviewers
- The history of the evaluation process for the different applications, which is stored and managed in the call platform.

2.1.4. WP4 – European and international interactions and training

There is no particular data produced in this work package.

2.1.5. WP5 – Scientific monitoring, dissemination and exploitation

This work package is dedicated to the dissemination of the public results of the OFFERR project. Data produced in this work package is as following:

- Report of the applications
- Raw data of the applications



Remark: in order to capitalise data produced by projects, and maximise the outcomes and exploitation of OFFERR sub-projects, we will store them at once single location on the call management platform.

2.1.6. WP6 – Project coordination, management and communication

This work package is dedicated to the management of the project. The data produced in this work package is the following:

- Project Quality Plan and Data Management Plan
- Minutes of the management meetings
- Technical and financial reportings for the European Commission

2.1.7. WP7 – Facilitating access to infrastructure

This work package is dedicated to the funding of the application projects. The data produced in this work package is the list of payments to the different infrastructure.

2.2. Use and Re-Use of the data

Data collected and generated by the consortium will be useful to the development of further activities related to other WPs within the project, to specific end-users, to some EU regulators and to other R&I projects.

More specifically, data produced by WP1, 2 and 3 is likely to be re-used in future R&I projects and will inspire the scope of the various OFFERR calls. Reports and data generated by the subprojects will be stored on a single database linked to the call platform. The SNETP website and social network will also aim at disseminating as much as possible data available for re-use beyond the OFFERR project duration.

3. FAIR Data Management in OFFERR

In compliance with the ORD pilot, every HORIZON-EURATOM project is required to draft a DMP in order to make the data Findable (1), Accessible (2), Interoperable (3) and available for Re-use (4).

3.1. Making data findable

This section will provide insight into how OFFERR intends to make it easier to find data collected or produced by the consortium. The ways to proceed in order to achieve this goal are the following:

- Documents
 - Store them in the TEAMS directory by using explicit names for subdirectories and files. Adding suffix to the file names to precise the versions of the documents.
 - As described in the Project Quality Plan, all OFFERR documents should be named as follows:
 - offer_wp<n>_T.<ii>_<shorts title>_v<j>

Where:



T is the document type: “d” for deliverables, “ms” for milestones, “minutes” for minutes, “r” for reports, “slides” for presentations and “o” for other documents

n is the work package number

ii is an incremental number for this WP and this type of document, which is delivered by the WPL (2 digits, except for deliverables, for which there is only 1 digit, as defined in Part B of Annex 1 to the GA)

j is an incremental number for the version of the document (starting at 0).

- European User Facility Network (EUFN) database
 - Store the content of the data in versioned excel files saved in a dedicated subdirectory of WP 1 subdirectory in TEAMS.
 - Consider the facility factsheet template as a prototype of the facility template
- Applications
 - Manage all the applications in the same platform, the call platform
 - Publish a document on the technical specification of the platform that describes the datamodel of the application.
 - Save the results (rawdata,reports, publications) of the applications in the call platform to be able to make a link between data (results) and metadata (context and history of the applications).
 - Due to the heterogeneity of the technical field of next coming OFFERR applications, the format used to save raw data will be .zip archives included, the more possible, text files with headers and comments.
 - The technical proposal below will be used to describe the experimental datasets.
- Servers
 - \$\$\$ define period of the incremental and full back up of the servers: TEAMS, application platform \$\$\$

Proposal to describe the experimental datasets

Each data set that will be collected, processed or generated within the project will be accompanied by a brief table description. The following detailed information sheet will be produced for every dataset to be produced/collected/curated in the project.

Name of the data setⁱ	Complete title of the data set
---	--------------------------------



Description	<ul style="list-style-type: none"> • A brief, easy to understand description of what the dataset contains and what it will be used for in the project • A list of institutions to whom the data set could be useful outside the project • Whether the dataset has been/will be used for a scientific publication (if yes, brief details about the content and journal) <p>If the dataset is collected, a brief description of its origin and how it was collected will be provided</p>
Media Type	The physical medium of the content representation, e.g., video, image, text, numerical data, n-grams, etc.
Language(s)	The language(s) of the resource content
Use & re-use	Foreseen use of the resource for which it has been produced
Size	Size of the resource with regard to a specific size unit measurement in the form of a number
Format/license	The format in which the data will be available (e.g. .xls, .csv, .txt) will be provided. The license to be used will also be provided.
Version Number	Specify the version number of the document

Table 1: Table specifying the content of an experimental dataset

If a dataset is directly collected, the origin of the data set will also be provided.

3.2. Making data openly accessible

Open accessibility of the data is the second key aspect for making data FAIR. This section will describe the type of data to be made available, its location and the procedure to obtain it.

Several degrees of accessibility are identified below, including both open access and restricted access.

3.2.1. Data licensing

Data licensing standards are used to layout the openness of data sets in concrete terms. There are many types of licenses to choose from, and this document will not cover them in depth. The table below provides a summary of common data licenses that will be considered for use in the project (based on definitions from opendefinition.org):

Name	Domain	Attribution	Share-alike*	Notes
Creative Commons CCZero (CC0)	Content, data	N	N	All rights (including those of attribution) waived



Open Data Commons Public Domain Dedication and Licence (PDDL)	Data	N	N	All rights (including those of attribution) waived
Creative Commons Attribution 4.0 (CC-BY-4.0)	Content, data	Y	N	Credit must be given, a link to the license must be provided, changes made must be indicated. If these terms are not followed, license may be revoked
Open Data Commons Open Database License (ODbL)	Data	Y	Y	Credit must be given, share-alike must be assured, data may be redistributed using DRM as long as a DRM-free version is also released

Table 2: Example of licenses

*Share-alike is the requirement that any materials created using the given dataset must be redistributed under the same license

3.2.2. Datasets that could be made openly accessible in OFFERR

Data producer	Brief description of dataset	Foreseen use & re-use	Possibility to share the data beyond the consortium
SNETP	List of research facilities of the EUFN Content of the factsheet of every facility belonging to the EUFN	Foreseen Use OFFERR applicants and reviewers Foreseen Re-use Next generation of OFFERR project	Re-use - Other EURATOM projects - SNETP Community
OFFERR applicants	Experimental dataset Raw data, technical reports	Foreseen Use OFFERR applicants Foreseen Re-use OFFERR applicants	Re-use - SNETP community - Other EURATOM projects

Table 3: Datasets that could be made openly accessible in OFFERR

3.2.3. Datasets to remain confidential

Data producer	Brief description of dataset
coordinator	List of the reviewers of the OFFERR applications A part of the database of the call platform
OFFERR applicants	Experimental datasets whenever applicants have decided to protect them for some IP reasons.



3.2.4. Data storage in OFFERR

After collection, data will be generally organised in database, Excel files and Word documents.

- Database: call platform content
- Excel files: European User Facility Network, management documents
- Word documents: management documents, deliverables, meeting minutes

3.3. Making data interoperable & increasing re-use

3.3.1. Making data interoperable

As described on 3.1, Standard vocabulary may be used on a case-by-case basis to make the data interoperable between researchers, institutions, organisations, countries, etc.

- A list of acronyms and/or abbreviations will be provided at the beginning of every report
- For surveys, standard definitions for entities such as trips, trip chains etc. will be adopted.

3.3.2. Restrictions for re-use

To be compliant with the new General Data Protection Regulation, data generated through interviews and surveys (e.g. as part of WP3) will not be re-used directly due to privacy concerns. To allow re-use and avoid loss of research data, two different techniques could be used to disseminate its data while abiding by regulations on privacy.

1) *Anonymization of data*ⁱⁱ

“Anonymization” of data means processing it with the aim of irreversibly preventing the identification of the individual to whom it relates. Data can be considered anonymised when it does not allow identification of the individuals it is related to, and no individuals can be identified from the data by any further processing of that data or by processing it together with other information which is available or likely to be available.

There are different anonymization techniques. Here are the two most relevant:

- Generalisation: generalising data means removing its specificity. For example, in the case of a table containing household income levels, with 4 figures mentioned: \$164,000, \$58,543, \$90,893, and \$232,234. One way of generalising this numbers would be to write that the values are “more than \$150,000, less than \$60,000, between \$90,000 and \$100,000, and more than \$225,000” respectively. Essentially it means taking exact figures, establishing a baseline category, and then obfuscating the data by assigning it to one of the categories in order to remove any sense of specificity from it.
- K-anonymity; A release of data is said to have the *k*-anonymity property if the information for each person contained in the release cannot be distinguished from the other individuals whose information also appear in the release. For instance, in a table composed of six attributes (Name, Age, Gender, State of Domicile, Religion and Disease), removing the name and the religion column while generalising the age is a way to effectively k-anonymise the data.



Other techniques, such as “masking” or “pseudonymisation”, which are aimed solely at removing certain identifiers, may also play a role in reducing the risk of identification. In many cases, these techniques work best when used together.

2) Pseudonymisation of data

"Pseudonymisation" of data means replacing any identifying characteristics of data with a pseudonym, or, in other words, a value which does not allow the data subject to be directly identified.

Although pseudonymisation has many uses, it should be distinguished from anonymization, as it only provides limited protection for the identity of data subjects in many cases as it still allows identification using indirect means. Where a pseudonym is used, it is possible to identify the data subject by analysing the underlying or related data.

PNO may use this technique if it decides to disseminate its databases for stakeholder analysis.

3.3.3. Archiving and preservation

It is of the outmost importance for OFFERR consortium to keep the data available for the partners after the end of the project.

As already mentioned in 3.2, to ensure medium-term preservation of the datasets, the consortium agrees to keep result datasets for a minimum duration of 5 years. Most relevant data will be publicly shared by the consortium will be shared and stored on the SNETP website and on Zenodo, which is a multi-functional open platform recognised by OpenAIRE and the European Commission.

These two types of storage have the advantage to be free of charge.



4. Human resources

Every Work Package Leader in OFFERR will be responsible for the data management within its own Work-Package.

5. Ethical aspects

5.1. GDPR

This Data Management Plan (DMP) was drafted and updated taking into account the General Data Protection Rules (GDPR) for the collection, storage and re-use of the data, in line with the following general principles.

Personal data shall be:

1. processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');
2. collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation');
3. adequate, relevant and limited to what is necessary for relation to the purposes for which they are processed ('data minimisation');
4. accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy');
5. kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation');
6. processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality')ⁱⁱⁱ.



-
- ⁱ Maria Koutsombogera & Stelios Piperidis (Athena RC). D3.1 Data Management Plan. Cracker, Cracking the language barrier. (June 2015). Retrieved from: <https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5a0203500&appld=PPGMS>
- ⁱⁱ Data Protection Commissioner. Anonymization and Pseudonymization. Retrieved from: <https://www.dataprotection.ie/docs/Anonymisation-and-pseudonymisation/1594.htm>
- ⁱⁱⁱ Intersoft Consulting. GDPR. Article 5. (2018) Retrieved from : <https://gdpr-info.eu/art-5-gdpr/>