



SUPER i

Plan and Recommendations for Ethical, Data and Risk Management

D6.3: Plan and Recommendations for Ethical, Data and Risk Management
WP 6, T 6.3

Authors: Martina Di Gallo



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 101028220.

Technical references

Project Acronym	SUPER-i
Project Title	Extended Public-Private Partnership for Investment in Smart Energy Efficiency Projects in a Social Housing context
Project Duration	September 2021 – August 2024 (36 months)

Deliverable No.	D6.3
Dissemination level*	Public
Work Package	WP 6 – Project Management
Task	T6.3 - Ethical, Data and Risk Management
Lead beneficiary	1 (CIVI)
Contributing beneficiary/ies	2 (UoY)
Due date of deliverable	31 December 2021
Actual submission date	11 January 2021

- * PU = Public
 PP = Restricted to other programme participants (including the Commission Services)
 RE = Restricted to a group specified by the consortium (including the Commission Services)
 CO = Confidential, only for members of the consortium (including the Commission Services)

v	Date	Beneficiary	Author
0.1	20/12/2021	CIVI	Martina Di Gallo
1.0	04/01/2022	Review (CIVI, UoY)	Angelo Giordano, Paola Zerilli



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 101028220.

Table of contents

Technical references	2
Table of contents	3
1.1. List of Tables	3
1.2. List of Figures.....	4
1.3. Table of Abbreviations and Acronyms	4
1. Executive summary	5
2. Introduction	6
3. Ethics	7
3.1. General Data Protection Regulation (GDPR) basics principles.....	8
3.2. Definition of sensitive data	8
3.3. Dealing with personal data	9
4. Data Management Plan	10
4.1. Background and context	10
4.2. Open Research Europe	10
4.3. Open Research Data Pilot and Compliance	11
4.4. FAIR data principles – EC Guidance.....	12
4.5. Data Management Plan Template.....	13
4.6. Data management in SUPER-i project	14
4.6.1. Data processing.....	15
4.6.2. Expected data type in SUPER-i.....	15
4.6.3. SUPER-i data management policy	16
4.6.4. Responsibilities and decision making	18
4.6.5. Data security.....	19
5. Risk assessment.....	20
6. Conclusions	23
7. References	24
8. Annexes.....	25
8.1. Dataset identification table	25

1.1. List of Tables

Table 1. SUPER-i Critical Implementation risks and mitigation actions	22
---	----



1.2. List of Figures

Figure 1. Steps to comply with the ORDP	11
---	----

1.3. Table of Abbreviations and Acronyms

DMP	Data Management Plan
DOI	Digital Object Identifier
EC	European Commission
EU	European Union
FAIR	Fundable, Accessible, Interoperable and Reusable Principle
GDPR	General Data Protection Regulation
OA	Open Access
ORDP	Open Research Data Pilot
POPD	Protection of Personal Data



1. Executive summary

The present deliverable defines the initial version of the Ethical, Data and Risk Management Plan of the SUPER-i project in month 4 of the project and is formed based on the H2020 guidelines regarding Open Research Data Pilot.

A Management Plan is an official document that describes how research data will be handled both during and after a research project. It identifies key actions and strategies to ensure that research data are of a high quality, safe, sustainable and – where possible – accessible and reusable.

The plan has been developed mainly based on two pillars:

- The Open Research Data Pilot Guidelines
- The F.A.I.R. Guidelines.

The document presents the appropriate methodology for the handling of the scientific outcomes of the project, the specification of data types that the project generates and/or collects, the standards that will be used, the process of how this data will be shared/made accessible for verification and re-use, the data preservation and maintenance processes etc.

For the characterisation of the research datasets within SUPER-i, the Data management template has been adapted and generic provisional guidelines concerning FAIR principles have been established. Additionally, even if the SUPER-i project data collection foreseen by SUPER-i does not involve sensitive personal data, provisional guidelines have been established following the requirements of the GDPR.

Moreover, D6.3 defines the steps to assess risks and identify and/or predict new ones.

The Plan and Recommendations for Ethical, Data and Risk Management is considered as a living document, which will be updated to incorporate further datasets generated/collected, and publications delivered throughout the project's duration.



2. Introduction

The present deliverable provides a plan for ethics, data management and data policies, including Open Research Data Pilot provisions, and risk management for the SUPER-i project. The management plan of a Horizon 2020 project is a living document, which shall be reviewed periodically and updated as new datasets are identified, as a methodology to ensure good scientific practice, good quality of the data used, appropriate security and confidentiality measures and freedom to operate regarding data used by the project both for its implementation and for the exploitation of its results.

The aim of the document is to establish an ongoing approach to ethics, data and risk management in the SUPER-i project. To do so, first Chapter 3 covers ethics aspects as well as privacy policy for personal data collected within the project. Chapter 4 describes the European context and background knowledge about the concept of the data management and open nature of research results within a H2020 programme. A general definition of the Data Management Plan (DMP) is followed by the guidance to comply with the Open Research Data Pilot (ORDP) and the guidance by the European Commission (EC) on making data Findable, Accessible, Interoperable, Re-usable (FAIR). In addition, Chapter 4 then highlights data management principles that apply in SUPER-i, first by identifying the necessary steps to be taken when a new research dataset is identified. The Chapter describes the main expected data types, the responsibilities and the decision-making process and the data security provisions. Finally, Chapter 5 covers risk assessment and guidance.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 101028220.

3. Ethics

SUPER-i will address and to all relevant guidelines and regulations on digital security, privacy and data protection. In addition, SUPER-i will comply with ethical principles and the applicable EU and national laws.

Some project activities involve personal data collection and/or processing, and therefore ethical aspects must be considered in this respect. It should be noted that data collection/processing foreseen by SUPER-i does not involve sensitive personal data, genetic information neither tracking nor observation of participants.

SUPER-i will comply with ethical principles and the applicable EU and national laws. The project has identified one ethic article applicable to the proposal: Protection of Personal Data (POPD), as per the common H2020 terminology. WP6 activities tackle these ethical issues and CIVI will be the main partner in handling and managing the monitored and generated data. Additionally, these issues will be tackled periodically in the Reporting Periods of the project with the potential support of an Ethics Advisor.

The European Union has developed the European Data Protection and Privacy legal framework¹ which aims at ensuring that personal data enjoys the highest standard of protection across the EU and that persons or organizations which collect and manage personal information must protect it from misuse.

The SUPER-i consortium adhere to the relevant data protection rules throughout the entire duration of the project and beyond. Personal data protection will be in-line with the EU General Data Protection Regulation (GDPR), which came into force on 25 May, 2018.

Any additional regulations at national level that do not fall under the GDRP and apply to data protection or any other sensitive information will also be taken into account. Data managed during the project will be processed only under the following preconditions, which need to be met:

- When the data subject has given her/his consent
- When the processing is necessary for the performance of or the entering into a contract
- When processing is necessary for compliance with a legal obligation
- When processing is necessary in order to protect the vital interests of the data subject.

To this end, personal data managed within SUPER-i will be anonymised and stored in a form that does not permit identification of users. Moreover, data processing will be done in respect to the purposes for which the data were collected or for which they are further processed, while ensuring appropriate protection for personal data stored for longer periods for historical, statistical, or scientific use. The majority of research data generated by the project will be made open and will be available in the Open Research Data Pilot. The Data Management Plan (DMP) details what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use and how it will be curated and preserved.

The DMP will be continuously updated until the end of the project.

¹ https://edpb.europa.eu/about-edpb/about-edpb/legal-framework_en



3.1. General Data Protection Regulation (GDPR) basics principles

When personal data is handled, it must be done under the requirements of the GDPR. The GDPR aims to enact the fundamental right of protection of personal data (Art. 8 of the EU Charter of Fundamental Rights). The GDPR is addressed at data controllers, i.e., public or private institutions and companies (not natural persons), who are either:

- established in the EU or
- established outside the EU, if:
 - processing personal data of individuals in the EU
 - to offer goods and services to individuals in EU or
 - monitoring behaviour in the EU.

The GDPR only covers any information relating to an identified or identifiable natural person (e.g. name, email, address, phone number, ID, picture, bank account, localisation data, physical characteristics). The GDPR does not cover data of companies or legal persons, nor anonymous or statistical data.

Beyond the data controllers, the GDPR outlines three further figures:

- **data processors**, who are (sub)contractors who act on behalf of the data controller
- **data subjects**, who own and provide the data (these can be everything from data controllers' and contractors' staff; to clients, visitors, experts and research participants...)
- the **Data Protection Officer**, who has an advising role.

Common personal datasets in H2020 project include:

- personal data of partners' staff (i.e. timesheets, salaries)
- personal data of event participants (i.e. attendance lists, signatures, photos, videos, usage data)
- email accounts of newsletter subscribers.

Personal data should be protected from breaches or leaks, i.e. Art 3 (12) GDPR: *"a breach of security leading to the accidental unlawful destruction, loss, alteration, unauthorized disclosure of, or access to, personal data transmitted, stored or otherwise processed."*

Breaches and leaks are always possible. As such, when handling personal data, each dataset should be scrutinised to minimise potential risk to data subjects. Mitigating measures should be defined ahead of a potential breach, to ensure rapid response. In case of high risks to data subjects, the controller shall inform them without undue delay.

3.2. Definition of sensitive data

Sensitive data is data that must be protected against unwanted disclosure. Access to sensitive data should be safeguarded. Protection of sensitive data may be required for legal or ethical reasons, for issues pertaining to personal privacy, or for proprietary considerations.

- **Personal data**: The term '*personal data*' is meant as any personal information which can be used to identify a person or an entity directly or indirectly, such a name, telephone number,



email address, place and date of birth, etc (REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, 2016).

- **Confidential data:** trade secrets, investigations, data protected by intellectual property rights Security: passwords, financial information, national safety, military information, etc.
- **Combination of different datasets** that can be combined into sensitive or personal data.

When handling and dealing with sensitive data, special attention should be given to collecting, processing, handling and storing data throughout the research process. In particular, research data that contains personal data with which a living person can, directly or indirectly, be identified has to be handled with care. This concerns both textual data and image and sound data. Examples of direct data are someone's name and address, but it could also be a photo or an interview. An indirect fact is, for example, someone's employer. For personal data fully informed consent should be given for collecting, processing and storing data.

3.3. Dealing with personal data

As abovementioned, SUPER-i data collection/processing does not involve sensitive personal data, genetic information neither tracking nor observation of participants. However, it is important to note that the European Commission ORDP data is data that is free to use, reuse and redistribute. The Pilot applies primarily to the data and metadata needed to validate results in scientific publications, as well as other data specified in the data management plan (DMP). Projects participating in the Pilot are required to deposit their research data in a research data repository and take measures to enable third parties to access, mine, exploit, reproduce and disseminate this data. However, the concept of the free use of research data within the Pilot may conflict with data protection rules if such data contain personal data.

Data protection rules always apply wherever personal data is being processed. Processing here includes any operation in connection with personal data – including collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.



4. Data Management Plan

4.1. Background and context

This chapter provides a background knowledge about the concept of the data management and open nature of research results within a H2020 programme. A general definition of the Data Management Plan (DMP) is followed by the guidance to comply with the Open Research Data Pilot (ORDP) and the guidance by the European Commission (EC) on making data Findable, Accessible, Interoperable, Re-usable (FAIR). Finally, the DMP template provided by the EC is described, as it is heavily used within the SUPER-i DMP.

A DMP is an official document that describes how research data will be handled both during and after a research project. It identifies the strategies to ensure that research data are of a high quality, safe, sustainable and, according to the type of data, accessible and reusable. More specifically, a DMP defines how the data will be collected and processed, the methodology and standards applied, who will be able to access it, where it is stored and by who, and whether and how it will be shared and preserved during and after the project ends (European Commission, 2021). A DMP is required for all projects participating in the extended Open Research Data Pilot (ORDP) and SUPER-i is one of them. The DMP will be to be updated over the course of the project every six months and whenever significant changes arise, such as (but not limited to):

- new data
- changes in consortium policies (e.g. new innovation potential, decision to file for a patent)
- changes in consortium composition and external factors (e.g. new consortium members joining or old members leaving).

The DMP should be updated as a minimum in time with the periodic evaluation/assessment of the project.

- If there are no other periodic reviews foreseen within the grant agreement, then such an update needs to be made in time for the final review at the latest.
- Furthermore, the consortium can define a timetable for review in the DMP itself (European Commission, 2021).

4.2. Open Research Europe

The EC published Open Research Europe, an open access publishing platform for the publication of research stemming from Horizon 2020 funding². All researchers funded by a Horizon 2020 grant can publish original research related to their project at no cost to them. It also enables reproducibility, transparency and impact, using an open research publishing model.

² [Open Research Europe | Open Access Publishing Platform \(europa.eu\)](https://europe.europa.eu/en/open-research-europe)



4.3. Open Research Data Pilot and Compliance

SUPER-i participates in the Pilot on Open Research Data launched by the European Commission (EC) along with the Horizon2020 programme.

From a high-level perspective, Open Access (OA) consists in providing online access to scientific information free of charge to the users to promote the reusability of the data (European Commission , Open access, 2021). In the context of the R&D actions, there are two main categories of data that OA addresses:

- the **peer-reviewed scientific papers** and
- the **research data** collected from the experiments conducted in the laboratories.

Research data refer in particular to facts or numbers, collected to be examined and considered as a basis for reasoning and discussion on project results, examples include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form.

In order to comply with the Open Research Data Pilot (ORDP), the conditions by the EC are the following:

- Develop a Data Management Plan and keep it up to date.
- Deposit the project data in a research data repository.
- Ensure third parties can freely access, mine, exploit, reproduce and disseminate the project data.
- Provide related information and identify (or provide) the tools needed to use the raw data to validate the project research.

The practical key steps to fulfil these compliance requirements are shown in the following Figure 1, adapted from the OpenAire website.

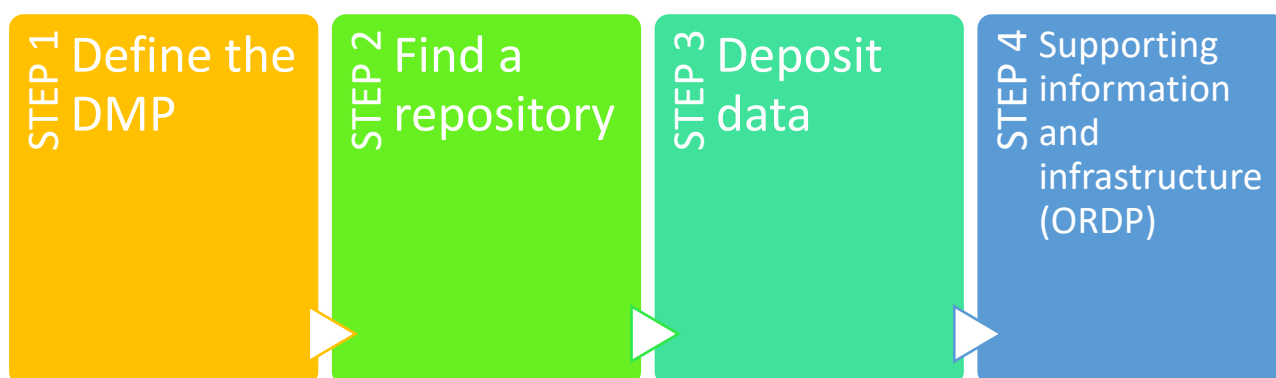


Figure 1. Steps to comply with the ORDP

Under Horizon 2020, each beneficiary must ensure open access to all **peer-reviewed scientific publications** relating to its results choosing between self-archiving (green open access) and open access publishing (gold open access).

However, the EC recognises that some research data cannot be made open and applies the principle of '*as open as possible, as closed as necessary*'. It is therefore possible to opt out of research data sharing at any stage - before or after the signature of the grant agreement - but reasons have to be



given e.g. for intellectual property rights (IPR) concerns, privacy/data protection concerns, national security concern, if it would run against the main objective of the project or for other legitimate reasons (European Commission, 2021).

The SUPER-i consortium is free to evaluate within the project any dataset which may be relevant to the ORDP. When such a dataset is identified, the compliance process should be applied.

4.4. FAIR data principles – EC Guidance

According to the Guidelines on FAIR Data Management in Horizon 2020 released by the EC (European Commission, H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020, 2016), beneficiaries must make their research data Findable, Accessible, Interoperable and Re-usable (FAIR) ensuring it is soundly managed.

The FAIR principles describe and can help understanding how to practically organise, create, store, share manage and preserve the data within H2020 project.

- **Findable**: discoverable with metadata, identifiable and locatable by means of a standard identification mechanism:
- **Accessible**: always available and obtainable; even if the data is restricted, the metadata is open;
- **Interoperable**: both syntactically parseable and semantically understandable, allowing data exchange and reuse between researchers, institutions, organisations or countries:
- **Reusable**: sufficiently described and shared with the least restrictive licences, allowing the widest reuse possible and the least cumbersome integration with other data sources.

As part of making research data FAIR, a DMP should include information on:

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

Additionally, the following questions should be answered for each of the datasets generated/processed within the project:

- What is the purpose of the data collection/generation and its relation to the objectives of the project?
- What types and formats of data will the project generate/collect?
- Will you re-use any existing data and how?
- What is the origin of the data?
- What is the expected size of the data?
- To whom might it be useful ('data utility')?

It is important to note that FAIR is a set of principles and not a standard. Following the FAIR principles does not mean that data has to be shared openly with everyone. Data can be FAIR but not open. For example, data could meet the FAIR principles, but be private or only shared under certain restrictions. Open data may not be FAIR. For example, publicly available data may lack sufficient documentation to meet the FAIR principles, such as licensing for clear reuse.



4.5. Data Management Plan Template

The EC provides a template for the Data Management Plan (European Commission, H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020, 2016), the first step towards ORDP compliance. The template consists in a set of guiding questions to be considered when addressing datasets in the project. The H2020 guidelines on FAIR data management introduce the DMP template as follows:

The template is a set of questions that you should answer with a level of detail appropriate to the project.

It is not required to provide detailed answers to all the questions in the first version of the DMP that needs to be submitted by month 6 of the project. Rather, 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 and when significant changes occur. Therefore, DMPs should have a clear version number and include a timetable for updates. As a minimum, the DMP should be updated in the context of the periodic evaluation/assessment of the project. If there are no other periodic reviews envisaged within the grant agreement, an update needs to be made in time for the final review at the latest.

1. Data summary

Provide a summary of the data addressing the following issues:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful

2. FAIR data

2.1 Making data findable, including provisions for metadata:

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- Outline naming conventions used
- Outline the approach towards search keyword
- Outline the approach for clear versioning
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

2.2 Making data openly accessible:

- Specify which data will be made openly available? If some data is kept closed provide rationale for doing so
- Specify how the data will be made available



- Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
- Specify where the data and associated metadata, documentation and code are deposited
- Specify how access will be provided in case there are any restrictions

2.3 Making data interoperable:

- Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.
- Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

2.4 Increase data re-use (through clarifying licenses):

- Specify how the data will be licenced to permit the widest reuse possible
- Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed
- Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why
 - Describe data quality assurance processes
- Specify the length of time for which the data will remain re-usable

3. Allocation of resources

Explain the allocation of resources, addressing the following issues:

- Estimate the costs for making your data FAIR. Describe how you intend to cover these costs
- Clearly identify responsibilities for data management in your project
- Describe costs and potential value of long-term preservation

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

4.6. Data management in SUPER-i project

In this section the data management procedures of SUPER-i project are described.

It is a common belief within the consortium that research and science have to be open, as there is a considerable potential for innovation and economic benefits when allowing reusing data at a larger scale. However, it should also be closed where necessary to respect confidentiality.



Therefore, selected datasets produced by the project may be published with open access, following certain guidelines as specified in the following paragraphs.

For the characterisation of the research datasets within SUPER-i the Data management template has been adapted which includes questions to be answered. Since it is still early in the project, many questions concerning FAIR principles cannot be answered, therefore we have established generic provisional guidelines concerning these principles.

4.6.1. Data processing

When handling data, it will go through three main stages: **raw**, **processed** and **analysed**. This is a crucial distinction to make, as the permissions and uses will vary between states (realistically, with access growing larger the more the data is far from the raw state). It is recommended that datasets in the SUPER-i project are reviewed with this distinction in mind (OpenAIRE, 2021).

4.6.1.1. Raw data

The original data that has been collected from a source and not yet processed or analysed. Raw data will provide the foundation for any downstream analyses. In many cases the captured or collected data may be unique and impossible to reproduce, such as time points in weather measurements and interviews. For this reason, they should be safeguarded from any possible loss. Moreover, raw data will typically be lossless - i.e. those file formats that are not compressed such as TIFF files for image data as opposed to compressed JPEG file format. Finally, in some cases, raw data may have additional information that may be specific to a brand and/or type of instrument used to capture the data.

4.6.1.2. Processed data

Data that has already undergone some kind of intervention. For instance, the data have been digitised, compressed, translated, transcribed, cleaned, validated, checked and/or anonymised.

4.6.1.3. Analysed data

Data already processed, interpreted and analysed. Analysed data can assume several representations (text, tables, graphs, etc.), in order to facilitate a better understanding and communication of the data.

4.6.2. Expected data type in SUPER-i

The SUPER-i project will collect data on energy efficiency in the social housing sector and the main sources of data to be used by the project are the activities taking place in WP1, WP2 and WP3.

4.6.2.1. On-line user data

Data collected through the SUPER-i portal will be used directly and/or by the SUPER-i Consortium for the following purposes:

- to produce statistics and reports in anonymous form (e.g. download counts, number of accesses, most visited pages, average visit duration, the country of origin of Users etc.)
- to provide information and answers about our services



- to communicate about the services offered by the SUPER-i portal by of online communication.

4.6.2.2. Off-line user data

Data collected on participants at SUPER-i regional roundtables and events will be managed via a written consent from the users, who wish to receive information from the project communication channels.

4.6.3. SUPER-i data management policy

4.6.3.1. Dataset Characterisation Process

The process described below will be followed whenever a new dataset is identified within the activities of the SUPER-i project:

- **Identify the dataset:** each dataset will have a unique ID and it will be described via the Dataset Identification Template³. Beneficiaries are asked to describe (a) the purpose of the data collection or generation and how this purpose reflects to the objectives set in the project as a whole, (b) the types and formats of data that will be generated or collected, (c) the origin of the data, (d) the expected size of the data, and also (e) whether existing data will be reused and (f) the usefulness of the described datasets.
- **List the purposes:** Describe how the dataset will be used within the project, the ownership and origin of data, who will use the data and for how long. Finally, this section requires the identification if there is data sensitivity and ORDP relevance.
- **Identify authorisations for data handling:** what authorisations are required? From whom? Are all authorisations already in place? If not, start the process to secure additional authorisations (for example, informed consent forms as used in WP1 for Regional roundtables).
- **Address personal data:** if needed, prepare the personal data protection provisions following the dataset template – Data sensitivity section. Information about Sensitive and Personal data handling is found in section 3.3.
- **Prepare data policy:** If needed (e.g. for sensitive data), outline a short informational document to explain how the data will be managed. This document will be provided to data subjects. Identify a periodicity to review the policy.
- **Select data processing and sharing techniques:** for project use, identify how the data will be handled, and to whom it will be provided.
- **Collect and distribute the data to partners for project purposes.**
- **Identify if the data set is relevant for ORDP:** if the data is relevant for the ORDP, identify the specific provisions on each principle by filling out the dataset template.
- **Maintain the dataset:** implement any long-term provisions of the data policy or the ORDP plans (e.g. remove data after a certain period).

³ Available in the SUPER-i repository. See Annex 1.



4.6.3.2. Provisional Fair Guidelines for SUPER-i datasets

As mentioned previously SUPER-i DMP is based on the EC's Guidelines on FAIR Data Management (European Commission, H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020, 2016). FAIR data management suggests that any research data collected/generated and processed through SUPER-i lifetime should follow the FAIR principles, i.e. be Findable, Accessible, Interoperable and Reusable.

The FAIR data principles provide guidance for data management by directly assisting data producers and data publishers to promote maximum use of their research, thus promoting and extending research data services, towards managing and disseminating the data outcomes of a research project. A description of the FAIR principles is presented in the following paragraphs.

Findable

Each dataset in SUPER-i that is selected for ORDP will get a globally unique and eternally persistent Digital Object Identifier (DOI), by depositing it to an open access repository like Zenodo⁴.

The dataset will also be associated with a meaningful name, description and keywords to make it easily findable for various researchers. A general ruleset may be applied to construct the name of the dataset:

“SUPER-i -data-set.datasetID.version. SUPER-i_controller”

Where:

- “SUPER-i -data-set” is the name of the dataset,
- “datasetID.version” is the unique ID and version number of the dataset, assigned by the “SUPER-i_controller”, a specific project partner, which is the owner of the data.

One or more metadata files are generated for each dataset. The metadata are identified by the same unique ID of the related dataset, with a different suffix/extension.

Accessible

This section of the DMP addresses and defines the Open Access (OA) type of data and if the data are to be kept closed, explaining the reason why.

The metadata related to a dataset selected of ORDP will be retrievable by their identifier using a standardised communications protocol. The protocol will be open, free, universally implementable and allows for an authentication and authorisation procedure, where necessary. Metadata will also be accessible, even when the data are no longer available.

The general rule in SUPER-i is that data related to publications will be made openly available. In addition, data collected through the SUPER-i portal will be used directly and/or by the SUPER-i Consortium. The Project Coordinator and the Project Scientific Coordinator will decide on a case-by-case basis which data can be released in order to avoid issues related to IP rights protection or access.

All data and metadata files will be uploaded onto a cloud storage (SUPER-i repository) and sharing facility specifically dedicated to SUPER-i project.

⁴ [Zenodo - Research. Shared.](#)



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 101028220.

Interoperable

This aspect of DMP focuses on the assessment of data interoperability by defining which data and metadata vocabularies, standards or methodologies should be followed to enable interoperability. In addition, it defines whether standard vocabulary will be used for all data types present in the dataset in order to allow inter-disciplinary interoperability.

Within the SUPER-i project, the following guidelines are to be followed for the various types of data towards ensuring interoperability:

- If the data are part of a domain with well-known open formats that are in common use, these formats should be selected.
- If the data does not fall in the previous category, an open and easily machine-readable format should be selected.

Reusable

Our plan is to make reusable as much data as possible under an Open Access licence. However, this will depend on the level of privacy and the IPR involved in the dataset or the scientific publication. The data will be available for re-use upon decision of the General Committee, in order to avoid issues related to IP rights protection or access. Within the strategy of development of DMP, the dataset that will be firstly available are those reported in publications originated from the consortium, thus intrinsically made for being reused.

4.6.4. Responsibilities and decision making

Prior to identifying the dissemination outputs of the SUPER-i project and a way to decide their Open Access, we define within this section the individual responsibilities of the SUPER-i consortium partner in regard to data management.

- Each individual partner should identify own project results which are suitable for publication.
- All datasets have to be created, managed and stored appropriately and in line with applicable legislation. Validation and registration of datasets and metadata is the responsibility of the partner that generates the data in the Work Package. Metadata constitutes an underlying definition or description of the datasets and facilitate finding and working with particular instances of data.
- The Backing up of data for sharing through Open Access repositories is the responsibility of the partner possessing the data.
- Quality control of these data is the responsibility of the relevant WP leader, supported by the Project Coordinator and the Project Scientific Coordinator.
- When datasets are to be updated, the partner that possesses the data has the responsibility of versioning control; ensuring that the latest version is available in the case of publicly available data.
- Prior to publishing data in the open domain, all partners must consult and request authorisation from the concerned partner(s), the Project Coordinator and the Project Scientific Coordinator.
- EC and project reviewers will be informed about related work done and publications provided in the project management reports.
- The Project Coordinator and the Project Scientific Coordinator should identify data collected by the project and any technical project outcomes suitable for publication.



4.6.5. Data security

Within SUPER-i, Data Collection security will be ensured by adequate (manual) procedures for collecting the data, by registering them into protected electronic systems and securely destroying the physical copies as soon as they are not required anymore. Both electronic data collection and integrity of data transfer, will be guaranteed by utilizing state-of-the-art software frameworks, libraries and protocols.

The provisions for data security and recovery should be endured by the partners running their respective databases. An encrypted storage and recovery policy will be applied by every partner involved in data storage, where all data will be anonymized as soon as possible, preferably before uploading them in any kind of electronic storage.

At the end of the project, the consortium will decide if a long-term preservation of the data is required. The selection of an adequate repository should consider the EU's OpenAIRE suggestions.

Since the datasets will be managed by the respective partners, Data Protection Officer duties will also be carried out by the Data Protection Officers of each partner separately. This policy may change in the future; if so, the updated policy will be included in the updated version of the Data Management Plan.

The project's compliance with the GDPR regulations will be managed by the respective partners' officers; this includes initial data protection impact assessments by the partners collecting personal data.



5. Risk assessment

The structure of SUPER-i and its management have been designed to routinely assess risks and identify and/or predict new ones.

The diverse quality control structures will ensure that risks are identified in advance, assessed and responses integrated into the work plan. Major risks identified so far include:

- **Project starts to exhaust funds:** A potential risk is that there will not be sufficient funds to perform all the proposed activities described in WPs 2-5 in the later stages of the project. The SUPER-i group has focused a high proportion of the resource on information management and the website, since these are crucial to the success of the project. The SUPER-i group anticipates that the demand for resources available in WPs 1-4 will exceed the budget available, given the nature of data collection. From the outset, the SUPER-i group will seek to address this issue by seeking additional funding as appropriate to augment funding available from SUPER-i. Opportunities for commercial sponsorship will be limited, however, as it is essential to maintain the actual and perceived independence of SUPER-i.
- **Components of the work plan prove suboptimal:** Evaluation of all project activities is explicitly built into the work plan of SUPER-i. Specifically, WP6 will provide on-going critical review of all project components, with formal reviews at months 12 and 36. The SAB will also review progress at the end of each periodic reporting period. Any weak areas will be identified by this mechanism, and the Work Package Leaders in consultation with the General Committee will take the appropriate remedial actions.

The Project Management Board will identify and monitor, during project implementation, internal and external risks as well as any other issues that might affect the Project progress towards its objectives, in order to carry out mitigation actions as early as possible. Risks and contingency plans have been identified in DoA. Each Partner has the responsibility to report immediately to their respective WP Leader and to the Project Coordination Team, any risky situation that may arise and may affect the project objectives or their successful completion.

The SUPER-i consortium has identified the potential risks and associated contingency plans in case any problem arises during the project execution. This contingency plan was elaborated at the beginning of the project and will be updated during the project execution if necessary.

As reported in the Grant Agreement, a set of potential risks has already been identified, as well as a list of potential risk mitigation measures.

RISK NUMBER	DESCRIPTION OF THE RISK	WP	PROPOSED RISK MITIGATION ACTION
1	Requirements will not be handed over in time or in needed quality to allow for delivery of portal in time	WP2	Move Go-Live of portal to M12 and cover stakeholder relationship through 1:1 interaction (WP5 is less effective as they could live with simple page few months longer as there are many other C&D channels available)



2	There are no resources to solve serious issues requiring IT support while the portal is in operation	WP2	In case needed, re-allocate budget from partners, Basis steering team decision according to governance structure as defined in WP6.
3	Disputes between work packages	WP6	Conflicts between work packages shall, in the first instance, be mediated by the Project Coordinator through the SC. If the SC is unable to reach consensus, the disagreement shall be referred to the General Committee (GC).
4	Beneficiary not responding to the requests of the coordinator	WP6	Provision in the Consortium Agreement in order to vote by majority the exclusion of the beneficiary from the Grant Agreement
5	Budget under/over-spending	WP6	Financial flow will be closely and detailed monitored during the project duration
6	Bad consortium communication	WP6	The long-lasting experience of the Project Coordinator will lead it to involve the partnership in a proactive way, in order to stimulate their participation and the sharing.
7	Inappropriate or insufficient development of dissemination materials	WP5	Development of a strong dissemination plan Review of all the materials previous to dissemination. Updating of materials as the project is developed. Active engagement of external agents and gathering of relevant feedback.
8	Inappropriate identification of the target groups dissemination activities	WP5	Active engagement by the project partners of interested parties and agents in the sector to define and identify the relevant target groups. Continuous monitoring of the indicators for dissemination. Adaptation of the dissemination plan as the project develops.
9	Number of participants in the events	WP5	Identify the best channels and sponsors for a good dissemination process. Take advantage of previous events for feedback about the process: in each country and for each stakeholder.



10	Lack of budget for special installations and/or technical material rental	WP5	Prepare strong media returns and targeted Comms packages to generate in kind contributions. Seek wide circle of partners and other events to combine projects and generate economies of scale
11	Risk of pandemic	WP1, WP6	All the activities of the 1st year of the project can be performed normally by remote. We would update the risk management plan, in case needed, in April 2022, if limitations caused by the pandemic will affect the best result of-actions (especially workshops, capacity building, roundtables planned in the project), proposing to organise virtual meetings and workshops.

Table 1. SUPER-i Critical Implementation risks and mitigation actions

Any change in time schedule of deliverables or in the allocated budget must be reported to the corresponding WP Leader and to the Project Coordination Team.

In case of problems or delays, the WP leader will be consulted and he/she may install task forces to take the necessary actions.

In case no resolution is reached, the General Committee will be consulted and will establish mitigation plans to reduce the impact of risk occurring.



6. Conclusions

The present deliverable (D6.3) documents the preliminary approach on the definition of the SUPER-i Ethical, Data and Risk Management Plan, by presenting the main ethical, data and risk management policy to be followed throughout SUPER-i duration, the main regulations and principles upon which the present management plan is aligned with, as well as the methodology deriving thereof.

The document presents the storage and archiving solutions and provides information on the datasets expected to be acquired and/or generated by SUPER-i consortium partners, thus providing a holistic view over the complete data management life cycle. Moreover, the document presents the datasets and public deliverables published by the time of writing this deliverable by utilising the EC's FAIR template as adjusted for the purposes of the SUPER-i project. Lastly, it should be noticed that as D6.3 is produced in the initial phase of the project, it is considered as a living document which will be further enriched with the project's future scientific outcomes/datasets and will be updated every six months.



7. References

- European Commission . (2021, December). *Open access*. Retrieved from Horizon 2020 Online Manual: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm
- European Commission. (2016). H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020.
- European Commission. (2021, December). *Data management*. Retrieved from Horizon 2020 Online Manual: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm
- European Commission. (2021, December). *Open access and Data management*. Retrieved from Horizon 2020 Online Manual: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm
- OpenAIRE. (2021, December). *Raw data, backup and versioning*. Retrieved from <https://www.openaire.eu/RAW-DATA-BACKUP-AND-VERSIONING>
- REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. (2016). Regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).



8. Annexes

8.1. Dataset identification table

SUMMARY

Name of the dataset	Provide a descriptive and short name of the dataset
Dataset description	What kind of data is this?
Work Package/Tasks	Specify within which Work Package and task (or tasks) the data is collected, processed, and/or generated.
Relation and use within the project	What specific SUPR-i objective does it address? How it will be used within the project?
Existing data re-used?	YES/NO
Expected Size of the data	Do you know roughly how large the dataset will be?
Data types and formats	Data types and formats often refer to the extension of the files in which the data is stored, but also any standard vocabularies, structures, or ontologies may be noted in this cell. Some common examples are: comma separated values (.csv) javascript object notation (.json) images (.jpeg, .png, etc.) presentations (.odp, .ppt, .pdf) models (.ods, .xlsx, .py) interviews (.txt, .odt, .pdf, .mp3) source code (e.g., programming language, build tools, deployment etc.) datasets (SQL, NoSQL, Timeseries, etc.)"

USE

Owner of the data	Who owns the data?
Origin of the data	Who collects the data? Who provides it to the partners?
Data user(s)	Who needs the data? Who uses the data? For what purpose?
Data access provisions	Are there any limitations for data use within the project?
Retention policy	How long will the data be kept within the project? Beyond the project?
Data sensitivity	"Does this data fall under the definition of sensitive data? Is it personal data? If the dataset is sensitive, fill out the Sensitive data section."
ORDP relevance	"Is this data relevant for the Open Research Data Pilot (Y/N)? If the dataset is relevant for ORDP, fill out the ORDP section."



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 101028220.

DATA SENSITIVITY

Confidentiality	Is the data confidential?
Confidentiality Protection	How is confidentiality protected? E.g. Non-disclosure agreement, Consortium Agreement etc.
General Data Protection Regulation (GDPR)	Is the data personal and fall under the GDPR?
Data Handling	How will sensitive data be collected and processed?
Data security	What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?
Data Controller	Who is the data controller? (for GDPR purposes)
Data subjects	Who owns and provides the data (for GDPR purposes?)
Consent	For personal data, how has consent from data subjects been secured?
Data Leaks	What would be the risk to data subject in case of a leak?
Withdrawal	How can personal data owners withdraw consent? How is consent withdrawal handled? How does it affect the use of data for research purposes?

OPEN RESEARCH DATA PILOT: FINDABLE

Discoverability	How is data discoverable? What metadata is used?
New metadata	How is metadata created (if no existing standard applies)?
Identifiability	Are persistent identifiers used (e.g. DOI)?
Searching	Are search keywords used? If so, how? Which are they?
Versioning	How is versioning implemented?
Naming Conventions	What naming conventions (if any) are used?

OPEN RESEARCH DATA PILOT: ACCESSIBLE

Open Access	Will the data be made openly available?
Closed Access	If the data is kept closed, provide a rationale
Access Path	How will the data be made available?



Prerequisites	What is needed to access the data (e.g. specific software)? How are prerequisites made accessible (documentation, software licence etc)?
Repository	Where will the data, metadata, documentation be deposited?
Restrictions	How will restricted access be provided?

OPEN RESEARCH DATA PILOT: INTEROPERABLE

Open Access	Will the data be made openly available?
Closed Access	If the data is kept closed, provide a rationale

OPEN RESEARCH DATA PILOT: REUSABLE

Licence	How will the data be licenced?
Timeline	When will the data be made available? Is there a need of an embargo period, and if so, why?
Third party use	Can the data be used by third parties? Can the data be used after the end of the project? If data reuse is restricted, explain why
Data Quality	How will data quality be assured?
Data Lifetime	For how long will the data remain reusable?

OPEN RESEARCH DATA PILOT: RESOURCES

Costs and Budget	How much will it cost to make this dataset FAIR? How will the costs be covered?
Responsibilities	Who will be responsible to make this dataset FAIR?
Long term plan	What are the costs and benefits of long term preservation?

ETHICAL ASPECTS

Ethical/Legal issues	Are there any ethical or legal issues that can have an impact on data sharing?
Responsibilities	Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

