D1.3 Data Management Plan

Date: July 27, 2023

Funded by
the European Union

This Project has received funding from the European Union's HORIZON Research and Innovation Programme under Grant Agreement number 101096698

Document Identification			
Status	Final	**Due Date**	31/07/2023
Version	1.0	**Submission Date**	27/07/2023

Related WP	WP1	**Document Reference**	D1.3
Related Deliverable(s)	D1.1, D1.2, D7.1	**Dissemination Level (*)**	PU
Lead Participant	FN	**Lead Author**	Ioannis Kitsos
Contributors	All partners	**Reviewers**	Hakki Aksoy
			Akshay Patil

Keywords:
data management plan, sustainable aviation, ethics, data protection policy

Document Information

List of Contributors	
Name	**Partner**
Ioannis Kitsos	FN
Edo Loenen	S&T
Elias Zea	KTH

Document History			
Version	**Date**	**Change editors**	**Changes**
0.1	20/06/2023	Ioannis Kitsos (FN)	Initial version
0.2	20/07/2023	Hakki Aksoy	Comments & Review
0.3	21/07/2023	Akshay Patil (TUDelft)	Comments & Review
0.4	24/07/2023	Edo Loenen (S&T)	Comments & Review
0.5	25/07/2023	Elias Zea (KTH)	Comments & Review
1.0	26/07/2023	Ioannis Kitsos (FN)	FINAL VERSION TO BE SUBMITTED

Quality Control		
Role	**Who (Partner short name)**	**Approval Date**
Deliverable leader	Ioannis Kitsos (FN)	26/07/2023
Quality manager	Gerardo Zampino (KTH)	27/07/2023
Project Coordinator	Ricardo Vinuesa (KTH)	27/07/2023

Table of Contents

List of Acronyms

Abbreviation / acronym	Description
DMP	Data Management Plan
DoA	Description Of Action
DPIA	Data Protection Impact Assessment
DPO	Data Protection Officer
CA	Consortium Agreement
CMD	Connected Medical Device
FAIR	Findable Accessible Interoperable Reusable
GA	Grant Agreement
GDPR	General Data Protection Regulation
HE	Horizon Europe
WP	Work Package
N/A	Not Answer
TBD	To Be Discussed

Executive Summary

This document constitutes deliverable D1.3 of the RefMap project, funded by the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No. 101096698. Its purpose is to outline the key aspects of data management and responsible research and innovation practices that will be followed throughout the project. While the information provided is preliminary and non-exhaustive, it will be refined and expanded upon as the project progresses. The Data Management Plan (DMP) highlights the expected data collection, generation, and processing activities that the consortium will carry out. It also addresses how such data and other research outputs will be handled and managed. This includes considerations of ethical and legal requirements that need to be adhered to during the research activities. It is important to note that the document acknowledges that the information presented is subject to change and will be updated as the project evolves. Other work packages and tasks within the project are expected to complement and influence the contents of this DMP.

As the project advances, more detailed concepts related to information exchange and data preservation will be developed. The consortium is fully aware of the evolving nature of the project and is committed to updating the DMP accordingly at a later stage. Any necessary changes or updates to the information presented in this document will be included in the periodic reports or during the project's lifespan. For a more comprehensive and updated version of the matters covered in this deliverable, D1.6 Data Management Plan Midterm (July 2024) and D1.7 Data Management Plan Final (January 2026) will be produced.

1 Introduction

1.1 Purpose of the document

This document, D1.3 of the project, plays a vital role as the initial version of the data management plan (DMP) for the project. Its primary objective is to outline the fundamental aspects of data management and responsible research and innovation practices that will govern the project's activities. While the information in this document is preliminary and not exhaustive, it will undergo continuous refinement and expansion as the project unfolds.

The DMP serves as a comprehensive guide that highlights the anticipated data collection, generation, and processing activities to be undertaken by the consortium. It delves into how the consortium will handle and manage the collected data and other research outputs, taking into account ethical and legal requirements that must be followed throughout the research activities. It is important to recognize that this document acknowledges the dynamic nature of the project, and therefore, the information presented within it is subject to change and updates. The contents of this DMP are expected to be complemented and influenced by other work packages and tasks within the project as they progress. As the project advances, more detailed concepts related to information exchange and data preservation will be developed and incorporated. The consortium is fully cognizant of the evolving nature of the project and is committed to updating the DMP accordingly in due course. Any necessary modifications or updates to the information in this document will be included in the periodic reports or addressed during the project's lifespan.

In essence, the purpose of this document is to establish an initial framework for data management and responsible research practices within the project. It serves as a solid foundation that will be built upon and refined to meet the evolving needs and requirements of the project as it progresses. The main purpose of this deliverable can be summarised as follows:

Description of Data Handling: The document aims to provide an overview of the general categories of data that the project expects to collect, process, and generate. It outlines how partners will handle the data during and, to some extent, after the project's completion. This includes detailing the processes used for data gathering, securing the data, and making it available. The document emphasises the adoption of the FAIR Guiding Principles for data management.

Implementation of FAIR Guiding Principles: The deliverable addresses how the project will make data "findable," "accessible," "interoperable," and "reusable", in line with the FAIR Guiding Principles. This involves ensuring unique and persistent identifiers for data, rich metadata descriptions, and indexing in suitable data

repositories. The aim is to facilitate data discovery and retrieval by humans and machines, adhering to legal and ethical guidelines for data access.

Data Management Policy: The document includes a discussion of the main elements of the data management policy that the project partners will use for all datasets generated by the project. This policy outlines the overarching guidelines and principles governing data management within the project, ensuring consistency and standardisation in handling the generated data.

Data Protection and Ethical Standards: The deliverable establishes procedures to ensure that all the project partners meet data protection and ethical standards. It introduces the purpose, core concepts, and general principles of the General Data Protection Regulation (GDPR) - Regulation (EU) 2016/679. The document proposes measures related to the data protection policy of the project and provides data management guidelines aligned with the requirements of the GDPR.

To enhance the readability of the document, we have compiled a list of useful terms:

- **Open Research Data**: Open research data refers to data made openly available for access, use, and reuse by anyone without restrictions. It promotes transparency, collaboration, and the sharing of research findings. Open research data encourages the reproducibility of research, allows for data-driven discoveries, and fosters innovation.
- **Research Data**: Research data refers to the information collected, observed, or created during the course of a research project. It can take various forms, including raw data, processed data, experimental results, survey responses, images, etc. Research data serves as evidence for scholarly investigations and can be analysed, interpreted, and used to generate new knowledge.
- **Secondary Data**: Secondary data refers to data collected by someone else for a different research purpose. Researchers can access and utilise secondary data to answer research questions or support their own investigations. Secondary data can come from sources such as previous research studies, surveys, public databases, or official records.
- **Open Access**: Open access refers to the practice of making publications, such as articles, papers, or conference proceedings, freely available to the public without paywalls or subscription barriers. Open access allows anyone to read, download, copy, distribute, and reuse the content, promoting widespread knowledge dissemination and facilitating research collaboration.

- **Metadata**: Metadata refers to descriptive information about a dataset or any other form of data. It provides essential details about the data's content, structure, format, and context. Metadata typically includes information like title, author, creation date, keywords, data sources, and data format. Metadata facilitates data discovery, understanding, and organization, enabling efficient search and retrieval of relevant data.
- **Research Data Repositories**: Research data repositories are platforms or databases specifically designed to store, manage, and provide access to research data. These repositories serve as centralized storage locations for researchers to deposit and share their data. Research data repositories often adhere to specific standards and guidelines for data management, ensuring long-term preservation, discoverability, and accessibility of research data. They may also provide features like metadata creation, version control, and data citation to enhance data usability and reproducibility.

The information provided in this document is based on the Guidelines on Data Management in Horizon Europe, Horizon Europe (HORIZON) Programme Guide (Version 2.0), EC Guidelines on FAIR Data Management in Horizon 2020, FAIR Guiding Principles for scientific data management and stewardship and the General Data Protection Regulation (GDPR).

1.2 Relation to other project work

This document is complementary to D7.1 OEI – Requirement No 1 and references D1.1 Project Management Handbook and D1.2 Dissemination and Communication Plan.

1.3 Structure of the document

This document comprises four sections. Following Section 1, Introduction, Section 2 lays out the first version of the RefMap DMP. The DMP refers to the data that will be handled by the RefMap partners, considerations regarding the "FAIR" usage of data, and touches upon the management of research outputs other than data. It also discusses data management responsibilities under the project and the allocation of resources. Section 3 presents the purpose and core concepts of the GDPR, as well as general principles of data protection and the rights of data subjects under this regulation. It also introduces proposed measures regarding the project's data

protection policy and discusses data management and measures. Section 4 summarises the document content offering some conclusions.

2 RefMap Data Management Plan

This section introduces the first version of the DMP of the RefMap project. In general, DMPs are documents outlining all aspects of the research data lifecycle. Consequently, the RefMap DMP will set out how data collected, processed and/or generated during the project should be handled during its lifecycle and will refer to the procedures relevant to making data FAIR. As indicated above, the DMP is a living document and it will be iteratively adjusted, updated, and enriched as the project progresses, taking into account, among others, the generation or use of new data, changes to the original planning, changes in data/output access provisions, or changes in consortium policies, practices or composition.

This section is divided into the following subsections:

- In subsection 2.1, the data summary is presented. Specifically, the data expected to be handled by RefMap partners in carrying out their research activities under the project work are listed.
- In subsection 2.2, matters related to data access, storage, and security during the project are discussed.
- In subsection 2.3, the FAIR Principles are presented and a preliminary identification of as opposed to on how the consortium will take them into account is made.
- In subsection 2.4 the management of research outputs other than data is discussed.
- In subsection 2.5 the data management responsibilities and allocation of resources are described.

2.1 Data Summary

This subsection provides a provisional overview of the data and information that the project partners will handle during the project. It outlines the anticipated data collection, generation, reuse, and processing activities that will take place within the research activities. It is important to note that this is an initial high-level overview based on early-stage information gathered from the partners.

During the initial phase of the project, partners were required to complete the "*Data Info Collector*" spreadsheet. This process involved documenting all the relevant data and information that would be utilised throughout the project. Partners provided

detailed information regarding the data's relationship to the project's work program, its utilisation within specific work packages and tasks, the format and type of data, and the potential end users. The collector also captured information about similar existing data, opportunities for integration and reuse, standards and metadata considerations, as well as plans for data sharing and archiving. It is worth noting that the "*DMP info collection*" is a dynamic document that will be continuously updated as the project progresses and will be maintained in the project repository. In general, the project is expected to involve the collecting, generating, and utilising various categories of research data and information.

Up until now, data provisions have been made internally to the consortium, resulting in tables stored in the project repository, which will serve as a starting point to identify the expected types of data and information that will be handled within the project. However, it is essential to acknowledge that this data is not exhaustive and will be further refined and expanded upon throughout the project's lifespan. The ongoing data collection, generation, and processing activities, along with the insights gained from the project's work packages and tasks, will contribute to a more comprehensive understanding of the data landscape.

2.2 Data storage, access and security

Following the previous subsection, this is also related to data collection, focusing on data storage, access, and security. Both subsections 2.1 and 2.2 cover the research data lifecycle.

While handling, storing and sharing data, the consortium shall consider and comply with requirements, obligations and standards set out in applicable legislation and guidelines, including – but not limited to – the GDPR. To the extent that personal data storage, sharing is involved, the fundamental data protection principles briefly outlined in the previous section should be respected.

2.2.1 Data storage, quality, and security

Data is expected to be stored by the project partner owning or providing each dataset. In case multiple partners are involved with any data set, data storage will also be provided by the partner leading or hosting the production environment used. Certain data and information will also be stored on the project's common repository provided by the project coordinator. Generally, it is anticipated that the partner owning the dataset will control access to it, and will be in charge of collecting, storing and deleting the data.

Partners handling data shall also adopt appropriate measures to ensure data integrity, quality and confidentiality. In addition, data security is imperative, and all

partners shall protect the data and information they hold by adopting the necessary security measures and mitigating any risks. Overall, ensuring data confidentiality, integrity and availability, in a balanced manner and in line with carrying out project activities is important. Therefore, measures shall be put in place, as necessary and appropriate, by the project partners to ensure that data access is restricted only to the intended audience and, to prevent intentional or accidental destruction or modification of data.

Data integrity, quality, confidentiality, and security measures recommended to be adopted when applicable by the project partners, are:

- Encryption at rest and encryption in transit methods/protocols
- Integrity file system checks
- Access controls with multi-factor authentication methods
- Data only being accessible from the organisation's cooperate network through the organisation's equipment
- Use of virtual private network ("VPN") to access data located on the organisation's servers
- Access to personal computers with password
- Definition of conditions and policies under which data, research infrastructure and related tools and applications can be used, instructing users to follow set guidelines on how to use the tools and services that handle data, and enforcement of security, trust management and acceptable use policies
- Storage and/or sharing of documents that do not contain personal or confidential data
- Storage of data on the partner's infrastructure only with appropriate access control and restrictions
- Firewalls to ensure network security
- Use of tools with appropriate security measures
- Validating input data to ensure accuracy and veracity of information regarding recorded values
- Regular backups (which will also ensure data recovery)
- Periodic recovery tests to ensure recoverability
- Appropriate measures for physical security
- Regular security checks and audit controls
- Opting for servers and services located in the EU to make sure that data handling is GDPR-compliant.

Furthermore, to ensure data integrity and quality, partners responsible for gathering information from other partners (e.g., use case providers) will communicate

frequently. In addition, confidentiality rules binding project partners as per the GA and the CA are relevant regarding data confidentiality.

2.2.2 Data availability and sharing between RefMap partners

Data availability and sharing between the project partners are expected to take place in order to carry out research activities under the project's work. Throughout the project's lifetime, all data will be available to all partners through the project's repository when possible. At this project stage there is an initial plan, but it cannot be conclusively specified which partners will have access to/use other partners' datasets and which datasets these will be. Some partners have also specified that their data will be shared through open-source platforms (e.g., GitHub) while others haven't specified yet if and by whom their data will be accessible.

2.2.3 Archiving, preservation and deletion of data

It is anticipated that data will be stored until it is clear that they will not be analysed again for project activities and/or until the project ends and the final review has been undertaken. Data will subsequently be deleted and/or discarded from the storage used during the project. Certain data and research results may be kept for a certain time and/or archived after the project's end, while taking necessary and appropriate measures to ensure their safety. Whether and which data will be treated in this manner will be further assessed as the project progresses. Datasets shared with third parties and published results will be kept after the project's end, and any public source code that will have been generated during the project by certain partners may remain open-source on such platforms with respect to the IPR strategy of the project. In addition, examples used to demonstrate the technology, e.g., in the project's website or project deliverables, are expected to remain public.

2.3 Making data FAIR

This Section gives a first overview of the FAIR principles and how the project will take them into account.

The FAIR Guiding Principles are high-level principles developed by a range of stakeholders from academia, industry, funding agencies and scholarly publishers, with the purpose of creating guidance for researchers wishing to increase the findability and, ultimately, reusability of their data. The FAIR Guiding Principles comprise four elements, which are related, but independent and separable: findability, accessibility, interoperability, and re-usability. These principles are meant to be followed "in any combination and incrementally", considering the context and

special circumstances of each case. These principles can be applied not only to data but also to non-data assets.

2.3.1 Making data Findable, including provisions for metadata

The first element of the FAIR Guiding Principles is “findability”. The following steps lead to data being findable:

1. (meta)data should be assigned a globally unique and persistent identifier
2. data should be described with rich metadata
3. metadata should include the identifier of the data it describes
4. (meta)data should be registered or indexed in a searchable resource

The project partners will consider measures related to the findability of their data, as appropriate and taking into account the specific circumstances of each case, although at this stage of the project, the concrete measures to be taken towards this goal cannot be identified with certainty. Measures that will be considered, and potentially taken, as appropriate, include assigning a persistent identifier to data and/or metadata, providing rich metadata (e.g., in terms of the type of data, timestamps, etc.), and searching keywords in the metadata to optimise the possibility for discovery and potential reuse. In addition, it is anticipated that specified naming conventions will be followed, version numbers will be provided, and metadata will be offered in a way that can be harvested and indexed. The use of trusted open-access data repositories will also be considered.

2.3.2 Making data Accessible

Once data has been found, the next step toward potential data reuse is to know how such data can be accessed. The FAIR Guiding Principles indicate that data will be “accessible” when:

- (meta)data can be accessed using a standardised communications protocol by their identifier. The protocol should meet the criteria of being open, freely available, and universally implementable. Additionally, it should support an authentication and authorization procedure, if required.
- metadata is accessible, even if the data are no longer available.

The current deliverable provides some information on the accessibility of the project's data, although this issue will be further addressed in the final version of the Data Management Plan (D1.7). As an example, certain data may be included in the technical documentation of solutions developed within the project. Such data is anticipated to be made accessible, possibly through a code repository or project deliverables. However, certain research data may not be made accessible to third

parties in its original form. Nevertheless, processed information, activities, analyses, models, and other results derived from the research are expected to be made available through public deliverables, reports, journal articles, conference proceedings, and other technical documentation related to the project solutions.

For code developed during the project, it is anticipated to be hosted on GitHub in a repository that allows reading access to any user. Some technologies developed may also be made available as open-source on GitHub. Research data supporting scientific publications may be shared with third parties at the time of publication and deposited in public repositories. When data is shared, it will typically be in a widely used format. Furthermore, relevant publications presenting project results may include metadata descriptions, providing information enabling users to access the accompanying data, if published alongside the research output.

2.3.3 Making data Interoperable

The “interoperability” principle entails that:

1. (meta)data utilise a formal, accessible, shared, and widely applicable language for knowledge representation. This enables effective exchange and interpretation of data among humans and, enables machines to read the data without requiring specialised algorithms, translators, or mappings. It eliminates the need for systems to be familiar with each other's data exchange formats
2. (meta)data employ vocabularies that adhere to the FAIR principles. This means the vocabulary used to describe the dataset is thoroughly documented and can be resolved using unique and persistent identifiers. The information within the dataset is easy to locate and accessible to those who utilise the dataset
3. (meta)data include qualified references that establish meaningful links to other data resources. These references enhance the contextual knowledge about the data by specifying relationships such as one dataset building upon another or indicating that complementary information can be found in another dataset. This principle contributes to technical interoperability by providing sufficient descriptions of these relationships between the data resources.

The project partners will seek to make data with a utility outside the project interoperable. Data held by the project partners is expected to generally use widely-known and community-endorsed vocabularies, standards, formats or methodologies, and necessary information to process the data and understand what it is by associating a publication with its metadata may be provided. Certain partners may also store data in an open data format, which can be opened with any programming language while making the format specifications.

2.3.4 Making data Reusable

To ensure that data is reusable we should follow the principles mentioned below:

1. (meta)data should be richly described with accurate and relevant attributes. This includes providing labels that allow users to determine its usefulness in specific contexts, including details such as its purpose, date, preparer, and software used.
2. (meta)data should be released with a clear and accessible data usage licence, indicating the conditions for reuse, and accompanied by detailed provenance information, including its origin, processing history, publication status, and incorporation of external data.
3. (meta)data should meet domain-relevant community standards, including using standardised organisation, sustainable file formats, common documentation templates, and a shared vocabulary, to facilitate easier reuse.

The project partners shall take a number of steps to increase the re-usability of their research data that may have a utility outside the project. Such steps will further be specified as the project progresses and may include, as necessary and appropriate, the following:

- Thoroughly documenting the provenance of data using appropriate standards;
- Including descriptions of the data in shared documents and describing formats, designs, and applications in technical documentations (e.g., methodology, fields used, purpose, etc.);
- Making available readme-files alongside the data, as well as notebooks to illustrate the use of data, instructions and guidelines
- Making available through publications results of the analysis carried out on the data, which will allow validating further the data analysis by comparing outcomes
- Publishing examples alongside technical descriptions of the solution developed through public deliverables, repositories and/or research publications.

2.4 Management of other research outputs

The project will produce digital research outputs other than data. Limited physical research outputs (printed samples) are also expected to be created. The project will also generate deliverables and other digital documents such as roadmaps, recommendations and scientific publications. Research outputs other than data will be managed to regard the FAIR Guiding Principles, as relevant and appropriate, with the aim of making research available as openly as possible.

Documents produced within the project's scope will follow standardised naming conventions, as specified in the Project Management Handbook (D1.1). In particular, deliverables shall follow a set nomenclature – namely Dx_y_TitleName_state_version – which corresponds to the project short name, the deliverable number as defined in the DoA, the name of the deliverable, the version number of the document, and optionally a suffix used to identify intermediate versions or contributions from partners to a draft version. Deliverables shall also include a "Document Identification" part, where, among others, the version number shall be indicated and, include keywords and abbreviations. The above will facilitate the findability of the project deliverables (most of which will be made openly accessible through the project website), as well as navigating through them.

Furthermore, the project partners will make (peer-reviewed) scientific publications openly available, using appropriate licenses and/or public repositories. Information regarding disseminating certain research outputs can also be found in deliverable D1.2. Disseminations and Communication Plan.

Overall, the management of research outputs other than data created during the project will take into account the FAIR Guiding Principles, while also considering the confidentiality of the information disclosed by partners during the project and ownership of outcomes stemming from project activities, the planned commercial utilisation of results, and the protection of IPRs (including patents), know-how and information related to the use of knowledge owned by a partner as a result of work carried out prior to the project.

2.5 Allocation of resources

Within the project, the project coordinator will act as Data Manager, administering the processes of the construction of the DMP and data management more broadly. The construction of the DMP, as well as data management within the remit of the project is a collaborative process, requiring input and contribution of all the project partners. To this end, the consortium has been asked to provide the information necessary to draft the first version of the DMP.

With regards to the handling of research data under the project, to the extent that a consortium partner collects and/or generates the data, that particular partner will be responsible for the proper collection, storage, processing and sharing of that data, and for ensuring that personal data is treated in accordance with the applicable legal framework, unless specified differently by data processing agreements concluded between partners in line with the applicable legal framework.

2.6 Ethics

The project partners shall carry out their project activities in accordance with the highest ethical standards and applicable EU, international and national law. Appropriate attention shall be paid to, among others, the right to privacy, the right to data protection, the right to the physical and mental integrity of the person, the principle of proportionality, and the need to ensure the protection of the environment. Any collection and processing of the personal data of data subjects shall be done in compliance with applicable EU, international and national law on data protection, particularly the GDPR. Respect for the fundamental principles of the processing of personal data enshrined in Article 5 of the GDPR shall be ensured.

3 GDPR

3.1 The purpose of the GDPR

The purpose of the GDPR is to protect fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data, while also ensuring the “free movement of personal data”. The GDPR sets out the EU regulatory framework for the processing of personal data. According to Article 3, the Regulation applies to the processing of personal data in the context of activities of an establishment of a controller or processor in the EU; to the processing of personal data of data subjects who are in the EU by a controller or processor not established in the EU, where the processing activities are related to offering goods or services to such data subjects in the EU or monitoring their behaviour as far as it takes place within the EU; as well as to personal data processing by a controller not established in the EU but in a place where EU Member State law applies on the basis of public international law.

In accordance with Article 1 of the GDPR, the Regulation lays down rules relating to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data. The Regulation protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data. The free movement of personal data within the Union shall be neither restricted nor prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data.

For a more effective understanding of the Regulation, the following definitions are considered indispensable, provided in Article 4 of the GDPR.

- “Personal data” means any information relating to an identified or identifiable natural person that any Project Partner and Policy Recipient processes during the execution of the Project.
- “Controller” means the owner of the personal data (usually the creator of the data itself) unless otherwise expressly clarified in this Policy or elsewhere in Project deliverables and/or reports.
- “Processor” means each Project Partner, unless otherwise expressly clarified in this Policy or elsewhere in Project deliverables and/or reports.
- “Consent” of the data subject means any freely given, specific, informed, unambiguous and in writing indication of the data subject's wishes by which they, by a statement or by clear affirmative action, signifies agreement to the processing of personal data relating to them.
- “Supervisory authority” means the competent Data Protection Authorities within the Project Partners’ jurisdictions.

The aim of the above definitions is to particularise and complement the definitions of Article 4 of the GDPR. Policy Recipients are advised to consult both texts in order to formulate the applicable definitions each time.

3.2 General principles of data protection and rights of the data subjects under the GDPR

Below are the key principles for processing personal data by data controllers and processors, according to Article 5 of the Regulation. These principles ensure data safety and must be respected by project partners during data processing:

- Lawfulness, transparency, and fairness: Personal data should be processed in a lawful, transparent, and fair manner. This includes obtaining consent, fulfilling contractual obligations, protecting vital interests, and pursuing the legitimate interests of the data controller.
- Purpose limitation: Personal data should only be collected for specified, explicit, and legitimate purposes. It should not be further processed in ways incompatible with those purposes.
- Data minimization: Personal data should be adequate, relevant, and limited to what is necessary for processing purposes.
- Accuracy of personal data: Personal data should be accurate and kept up to date. Measures should be taken to rectify or erase inaccurate data promptly.

- Storage limitation: Personal data should be retained in a form that allows identification of data subjects for no longer than necessary for the intended purposes of processing
- Integrity and confidentiality: Personal data should be processed securely, protecting it from unauthorised access, unlawful processing, accidental loss, destruction, or damage through appropriate technical and organisational measures.
- Accountability: Data controllers and processors are responsible for and must demonstrate compliance with the principles of personal data processing.
- Proportionality: There should be a connection between the collected data and the purpose of which it is gathered.

3.3 Data protection policy

To demonstrate compliance with data protection regulations, the organisation must implement a data protection policy that considers the nature, scope, context, and purposes of the data processing, as well as the potential risks to the rights and freedoms of individuals. The data protection policy serves as a framework of principles, rules, and guidelines that guide the organisation in maintaining ongoing compliance with data protection laws. These policies should align with the data protection principles and the rights of individuals outlined in the GDPR, and explain how they are practically implemented in relation to the organisation's data processing activities. The project partners should be adequately informed about the fundamental rules and principles of the GDPR, possess the necessary knowledge to address any data protection issues that may arise and be willing to seek guidance or assistance in handling any data protection-related matters.

3.3.1 Data protection officers

GDPR provides guidelines on appointing a data protection officer (DPO) and outlines the situations where such appointment is recommended or required. A DPO is necessary when data processing is conducted by a public authority, when the core activities involve large-scale processing with regular monitoring of data subjects, or when the core activities involve large-scale processing of special categories of data. It is highly recommended that the partners thoroughly evaluate the need for a DPO to ensure compliance with the requirements outlined in Articles 37-39 of the GDPR. For this project, DPO is the Data Manager, the project coordinator.

3.4 Data management and measures

In this section, we will discuss the collection, management, description, analysis, storage, and sharing/preservation mechanisms of personal data throughout the project.

3.4.1 Data processing principles

Data collection, management, and in general processing throughout the project; will be guided through the following principles/measures:

- Whenever possible, data should be collected and processed in an anonymized form to ensure that participant identities cannot be derived from the collected information.
- In cases where pseudonymization is necessary for specific tasks, partners must provide justification, inform the data subjects, and obtain their consent in accordance with Article 6 of the GDPR. Anonymization should be applied once data processing is complete.
- The legal basis for processing the personal data of project partners (e.g., names, contact information) for project-related purposes, such as communication and cooperation, may be the performance of the contract or another appropriate legal basis.
- For participants in the project (e.g., stakeholders, citizens involved in pilot activities), the legal basis for processing their personal data may be their consent as outlined in Article 6 of the GDPR.
- Project partners are strongly advised to avoid processing special categories of personal data defined in Article 9 or 10 of the GDPR. If processing such data is necessary for the project, partners should obtain explicit consent from the data subjects and inform them about the necessity of such processing.
- Anonymity and privacy of participants should be respected whenever possible. Personal information should be kept confidential and shared with other project partners only when necessary for specific project-related tasks. Data subjects should be informed about the possibility of such processing and provide their consent. Confidentiality and anonymity commitments made to participants must be upheld unless there are compelling reasons to do otherwise.

- Data subjects have control over their data, even though it is provided and processed for the project. If data subjects request the deletion of their data, it should be promptly carried out.
- Researchers have a responsibility to maintain the confidentiality of the collected data.
- Researchers and the project consortium must ensure the integrity of stored, processed, and published data.

3.4.2 Security of processing

The GDPR provides minimum recommendations for data security without specifying specific methods. It encourages data controllers to assess the state of the art, implementation costs, risk likelihood, and the importance of safeguarding fundamental rights and freedoms when determining which technical and organisational measures to adopt. These measures, outlined in Article 32 of the GDPR, include:

- Pseudonymization and encryption of personal data involve replacing identifying information with pseudonyms or encoding the data to prevent direct identification of individuals. Pseudonymization is a measure to ensure non-attribution to an identified or identifiable person. Under the GDPR, pseudonymized data is still considered "personal data" as it can potentially be linked to the data subject using additional information. On the other hand, anonymized data has had all personal identifiers removed, making it unable to identify the data subject and thus not classified as "personal data" according to the Regulation.
- It is important to have the capability to promptly restore the availability and access to personal data in the event of physical or technical incidents that may compromise its availability or accessibility.
- There should be a process in place for regularly testing, assessing, and evaluating the effectiveness of technical and organisational measures implemented to ensure the security of the data processing. This is necessary to verify that the measures in place are working as intended and providing adequate security.

Partners should establish policies and measures that align with data protection principles. Data protection by design involves implementing effective measures like

pseudonymization, while data privacy by default means processing only necessary personal data. The project will consider techniques like directory replacement, scrambling, masking, personalised anonymization, and blurring. The consortium will compare and implement relevant techniques based on project needs.

3.4.3 Data minimisation

Data should be minimised to what is necessary for its intended purposes. This means that:

- Data should not be held or processed unless essential and previously stated.
- Only collect and process the minimum data required to answer research questions.
- Repurposing data for other purposes is allowed only under strict restrictions.
- Further processing should be compatible with the initial purpose, without requiring a separate legal basis.
- Archiving, research, and statistical purposes may be considered compatible processing operations.
- Factors to consider for compatibility include the link between purposes, data subject expectations, nature of data, consequences for data subjects, and appropriate safeguards in both original and further processing.
- In the project, the following guidelines are provided to the partners for them to conclude whether the data minimization principle is respected:
    - When collecting personal data, consider the purpose and necessity of the data, exploring alternative options to minimise data collection.
    - Collect only the personal data that is strictly necessary to address the research question(s) and avoid gathering incompatible data.
    - Store personal data for no longer than necessary to fulfil research objectives, comply with legal requirements, and validate research outcomes.
    - De-identification techniques, such as anonymization and pseudonymization, can help reduce the risk of identification.
    - Anonymized data falls outside the scope of GDPR, while pseudonymized data remains subject to GDPR regulations.

- When obtaining consent, ensure the language allows for flexible use within research purposes, considering the potential for repurposing data in the future.
- Data processing should always align with specified and explicit purposes outlined in Article 5 of the GDPR.

3.4.4 Data breaches notification obligation

According to Article 33 of the GDPR, in case of a personal data breach, the controller must notify the supervisory authority within 72 hours unless the breach poses minimal risks. The processor should inform the controller promptly. The notification should include:

- Description of the breach, affected data subjects, and personal data records.
- Contact details of the DPO or relevant contact point.
- Assessment of the breach's likely consequences.
- Measures taken or planned to address the breach and mitigate its effects.

3.5 Data protection impact assessment

The GDPR requires performing Data Protection Impact Assessments (DPIAs) in specific situations. DPIAs help assess the risks to individuals' rights and freedoms resulting from personal data processing and determine necessary measures. The data controller is responsible for conducting DPIAs, with assistance from processors if applicable. DPIAs are required for projects posing a high risk to personal data and involving new technologies, monitoring, sensitive data, automated decision-making, or other specified conditions. The minimum features of a DPIA include describing processing operations, assessing necessity and proportionality, evaluating risks to data subjects, and proposing measures to address risks and ensure compliance. DPIAs can cover individual or similar processing operations and may assess the impact of technology products. Although not mandatory for all processing operations, conducting DPIAs periodically is advisable to maintain compliance and mitigate risks.

4 Conclusions

This document is the initial version of the Data Management Plan deliverable for the RefMap project. It provides a comprehensive overview of how project partners will handle conventional research data, software research data, and other research

outputs. The guidelines outlined in this document cover aspects of ethics and data protection. It applies to all work packages and tasks within the project, and all the partners must adhere to its content during their research activities. To ensure transparency and accessibility, project partners must ensure that the data they utilise is openly available and easily accessible to humans and machines. The data should be structured using standardised, widely accepted, and machine-readable formats. It is important to utilise shared vocabularies and ontologies, enabling the combination and integration of data from diverse sources. Furthermore, the data should be well-documented, providing comprehensive information about its origin, processing, and context. Additionally, the shared data should be accompanied by clear usage licences and permissions that allow others to understand and reuse the data for various purposes. The principles of data protection must be strictly followed, respecting the rights of data subjects. Project partners are required to implement a data protection policy and take appropriate measures for data management to align with the guidelines set forth by the GDPR. Regarding the storage and preservation of data, the project partners will retain and/or preserve the generated data for an extended period. However, the specific duration will be defined in greater detail as part of the project's sustainability strategy. An updated version will be provided in D1.6 Data Management Plan Midterm (July 2024) and D1.7 Data Management Plan Final (January 2026).