\\\

Grant Agreement number: 101147454

Project acronym: DEMOQUAS

Project title: DEsigning, Manufacturing and Operating Quantification of Uncertainties to increase Aviation Safety

Type of action: HORIZON Research and Innovation Actions

WP1 "Coordination, Administration & EU Synergies"

T1.3 "Data management and open-access establishment"

Data Management Plan [D1.2]

Delivery type:	**Report**
Lead beneficiary:	**AUTh**
Lead author(s):	**George Lyssas, Vasileios Gkoutzamanis**
Contributions:	**All**
Contractual delivery date:	**31/10/2024**
Delivery date:	**29/10/2024**
Dissemination level:	**PU - Public**

Funded by the European Union

Funded by European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.

Information Table

Project Title	DEsigning, Manufacturing and Operating Quantification of Uncertainties to increase Aviation Safety
Project Acronym	DEMOQUAS
GA n.	101147454
Project Coordinator	Aristotle University of Thessaloniki {AUTH}
Project Duration	36 months
Deliverable n.	D1.2
Deliverable title	Executive R&I action plan and project risk management
Deliverable version v.	04 (Submitted)
Deliverable description	Includes data type description, short/medium/long term strategy, provisions for data security, identification of issues on data sharing and GDPR
Dissemination level	PUBLIC
Work Package	1
Task(s)	1.3
Lead Beneficiary	AUTH
Contributing beneficiary/ies	All partners
Due date of deliverable	31/10/2024
Submission date	**29/10/2024**

History of Changes

Version	Date	Author/Contributor	Changes
01	01/08/2024	(AUTH) George Lyssas, Vasileios Gkoutzamanis	Table of Contents
02	17/10/2024	(AUTH) George Lyssas, Vasileios Gkoutzamanis	Version submitted for internal review
03	18/10/2024	(STAM) Giulia Barbagelata, Giannicola Loriga	Internal review
04	25/10/2024	AUTH	Final version

Abbreviations and acronyms

Abbreviation	Definition
DM	Data Management
GDPR	General Data Protection Regulation
WP	Work Package

Disclaimer

Table of Contents

Executive Summary

This document is the Initial version of the Data Management Plan (DMP) of the DEMOQUAS project. Its purpose is to set the stage and outline the key aspects of data management and the research and innovation practices that will be implemented throughout the project. The DMP highlights the expected data collection, generation and processing activities that the consortium will carry out throughout the lifetime of the project. It also addresses how the data and other research outputs are going to be handled and managed. This includes the ethical and legal requirements that need to be adhered to during the research activities. It is important to note that the information presented in this version of the deliverable is subject to changes and will be updated as the project evolves (follow-up deliverables D1.4, D1.6). The work packages and tasks within the project are expected to complement and influence the contents of this deliverable.

As the project advances, more detailed concepts related to information exchange and data preservation will be developed. This consortium is fully aware of the evolving nature of the project and as an extension of this deliverable and is committed to provide updates to this document in the periodic reports during the project's lifespan.

1. Introduction

1.1 Purpose of this document

This DMP is designed to be dynamic, allowing for ongoing updates and refinements as the project progresses. The DMP serves as a comprehensive guide that highlights the anticipated data collection, generation and processing activities that are going to be implemented by the consortium. It explores how the consortium will handle and manage the collected data and other research outputs, considering ethical and legal requirements that must be met throughout the research activities. It is important to recognise that this document constitutes an initial version of the project's DMP and therefore, the information presented within it is subject to change due to the dynamic nature of the project. The contents of this DMP are expected to be complemented and influenced by other work packages and tasks within the project as they progress. With the advancement of the project more detailed concepts related to information exchange and data preservation will be developed and incorporated. Any necessary modifications and updates will be implemented in the periodic reports of the DMP.

The main purpose of this deliverable is summarised as follows:

- **Description of Data Handling:** It provides an overview of the general categories of data that the project is expected to collect, process and generate. It also 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 emphasizes the adoption of the FAIR Guiding Principles for data management[1].
- **Implementation of FAIR Guiding Principles:** It 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, conforming to legal and ethical standards for data access.
- **Data Management Policy:** It includes a discussion of the main elements of the DM policy that the project partners will use for all the data used and generated from 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 for all the project partners to meet the General Data Protection Regulation (GDPR) – Regulation (EU) 2016/679

A list of useful terms is provided below for increasing the understanding of this document:

- **Open Research Data:** Open Research Data refers to data made available for access to, use and reuse by anyone without restrictions. It promotes transparency, collaboration and sharing of research findings. Open research data encourages the reproducibility of research, allows for data-driven discoveries, and fosters innovation.

[1] Wilkinson, M. D. et al. (2016), Comment: The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data, 3, 160018.

- **Research Data:** Research data refers to the information collected, observed, or made during a study activity. It can be in the form of processed or raw data, photographs, survey replies, experiment findings, etc. Scholarly studies rely on research data, which can be analysed, evaluated, and used to produce new knowledge.
- **Secondary Data:** Information gathered by another party for an alternative study goal is referred to as secondary data. Scholars can obtain and employ secondary data to address research inquiries or bolster their studies. Secondary data may be obtained from publicly accessible databases, official records, surveys, and earlier research projects.
- **Open Access:** This is the process of releasing publications—like papers, articles, and conference proceedings—to the general public without the need for a paywall or other subscription restrictions. Anyone can read, download, copy, distribute, and reuse the content thanks to open access, which encourages extensive knowledge distribution and makes collaborative study easier.
- **Metadata:** Descriptive data about a dataset or any other type of data is referred to as metadata. It gives important information on the context, structure, format, and content of the data. Information such as title, author, creation date, keywords, data sources, and data format are frequently included in metadata. Metadata makes it easier to find, comprehend, and organize data, making it possible to search for and retrieve pertinent info quickly.
- **Research Data Repositories:** These are databases or platforms created expressly to hold, handle, and make research data accessible. These repositories act as central places where researchers can store and distribute their data. Research data preservation, discoverability, and accessibility are guaranteed by research data repositories that frequently follow particular standards and criteria for data management. To improve the usability and repeatability of data, they might additionally offer capabilities like version control, metadata creation, and data citation.

The information provided in this document is based on the Guidelines on Data Management in Horizon Europe[2], Horizon Europe (HORIZON) Programme Guide (Version 2.0)[3], EC Guidelines on FAIR Data Management in Horizon 2020[4], FAIR Guiding Principles for Scientific Data Management and stewardship[5] and the General Data Protection Regulation (GDPR)[6].

1.2 Relation to other project work

This document is referencing the D1.1, entitled 'Executive R&I action plan and project risk management'.

[2] European Commission (2022), Horizon Europe (HORIZON) Programme Guide (Version 2.0), https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf, retrieved on 04-24-2024.
[3] European Commission (2021), Horizon Europe Data Management Plan Template (Version 1.0),https://enspire.science/wp-content/uploads/2021/09/Horizon-Europe-Data-Management-Plan-Template.pdf, retrieved on 04-24-2024.
[4] Science Europe (2021), Practical Guide to the International Alignment of Research Data Management, DOI: 10.5281/ZENODO.4915861.
[5] Consortium of European Social Sciences Data Archives (CESSDA) (2019), Adapt your Data Management Plan: A list of Data Management Questions based on the Expert Tour Guide on Data Management, https://static-archive.cessda.eu/content/download/4302/48656/file/TTT_DO_DMPExpertGuide_v1.2.pdf, retrieved on 04-24-2024.
[6] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 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), OJ L 119, 4 May 2016, pages 1-88.

1.3 Structure of the document

The objective of this subsection is to present readers with the structure of the document, while presenting an insight into the document.

This document is structured in 4 major chapters and four sections. Following:

- **Section 1:** Introduction
- **Section 2:** DEMOQUAS DMP referring to the data that will be handled by the project partners, considerations regarding the FAIR usage of data and focusing on the management of the research outputs other than the data. It also discusses data management responsibilities under the project and the allocation of resources.
- **Section 3:** GDPR 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 measure regarding the project's data protection policy , data management and data protection.
- **Section 4:** Conclusions summarises the document content offering some conclusions.

2. DEMOQUAS Data Management Plan

This section introduces the initial DMP for the DEMOQUAS project. It outlines the management and ensuring the adherence to FAIR principles of the collection, processing and generation of data during the project's lifecycle. The DMP will be continually revised to reflect new data, changes in project plans, access provisions and consortium policies, due to this document being a living document and can be altered throughout the project. This section is also, divided into the following sections:

- In subsection 2.1, the data summary is presented. Specifically, an overview of the data that are expected to be handled by the DEMOQUAS partners in their research activities for the project will be presented.
- In subsection 2.2, matters related to data access, storage and security are discussed.
- In subsection 2.3, the FAIR Principles are presented and how the consortium will take them into account.
- In subsection 2.4 the management of research outputs other than the data is discussed.
- In subsection 2.5 the data management responsibilities and allocation of resources are described.
- Finally, in subsection 2.6 the ethical principles of data usage in this project are described.

2.1 Data Summary

This subsection provides an initial overview of the data and information that the project partners will use and handle during the project. It outlines the anticipated data collection, generation, reuse and processing activities that will take place during the project research activities.

In this initial phase of the project the process involved documenting all the relevant data and information that will be utilized throughout the project. Partners are required to provide a detailed report regarding the data that will be utilized throughout the project. Due to the public nature of this document, a complete and detailed report regarding information about the description of the data, the relation of the data to specific WP(s), details about the existence of similar data are sensitive and cannot be described fully. The following table includes a preliminary description of data types at high level (Table 1).

Internally, the project partners were required to submit this information in a Data Collector Spreadsheet, so that this information can be used, and the partners can keep track of the data they use and produce.

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 landscape.

Table 1. Preliminary information on the data types that will be used within the project.

#	Data type/purpose of data	Origin	WP	Format extension
1	State-of-the-art reviews and open literature/databases	Open literature; open databases and libraries	2, 3, 6	*To be defined in updates of DMP (D1.4, D1.6)*
2	Datasets related to models developed and simulation	Primary data (i.e. source code) using the consortium's models; publicly available data	2, 3, 6	
3	Datasets related to design of components	Primary data (i.e. geometrical) using the consortium's models; publicly available data	2, 3, 4, 6	
4	Datasets related to test campaign, validation, verification	Primary data emerging from manufacturing and testing activities	4, 5	
5	Datasets related to safety risk assessment	Primary data using the consortium's models; publicly available data	6	
6	Questionnaire(s) and quantitative surveys	Data emerging from the target groups of DEMOQUAS	6	
7	Advisory Board feedback	Primary data; publicaly available data	All	
8	Stakeholder contact lists	Primary data	1, 7	
9	Feedback to policies	Primary data	All	
10	Scientific and general communication	Primary data emeging from dissemination/communication activities (i.e. deliverables, reports, papers)	7	

2.2 Data storage, access and security

Following the previous subsection, this is also related to data collection, focussing on data storage, access and security.

During handling, storing and sharing data, the partners of the project shall consider and comply with the requirements, obligations and standards that are described in applicable legislation and guidelines, such as but not limited to the GDPR.

2.2.1 Data storage, quality and security

Data is expected to be stored by the consortium partner that owns or provides each dataset. In the case that multiple partners are involved with any dataset the leader of this task or work package is tasked to store the data. Certain data will also be stored in the project's common repository that is provided by the Coordinating organization. Generally, it is anticipated that the owner partner of the dataset will control the access to it and will oversee collecting, storing and deleting data.

Partners handling data shall also adopt appropriate measures to ensure the data integrity, quality and confidentiality. Data security is imperative, and all partners are required to protect the data and information they hold by adopting the necessary security measures and mitigating any risks.

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

- Encryption in transit methods/protocols and storage.
- Integrity file system checks.
- Access controls with multi-factor authentication methods.
- Data access only through the organisation's cooperate network/equipment.
- Use of virtual private network (VPN) to access data located on the organisation's server.
- Access to PCs with password.
- Users to follow set guidelines on using tools and services that handle data.
- Storage and/or sharing of documents without personal or confidential information.
- Storage of data only with appropriate access control and restrictions.
- Use of Firewalls for ensuring network security.
- Validation of data before and after each use.
- Regular backups (to ensure data recovery)
- Periodic recovery tests to ensure recoverability of data.
- Measures for physical security.
- Regular security checks and audit controls.
- Opting for servers and services located in the EU to make sure the compliance with GDPR.

Furthermore, to ensure data integrity and quality partners responsible for gathering information from other partners will communicate frequently.

Confidentiality rules binding project partners as per the GA and the CA are relevant to data confidentiality.

2.2.2 Data availability and sharing

To conduct research as part of the project's work, data availability and sharing between the project partners are anticipated. Every partner will have access to all data during the project's duration via the project repository whenever it is practical. At this point in the project, a midterm plan exists, but it is unclear exactly which partners will have access to and use the datasets of other partners and which partners these will be. Certain partners have additionally indicated that their data will be made available via open-source platforms like GitHub, but others have not yet indicated whether or by whom their data will be available.

2.2.3 Archiving, preservation and deletion of data

Data will likely be kept on file until it is certain that they will not be examined for project purposes again, or until the project is over and the last review has been completed. After the project, the data will be removed from the storage and/or disposed of. After the project is finished, some data and research findings may be preserved for a predetermined amount of time and/or archived, with the required precautions taken to guarantee their security. As the initiative moves further, more evaluation will be done to determine whether and which data will be handled in this way.

After the project is over, datasets that were made available to third parties and results that were published will be preserved. In accordance with the project's intellectual property rights strategy, any public source code that selected partners created during the project may continue to be accessible on these platforms. Furthermore, it is anticipated that examples of the technology—such as those found on the project website or in project deliverables—will continue to be accessible to the public.

2.3 FAIR Data

An outline of the FAIR principles and how the project will apply them is provided in this section. In order to provide advice for research seeking to improve the findability and, eventually, reusability of their data, a number of stakeholders from academia, business, funding agencies, and scholarly publishers developed the high-level principles known as the FAIR Guiding Principles. The four interrelated, independent, and separable components that make up the FAIR Guiding Principles are findability, accessibility, interoperability, and reusability. Taking into account the unique conditions and context of every instance, these guidelines are intended to be applied "in any combination and incrementally." These guidelines can be used with both data and non-data items.

2.3.1 Make data Findable, including provision for metadata

"Findability" is the first of the FAIR Guiding Principles. The steps listed below make data accessible:

1. (Meta)data ought to be given a persistent, globally unique identification.
2. Rich metadata should be used to describe the data.
3. The identifier of the data that the metadata describes should be included.

4. A searchable resource should register or index (meta)data[7].

The project partners have taken adequate steps to ensure that their data is easily found while taking into account the unique circumstances of each individual case. The following actions are taken into consideration: giving data and/or metadata a persistent identifier; offering rich metadata (such as timestamps, data types, etc.); and looking for keywords in the metadata to maximize the likelihood of discovery and possible repurposing. Furthermore, version numbers are supplied, specific naming rules are adhered to, and metadata is presented in a manner that facilitates harvesting and indexing. Trusted open-access repositories are being used[8],[9],[10].

2.3.2 Make data accessible

Finding out how to obtain data is the next step towards perhaps reusing it once it has been located. According to the FAIR Guiding Principles, information is considered "accessible" when:

- It can be retrieved by its identification using a standardized communications protocol for (meta)data. The protocol must satisfy the following requirements:
  - it must be open,
  - publicly accessible,
  - implementable everywhere.

  If necessary, it should also provide an authorization and authentication process.
- Metadata is accessible, even if the data is no longer available.

For instance, specific information might be incorporated in the technical documentation of the project's developed solutions. It is expected that such data will be made available, perhaps via project deliverables or a code repository. On the other hand, some study data might not be available to outside parties in its original format. However, it is anticipated that processed data, investigations, analyses, models, and other research outcomes will be made public through deliverables, reports, journal papers, conference proceedings, and other technical materials pertaining to the project's resolutions.

It is expected that code created during the project will be stored on GitHub in a repository that is open to anybody with reading access. Certain technologies that are developed might also be posted as open source on GitHub. At the time of publication, research data that supports scientific articles may be placed in public repositories and shared with other parties. In most cases, data that is exchanged is in a format that is commonly utilized. Additionally, if published alongside the study output, pertinent articles reporting project outcomes may include metadata descriptions, giving users access to the corresponding data.

[7] GO FAIR, F1: (Meta) data are assigned globally unique and persistent identifiers, F1: (Meta) data are assigned globally unique and persistent identifiers, https://www.go-fair.org/fair-principles/f1-meta-data-assigned-globally-unique-persistent-identifiers/, retrieved on 04-24-2024.
[8] GO FAIR, F2: Data are described with rich metadata, https://www.go-fair.org/fair-principles/f2data-described-rich-metadata/, retrieved on 04-24-2024.
[9] GO FAIR, F4: (Meta)data are registered or indexed in a searchable resource, https://www.go-fair.org/fair-principles/f4-metadata-registered-indexed-searchable-resource/, retrieved on 03-01-2023.
[10] GO FAIR, FAIR Principles, https://www.go-fair.org/fair-principles/, retrieved on 04-24-2024.

2.3.3 Make data Interoperable

According to the "Interoperability" concept[11], (Meta)data must:

1. Use a formal, open, common, and broadly applicable language for knowledge representation. As a result, data may be interpreted and exchanged between people more effectively, and robots can read the data without the need for specialized algorithms, interpreters, or mappings. It removes the requirement that systems understand each other's data interchange formats[12].
2. The (meta)data use vocabulary that follow the FAIR guidelines. This indicates that the dataset's vocabulary is well-documented and resolvable with the use of distinct, long-lasting identifiers. To anyone who uses the dataset, the information is easily located and available[13].
3. Qualified references that provide significant connections to additional data resources are included in (meta)data. By describing relationships between the datasets—for example, how one dataset builds upon another or suggests that complementary information might be discovered in another dataset—these references improve the contextual knowledge about the data[14]. Because this principle gives adequate representations of these linkages between the data resources, it aids in technical interoperability.

The project partners want to achieve interoperability of data with a utility that is not part of the initiative. Data that is kept by the project partners should typically make use of commonly used and accepted vocabularies, standards, formats, or procedures. It may also be required to supply the information needed to analyse the data and identify what it is by linking a publication to its metadata. While creating the format standards, some partners might also store data in an open format that can be opened with any programming language.

2.3.4 Make data reusable

The guidelines listed below should be adhered to guarantee that data is reusable:

1. Accurate and pertinent characteristics should be used to provide a rich description for (meta)data. This entails labelling information such as the purpose, date, preparer, and program utilized so that users can assess its applicability in particular scenarios[15].
2. When (meta)data is released, it should come with a transparent and easily readable data usage license that outlines the terms of reuse. It should also be accompanied by

[11] GO FAIR, I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation, https://www.go-fair.org/fair-principles/i1-metadata-use-formal-accessible-shared-broadly-applicable-language-knowledge-representation/, retrieved on 04-24-2024.
[12] GO FAIR, I2: (Meta)data use vocabularies that follow the FAIR principles, https://www.go-fair.org/fair-principles/i2-metadata-use-vocabularies-follow-fair-principles/, retrieved on 04-24-2024.
[13] GO FAIR, R1: (Meta)data are richly described with a plurality of accurate and relevant attributes, https://www.go-fair.org/fair-principles/r1-metadata-richly-described-plurality-accurate-relevant-attributes/, retrieved on 04-24-2024.
[14] Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021 establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, and repealing Regulations (EU) No 1290/2013 and (EU) No 1291/2013, OJ L 170, 12.5.2021, pages 1–68.
[15] European Union Agency for Cybersecurity, ENISA proposes Best Practices and Techniques for Pseudonymisation, https://www.enisa.europa.eu/news/enisa-news/enisa-proposes-best-practices-and-techniques-for-pseudonymisation, retrieved on 04-24-2024.

comprehensive provenance information that details the data's origin, processing history, publication status, and external data incorporation.

3. To enable simpler reuse, (meta)data should adhere to domain-relevant community standards, which include employing standardized organization, sustainable file formats, standardised documentation templates, and a common terminology[1].

In order to improve the reusability of their research data that might be used outside of the study, the project partners will take a number of actions. As the project develops, these actions will be further detailed and may include the following, if necessary and suitable:

- Carefully recording the data's provenance in accordance with the relevant requirements.
- Including data descriptions in shared documents and providing technical documentation that describes formats, designs, and applications (e.g., purpose, technique, fields used, etc.).
- Including readme files and notebooks that demonstrate how to use the data, along with guidelines and instructions.
- Publishing the findings of the data analysis, which will enable additional validation of the data analysis through outcome comparisons.
- Disseminating technical details of the solution together with examples in academic articles, public deliverables, and/or repositories.

2.4 Management of other research outputs

This project will produce digital research outputs other than data. Limited physical research outputs 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 as openly as possible. Documents produced within the project will follow standardised naming conventions and deliverables shall follow a set nomenclature.

Furthermore, the project partners will make peer-reviewed scientific publications openly available, using appropriate licences and/or public repositories. Overall, the FAIR Guiding Principles will be considered in the management of research outputs other than data created during the project, along with the confidentiality of information disclosed by partners during the project, ownership of outcomes resulting from project activities, the intended commercial use of results, and the protection of intellectual property rights (including patents), know-how, and information pertaining to the use of knowledge owned by a partner as a result of work completed prior to the project.

2.5 Allocation of resources

The project coordinator will oversee the procedures for building the DMP and, more generally, data management within the project as the Data Manager. All project partners must provide feedback and contribute to the collaborative process of building the DMP and managing data

under the project's purview. To accomplish this, the consortium has been requested to supply the data required in order to create the DMP's second iteration.

In terms of how research data is handled under the project, if a consortium partner generates or collects data, that partner will oversee making sure that personal data is handled in accordance with the project's guidelines and that the data is properly collected, stored, processed, and shared.

2.6 Ethics

The project partners will conduct their business in compliance with all applicable national, international, and EU laws as well as the highest ethical standards. The right to privacy, the right to data protection, the right to one's own bodily and mental integrity, the principle of proportionality, and the necessity of ensuring environmental preservation will all receive the appropriate attention[16]. Any gathering and handling of data subjects' personal information must adhere to all applicable national, international, and EU data protection laws, especially the GDPR. Article 5 of the GDPR enshrines essential standards for the processing of personal data, which must be respected.

3. GDPR

3.1 The purpose of the GDPR

The General Data Protection Regulation (GDPR) aims to safeguard individuals' fundamental rights and freedoms, namely their right to data protection, while simultaneously permitting the "free movement of personal data." The EU's legal framework for handling personal data is outlined in the GDPR. Article 3 states that the Regulation applies to the processing of personal data in connection with the activities of a controller or processor's establishment in the EU; to the processing of personal data of individuals residing in the EU by a controller or processor that is not established in the EU, when the processing activities are connected to providing such data subjects with goods or services inside the EU or tracking their behaviour within the EU; additionally, when a controller not based in the EU processes personal data in a location where public international law applies and EU Member State law is applicable.

Article 1 of the GDPR states that the Regulation establishes guidelines for the free flow of personal data as well as guidelines for the protection of natural persons while processing personal data. The Regulation safeguards natural persons' fundamental liberties and rights, especially their right to the privacy of their personal information. For reasons related to the protection of natural persons regarding the processing of personal data, there shall be neither restrictions nor prohibitions on the free movement of personal data throughout the Union.

The definitions listed below are deemed essential for a better comprehension of the Regulation and are included in Article 4 of the GDPR.

- Any information pertaining to an identified or identifiable natural person that is processed by a project partner or policy recipient while the project is being carried out is referred to as "personal data."
- ‘Controller’ means the entity which, alone or jointly with others, determines the purposes and means of the processing of personal data
- ‘Processor’ means the entity which processes personal data on behalf of the controller.
- "Supervisory authority" refers to the appropriate Data Protection Authorities within the Project Partners' jurisdictions.
- "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.

The definitions provided above are meant to supplement and particularize those found in Article 4 of the GDPR. Policy Recipients are recommended to refer to both texts in order to determine the appropriate definitions on each occasion.

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

The main guidelines for how data controllers and processors should handle personal data, as per Article 5 of the Regulation, are listed below. Project partners must adhere to these guidelines in order to maintain data security when processing data:

- Lawfulness, openness, and justice: Processing of personal data must be done so in a way that is fair, transparent, and compliant with the law. Consent must be obtained, contracts must be followed, vital interests must be safeguarded, and the legitimate interests of the data controller must be pursued.
- Purpose limitation: Only clear, defined, and justifiable purposes should be given for the collection of personal data. It shouldn't be processed in a way that is not appropriate for those goals.
- Minimization of data: Personal information should only be as much as is required for processing and should be sufficient and pertinent.
- Accuracy of personal information: Personal information must be current and accurate. Prompt action should be taken to correct or remove erroneous data.
- Limitation on storage: Personal information should only be kept for as long as is required to fulfil the intended processing purposes in a format that permits data subjects to be identified.
- Integrity and confidentiality[16]: Through the use of suitable organizational and technical safeguards, personal data should be processed securely to prevent unauthorized access, improper processing, unintentional loss, destruction, or damage.
- Accountability: Data processors and controllers bear accountability for adhering to the rules governing the processing of personal data.
- Proportionality: There need to be a relationship between the information gathered and its intended use.

3.3 Data Protection Policy

The lead organization (AUTH) will create a data protection strategy that takes into account the nature, scope, context, and goals of data processing in addition to any risks to individuals' rights and freedoms in order to demonstrate compliance with data protection laws. The organization is guided in ensuring continuous compliance with data protection legislation by the framework of principles, rules, and guidelines provided by the data protection policy. These policies should describe how the organization's data processing activities realistically apply the GDPR's individual rights and data protection principles. They should also be in line with those principles. The project partners should be well-versed in the core regulations and guidelines of the General Data Protection Regulation (GDPR), have the know-how to handle any potential data protection issues, and be prepared to ask for advice or help when needed.

[16] GitHub, Confidentiality, Integrity, Availability (CIA), https://github.com/0xsanny/guides/blob/master/src/pages/security/confidentiality-integrity-availability/index.md, retrieved on 04-24-2024.

3.3.1 Data Protection Officer

The GDPR describes the circumstances in which designating a data protection officer (DPO) is advised or necessary, as well as the procedures for doing so. A DPO is required when processing data on behalf of a public authority, when processing data on a big scale while regularly monitoring data subjects, or when processing data on a large scale involving particular categories of data. To guarantee adherence to the regulations delineated in Articles 37–39 of the GDPR, it is strongly advised that the partners do a comprehensive assessment of the necessity for a DPO. DPO is the project coordinator and data manager for this particular project.

3.4 Data Management and Measures

This section will cover the methods used to gather, manage, describe, analyse, store, share, and preserve personal data over the course of the project.

3.4.1 Data Processing Principles

The project's data collection, administration, and overall processing will be guided by the following principles and/or measures:

- Data should be gathered and processed in an anonymised way whenever possible to guarantee that participant names cannot be deduced from the information gathered.
- Under Article 6 of the GDPR, partners are required to give reasons, notify data subjects, and secure consent when pseudonymization is required for particular tasks. After data processing is finished, anonymization should be implemented.
- The legal justification for using project partners' personal data (names, contact details, etc.) for project-related activities such collaboration and communication, contract fulfilment, or any other relevant legal justification.
- According to Article 6 of the GDPR, consent from project participants (such as stakeholders and citizens taking part in public activities) may be the legal basis for processing their personal data. Project partners are strongly encouraged not to process any of the special categories of personal data listed in GDPR Articles 9 or 10. Partners must have express consent from data subjects and notify them of the need for processing their data if it is required for the project.
- Participants' privacy and anonymity should always be protected. Only when necessary for certain project-related duties should personal information be shared with other project partners; otherwise, it should be kept private. The possibility of such processing should be disclosed to data subjects, and they should give their consent.
- Even though their data is supplied and processed for the project, data subjects retain control over it. Data subjects' requests to have their data deleted should be honoured right away.
- It is the duty of researchers to protect the privacy of the information gathered.
- The integrity of data that is processed, published, and stored must be guaranteed by researchers and the project consortium.

3.4.2 Security of processing

The GDPR does not outline particular techniques for data security; instead, it offers minimal suggestions. When choosing the organizational and technical measures to take, it encourages data controllers to consider the state of the art, implementation costs, risk likelihood, and the significance of protecting basic rights and freedoms. These actions are listed in GDPR Article 32 and consist of:

- Pseudonymization and encryption of personal data entail the use of pseudonyms in place of identifying information or data encoding to prevent individual identification. One way to make sure that no identifying or identifiable person is mentioned is to use pseudonymization. Pseudonymized data is nonetheless regarded as "personal data" under the GDPR because it may be possible to connect it to the data subject with supplementary information. However, all personal identifiers have been removed from anonymized data, rendering it impossible to identify the data subject and disqualifying it from being considered "personal data" under the Regulation.
- In the case of physical or technical incidents that may jeopardize the availability or accessibility of personal data, it is critical to have the capacity to quickly restore that data's availability and access.
- To guarantee the security of the data processing, a procedure should be in place for routinely testing, reviewing, and assessing the efficacy of organizational and technical safeguards. This is required to confirm that the security mechanisms in place are sufficient and functioning as intended.

It is recommended that partners have policies and procedures that conform to data protection principles. While data privacy by default refers to processing only necessary personal data, data protection by design entails putting in place practical protections like pseudonymization. The research will take into account methods such as blurring, masking, customized anonymization, scrambling, and directory replacement. Depending on the requirements of the project, the consortium will evaluate and apply pertinent techniques.

3.4.3 Data Minimisation

Less data than is required for its intended uses should be used. This implies that:

- Unless explicitly indicated as necessary, data should not be processed or kept.
- Only gather and handle the minimal amount of information needed to address research questions.
- Reusing data for other purposes is only permitted under very specific guidelines.
- Any additional processing ought to be consistent with the original intent and not necessitate the creation of a new legal foundation.
- Processing processes for statistical analysis, research, and archiving may be deemed compatible.
- The relationship between goals, data subject expectations, data type, data subject repercussions, and suitable protections in both original and subsequent processing are factors to take into account for compatibility.
    - The partners in the project are given the following instructions so they can determine whether the data minimization principle is being followed: Take into account the need

and purpose of the data while gathering personal information and look into other ways to minimize the amount of information collected.

- Avoid collecting inconsistent data by just collecting the personal information that is absolutely required to answer the study question or questions.
- Preserve personal information for as long as is required to meet legal obligations, accomplish research goals, and verify study findings.
- The risk of identification can be decreased with the use of de-identification techniques like anonymization and pseudonymization.
- While pseudonymized data is still subject to GDPR requirements, anonymized data is not covered by them.
- When seeking consent, make sure the wording takes future data repurposing into account and permits flexible use within research aims.
- The GDPR's Article 5 specifies and explicitly states the reasons for which data processing must always be carried out.

3.4.4 Data breaches notification obligation

Unless the breach poses minor risks, the controller is required by GDPR Article 33 to notify the supervisory authority of a personal data breach within 72 hours. The controller should be instantly notified by the processor.

- A description of how the breach impacted the data subjects and their personal records should be included in the notification.
- An evaluation of the possible repercussions of the breach.
- The DPO's or the pertinent contact point's contact information.
- Actions made or scheduled to rectify the breach and lessen its consequences.

3.5 Data protection impact assessment

In certain circumstances, the GDPR mandates the completion of Data Protection Impact Assessments (DPIAs). DPIAs assist in identifying the appropriate actions and evaluating the dangers that processing personal data poses to an individual's rights and freedoms. DPIAs must be carried out by the data controller, with assistance from processors as needed. Projects utilizing new technology, monitoring, sensitive data, automated decision-making, or other conditions that pose a high risk to personal data must have DPIAs. A DPIA must at a minimum describe processing activities, evaluate risks to data subjects, determine need and proportionality, and provide remedies to mitigate risks and maintain compliance. DPIAs can evaluate the effects of technological items and encompass specific or comparable processing activities. While not required for every processing transaction, frequently carrying out DPIAs is advisable to maintain compliance and mitigate risks.

4. Conclusions

This document is the first draft of the deliverable for the DEMOQUAS project's data management plan (DMP). Two additional versions will follow: D1.4 as the intermediate DMP version and D1.6 as the final DMP version. The first DMP provides a preliminary rundown of how the project's collaborators will manage software research data, conventional research data, and other research outputs. This document also contains the first set of guidelines that address data protection and ethics. They will be used in all project work packages and tasks, and all partners need to follow its guidelines when doing research. Project partners must make sure that the public data they use is freely available and simple for both people and machines to use, in order to guarantee openness and accessibility. It is recommended that the data be organized using commonly used, machine-readable formats that are defined. Using ontologies and common vocabularies is crucial for combining and integrating data from many sources. The data must also be thoroughly documented, offering detailed details about its source, processing, and context. Clear use licenses and rights that enable people to comprehend and utilize the given data for a variety of purposes should also be included with it. Respecting the rights of data subjects, rigorous adherence to data protection principles is required. Project partners must put in place a data protection policy and take the necessary steps to manage data in accordance with the GDPR's rules. In terms of data storage and preservation, the generated data will be retained and/or preserved for a considerable amount of time by the partners.

Finally, it is noted that the follow-up two versions of the DMP will further emphasize at:

- Data type summary.
- Data protection strategy.
- Data flows from one WP to another.