SEAMLESS

Project title	SEAMLESS: Safe and Autonomous: Multimodal Library of European Shortsea and inland Solutions		
Project acronym	SEAMLESS		
Project number	101096923		
Project start date	01/10/2023	Duration	48 months

D1.3 – DATA MANAGEMENT PLAN

Due date	30/06/2023	Delivery date	03/07/2023
Work package	WP1		
Responsible Author(s)	Børge Rokseth (NTNU)		
Contributor(s)	Alexandros Rammos (NTUA), Vassileios Podimatas (NTUA), Odd Erik Mørkrid (SO)		
Reviewer(s)	Håvard Nordahl (SO), Konstantinos Louzis (NTUA)		
Version	1.0		
Dissemination level	Public		

VERSION AND AMENDMENTS HISTORY

Version	Date	Created/Amended by	Changes
0.1	26/06/2023	Børge Rokseth (NTNU)	Initial draft
0.8	27/06/2023	Alexandros Rammos (NTUA), Odd Erik Mørkrid (SO)	Comments and changes to initial draft
0.9	28/06/2023	Vassileios Podimatas (NTUA)	Inserted Introduction and Executive Summary
1.0	30/06/2023	Konstantinos Louzis (NTUA)	Review Document

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This deliverable constitutes the first version of the Data Management Plan (DMP) for the SEAMLESS project. The DMP will follow an iterative approach and will be regularly updated throughout the project's duration. The DMP describes the types of datasets that are expected to be used, the datasets that are expected to be generated during the project (incl. a general description of metadata), as well as procedures for data sharing, archiving, and preservation. It also provides preliminary information on allocation of resources, data security, and ethics. To ensure compliance, the project will assign a dedicated Legal, Ethical, and Security Issues Manager (LSEIM), or data steward, responsible for assessing the project's adherence to ethical requirements in accordance with applicable national and European legislation.

1 INTRODUCTION

1.1 PURPOSE OF THE DOCUMENT

Effective management of data is a crucial aspect of any project, by outlining the lifecycle of data, covering its collection, processing, and generation. The SEAMLESS DMP adheres to the EU guidelines on FAIR data management, ensuring that research data is findable, accessible, interoperable, and reusable. The SEAMLESS DMP shall follow an iterative approach with periodical updates during the project's implementation period.

1.2 INTENDED READERSHIP

The deliverable is open to the public and in this context the intended audience is whomever is likely to be interested in the SEAMLESS research data management.

1.3 DOCUMENT STRUCTURE

A summary of the document's structure is as follows: Section 2 provides a summary of the datasets that shall be collected, analysed, and generated in the context of the SEAMLESS project. Section 3 outlines the general and guiding principles of the SEAMLESS DMP, where FAIR data objectives. Section 4 lists other research outputs that may be generated during the project. Section 5 provides a brief description of the allocation of resources for effective data management throughout the project's implementation period. Section 6 gives a brief outline of the approach that will be followed and implemented to ensure data protection and security. Section 7 is dedicated to the ethical background for SEAMLESS DMP, and Section 8 concludes with a reference to other procedures that may be used for data management in the project.

2 DATA SUMMARY

The project will collect and generate data by means of:

- Questionnaires and survey forms distributed to stakeholders.

- Cargo handling system sensors.

- Sensors aboard autonomous vessels to be used in the demonstrations.

- Retrieving existing data sets regarding waterborne safety incidents.

- Collecting particulars of existing inland and coastal cargo vessels.

- Collecting cargo flow traffic.

- Retrieve existing data regarding emission produced by waterborne transportation.

- Retrieve existing data regarding security incidents on coastal and inland waterway shipping.

- Through reports, presentations, scientific publications and technical reports.

- Collecting contact information for relevant stakeholders and informants.

Data will be generated in various formats and datasets, all of which will be accessible using common software allowing easy access and long-term validity during and after the project. Specifically, expected collected and generated datasets include the following:

- Evaluation of the transferability use cases from different stakeholders.

- Data sets regarding the use cases, relating to the SEAMLESS key building blocks.

- Emissions from existing waterborne transportation as well as for trucks (for comparison).

- Overview of relevant safety and security incidents for coastal and inland waterway shipping.

- Cargo flow traffic and data related to the cargo items for transferability evaluation and EU-wide impact analysis.

- Operational sensor data from autonomous vessels and cargo handling systems.

In general, survey data will be collected and stored through software such as Microsoft Forms. Processed data from surveys may be stored in Microsoft Access databases, Microsoft Word documents or Microsoft Excel files. Other data such as emission data and sensor data will be stored in Microsoft Excel files or similar formats.

The size of the collected data will be in the range that is manageable without special considerations (i.e. in the GB order of magnitude).

Data may be useful outside the project for recreating published results.

3 FAIR DATA

3.1 MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA

The project will utilize Open Research Europe and OpenAire in order to make project data, metadata and other research results and public dissemination fully open, freely available and findable. For scientific publications we will prioritize the data sharing platform of the Publisher, as it will allow associating the data repository with a Digital Object Identifier.

3.2 MAKING DATA ACCESSIBLE

Through Open Research Europe and OpenAire both data and metadata will be freely available. Whenever possible, data will be made openly available to the community by deposition in trusted, free-of-charge data-sharing platforms offered by partners (e.g., MS Teams-NTUA and NTNU open research data), and on the project's website. Datasets stored in the data-sharing platforms offered by partners will, when applicable be issued DOI through the archives. Collaboration and sharing within the project are achieved through Microsoft 365 collaboration tools like Teams and Outlook.

Data sets classified as restricted will not be further transferred, shared or published. Confidentiality is regulated by the consortium agreement. In order to ensure general and publishable results, the project will seek to keep the number of restricted data sets at a minimum.

The Consortium Agreement, signed by all members of the consortium, specifies the confidentiality obligations related to dissemination activities and confidential information. The Consortium Agreement also contains an attachment specifying the intellectual property and data brought into the project. Intellectual property generated by the project is also regulated by the Consortium Agreement.

In terms of software licenses for algorithms, components and modules to be utilized, the consortium shows preference for the business-friendly Apache Software License (ASL), which allows the redistribution of the program's source code in any form (compiled binary or plain text). ASL is used by a wide array of Open-Source Software projects such as the Apache web server, and the Lesser General Public License (LGPL) software licenses.

A README-file will provide details regarding software or procedures for access for each data set.

3.3 MAKING DATA INTEROPERABLE

Whenever applicable, the data will include qualified reference to other data. Specifically, scientific publications and other public disseminations presenting project results will refer to datasets whenever they are based on collected or generated data in the project as references to public repositories or earlier publications.

3.4 INCREASE DATA RE-USE

Examples of metadata can be data owner, file name, version, file type, classification (unrestricted/restricted), project code, acronym and timestamp. In cases where not all metadata are defined by the file name, additional metadata will be documented as text in a separate document (ReadMe-file), such as details regarding software or procedures for access for each data set.

Routines for high quality research in the project includes external/internal evaluation of data and reports, as well as peer review of research papers.

The provenance of data will be documented in readme-files or in the event that data is reported in a scientific publication or public dissemination, the provenance of the data will be documented through the publication/ dissemination.

The data will be available under appropriate open licences for reuse after the end of the project.

4 OTHER RESEARCH OUTPUTS

The project aims to publish any software generated as open-source software according to appropriate licences such as ASL. Scientific publications will be available through publisher´s channels and indexed in google scholar, Scopus and similar databases. Other research outputs such as public disseminations in the form of reports, will be available through the project´s website.

5 ALLOCATION OF RESOURCES

The role of Legal, Ethical, and Security Issues Manager (LSEIM), or data steward, has been assigned to the author of this document. The responsibility will be fulfilled in collaboration with the project management. The data steward has been allocated 3PM for this task. For the duration of the project, partners will cover curation and storage costs through offering their own infrastructure. However, a cost-predictive model will be developed as part of the business plans (WP6) considering digital curation and preservation as promoters of data policies in the long-term.

6 DATA SECURITY

For the duration of the project, partners will use their own infrastructure for data storage. In general, it is the policy of the project that project data should be stored safely with automatic backup. Collaboration and sharing within the project are achieved through Microsoft 365 collaboration tools like Teams and Outlook.

Data will be made available long-term by deposition in trusted, free-of-charge data-sharing platforms offered by partners (e.g., MS Teams-NTUA and NTNU open research data), and on the project's website.

Data shared or stored through any of the Microsoft 365 application, such as Teams or Sharepoint, benefit from Microsoft Defend, which provides protection against threats that arrive in email, links (URLS), attachments, or collaboration tools like SharePoint, Teams, and Outlook.

7 ETHICS

Collection of personal data during specific activities such as workshops, or through surveys of stakeholders, will be carried out in compliance with international & national legislation but also the European Charter of Fundamental Rights and several of its principles relevant to this project (Art. 3, 7, 8, GDPR regulation & Regulation 2016/67).

The project will not collect data from animals. There may be ethical issues related to data collection, generation and utilization of data for training and deployment of AI. *SEAMLESS will ensure ethical AI deployment based on 6 pillars:*

1. Human agency and oversight: AI systems will support human autonomy and decision-making – the project will be based on explainability and SEAMLESS will allow users to make informed autonomous decisions regarding the AI systems;

2. Privacy and data governance: SEAMLESS autonomous systems will guarantee trustworthiness and data privacy;

3. Transparency: All data sets and processes associated with AI decisions will be well communicated and appropriately documented (Sect.1.2.6);

4. Fairness, diversity and non-discrimination: Although the project does not refer to personal data for AI, SEAMLESS will ensure to avoid unfair bias with respect to the developed interfaces;

5. Societal and environmental well-being: The impact of the developed AI system will be significant for the EU transport industry ecosystem and the environment and this will be validated in numerous complementary experiments

6. Compliance with the "do no harm" principle for activities carried out during course of the project as well as for the expected life cycle impact of the project.

8   OTHER ISSUES

The project will not make use of other national/funder/sectorial/departmental procedures for data management.