AIRSHIP

D1.1 Data Management Plan
WP1

Disclaimer

The content of this deliverable reflects only the author's view. The European Commission is not responsible for any use that may be made of the information it contains.

Document Information

Grant Agreement / Proposal ID	101096487
Project Title	AUTONOMOUS FLYING SHIPS FOR INTER-ISLAND AND INLAND WATERS TRANSPORT
Project Acronym	AIRSHIP
Project Coordinator	Universidad Politécnica de Madrid
Project starting date (duration)	1st January 2023 (48 months)
Related Work Package	WP1
Related Task(s)	T1.3
Lead Organisation	UPM
Contributing Partner(s)	TAU
Due Date	30.06.2023
Submission Date	30.06.2023.
Dissemination level	Confidential

Revision History

Revision	Revision Date	Name	Position
1	28/06/2023	Jussi Aaltonen (TAU)	Research manager
2			

TABLE OF CONTENTS

Abbreviations and Acronyms

Acronym	Description
WP	Work Package
CA	Consortium Agreement
CFD	Computational Fluid Dynamics
GDPR	General Data Protection Regulation
FAIR	Findable, Accessible, Interoperable, and Re-usable
SC	Steering Committee
IPR	Intellectual Property Rights

Table 1. Abbreviation and Acronyms

EXECUTIVE SUMMARY

This document is the first version of the Data Management Plan (DMP) which will evolve during the execution of the project to support the collection, curation, preservation and access of the data produced by AIRSHIP.

This DMP describes the data management life cycle for all datasets to be collected, processed, or generated during the AIRSHIP project. It covers:

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

Reviews, updating and adjustments of the DMP will take place during the execution of the project at meetings/workshops organized in the project (months 12-48). The DMP will be updated as required.

This document is prepared according to the requirements for FAIR data (making data findable, accessible, interoperable, and re-usable) under the ORDP provisions, thus following the principle follows the principle "as open as possible, as closed as necessary".

The largest volume of data generated by the project will be scientific data obtained during the computer simulations, design and prototype development work, and laboratory studies. It is intended that these data will form the basis for a series of scientific publications by AIRSHIP consortium partners, but it is intended to make them available as soon as possible, and in any case by the end of the project (M48) at the latest.

1   Objectives of the Data Management Plan

The purpose of this DMP is to support the management, by the project partners, of the data by the project partners which will be collected, generated and/or processed during the execution of the AIRSHIP project, and enhance the re-use, availability, and survivability of these data.

2   Data Summary

2.1   Types of data produced

There are three main categories of data which will be collected.

A.   First is the data relevant to the AIRSHIP vehicle design and performance, which range from stakeholder requirement evaluations, simple conceptual descriptions, general drawings to elaborated blueprints of all the parts and their test and performance reports from simulations to real hardware tests in relevant environments. Also included in this category is the AIRSHIP digital twin, data obtained from Computational Fluid Dynamics (CFD) simulations, wind tunnel and water channel tests. This will be considered as data type "A".

B.   Second is the data on the business cases and evaluation of the European macro-economic potential for possible commercial exploitation by the technology to be developed in AIRSHIP. This will be considered as data type "B".

C.   Third is the data type "C", recorded both in the set of deliverable documents and in published and unpublished papers, on the scientific interpretation and results from the project.

2.2   Re-use of existing data

Where relevant, existing data on AIRSHIP components and sensors to be used will be compiled and used in the design/development process. Similarly, for software, existing products will be used where appropriate as the basis and starting point for development of on-board and post-processing systems. Where possible and appropriate, open-source solutions will be used: thus the Robot Operating System (ROS) based on the open-source LINUX operating system will be used.

3 FAIR Data

All data handling will be GDPR and FAIR compliant.

3.1 Making data findable, including provisions for metadata

3.1.1 Data set reference and name

All data files produced in the project will include in their names the name of the project (AIRSHIP) and a short information on the type of data (e.g., "AIRSHIP_camera_tests_Madrid_15052023") or, in cases where this may be impractical, will be held within folders that themselves follow this convention.

3.1.2 Version numbers

Version numbering is of importance to documents which are updated and to software developed by the project. In both cases version numbering will be simple and clear.

For post-processing applications (data conversion, database access, data analysis, and visualisation) software, the project uses git, SourceTree and bitbucket for version control.

3.1.3 Standards and metadata

When compatible with the themes of the INSPIRE Directive, the data and metadata added in the project will follow the format recommended by the INSPIRE Directive.

The quality of the data will be discussed between partners, who will decide which data are of sufficiently high quality to be used for publication of the research done in the project. Only these data will be kept in an open data repository.

Metadata will be created following the model of the INSPIRE Directive (they will offer the same types of information). As the provisions in the INSPIRE Directive are compulsory for EU members, this will ensure a high degree of compatibility with other data sets in the EU.

3.2 Making data openly accessible

3.2.1 Policies for access and sharing

An important issue is to monitor the progress of the novel technology development, carefully distinguishing among confidential data (like trade secrets e.g., confidential know-how) and Open Research Data in accordance with the HE directives, and if relevant seek patent protection to proprietary technology and methods, if such new discoveries evolve through the research in the project. The responsibility to initiate the patenting will be upon the respective WP leader participants, with ownership proportion agreed within the actively participating research team members.

3.2.2 Data sharing

In addition to the various scientific methodologies and concepts AIRSHIP will develop technologies that will be suitable for commercial exploitation in the future. This will require very careful management of Intellectual Property Rights (IPR). The principles governing IPR are outlined in the CA.

As a general rule, the data representing syntheses of the knowledge from previously published research will be open data from the moment they are put in the repository. The unpublished data (produced in the project and older unpublished data provided by partners) will be confidential for a 3-year period for the purpose of ensuring the novelty of the data used in scientific articles published by the partners. The unpublished technological data will remain confidential over a period that is decided by their owners.

For components of the project that are sufficiently unique and innovative, the developing WP shall give early warning to the steering committee about the confidentiality relevant to this component to all consortium members, not to disclose details until protection is in place. Consideration will be given to patent protection before publication of details of these components.

All the metadata and data of all types will be stored in a repository, accessible from the AIRSHIP project website. The publicly available Open Research Data will be available by any simple web browser. The search, browsing, displaying, and downloading the data will be available freely, without any registration.

The confidential data will be available only for the authorized personnel.

In accordance with ORDP requirements, each beneficiary will ensure open access (free of charge, online access for any user) to all peer-reviewed scientific publications relating to its results.

In particular, all beneficiaries will:

(a) as soon as possible and at the latest on publication, deposit a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication in a repository for scientific publications;
Where practicable, the beneficiary will also deposit at the same time the research data needed to validate the results presented in the deposited scientific publications (if these data are derived from the pilot surveys, they will be stored and made available as described in Section 3.2.4 below).

(b) ensure open access to the deposited publication — via the repository — at the latest:
   i. on publication, if an electronic version is available for free via the publisher, or
   ii. within six months of publication in any other case.

(c) ensure open access via the AIRSHIP repository (website) to the bibliographic metadata that identify the deposited publication. The bibliographic metadata will be in a standard format and must include all of the following:
   • the terms ["European Union (EU)" and "Horizon Europe"];
   • the name of the action, acronym and grant number;
   • the publication date, and length of embargo period if applicable, and
   • a persistent identifier.

3.2.3 Methods or software tools needed to access the data

Full descriptions of data analysis and visualisation tools which will be used in the preparation of scientific publications will be included in those publications and they will be listed in the metadata related to each publication.

3.2.4 Depositing data, metadata, documentation and code

All the metadata and data of all types will be stored in a repository, accessible from the AIRSHIP project website. Type "A" data of large volumes generated by simulation for the purpose of training AI systems of the AIRSHIP software will be held in offline storage with multi-site backup for security, and made available to users on request. Examples of these data will be put to the online accessible storage, to represent what kinds of data are available upon request. These data will be carefully selected by the consortium to have good quality, contain relevant information, but do not contain any confidential, or ethically questionable information.

Otherwise, all master copies of data/metadata will be held by the AIRSHIP consortium with additional copies held by relevant participants (in multiple locations) to ensure long-term data security.

Security of data will be addressed as an integral part of the task of developing online and offline data storage methods. Security of data is described in Section 5 below.

3.2.5 Access conditions

Scientific data will be made available openly. Licensing is not required beyond the standard EC requirements for acknowledgment of EC funding of the project.

Licensing of software will be decided individually by the software developers in line with their own standard software licensing policies. Open source access conditions do not require the identity of the person accessing the data to be disclosed. There is therefore no need to ascertain this identity. However, if software licensing terms of individual consortium members require this, their own standard procedures will be followed.

Type "A" data that cannot be patented (or would be too expensive to patent) should by default be considered "trade secrets" unless agreed by all the owners of the data and/or the SC to be eligible to be placed in the public domain.

3.3 Making data interoperable and re-usable

Metadata will be created following the model of the INSPIRE Directive (they will offer the same types of information). As the provisions in the INSPIRE Directive are compulsory for EU members, this will ensure a high degree of compatibility with other data sets in the EU.

3.3.1 Interoperability of project data

The data are held in open-source ROS bags and can be easily exported in CSV format which is readable by a very wide range of standard software applications.

3.3.2 Increase data re-use (through clarifying licences)

Access and licensing conditions are as described in Section 3.2.5 above. Data will be freely available, without licensing restrictions beyond a requirement to acknowledge its source as the AIRSHIP project and the standard statement on funding by the EC.

3.3.3 When will the data be made available for re-use?

The data will be made available for reuse not later than the end of the project. By agreement with consortium partners it could be made available earlier. Data will not be made available until it has been thoroughly validated for correctness and consistency.

3.3.4 Restrictions on data use by third parties

Some of the technical data on hardware and software will be defined as Trade Secrets and its use restricted to beneficiaries. The scope of such data will be kept as limited as possible.

3.3.5 Data quality assurance processes

No formal processes are defined for data quality assurance, but validation will be carried out during the construction and population of the databases. This will include such processes as outlier (rogue point) detection and elimination, and other statistical verification of data sets.

4 Allocation of resources

4.1 Costs for making data FAIR

No additional costs for making data FAIR have been identified. The processes required are an integral part of the project itself.

For third parties who request mission data which (because of its volume) requires the use of special media, it will be expected that they will supply their own media for transcription of the data.

4.2 Responsibility for data management

The project coordinator is ultimately responsible for data management, but this responsibility will be automatically devolved to the participant managing each phase of the project (design, development, testing, computer simulations).

4.3 Long term preservation

As the Open Research Data will be kept in multiple copies by relevant project participants, the archiving and preservation will follow the general procedures used by the project participants. The partners will assess the archiving and preservation procedures and decide whether special procedures are needed for the data produced in the project and also for how long the data are going to be kept in the repository.

The online open research data will be available for the public for minimum of five years after the end of the project. As a contingency plan, all the data, and the responsibility with them will be transferred to a permanent organization in the AIRSHIP consortium.

5 Data security

Responsibility for security of design and development data rests with the relevant participants, all of whom follow standard industry practice.

For software under development, version control systems (git / SourceTree / bitbucket) also provide a high level of security. Additionally, the post-processing software developers use normal grandfather / father / son backup systems with backups held in separate locations.

The physical security of AIRSHIP data may be summarised as:
- Secure backups of data before any processing
- Storage for most efficient transfer on external hard disks
- Audit trail to be maintained – log of database operations
- Need to establish a standard "chain of custody" for data security
- Data locking: read-only access to any copies of the primary database

For long-term security, all data will be held centrally by AIRSHIP with at least two further copies at different safe locations.

6 Ethical aspects

There are no ethical or legal issues which have an impact on sharing of data acquire during the project. For technical design data, the only legal issues are as defined above, related to possible patenting or definition as Trade Secrets of some parts of the design of hardware and software components.

The project does not handle or store personal data.

7 Other issues

The AIRSHIP project does not make use of other national, funder, sectorial, or departmental procedures for data management.