D1.2 -

SELF-ASSESSMENT & DATA MANAGEMENT PLAN V1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 883302.

Deliverable Information

Work Package: WP1

Deliverable Number: D1.2

Date of Issue: 28/022021

Version Number: 1.0

Nature of Deliverable: ORDP

Dissemination Level: Public

Author(s): CENTRIC (Responsible),

Keywords: data management plan

Abstract: This deliverable is the first iteration of the ISOLA self-assessment and data management plan. This document sets forth the principles for managing research data within the project and provides an initial overview of the datasets that are used, produced or generated for the purposes of carrying out the ISOLA project. This deliverable will be updated prior to any period reporting, in D1.4 due at M24 and at the culmination of the project.

Document History			
Date	**Version**	**Stage – remarks**	**Contributors**
13/01/2021	0.1	ToC	Helen Gibson (CENTRIC)
25/01/2021	0.2	First draft of the doc	Helen Gibson (CENTRIC)
02/02/2021	0.3	Version for review	Helen Gibson (CENTRIC), Katie Bailey (CENTRIC)
21/02/2021	0.4	Update following review	Helen Gibson (CENTRIC), Reviewers: Philippe Chrobocinski (ADS), Georgios Potamos (CYMOD)

Executive Summary

This is the first version of the ISOLA data management plan. Data is an integral part of the ISOLA system and will underpin several machine and deep learning models, be used to inform rapid assessments of security incidents, and help improve the safety onboard for all crew members and passengers.

ISOLA has opted into the Open Data Research Pilot and is committed to making available research data and other data including the supporting metadata where this does not infringe rights for the protection of personal data, security, intellectual property, and confidentiality obligations. ISOLA follows the process for ensuring FAIR (findable, accessible, interoperable, and reusable) data and the details laid down in this deliverable will support all beneficiaries to follow a FAIR process.

ISOLA is keenly aware of the ethics, data protection and security considerations that affect all open access decisions for all research data but are particularly relevant in security focused projects that not only have to protect the rights of any data subjects but also have obligations to not expose specific security requirements that are necessary to keep the maritime and border security industry operating safely. The majority of ISOLA deliverables contain EU Classified Information; therefore, this data management plan presents a process for data management alongside an initial high-level view of the expected datasets in ISOLA that will be refined in future versions.

Table of Contents

List of Tables

List of Figures

No table of figures entries found.

List of Acronyms

Acronym	Meaning
AI	Artificial Intelligence
DMP	Data Management Plan
DPIA	Data Protection Impact Assessment
EUCI	EU Classified Information
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
PSO	Project Security Officer
RDF	Resource Description Framework
SAB	Security Advisory Board
TBC	To Be Confirmed
WP	Work Package

Table 1. List of acronyms.

1 Introduction

1.1 Overview

ISOLA is conceived as a project to develop and enhance the situational awareness and therefore the safety of passengers onboard ships. ISOLA brings together a number of state-of-the-art technologies to deliver solutions that will aid the execution of the ship's security plan and support the Ship Security Officer and Crew to maintain the safety and security of all those aboard. ISOLA will incorporate the use of innovative technological applications such as sensing, monitoring, data fusion, alerting and reporting in real-time to swiftly resolve security incidents. ISOLA will evaluate the develop solution across five core use cases:

1. Incidents that occur due to the consumption of drugs or alcohol
2. Incidents of theft on board the ship
3. Attack of the vessel in relation to piracy
4. Detection and discovery of stowaways or illegally boarded passengers
5. Cyberattacks on the ships systems.

Data will form an integral part of the ISOLA system; thus, it is important that a clear structure and plan for the management of data within the project is available from the outset and is regularly updated throughout the project's duration.

Open access to research data in ISOLA is provisioned through Article 29.3 of the Grant Agreement[1]. This states that with the digital research data generated in ISOLA partners must:

> *a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:*
>
> *a. the data, including associated metadata, needed to validate the results presented in scientific publications, as soon as possible;*
>
> *b. other data, including associated metadata, as specified and within the deadlines laid down in the 'data management plan'*
>
> *b) provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves).*

Exceptions apply concerning the protection of results, confidentiality and the need to protect personal data. Furthermore, ISOLA will also consider the security implications of making any data open. Assessment of this eventuality forms part of the data management planning process. This initial version of the DMP focuses on identifying where data will exist in the ISOLA project, who has ownership over that data and what processes need to be in place to manage that data over the duration of the project.

[1] European Commission – H2020 Programme, Annotated Model Grant Agreement https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf

1.2 Deliverable positioning

D1.2 is the first iteration of the data management plan within ISOLA, it is due in M6 of the project and will be updated formally in M24 in D1.4. The final activity report due at the end of the project will set out the long-term legacy data management plan for any data created over the project. Other updates will occur as necessary when new datasets, data requirements or standards emerge over the course of the project as well as being updated prior to the periodic and final reviews. Partners will be contacted regularly to remind them to update the DMP as their tasks and the project evolves.

The aim of this specific deliverable is to identify the processes around data management, and the expected and potential data sets to be created and used within the project. This deliverable is due prior to the delivery of the use cases (D2.3) and both the technical requirements (D7.1) and architectural definition (D7.2); therefore, it is expected that a number of aspects relating to data management are still to be determined and will emerge both through future stakeholder and technical consultations as well as wider progress within the field. Similarly, as we move towards the first piloting exercise and as the ISOLA research begins to generate findings suitable for scientific publication further data requirements will become apparent.

Given these constraints, we expect this deliverable to act as a living document to ensure good data management practices throughout ISOLA.

1.3 Deliverable structure

This deliverable is set out as follows: firstly, we address the overall principles for managing data within the ISOLA project considering what is FAIR data management, what are the types and forms of data likely to be encountered during the project, additional considerations for data relating to personal data and any special categories of personal data, further ethical and security requirements; and finally placing ISOLA in the context of the open research data pilot.

Following this we consider, via each work package, the potential data considerations relating to each task and the likely data management implications. This forms the core basis of the data management plan. Finally, we review the status of data management at M6 and set out the plan for data management throughout the remainder of the project.

2 Principles of good data management

The development of this data management plan (DMP) will follow the approach as set out in the European Commission's (EC) 'Guidelines on FAIR Data Management in Horizon 2020'[2]. This approach is taken to ensure a consistent and replicable approach to data management that will evolve throughout the project based on standardised principles. The guidelines cited above suggest an approach to data management that is motivated by the following overarching questions:

- How should research data be handled during and after the project?

[2] European Commission (2016) Guidelines on FAIR Data Management in Horizon 2020 https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

- What data will be collected, processed and/or generated?
- Which methodology and standards apply?
- To what extent can data be shared or made available for open access?
- How should research data be curated and preserved during and after the project?

In the first version of the DMP, the goal is to set out an overarching plan for the management of data and develop an awareness within the project of the scope of data that will be present. This includes data that will be collected, used and generated. It should also motivate partners to consider data management as an integral aspect of their research activities. As per the EC's guidelines the following six (paraphrased) questions are to be considered for each possible dataset. We consider each of these questions in turn and place them in the context of the ISOLA project.

a) *What is the purpose of the data collection/generation and its relation to the objectives of the project?*

In the context of ISOLA, this includes identifying which work package(s) and task(s) the dataset appears and how it related to the goals, development and evaluation of that task and the wider project.

b) *What types and formats of data will the project generate/collect?*

As we will detail further in Section 2.2 below, this considers whether the data is collected, used or generated; the format and filetypes of the dataset; and considerations for data that resides inside the ISOLA system itself.

c) *Will you re-use any existing data and how?*

In ISOLA, this includes existing data that is used for training specific machine learning applications; data provided by the project's end users in support of the project's objectives, and data generated by one module that is then consumed by another module.

d) *What is the origin of the data?*

Data used within ISOLA may come from multiple sources, for example, data can provided directly to the consortium (e.g., from end-users) or from existing openly accessible repositories. Partners may also collect data in order to carry out their tasks in ISOLA, or data may result from specific piloting, demonstration and evaluation activities in ISOLA.

e) *What is the expected size of the data?*

This could be described in KBs/MBs/GBs; number of rows/features, or any other appropriate metric given the data type but this may well be unknown at this early stage.

f) *Who else could the data be useful to?*

This considers both other stakeholders and researchers working in other aspects of the project, those working in similar domains as well as wider research applicability. Furthermore, collaborations with other H2020 projects funded under similar calls may also be considered.

To help answer these questions a process known as FAIR data management is proposed.

2.1 What is FAIR data management?

FAIR data management refers to the goal of making research data **F**indable, **A**ccessible, **I**nteroperable and **R**eusable. By adhering to these principles, researchers are able to support the wider research ecosystem as well as ensuring transparency on their decision making around data management. As emphasised in the OpenAIRE[3] guide to FAIR data, it is possible for data to adhere to all four FAIR principles without the data itself being openly accessible; thus, even though many aspects relating to the ISOLA project may raise security considerations, users and owners of datasets in ISOLA should still follow a FAIR approach to data management and ensure that all datasets are appropriately archived and documented. This is reflected in the principle 'as open as possible, as closed as necessary'[4].

In the next four sections, we review the core questions that should be considered when making data FAIR. These questions are motivated by combining recommendations for FAIR data management from the European Commission[5], the FAIR principles from the Go FAIR initiative[6] and the initial conceptualisation of the data management problem and resulting FAIR principles from Wilkinson et al.[7] Naturally, several of these principles are interlinked and so, for example, aspects of making data accessible will impact on the extent to which data is reusable.

2.1.1 Making data findable

The first core principle of FAIR data is to make data ***findable***. Data can only be reused if other researchers are able to discover datasets that are relevant to their work. To make data findable the following aspects should be considered:

- *Provisions to support the discovery of data.*

This includes the use of

  - Standardised metadata formats that are relevant and appropriate to the domain or academic discipline
  - Identification mechanisms such as persistent and unique identifiers (e.g., using digital object identifiers (DOIs).

- *The use of appropriate naming conventions.*

This includes the use of industry or academic standards, ensuring consistency across datasets, using suitably descriptive language representative of the information being

[3] OpenAIRE (n.d.) How to make your data fair. https://www.openaire.eu/how-to-make-your-data-fair

[4] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

[5] European Commission (2016) H2020 Programme: Guidelines on FAIR data management in Horizon 2020. Version 3.0. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[6] Go FAIR Initiative (2021) FAIR principles. https://www.go-fair.org/fair-principles/

[7] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, *3*(1), 1-9. https://www.nature.com/articles/sdata201618

described. For ISOLA, this could mean aligning all data naming conventions with the ISOLA ontology as well as reflecting the conventions used in maritime applications.

- *Defining keywords that optimise dataset discovery and re-use.*

Researchers working in specific fields have their own vocabulary; assigning relevant keywords appropriate to the domain will facilitate easier discovery of the data, especially when working across multiple domains. The maritime and shipping industry is no different; however, it is also important for ISOLA to consider keywords in areas such as artificial intelligence or decision support to ensure that data reaches researchers working in parallel domains.

- *Clearly versioning of iterations of datasets*

As models are improved and iterated upon elements of the dataset may evolve. Further, new data may be collected that extend the original dataset, or corrections could be made where errata are identified. Each new update should be clearly versioned along with appropriate documentation that describes how the dataset has changed.

- *Identifying and applying specific standards for the use and creation of metadata*

This includes outline approaches for metadata creation if no current standards exist. The use of resources such as those found at FairSharing[8] can support the discovery of existing standards.

2.1.2 Making data openly accessible

The second core principle of FAIR data is to make data ***accessible.*** The accessibility of data considers the actual usage of data after it has been discovered. The main initial concern about the accessibility of data is the extent to which it should be made openly accessible, i.e., *who can use the data?* Options for ISOLA include (but are not necessarily limited to):

- The data is fully openly accessible by default.
- The data is available but certain restrictions are imposed on who can access it; how it is accessed or what it can be used for - these restrictions should be fully justified and transparent.
- The data is not available; and should be accompanied by a full justification in the data management plan (reasons include but are not limited to legal, contractual, security, protection of intellectual property).

Furthermore, the decision to not make the data available does not absolve the researcher of following proper data management processes. Once the initial decision on accessibility has been made and justified several further considerations flow from this position; specifically,

- *How and where will the data be made accessible (e.g., in a repository)? Will the metadata, documentation and code also be deposited? Are there any specific arrangements that need to be put in place?*
- *What methods or software tools are needed to access the data? Are these methods documented and, if the software is specialised, can this also be made available?*

[8] https://fairsharing.org/

- *If there are restrictions on use, how will access be provided? Who makes the decisions on access (e.g., data access committee) and is this well-described upfront (e.g., in a machine-readable license)?*
- *How will the identity of the person accessing the data be ascertained?*

In ISOLA, partners can opt to make data available through institutional or national repositories (e.g., CENTRIC as part of Sheffield Hallam University has procedures to deposit data in the Sheffield Hallam University Research Data Archive)[9] or those approved data hosting repositories (such as Zenodo[10] or those listed in curated by Open Research Europe[11]). Any deposited data should clearly refer the ISOLA project and include the appropriate funding statement.

2.1.3 Making data interoperable

The third core principle of FAIR data management is ***interoperability.*** Research, and therefore the use of data, does not exist in a vacuum and thus it is imperative that due consideration is given to the interoperability of research data. Interoperability should facilitate both data exchange and data re-use. Interoperability will be supported by the application of standard formats, compliance with standard and preferably open software applications, and ease of combining with other datasets. Specific considerations for interoperability should also include:

- *Data and metadata vocabularies, standards or methodologies that enhance interoperability.*
- *Use of standard vocabularies for all data to allow inter-disciplinary interoperability.*
- *Provision of mappings to more commonly used ontologies if nonstandard ontologies/vocabularies are used.*

For example, the project will develop the ISOLA ontology, the mapping of this ontology to existing standards as well as adherence to the ontology's structure within the project will support long-term interoperability; while the standardisation activities in the project will identify the standards utilised within the project that will also inform and improve interoperability.

2.1.4 Making data reusable

The final principle for FAIR data is to promote the ***re-usability*** of the data. Researchers and other organisations can only reuse data legitimately if they (1) understand the context in which that data was created and (2) are permitted to use the data in their activities. Uncertainty in this matter can reduce uptake as, for example, some data may only be available for non-commercial uses or only be appropriate for a specific application in a given context.

- *How will licences be applied to facilitate data re-use?*

[9] https://shurda.shu.ac.uk/

[10] https://zenodo.org/

[11] https://open-research-europe.ec.europa.eu/for-authors/data-guidelines#hosting

The use of accessible, standardised and widely known licences should support this. If licences impose specific restrictions on data reuse these should be fully justified. Appropriate licences may include, but are not limited to, Creative Commons licences[12].

- *What are the timescales for making data available for re-use*?

This consideration may include publications restrictions (e.g., embargoes) or requirements (e.g., reproducible research); delays may be sought also to facilitate publication or patent applications and the accepted timescales on these activities should be made known to other partners so as not to hinder their own work. Given some research data is relevant only for specific time periods or in a specific context consideration should also be given to the speed in which the data can be made available. In particular, when data underpins a scientific publication it should be made available as soon as possible after publication.

- *Will the data be relevant and useful for third parties to use and for how long?*

Some data may be project specific, or useful only in the specific context of a module, pilot use case, or configuration of an end-user's processes. The appropriate shelf-life of the data may have to be considered, as should clear explanations of how the data should or should not be interpreted in a context outside that of the project.

- *Will quality assurance processes be put in place and how will they be applied?*

Data that is made available to other researchers should be able to be used without having to contact the original researchers to ask questions and make clarifications. The package of the data, metadata, documentation and other supporting material should provide enough information to ensure correct reuse and prevent misinterpretation. Nonetheless, contact details should be provided with any dataset, especially in case errata are identified either by the original researchers or by those re-using the data so that corrections or notifications can be published alongside the original data.

2.2 Types of data in ISOLA

Within the DMP of ISOLA we will consider multiple types of datasets that vary in purpose and scope that will impact upon how they are considered and managed. For the purposes of the ISOLA DMP, we explicitly define six 'types' of data that will be encountered within the project.

1. **Project management and administrative data** – datasets maintained that support the administration, management and dissemination activities of the ISOLA project.
2. **Primary data collected by partner in ISOLA** - data collected by ISOLA partners for the purposes of carrying out a specific task with ISOLA. This is a new dataset.
3. **Secondary data (not publicly available)** – previously collected data provided to the ISOLA project that is not publicly available (e.g., datasets provided by end-users)
4. **Derived data** – data created from the output of processing by ISOLA module.
5. **Publicly available dataset** – data that is openly available for research purposes such as training / benchmark data for machine learning and research applications.
6. **Synthetic / generated data** – data that has been specifically created for the ISOLA project that is representative of real-world data.

These types will be monitored as the project progresses and new types could be introduced or clarified in later versions of the DMP. Any data that forms the basis of any analysis for a

[12] https://creativecommons.org/about/cclicenses/

deliverable or publication should automatically be registered within the data management plan.

2.2.1 Personal data and special categories

Management of personal data as considered through the lens of GDPR and from a research ethics standpoint is particularly important in the data management process. The first of these issues is to ensure that any data collected during the project is compliant with the GDPR for the purposes of carrying out the project. Other deliverables in ISOLA will include more extensive guidance on the implications for the GDPR in the ISOLA project; however, here we include a high-level overview.

What is personal data?

Personal data is data that relates to an identified or an identifiable individual. This can include obvious personal data such as a name, identification number, or username all the way to location data, financial data, photographs and many more.

The processing, including collection, analysis and storage, of personal data is covered by the General Data Protection Regulation (GDPR)[13] noting that any personal data should be processed "lawfully, fairly and in a transparent manner in relation to the data subject". A core pillar of the GDPR is Article 6 which establishes the legal basis for lawful processing activity. All activity in ISOLA should strive to use informed consent for the legal basis for any processing of personal data noting that possible other bases are:

a) *the data subject has given consent to the processing of his or her personal data for one or more specific purposes;*

b) *processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;*

c) *processing is necessary for compliance with a legal obligation to which the controller is subject;*

d) *processing is necessary in order to protect the vital interests of the data subject or of another natural person;*

e) *processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;*

f) *processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.*

With respect the DMP in ISOLA, here we assume, that a lawful basis has been established for the purposes of carrying out activities within the project itself. Therefore, in the context of making data FAIR, it is necessary to discuss the impact and provisions that should be made for making any data available in other contexts. For this further processing of personal data, we can rely on four possible options:

[13] European Parliament and Council of European Union (2016) *Regulation (EU) 2016/679*. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN

1. Remove all personal data before making the data available.
2. Include the possibility for future processing within the informed consent procedures.
3. Anonymise the dataset completely.
4. Rely on the provisions set out in Article 5(1)(b) that states: *"further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation')*.

Preference is given to complete anonymisation of the dataset where possible.

Considerations for special categories of personal data

Special categories of personal data may include data related to racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation.

In this case, processing of data should rely explicitly on informed consent as a lawful purpose for processing and in the context of making data FAIR such data should again be anonymised where possible.

2.2.2 Research data and 'system data'

The data used in ISOLA will be collected, created, and processed for different purposes. Thus far this deliverable has considered data that specifically underpins research activity. ISOLA will also collect, create and use data that is more transient and the processes for monitoring the usage of this data should also be considered. In this case, data that, for example, is used or created on an ad-hoc basis for testing whether two modules are successfully connected together must still comply with all ethical and data protection considerations for its use but if it is not maintain following the completion of the test, then it is not subject to the processes set out in the data management plan. Data that is utilised repeatedly for this kind of testing however, should form part of the DMP as it can be considered as a dataset that facilitates the evaluation of ISOLA.

2.3 Ethical data management

When collecting, managing and processing data it is not only the legal principles of the GDPR that should apply, but the ethical management of data should also be considered; nevertheless there are many overlaps between the GDPR and ethical data management. The ALLEA European Code of Conduct for Research Integrity[14] provides four key principles that should underpin all good research practices: reliability, honesty, respect and accountability. Each of these can be applied to ethical data management processes through ensuring data is collected and managed is a reliable and transparent manner with respect for all organisations, data subjects and domains following clear and repeatable processes. The ALLEA code promotes FAIR data management practices as well as clearly aligning with the goals of the data management plan for stewardship of data and research materials, fostering

[14] All European Academies. (2017). *The European code of conduct for research integrity*. European Science Foundation. https://allea.org/portfolio-item/the-european-code-of-conduct-for-research-integrity

transparency on data access and reuse, ensure that data are considered legitimate and citable products of research and fulfilling any contractual obligations.

Other ethical data management processes include considering how data that will be used for training and developing artificial intelligence (AI) models follow the guidance set out in the European Commission's Ethics Guidelines for trustworthy AI[15].

If collected data contain personal data, then it may also be necessary to carry out a data privacy impact assessment (DPIA) based on the guidelines set down in GDPR for any data processing that is considered high-risk. This is regardless of whether the lawful basis for processing is informed consent. CNIL[16] provide extensive guidance for the needs and requirements for carrying out a DPIA, furthermore this guidance will also be elaborated in other outputs of the project.

2.4 Security considerations

ISOLA is funded under the European Commission's Horizon 2020 Secure Societies programme under the call SU-BES02 Technologies to enhance border and external security specifically focused on the sub-topic of security on-board passenger ships. Many aspects of the ISOLA project will use or include EU Classified Information (EUCI). In light of this, any security implications associated with the data should be keenly considered and advice should be sought from the Project Security Officer (PSO) or the Security Advisory Board (SAB) if doubts occur.

All facets of the data should be considered for security concerns, this includes but is not limited to, the

- Dataset (or particular aspects of).
- Metadata provided that describe the data.
- Documentation on how the data was collected/created.
- Processes for parsing the data.
- Inferences that could be made from the data.

This is particularly important when considering data relating to the piloting and demonstration activities that may include multiple datasets linked together that provide an overall picture of aspects of the ship's security plans, procedures and response mechanisms.

2.5 Participation in the open data research pilot

As mentioned in the introduction, ISOLA has opted-in to the open research data pilot. This choice imposes specific data management requirements upon the project, not least the obligation to publish data underpinning any scientific publications (unless publication would conflict with intellectual property, confidentiality or security obligations). Nonetheless, when partners are preparing scientific publications appropriate thought should be given to following the FAIR principles and how data may be made available, even with some restrictions.

---

[15] European Commission (2019) Ethics guidelines for trustworthy AI. *High-Level Expert Group on AI.* https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[16] CNNIL (2020) Privacy Impact Assessment. https://www.cnil.fr/en/privacy-impact-assessment-pia

Given this requirement, all publication material should be reviewed by the security advisory board (SAB) prior to publication and this should include the datasets that support such publications and/or the justifications on the decision to publish (or not) any data.

3 Data management in ISOLA

This section will consider the main activities within the ISOLA project, the associated data that may be processed during those activities and the plans for the management of that data. As the outcomes of the majority of tasks in WP2-WP8 are considered as EU classified information at the restricted level the discussion of the data in these WPs have been moved to Confidential Annex that is not present in the public version of the document.

It should be noted that the information presented in this section and in the Annex has been produced prior to the finalisation of the project use cases, the user requirements and where many of the technical tasks are still in their infancy. Thus, aspects of these proposed datasets may evolve over the course of the project. Nonetheless, each partner was asked to consider their tasks within the project, their expectations around the data to be collected, created, processed and used within those tasks and complete a version of the template in Annex.

3.1 WP1 - Project Management and Coordination

WP1 is concerned with the project management and the administrative functions of the project as well as supporting ethics and data protection activities in combination with WP10.

No research datasets are foreseen within the scope of WP1; however, data relating to the management of the project will include, but is not limited to,

- names and email addresses of project participants within the consortium
- names and email addresses of advisory board members; and any other external advisors
- data relating to the management of the project – e.g., consumed PMs per partner, financial expenditure, etc.

Such data is managed by the coordinator with support from the Scientific and Technical management where necessary.

3.2 Overview of data management in WP2-WP8

As discussed above, due to the security requirements of ISOLA it is not possible to review in detail in a public document the specifics of the datasets that will form part of activities related to EU classified information. Nonetheless, in this section we set out the key elements of data management for such data.

In these WPs we expect there to be four key themes within the data

1. Data produced by the end users (i.e., ship owners)

This data related specifically to existing data already collected by passenger ship operators. This type of data is highly sensitive and is used to inform model development and support testing of different technologies as well as ensuring that ISOLA is able to integrate with existing ship systems with the view of facilitating long-term exploitation of ISOLA products and services. This type of data will only be available to partners in the consortium who require it for the development of their module prototypes.

2. Sensor data

ISOLA proposes to improve ship security partly through the deployment of additional sensors that can provide early warnings of incidents when combined with other data processing activities. Data related to sensors will have different levels of potential availability within ISOLA. Some activities will use existing publicly available datasets that will inform detection activities; however, some of this data maybe be combined with information from end-user partners which will be restricted to the consortium.

3. Passenger data

ISOLA will not use actual passenger data but will develop datasets that a representative of data relating to the information a ship operator or border security authority would hold on passengers as well as some passenger activity data relating to ISOLA system development. Such data will either be completely simulated or be collected through informed consent procedures as part of testing or piloting or demonstration activities within the project. It is likely some of this data will be personal and thus will be managed accordingly.

4. Border authority data.

As well as ship operators, border control authorities are also a key stakeholder within the project, specifically regarding the embarking and disembarking procedures. User of such data will be entirely simulated but will remain confidential to the consortium.

5. Partner data

Partners' data covers all other datasets relating to data collected by partners in order to carry out their specific tasks within the project. Data for the training and testing of AI-based models is one such type of data but existing datasets held by partners also contribute to this type of data. In some cases, this data may be able to have a wider distribution than the project; however, this is both dependent on the task and guidance from the security advisory board.

6. Evaluation data

Testing and demonstrating ISOLA is an essential part of the project. Several datasets will be created, collected and processed within these activities. While evaluations from participants in the demonstration activities will also contribute further data. All participants will be involved in such testing only after fully understanding the scope of the activities and giving informed consent. Such datasets cannot easily be made available outside the scope of the project due to the likelihood that they will expose specific security procedures.

3.3 WP9 - Impact Creation, Dissemination and Exploitation

WP9 monitors the communication and dissemination activities of the project. This will create several administrative datasets in support of monitoring KPIs related to the project.

Table 2: Communication and dissemination data

Overview	
Dataset ID	T9.1
Dataset Title	Communication and dissemination actions
Work Package	WP9
Task/ Deliverable	D9.1
Partner(s)	PROMETECH, ADS, CERTH, ACCELI, ANEK, CSM, PRISMA, SIVECO, T4I, IDM, NTUA, AVR, BD, CENTRIC, CLS, ZEUS, MST,

	DBS, IBM, CY-MOD, EN, CELESTYAL
Data Type	Primary data collected by partner in ISOLA Derived data (e.g., output from processing by ISOLA module) Publicly available dataset (e.g., training / benchmark data)
Format	Text, image, video
Details	
Description	Approved for public dissemination of ISOLA partner data.
Use in ISOLA	To increase awareness of the project
Use beyond ISOLA	To show the progress and lessons learned from the ISOLA project
Open Data	
Is the data open?	Yes – Public
Explanation	The Security Advisory Board needs to approve all publicly disseminated information from the project
Storage location	Isola-project.eu, facebook.com, twitter.com, linkedin.com, youtube.com and the ISOLA wiki
Who	Prometech, CERTH and the mentioned social media websites
Metadata	Dissemination submission email and time
How	TBC
Ethics and Data Protection	
Personal data	TBC
Security considerations	No

4 Conclusion and next steps

This deliverable has set out the requirements and processes for the FAIR management of data during the ISOLA project. The specific requirements related to the GDPR, ethical data protection, and security considerations have been reviewed and highlighted to ensure they form a core part of all aspects of the data management process from collection, to processing, to evaluation.

For each WP the possible datasets that will be created have been considered and initial descriptions of the data structure and use within ISOLA have been presented. In the majority of cases, the extent to which data can be made available is subject to further review and will be more concrete once the use cases, user requirements and initial technical requirements have been finalised. Furthermore, given the dissemination level of this deliverable is 'Public' some details have had to remain necessarily brief to ensure no confidential or restricted information is exposed. The consortium may explore maintaining a secondary version of information in the DMP that contains more extensive information to support data management processes.

This deliverable will be updated prior to the first project review and formally as another deliverable at M24.

5 Annex 1: Data Collection Template

Overview	
Dataset ID	
Dataset Title	
Work Package	
Task/ Deliverable	(if applicable)
Partner(s)	
Data Type	*Select as appropriate* • *Primary data collected by partner in ISOLA* • *Secondary data (not publicly available)* • *Derived data (e.g., output from processing by ISOLA module)* • *Publicly available dataset (e.g., training / benchmark data)* • *Synthetic / generated data*
Format	xls/csv/json/etc
Details	
Description	*Brief description of the dataset and how it was collected / purpose*
Use in ISOLA	*How the data is/will be used in ISOLA*
Use beyond ISOLA	*How the data could be useful to other researchers beyond ISOLA*
Open Data	
Is the data open?	*Select as appropriate:* • *Yes – Public* • *Yes – but restricted access* • *No*
Explanation	*Justification and explanation of open access decision*
Storage location	*Where will it be stored; (if open, specify repository if known)*
Who	*Who is responsible for storing the data*
Metadata	*What metadata has been created and how is this managed*
How	*How can the data be accessed? (software, techniques)*
Ethics and Data Protection	
Personal data	*Does the dataset contain personal data?* *Are any special categories of personal data present?* *Was informed consent given for use/reuse?*
Security consideration	*Are there any security / classified information considerations?*