D7.2 MISSION

Data management plan

M6

MISSION

Funded by the European Union

Funded by the 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 CINEA can be held responsible for them.

Deliverable Information Sheet

Version	I
Grant Agreement Number	101138583
Project Acronym	MISSION
Project Title	Maritime just-in-time optimization interoperable port call and voyage optimization tool
Project Call	HORIZON-CL5-2023-D5-01
Project Duration	42 months: 1 January 2024 – 30 June 2027
Deliverable Number	D7.2
Deliverable Title	Data management plan
Deliverable Type	DMP
Deliverable Dissemination Level	PU - Public
Work Package	WP7
Lead Partner	SDU
Authors	Helle Lyngborg
Contributing Partners	MISSION consortium
Reviewers	Jussi Poikonen
Official Due Date	M6
Delivery Date	M6

List of Acronyms

DMP	Data management plan
FAIR	Findable, Accessible, Interoperable, Reusable
HE	Horizon Europe

KPI	Key performance indicator
STM	Management of sea traffic
WP	Work package

List of Tables

Keywords list

- Just-in-time (JIT)
- Shipping
- Port Call Optimisation
- Port traffic
- Environmental
- Maritime safety
- Digitalisation
- Call To Action (CTA)

Disclaimer

Table of Content

S

1. Executive summary

The purpose of this document is to present the initial data and risk management plan. This includes information on how to deal with data and potential associated risks during the project. The plan relates to quality, security, validity standards as well as to knowledge management, sharing and protection.

This data and risk management plan is the first of three in the MISSION project, since it will be updated in month 14 as well as in month 30 (final version).

2. Introduction

The MISSION coordinator team at SDU is lead of work package 7 regarding project management and administration, hence dedicated to the efficient and timely management of the project and supports the effective integration, management, and delivery of all the other work packages.

According to the project task 7.3, the Data management plan will be created in compliance with the Horizon Europe Data Management Plan Template and the FAIR principles and open-access policies to ensure a safe and clear data handling strategy any project dissemination outcome. SDU will develop the DMP in compliance with the HE guidelines and with the support of all partners. It will set out procedures and measures to ensure an efficient data management process and will describe the data management life cycle in terms of their collection, processing and/or generation by the MISSION project (see Section 1 for initial DMP). It will be updated over the course of the project whenever significant changes arise (e.g., new data, changes in consortium policies or composition, or external factors).

In accordance with the project task 7.4, risk management will be carried out throughout the project lifetime. For critical risks which are likely to occur, risk-mitigation measures will be proposed.

3. Objectives

MISSION aims to digitalise the maritime sector to enhance maritime safety and protect people's health by increasing port traffic safety, associated costs, and GHG emissions.

MISSION will communicate about how it is enabling collaboration and transparency among different stakeholders with the main objective of enhancing operations efficiency and reducing fuel consumption by 10-20%, and the related GHG emissions. These benefits will help in emphasising the importance of MISSION to industry, business, and political stakeholders.

General
MISSION's primary objective is reducing GHG emissions up to 20% as required by the call through just-in-time (JIT) port call (process) optimisation, including the end-to-end orchestration of maritime traffic and port operations extending to hinterland connections through communication and information sharing assisted by digital solutions.

Economic
Maritime transport accounts for over 80% of the world's trade, therefore being a key player in the global economic system. However, most ports in the world serve ships on a first-come first-served basis, leading to inefficiencies and negative environmental impacts. Moreover, this creates unbalanced port resource utilisation with high peaks leading to severe congestion and logjams that cause rippling effects across industries.

Improving the efficiency of vessels operations arriving at and departing from ports will have a huge impact in the European maritime sector, translating in economic gains for the actors involved (e.g., port authorities, shipping companies, terminal owners, and operators, etc.) It will also have a positive effect in jobs retention and creation as well as important micro-economic effects since they result into overall welfare and economies of scale. Moreover, the development of green business models will ensure extensive market penetration and maximize benefits realization by aligning the incentives for both primary and secondary stakeholders in the sea transportation ecosystem.

Health & Safety
One of the European Commission's objectives is minimising the risks of serious maritime accidents in Europe by enforcing safety rules. This is aligned with IMO's e-navigation strategy to improve marine safety through the establishment of open digital data exchange standards.

Decreasing the waiting time of vessels in ports and port areas, streamlining their operations, and thus reducing their fuel consumption by 10-20% will have considerable benefits for the health and quality of life of near-port communities. In the long term, this will also have positive economic repercussions connected to fewer money spent for healthcare.

Environmental

By enhancing operations efficiency and reducing fuel consumption by 10-20% - and related GHG emissions required by the current topic – and support by International Maritime Organisation (IMO) and the Maritime Labor Convention, MISSION will significantly improve the climate and environmental footprint of the shipping industry.

4. Background

The MISSION Data Management Plan is detailing what data the project will generate, whether and howthey will be made accessible for verification and re-use and how they will be curated and preserved.

The DMP will cover principles for data collection, data transfer, data processing, data storage, and data retention and sharing, and will be regularly reviewed and updated throughout the duration of the MISSION project.

The DMP will ensure that data management and protection are compliant with EU principles and standards, and with relevant national data protectionlaws and institutional data management policies. All produced data will be treated according to the management guidelines. The DMP will address the points below and will observe FAIR principles.

Type of data: A wide varety of data will be generated during the course of the project, including text, numerical data, qualitative data, images, and models.

Standards to be used: To ensure the interoperability of the underlying, collected, and generated data,the project will rely on data standards to represent and describe the data.

Data exploitation, sharing and retention: Details on data management (procedures, responsibilities, data corrections) will be described in the DMP.

FAIR data principles: We will prepare the data following the FAIR data principles.

FAIR data means that they are:

- Findable
- Accessible
- Interoperable
- Reusable

Data security: For managing the storage of data obtained and shared through the system, all partners must comply with relevant European, national and regional regulations, and practice standards.

Ethics and Legal Compliance: Ethical standards and requirements, as well as data protection regulations applicable to the project's content and procedures, are maintained and evaluated throughout the project's duration.

5. DMP

In the project, we are required to maintain the Data management plan. We are also required to Manage data responsibly, in accordance with legislation, contracts, information security requirements, etc. Furthermore, we will manage data according to the FAIR principles and make data accessible by default, but “as open as possible, as closed as necessary”.

5.1. Methodology

In order to ensure optimal project implementation, task leaders are approached to provide input regarding the following questions on the data that are seen as relevant for the project implementation.

The following questions have been applied to get an understanding of the data that will be used in the MISSION project.

1. Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.
2. What types and formats of data will the project generate or re-use?
3. What is the purpose of the data generation or re-use and its relation to the objectives of the project?
4. What is the expected size of the data that you intend to generate or re-use?
5. What is the origin/provenance of the data, either generated or re-used?
6. To whom might your data be useful ('data utility'), outside your project?

5.2. Initial feedback

Task leaders responded to the requests which provided a preliminary overview of the data that will be applied to the project as well as the purpose of it.

5.2.1. Overview of data

The table below summarises the initial feedback provided from task leaders in the MISSION project month 4 of the project.

Table 1. Data use in the MISSION project

MISSION data use						
	1. Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.	**2. What types and formats of data will the project generate or re-use?**	**3. What is the purpose of the data generation or re-use and its relation to the objectives of the project?**	**4. What is the expected size of the data that you intend to generate or re-use**	**5. What is the origin/provenance of the data, either generated or re-used?**	**6. To whom might your data be useful ('data utility'), outside your project?**
T1.2	Task T1.2 deliverables will be based on information residing in the public domain. No external data will be ingested, and no data	No data will be generated or re-used by task T1.2.	No data will be generated or re-used by task T1.2.	No data will be generated or re-used by task T1.2.	No data will be generated or re-used by task T1.2.	No data will be generated or re-used by task T1.2.

	output will be generated by this task.					
T2.2 Implement seamless data sharing protocols and APIs for port stakeholders	No data re-use is expected.	The proposed data sharing protocols will make use of the following formats: • NGSI-LD (interface and standard upon JSON files) • Protobuf (for Redpanda interfaces – message broking) • TIC4.0 messages and data Also, the technological tool will have the capacity to adapt to other formats in data coming from the pilots. Conversion/translation/annotation mechanisms will be used.	Purpose of data generation is to facilitate secure, real-time data exchange among port stakeholders. The solution will be further implemented to integrate stakeholders' systems with the port call optimization solution using the developed APIs.	This is still unknown.	The origin will be: • The port stakeholders in MISSION pilots • External data sources necessary to complement the JIT algorithms (e.g., Lloyd's register, MarineTraffic, VesselFinder, OpenMeteo…). Particular sources to be used is still to be discussed.	If all or part of the data that will circulate through the data sharing platform would be classified as "public", therefore exposable to external entities, this will be accessible via a protected API. Parties that could be benefit from this information might be: • Shipping lines • Environmental entities • Port Authorities worldwide • Transport carriers • Municipalities and other public government • Police and other (e.g., custom)

						maritime authorities
T3.1 and 3.2	We may use existing data on the following: global vessel positioning, vessel particulars, vessel cargo planning, vessel fuel consumption and emissions, global weather, port locations, port geoareas, portcall planning information, port cargo operations planning and tracking. These data are used to reach the objectives of Tasks 3.1 and 3.2 as specified in the project plan; this may include training machine learning models, applying such models for inference, and producing predictions, prescriptions, and descriptive analytics	The data used in these tasks will be primarily in tabular or text-based data interchange formats such as Microsoft Excel tables, csv files, or json files. Services developed in the project may provide data through APIs e.g. in json format, and internal results of data processing will be stored in compressed parquet format or in dedicated databases.	Tasks 3.1 and 3.2 focus on data analysis and providing prediction and optimization services according to Task objectives based on the available data.	Needs to be evaluated in detail, pending.	The outlined source data are acquired from multiple sources, including project partners, commercial sources, and public sources. Analytics data may be produced in the project by Task participants, e.g. Awake.AI, Napa, CETENA, and Blue Visby.	A large part of the existing source data cannot be shared outside the project, or even internally to all partners (sharing may be subject to NDAs between parties). The models trained in the Tasks, and outputs of analytics services developed in the Tasks are planned to provide benefit to project partners by enabling schedule optimizations, and results will also be applied in providing similar optimization functionalities and services after the project for external parties operating in maritime logistics.

	related to the source data.					
T3.5	During the project we will re-use existing ship position (AIS) and Metocean data. The data will be used as part of ship context evaluation in the orchestration tasks. Both data types are well standardized and can be sourced from various commercial sources. The post-project use of the said re-used data is not possible due to commercial conditions of the original data provider. Separate commercial contract will be needed.	The task will foreseeably generate data that is related to overall condition of the orchestrated queue of ships in any given moment of time. This overall summary of the queue will be produced individually for each destination port, and it will evolve over time. These snapshots might contain information related to the other tasks or link to those in an agreed way. The generated data will be stored with widely used formats and with sufficient metadata and naming conventions. Most probable storage formats will be (a) JSON for object serialization and audit trail and standard SQL database for relational data.	The Re-used data is used to make the project execution effective. The generated data is only done to an extent that is necessary for reaching the T3.5 objectives. The generated data will act as a single-source-of-truth for the given queue status and act as the basis for communication and end-to-end orchestration.	The re-used data will be in magnitude of 100+GB – 1+TB per year. The generated data will be in the magnitude of 20MB per port per day.	The re-used data (research purpose) is now belonging to Napa Oy, originally sourced from Spire (AIS) and Metocean (Tidetech).	The generated data could be useful to various stakeholders working with ports, terminals, shipping, or logistics. There should be specific effort put to resulting API definitions and design so that the data will be easily re-usable in emerging use cases.

T4.1 and 5.5	There will not be any re-use of existing data that comes from another project. The Financial and Socio-Economic analysis could use overall information from datasets and methodologies or handbooks according to the guidelines to carry out the work. The use of statistics coming from official organisms are part of the usual calculations in this kind of exercises.	The project for task 5.4 will use data extracted from the demo use cases analyzed, mainly calculations using Excel. It is possible that EDIFACT format data to be exchanged or transmited is needed.	The only purpose is to evaluate the financial and socioeconomic perspective of the demos within the project and the possible scalability of future solutions for the sector.	It will depend on the definition of the datasets expected from WP2 and WP3. In this sense, we are not talking about big datasets. Expected size below 100MB	Data will be generated during the analysis. Some statistics will feed the analysis coming from different open-source datasets. For 4.1, the possible origen from the end users legacy systems is expected.	Profile of possible stakeholders that want to replicate the results of the project. Finally, from the point of view of 4.1 and the development of the pilots from COSCO, ERSHIP and PAULA: • COSCO used to request a NDA in order to share their data with the project. • ERSHIP, we think that the same situation. • PAULA uses a lot of information from the Port Community System (PCS). It is important to take it into account in order to reflect this data as confidential. This last point has an impact on the possible deliverables that includes

						anything related to those entities, trying to anonymize the info
Task 4.2 - Container Shipping Demo Case: Planning, Integration Activities, and Implementation	Yes, we will re-use data that is being employed by Port of Valencia for different internal actions and research projects. There is no risk, however, of re-use of information containing sensitive personal information.	In general, the information will be expressed in JSON, following TIC4.0 standard specifications. However, NGSI-LD conversions are expected (in order to align with T2.2 provisions – data sharing platform of MISSION). The grammar is compatible with the main protocols, included and not limited to: • MQTT • JSON • OPC-UA • CanBUS • ModBus • Profinet (DB fix) • Plain text (email, handwritten document, etc.)	Purpose of data generation is to facilitate secure, real-time data exchange among port stakeholders. The solution will be further implemented to integrate stakeholders' systems with the port call optimization solution using the developed APIs.	Still unknown.	The origin of data is the Port of Valencia enterprise information systems, such as NAVIS and PAULA.	It will be useful to apply the JIT advances that MISSION will perform. It is expected to be helpful to the whole port ecosystem. In particular, stakeholders that will benefit from the original data (and from the results of the project) will be: • Terminal operators • Port Authority • Innovation center associated to port activities • Hauliers • Shipping lines operating in the port

T5.2	Task T5.2 activities will be based on information residing in the public domain. All regulations and standardisation literature - that will be used for the compliance evaluation - are either publicly available or will be provided by consortium partners. No external data will be ingested, and no data output will be generated by this task.	No data will be generated or re-used by task T5.2.	No data will be generated or re-used by task T5.2.	No data will be generated or re-used by task T5.2.	No data will be generated or re-used by task T5.2.	No data will be generated or re-used by task T5.2. The information that will be generated with respect to the deliverable D5.2, since it will be publicly available, will be used by relevant stakeholders, e.g., policy makers and end users, to facilitate the adoption of the MISSION platform.
T5.3	We will use characteristic data from CETENA's operational experience related to the type of port and ship that will be studied.	Data on the physical models of ships and ports required for simulator operation and descriptive data on the risk models of the ships and ports under analysis will be used.	Data concerning physical models will have the size of GB. Data relating to risk models will be kB in size.	Most of the data is the result of CETENA's experience and database, others will be generated	At least ship owners/operators, Port Authorities and designers.	

				within the project.		
T5.4	We will need to get access to data collected from the vessels once the trades/routes/vessels are chosen. I expect this data to be collected in WP3 or WP4.	The output of task 5.4 will be graphs and tables with calculation results. Depending on how the vessel data is available, these results will be obtained from a python script or directly in excel.	We will use data collected in other work packages and the main purpose of the analysis will be to evaluate the CO2 emissions reduction post deployment of the JiT solution	The data generated from 5.4 will be small, since it will just be analysis results. However, the data from the vessels collected in other work packages could range from 100-200MB to 1-2 GB depending on if it's noon reports or autologged data.	The data we will use in 5.4 for the baseline calculations will come from the vessel's own vessel performance systems. This means, there will be different formats, collection frequency, etc… for each vessel owner/operator and that will need to be streamlined for the data to be used in the project. For the evaluation of CO2 savings, I'd assume the data may be collected directly from the JiT solution developed by the project or a hybrid model with some data from the vessel's performance system and some from the JiT solution.	The analyzed results will be relevant for other work packages in the project such as 4.5, 5.1, 5.2, 5.5, 5.6, 6.2, 6.3 – but could be others. Outside the project, the results will be made available through the technical report and scaling/advocacy efforts in presentations. I don't expect raw data to be shared outside of the project. These results will likely be used/referred to by ship owners, operators, digital solutions providers in shipping, class societies and regulators

WP6	Project objectives, outcomes, timeline; Information about project partners and meetings. All data shared and re-shared will be public. Reasons for re-use: increasing project's visibility and outreach.	Public datasets/results, which will be converted to communication materials. Those materials will be tailored to each one of identified stakeholder groups and adapted to the communication channels used for such purpose.	Dissemination and Communication of our interdisciplinary contributions over the full duration of the project. Public data will be used to make the project highly visible and raise awareness about the action and its wider objectives. To widely disseminate the public results of the project for use by all relevant stakeholders.		Provided by partners (updates on the WP progress, attendance at the external events, articles/publications written etc.)	Researchers & scientific community; policy makers & public authorities; businesses, entrepreneurs & investors; industry associations & civil societies, and general public. These groups can be subdivided into, for example, port and terminal operators, cargo owners or charters, private logistic operators, shipping lines and operators, EU policymakers, regional and maritime authorities, an academia.

5.3. Further development

Having a more concrete understanding of the data that will be used in the project, we will iterate some of the points related more generally to data sharing infrastructure etc

6. Risks

6.1. Identification

For the MISSION project, proposed risk-mitigation measures will be developed and listed. In that way, any possible risk to proper project implementation will be identified, and, subsequently, the likelihood of success in the project is higher.

6.2. Identified risks

As it appears from the list below, some risks are related to a specific work package, and other risks may appear in any of the work packages.

In the description of the risks respectively, the likelihood as well as the severity of the risk is indicated as low, medium or high.

Table 2. Critical risks for implementation

Description of risk (level of (i) likelihood, and (ii) severity: Low/Medium/High)	**WP(s) involved**	**Proposed risk-mitigation measures**
Harmonized set of standards does not contain up-to-date trends and latest developments. (i): low, (ii): high	WP1	WP-L and task leaders will attend important meetings with standard-defining bodies, e.g., TIC4.0, DCSA, IALA, etc. and hold close contact, so that selected and harmonized set of standards for MISSION are based on latest and acknowledged standards
Data sharing and IT-infrastructure not compatible with existing systems and requirements due to outdated IT-infrastructure. (i): low, (ii): medium	WP2	APIs and data sharing protocols will be made available for existing systems
Data to train and evaluate AI/ML models is not sufficiently available. (i) low, (ii): medium	WP3	The AI/ML models for voyage, fleet, and port call analytics can be trained on AIS data available ex-post for all ships
End users are unavailable to sufficiently participate in	WP4	Keep close contact to key people in the organization

the DCS. (i): medium, (iii): high		to keep up motivation; keep sufficiently and continuously informed about progress and involve them in the scenario definition so that usefulness and benefits are clear and motivating.
Methodology to measure emissions reductions does not fit the different traffic types. (i) low, (ii) medium	WP5	Make sure that end user shipping companies are involved in the process, so that specificities are captured for correct measurement further consulting technical guidance
Lack of impact. (i): low, (ii): medium	WP6	Booster activities will be considered (e.g., renewed social media engagement; engagement with local press, institutional communications channels, academics personal social media pages, maritime/IT/oceanographic conferences etc.). Standardization organizations and policy makers will be approached and informed about the cost and benefits of MISSION's solutions for end users and stakeholders.
A consortium member leaves the project. (i): low, (ii): medium	WP6	The project management team will analyze: (1) substitution of the partner by another one with similar expertise; (2) redistribution of tasks among project partners.
Delay in any individual WP or key milestones and non-performance of partners. (i): low, (ii): medium	All	The project coordinator will call for an emergency meeting to re-establish the terms of the consortium agreement. If necessary, work will be redistributed from non-performing partners to other partners with appropriate competences.
Loss of critical competencies or key participants in the project. (i): low, (ii): low	All	The project is led and anchored by partners with considerable organizational capacity. In this unlikely event, individuals will be replaced using the different networks of the involved partners.

7. Responsibilities and Resources

The present DMP will ensure that data management and protection are compliant with EU principles and standards as well as with relevant national data protection laws and institutional data management policies. All produced data will be treated according to the management guidelines. The DMP will regularly be reviewed and updated throughout the project.

MISSION pays detailed attention to data management and will take all necessary steps to ensure the safe storage of the data, as well as that any data sharing and publication do not breach the participants' privacy.