

DATA MANAGEMENT PLAN

Grant Agreement number 23IND09

Project short name MaritimeMET

Project full title Metrology for green maritime shipping: Emission control through traceable measurements and machine learning approaches

Data management plan 1st 2nd

Confidentiality status PU - Public, fully open

TABLE OF CONTENTS

1	Data management plan.....	2
1.1	Data summary.....	2
1.2	Findable, Accessible, Interoperable and Re-usable (FAIR) Data.....	4
1.2.1	Making data findable, including provisions for metadata	4
1.2.2	Making data accessible	4
1.2.3	Making data interoperable.....	6
1.2.4	Increase data re-use	6
1.3	Other research outputs	8
1.4	Allocation of resources	8
1.5	Data security	9
1.6	Ethics	10
1.7	Other issues.....	10

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 EURAMET. Neither the European Union nor the granting authority can be held responsible for them.

European Partnership  Co-funded by the European Union

Data Management Plan

Issued: February 2025

The project has received funding from the European Partnership on Metrology, co-financed from the European Union's Horizon Europe Research and Innovation Programme and by the Participating States.



1 Data management plan

1.1 Data summary

Questions	Answers
1 Will you re-use any existing data and what will you re-use them for? State the reasons if re-use of any existing data has been considered but discarded.	<p>Yes, the project will use existing data that is available in the literature on the following aspects</p> <ol style="list-style-type: none"> 1. Use of validation data from fundamental reactors for selection of suitable chemical kinetic mechanism for methanol and ammonia (Task 3.1) 2. Machine learning models developed as part of the project will also look for opportunities to include data from literature to extend the database against which the model can be developed and validated (Task 3.2-3.4). 3. Spectral data used in spectroscopy methods available in various databases. (Task 1.1 – 1.2) <p>Reasons: the data is primarily used for extending the validation database. In the aspect of ML, it is used to make the models more robust.</p>
2 What types and formats of data will the project generate or re-use?	<p>Data Generated in the Project</p> <ol style="list-style-type: none"> 1. Types: Spectral line parameters, Laboratory Measurements – Spectroscopy, Emission measurements – Exhaust gas, On-Board Ship Measurement – PAS Sensor, Low-cost sensors performance comparison Formats: ASCII and/or excel and/or .csv format Formats: ASCII and/or .csv format Formats: ASCII and/or .csv format Formats: ASCII and/or .csv format 2. Types: Dynamic Pressure Measurements/calibration methods Formats: excel, and/or .csv format and/or .txt format 3. Dynamic Temperature Measurements/calibration methods Formats: excel and/or ASCII format 4. Types: Engine Measurements – MISC parameters at Wärtsilä and DTI Formats: ASCII and/or .csv format 5. Types: Machine Learning Models and Virtual Sensors (Task 3.2 – 3.4) Formats: ASCII, .csv formats <p>Data planned to be Re-used:</p> <ol style="list-style-type: none"> 1. Types: Spectral line parameters Formats: ASCII and/or excel and/or.csv format 2. Types: Ignition delay data and speciation measurements in fundamental reactors Formats: Excel, .csv, .txt 3. Types: Data intended for ML Formats: ASCII, .csv formats
3 What is the purpose of the data generation or re-use and its relation to the objectives of the project?	<p>Purpose of data generation in relation to objectives</p> <p><u>Objective 1:</u> The spectroscopy data generated in WP1 are used to obtain line parameters, which, in turn, are used to measure absolute concentrations for target species. The data is used to establish the spectroscopic measurement method in this process. The spectroscopic method will validate and provide traceability to the developed low-cost photoacoustic sensors (e.g., CH₃OH, NH₃ and CO). The PAS sensor will be then tested on a test engine and on-board a ship, and the data produced will provide an understanding of the sensor's performance at varied conditions (i.e., lab, test conditions, on-field conditions)</p> <p><u>Objective 2:</u> The data generated in WP 2 on dynamic pressure calibrations will be used for interlaboratory comparisons and eventually establish a Coordinated measurement capability (CMC) under the KCDB key comparison database. Data generated on dynamic temperature will be used to establish calibration methods. The data from the characterisation and validation of established methods in engine/engine-like conditions will be used to train and validate Machine Learning</p>

	<p>models in WP 3.</p> <p><u>Objective 3:</u> The data generated in WP 3 are primarily validated models that are outcomes from machine learning approaches. However, experiments performed in fundamental reactors would be used to validate chemical kinetic mechanisms. The mechanisms in turn, will be used to determine physico-chemical properties essential in Hybrid ML models.</p>
4 What is the expected size of the data that you intend to generate or re-use?	The estimated overall size of the data generated in different work packages /objectives will be 200 GB to 1TB.
5 What is the origin/provenance of the data, either generated or re-used?	<p><u>Data generated in the project</u></p> <p>The data generated will be from measurements, calibrations, and validation of methods developed as part of the three technical objectives.</p> <p><u>Re-used data</u></p> <p>The existing data on spectroscopic parameters will originate from several sources: participant pre-existing data, scientific literature, and data from databases (HITRAN, etc.). Similarly, additional data from the literature and pre-existing data from participants will also be used to train machine-learning models in WP3.</p>
6 To whom might your data be useful ('data utility'), outside your project?	<p>The data will be suitable for use by other research groups working on the following topics: measurement methods in ship engines (e.g., emission and in-cylinder quantities), spectroscopy, machine learning approaches, and reaction kinetics.</p> <p>It will also be helpful for standards committees, including</p> <ul style="list-style-type: none"> • International Organisation for Standardisation /Technical Committee 158 - Analysis of gases (ISO/TC 158) • ISO/TC 108 Working Group 34/ WT19666 Dynamic Pressure Calibration • ISO/TC 24/ Sub Committee 4/ Working Group 12 • CEN/TC 264/ WG 40 • BIPM Metrology in Chemistry and Biology (CCQM) Gas Analysis Working Group BIPM CCQM-GAWG • BIPM Consultative Committee for Thermometry (CCT) • BIPM Consultative Committee for Mass and Related Quantities (CCM) Working Groups on Pressure and Vacuum (BIPM CCM-WGPV) <p>The data might be useful to:</p> <ul style="list-style-type: none"> • Stakeholders from the industry: Shipping industry, Engine testing and consulting agencies, emission monitoring, reporting and verification (MRV) entities, sensor manufacturers, emission system developers, and Ship operators. • NMIs/DIs: Working in emissions and environmental monitoring, dynamic pressure and temperature calibrations, and institutes working on machine learning and estimating uncertainties on ML models. • Other scientists working in the field: Spectroscopy, engine and emission measurements, machine learning approaches • Economic actors: EU-Emission trading system (EU-ETS) – Agencies addressing aspects related to emission monitoring, verification and reporting

	<p>(MVR)</p> <ul style="list-style-type: none"> Decision makers: International Maritime Organization (IMO) and its affiliated entities / national bodies.
--	--

1.2 Findable, Accessible, Interoperable and Re-usable (FAIR) Data

1.2.1 Making data findable, including provisions for metadata

Questions	Answers
7 Will data be identified by a persistent identifier?	Yes, the project's deposited dataset will be identified by a DOI and handled either through Zenodo or PTB Open Access Repository (PTB-OAR)
8 Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.	Yes, the metadata created for all of the project's deposited datasets will be open under a Creative Commons Public Domain Dedication (CC 0) or equivalent (to the extent legitimate interests or constraints are safeguarded), in line with the FAIR principles (in particular machine-actionable) and provide information at least about the following: datasets (description, date of deposit, author(s), venue and embargo); the European Partnership on Metrology funding; grant project name, acronym and number; licensing terms; persistent identifiers for the dataset, the authors involved in the action, and, if possible, for their organisations and the grant. The metadata will include persistent identifiers for related publications and other research outputs where applicable.
9 Will search keywords be provided in the metadata to optimise the possibility for discovery and then potential re-use?	Yes, the following search keywords will be provided in the metadata to optimise the discovery and potential re-use of the deposited datasets: MaritimeMET, spectroscopic parameters, dynamic pressure, dynamic temperature, machine learning approaches, calibration, traceability, uncertainties, virtual sensors.
10 Will metadata be offered in such a way that it can be harvested and indexed?	<p>Yes, both PTB-OAR (https://oar.ptb.de) and Zenodo (https://zenodo.org) comply with FAIR principles. The metadata are indexed in a searchable resource selected by the consortium for each data set based on the topic.</p> <p>Metadata is licensed under CC0, except for email addresses. All metadata are exported via the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and can be harvested.</p>

1.2.2 Making data accessible

Questions	Answers
Repository:	
11 Will the data be deposited in a trusted repository?	To ensure that the data is accessible, the project data will be made available to target end users (question 6) via PTB – OAR or Zenodo
12 Have you explored appropriate arrangements with the identified repository where your data will be deposited?	Yes, separate arrangements with the host of the repository are necessary because of the data's large size and formats for some of the data generated in the project.
13 Does the repository ensure that the data are assigned an identifier? Will the repository resolve the identifier to a digital object?	Yes, PTB-OAR Repositories using globally unique and persistent identifier e.g. DOI will be used as it allows data to be easily retrievable. The same is true for Zenodo repository
Data:	
14 Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access	All of the data that are needed to validate the results presented in scientific publications will be made openly available as the default unless there is a specific reason not to publish the data.

Questions	Answers
<p>conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.</p>	<p><u>Datasets which cannot be shared – voluntary restrictions</u> Other data may be made available on a case-by-case basis.</p> <p>The following data will not be made publicly available:</p> <ul style="list-style-type: none"> • Data obtained with the permission of third parties, but the third parties have not agreed to make the data publicly available. • Data that discloses the identity of a manufacturer. • Data that compromises the protection of a participant(s) intellectual property. <p>The level of data made available will also be considered, for example, pre-processed data will not be provided unless there is a clear reason for doing so.</p> <p><u>Datasets which cannot be shared - legal/contractual reasons</u></p> <p>The engine test data from Wärtsilä will be made available to project partners who intend to use the data to train and validate their ML models and understand the sensor's performance (dynamic pressure and emission sensors). However, any measurements done on Wärtsilä engines shall be reviewed by Wärtsilä before publishing.</p>
<p>15 If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</p>	<p>The data used in scientific publications, posters, and oral communications will be available for re-use as soon as possible. Some of the data are expected to be subject to an embargo period of 18 months whilst a patent application is pending.</p>
<p>16 Will the data be accessible through a free and standardised access protocol?</p>	<p>Both PTB-OAR and Zenodo provide well-described conditions for free and standardised access.</p>
<p>17 If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?</p>	<p>There are no restrictions on the use of the published data. Still, users will be required to acknowledge the project and the data source in any resulting publications, according to the latest version of the CC-BY license.</p>
<p>18 How will the identity of the person accessing the data be ascertained?</p>	<p>For better knowledge of the data end users, users are required to register to use the repository. The data provider will get an update on the background of the users and voluntarily contact information.</p>
<p>19 Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?</p>	<p>This MaritimeMET consortium will have a Data Access Committee. Their remit will be to select the data that will be openly accessible on a case-by-case basis. Ethical aspects and data security, including intellectual property requirements, will be considered, as will access requests to personal/sensitive data. Some or all of a potential publication's data will be withheld if necessary. This will be decided in consultation with the relevant participant(s).</p>
<p>Metadata:</p>	
<p>20 Will metadata be made openly available and licensed under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?</p>	<p>Yes, metadata will be made openly available and licensed using an appropriate open-access license, such as CC0. All metadata are exported via OAI-PMH and can be harvested.</p>

Questions	Answers
21 How long will the data remain available and findable? Will metadata be guaranteed to remain available after data are no longer available?	<p>The data will remain available and findable for the lifetime of the Zenodo repository, which is expected to be a minimum of 20 years. If data are withdrawn from Zenodo, the DOI and the URL of the original object are retained. In case of closure of the Zenodo repository, best efforts will be made by Zenodo to integrate all content into suitable alternative institutional and/or subject based repositories.</p> <p>In the case of PTB-OAR, the data will remain available and findable for the lifetime of the repositories unless the data are withdrawn with strong specific reasons given.</p>
22 Will documentation or reference about any software be needed to access or read the data and will this be included? Will it be possible to include the relevant software (e.g. in open source code)?	Majority of the cases, the data are in a common format (see question 2) and can be read using widely available software (open source or commercial). However, for the models developed as part of the ML approaches, format description along with instructions about how to use the model will be provided.

1.2.3 Making data interoperable

Questions	Answers
23 What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?	<p>The datasets will be interoperable as Zenodo's basic metadata requirement is compliant with the DataCite's Metadata schema and OpenAIRE</p> <p>Standard vocabularies will be used to ensure inter-disciplinary interoperability and re-use.</p>
24 In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow their re-use, refinement or extension?	<p>Yes.</p> <p>The compatibility of our project-specific ontologies and vocabularies will be guaranteed through appropriate mappings (glossary, alignment tables, etc.) to more commonly used ontologies. The generated ontologies and vocabularies will be published.</p>
25 Will your data include qualified references ¹ to other data (e.g. other data from your project, or datasets from previous research)?	Yes, the project's datasets that will be deposited in the chosen repository (e.g. Zenodo or PTB-OAR) will include qualified references to other datasets from the same project.

1.2.4 Increase data re-use

Questions	Answers
26 How will you provide documentation needed to validate data analysis and	A short README file (e.g. Markdown) will be provided together with the data, to enable data analysis and to facilitate data re-use.

¹ A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: <https://www.go-fair.org/fair-principles/13-metadata-include-qualified-references-metadata/>)

Questions	Answers
facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?	
27 Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard re-use licenses, in line with the obligations set out in the Grant Agreement?	The data will either be licensed under the latest available version of the creative Commons Attribution International Public License (CC BY) or a license with equivalent rights as set out in the Grant Agreement. Users will be required to acknowledge the consortium and the source of the data in any resulting publications. Alternatively, the Creative Commons Public Domain Dedication License (CC 0) will be used.
28 Will the data produced in the project be useable by third parties, in particular after the end of the project?	Any data published in open-access journals will be usable by third parties after the datasets have been deposited in Zenodo. The data that do not relate to peer-reviewed publications will be made available for re-use on a case-by-case basis.
29 Will the provenance of the data be thoroughly documented using the appropriate standards?	Yes, the provenance and context of the data will be thoroughly documented to meet relevant standards using the Provenance and Context Content Standard (PCCS) Matrix. Data will be accompanied by information on how they were captured, processed, analysed, and validated. Other information that enables interpretation and use will also be provided.
30 Describe all relevant data quality assurance processes.	<p>Data quality will be assured through several quality assurance procedures:</p> <ul style="list-style-type: none"> • Repeated and comparison measurements. • Adherence to standards for data recording. • Use of controlled vocabularies and standard terminology. • Metrological characterisation of the measurement set-ups. • Validation of the data collected. • Provision of test results along with the data. • Peer-review of publications based on the data
31 Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.	<p>Allocation of resources</p> <p>The estimated costs for making the (data and) other research outputs FAIR are 1 000 € (personnel costs) (see question 34). The costs for making other research outputs FAIR are included in the project's budget and will be claimed if compliant with the Grant Agreement's conditions. The consortium's Data Access Committee will also have overall responsibility for managing other research outputs (see question 36). Where feasible, long-term preservation will be ensured by depositing the other research outputs in repositories. The Data Access Committee will decide on a case-by-case basis on which other research outputs will be deposited and for how long.</p> <p>Security of other research outputs</p> <p>All participants are either accredited to, or work in compliance with, the ISO 17025 standard on the "General requirements for the competence of testing and calibration laboratories". The participants will store other research outputs on their organisations' networks, which are protected by firewall, backups etc. Other research outputs will also be stored in the project's SharePoint environment, with a password-protected login. Deposition in public repositories will provide additional security as they have multiple replicas in a distributed file system which is backed up on a nightly basis. This project will not generate sensitive other research outputs. The other research outputs will be safely stored in open access repositories.</p> <p>Ethical aspects</p> <p>There are issues that could impact on the sharing of other research outputs.</p> <ul style="list-style-type: none"> • Information relating to other research outputs acquired from third parties, e.g. manufacturers, will not be shared without their explicit consent.

Questions	Answers
	<ul style="list-style-type: none"> Information relating to other research outputs collected by the consortium at commercial sites will not be shared without the site owner's explicit consent. Ethical issues will be addressed as the project will prepare an ethics report. <p>The project will not share other research outputs with identifiable personal information. Sensitive information relating to the other research outputs will be collected, separated as soon as possible and kept secure. Please also see the information provided in section 1.3 below</p>

1.3 Other research outputs

Questions	Answers
32 In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).	<p>The models developed in the project will be released as open access in peer-reviewed publications.</p> <p>The new calibration methods, and protocols produced by the project will be stored in a repository.</p>
33 Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.	As far as possible, the FAIR data approaches specified in questions 7-30 above will be applied to the management of this project's other research outputs. This commitment will be met by releasing the new software that will be developed in the project under license, by placing the new calibration methods, and protocols, in a trusted repository in line with the requirements of the project's consortium agreement

1.4 Allocation of resources

Questions	Answers
34 What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.) ?	The estimated costs for making the data and other research outputs Findable, Accessible, Interoperable and Re-usable (FAIR) are 1 000 € (personnel costs). These costs have been kept to a minimum by using a free repository (Zenodo) and by making only relevant data and other outputs FAIR.
35 How will these be covered? Note that costs related to research data/output management are eligible as part of the European partnership on metrology grant (if compliant with the Grant Agreement conditions).	The costs for making the data FAIR are included in the project's budget and will be claimed if compliant with the Grant Agreement's conditions.

36 Who will be responsible for data management in your project?	<p>The consortium's Data Access Committee will have overall responsibility for data management.</p> <p>The coordinator will lead this committee and will be responsible for coordinating updates to the data management plan. The committee will be responsible for organising data backup and storage, data archiving and for depositing the data within the repositories (Zenodo, PTB's Open Access Repository)</p>
37 How will long-term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?	<p>Long term preservation will be ensured by depositing the data within repositories (Zenodo, PTB's Open Access Repository). There are no costs associated with the long-term preservation of the data in these repositories.</p> <p>The data will increase in value over time because of its fundamental impact in a wide range of applications. It will enable the technologies developed in the project to be taken up by the measurement supply chain and by standards bodies including Technical Committee 158 - Analysis of gases (ISO/TC 158), ISO/TC 108 Working Group 34/ WT19666 Dynamic Pressure Calibration, ISO/TC 24/Sub Committee 4/ Working Group 12, CEN/TC 264/ WG 40, BIPM Metrology in Chemistry and Biology (CCQM) Gas Analysis Working Group BIPM CCQM-GAWG, BIPM Consultative Committee for Thermometry (CCT), BIPM Consultative Committee for Mass and Related Quantities (CCM) Working Groups on Pressure and Vacuum (BIPM CCM-WGPV), Joint Committee for Guides in Metrology (JCGM) Working Group 1 on the expression of Uncertainty in Measurements (JCGM-WG1:GUM), and relevant Euramet TCs. These standards bodies will need access to the data to justify the robustness of future standards. The data will also be valuable as it underpins published datasets' results.</p> <p>The Data Access Committee will decide on a case-by-case basis what data will be kept and for how long.</p>

1.5 Data security

Questions	Answers
38 What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?	<p><u>Data recovery and secure storage</u></p> <p>All participants are either accredited to, or work in compliance with, the ISO 17025 standard on the "General requirements for the competence of testing and calibration laboratories". The participants will store data on their organisations' networks, which are protected by firewall, backups etc. Data will also be stored in the project's SharePoint environment, with password protected login.</p> <p>Deposition in the Zenodo public repository will provide additional security as it has multiple replicas in a distributed file system which is backed up on a nightly basis.</p> <p><u>Transfer of sensitive data</u></p> <ul style="list-style-type: none"> • This project will not generate sensitive data.
39 Will the data be safely stored in trusted repositories for long term preservation and curation?	<p>Yes, the data will be safely stored in the Zenodo open access repository. Zenodo and the underlying Invenio Framework for digital repositories were designed according to the Open Archival Information Systems (OAIS) reference model. Zenodo is working towards ISO 16363 certification.</p> <p>OR</p> <p>Yes, the data will be safely stored in PTB's Open Access Repository, which is stored on two physically and geographically separated servers that are regularly backed up. PTB is working towards German Initiative for Network Information (DINI) certification.</p>

1.6 Ethics

Questions	Answers
40 Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics report(s) and the ethics section in the Annex 1.	<p>There are issues that could impact on data sharing</p> <ul style="list-style-type: none"> • Data acquired from third parties, e.g. collaborators, will not be shared without their explicit consent. • Data collected by the consortium at industrial partner's facilities will not be shared without the owner's explicit consent. • The data from the market surveys will be made anonymous to comply with the General Data Protection Regulation (GDPR).
41 Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?	Informed consent for data sharing and long-term preservation will be included in the market and customer surveys, but the project has no plans to share data with identifiable personal information. If any sensitive data are collected, they will be separated as soon as possible and kept secure.

1.7 Other issues

Questions	Answers
42 Do you, or will you, make use of other national / funder / sectorial / departmental procedures for data management? If yes, which ones (please list and briefly describe them)?	<p>Data management will be compliant with:</p> <ul style="list-style-type: none"> • The research data policy of the European Partnership on Metrology; • European laws about data security and the protection of privacy (e.g. GDPR); • Institutional guidelines; • Scientific community guideline

2 Open science: research data management

Statement	Put an X in the box to confirm	Or, list any exceptions to this
All participants have adhered to the requirements of the project's GA and CA with respect to open science: research data management (GA Article 17 and its Annex 5) for this reporting period	<input checked="" type="checkbox"/>	