EMPIR Reporting Guidelines
Part 9 – Preparing data management plans

EURAMET EMPIR MSU
Phone: +44 20 8943 6666 Email: EMPIR.msu@euramet.org
msu.euramet.org

The EMPIR initiative is co-funded by the European Union's Horizon 2020 research and innovation programme and the EMPIR Participating States

1 Introduction

EMPIR Reporting Guidelines Part 9 provides guidance for the data management plans (DMPs) produced by Joint Research Projects (JRPs), Support for Impact Projects (SIPs) and Joint Network Projects (JNPs). Reporting Template 9 is linked to this part of the EMPIR Reporting Guidelines.

2 Overview of the Data Management Plan

2.1 Purpose

The purpose of data management is to ensure that researchers can find, access and re-use each other's data, thus maximising the effectiveness and reproducibility of the research undertaken. In EMPIR, this is facilitated through the use of a Data Management Plan (DMP) which is designed to make research data Findable, Accessible, Interoperable and Reusable (FAIR) (see https://www.go-fair.org/fair-principles/) and to ensure that it is soundly managed. It will describe the data management plans for all of the datasets that will be collected, processed or generated by a project, and it must cover the following aspects: the handling of research data during and after the end of the project; specification of the data that will be collected, processed or generated; the methodology and standards that will be applied; plans for data curation and preservation (including after the project). The DMP is updatable and it provides clarity and awareness of the data that projects will create and make openly accessible, where and how data will be stored, who owns the data, and who will have access to the data. It is flexible as data can be open or closed, and projects can opt-in / opt-out at any time (if this is justified), however an amendment will be required.

Please note that projects are not required to make all of their datasets openly accessible. The DMP applies primarily to the data needed to validate the results presented in scientific publications ('underlying data') and also to any other data that the project specifies in their DMP (i.e. projects should specify in their DMP which datasets will be openly accessible).

For data management, EMPIR has to align with Horizon 2020 requirements. The DMP provides a clear plan on how each project intends to manage its research data and EURAMET uses this to monitor the progress and delivery of each project's data management objectives. The DMP also provides an insight into the types and volumes of data that each project will create, as well as contributing to fulfilling the Strategic Research Agenda for Metrology in Europe.

2.2 Opt out

Prior to the issue of the EMPIR Grant Agreement, projects can "opt-out" of the requirement to submit a DMP if any of the following are applicable:

a) Incompatibility with the EMPIR obligation to protect results that are expected to be commercially or industrially exploited

b) Incompatibility with the need for confidentiality in connection with security issues

c) Incompatibility with rules on protecting personal data

d) Incompatibility with the project's main aim

e) If the project will not generate / collect any research data, or

f) If there are other legitimate reasons not to provide open access to research data

The use of a DMP is obligatory for all projects that do not opt-out. Projects that opt-out are also encouraged to prepare a DMP if relevant for their planned research, but the DMP does not need to be submitted to the EURAMET MSU.

2.3 Opt in

2.3.1 Scope

The first DMP should provide an outline of the project's research data management plans. It is expected to evolve during the lifetime of the project with updates being made whenever significant changes arise, such as the generation of new data, changes in data access provisions or curation policies, attainment of tasks (e.g. datasets deposited in a repository, etc.), changes in relevant practices (e.g. new innovation potential, decision to file for a patent), and/or changes in the composition of the consortium. The final DMP is expected to include more detailed answers and to include information on the datasets produced in the project.

2.3.2 When required

The first DMP shall be provided to EURAMET in month one of the project (see the example first DMP in section 4.2). The first periodic DMP shall be provided to EURAMET 60 days after the end of the reporting period (see the example DMPs in sections 4.3-4.4). The final DMP shall be provided to EURAMET 60 days after the end of the project.

2.3.3 Format

The DMP should be in English and submitted to the MSU as a Word file. The font for a DMP is Arial 10. Reporting Template *09 - Data management plan template* should be used.

3 Research data management

3.1 Digital research data

Research data management relates to digital research data, which the EC define as "information in digital form (in particular facts or numbers), collected to be examined and used as a basis for reasoning, discussion or calculation". It includes statistics, results of experiments, measurements, observations from fieldwork, survey results, interview recordings and images.

For EMPIR, data management only applies to digital research data. There are 2 types of digital research data: 1) it is primarily the data needed to validate the results presented in scientific publications and associated metadata; 2) any other data, associated metadata and developed tools/instruments (e.g. software and hardware) specified in the DMP (e.g. unpublished or raw data). Data embargo periods can be specified.

3.2 Data access

3.2.1 Open data

Access to research data should be 'as open as possible, but as closed as necessary' (see section 3.2.2 for exceptions to open access to research data). Open data, and any other tools and instruments needed to re-use or validate the data, must be deposited in a repository (see section 3.6), which provides open access, as soon as possible, after data production/generation or after adequate processing and quality control have taken place. This must be within the deadlines set out in the DMP. In exceptional cases, where specific constraints apply (e.g. security rules), deposition can be delayed. Data underpinning a scientific publication should be deposited as soon as is reasonably possible, and in line with standard community practices.

3.2.2 Closed data

Data can remain closed if it would be against a partner's or the consortium's legitimate interests, including commercial exploitation, or if it would be contrary to any other constraints such as data protection rules, privacy, confidentiality, trade secrets, the EU's competitive interests, security rules, intellectual property rights or if it would be against other obligations under the Grant Agreement. **If open access is not provided, to some or all of the data (e.g. confidential and restricted datasets), this must be explained and justified in the DMP.**

Why datasets may need to remain closed:

- Data which is commercially valuable may be kept closed if making the data open would undermine the exploitation of the data or other results (e.g. if it could endanger trade secrets), or if it would make IP protection of the results more difficult.

- Data protection/privacy rules may mean that certain (sensitive) personal data cannot be made open.
- Security rules may also require closed data.

3.3 Licensing

To enhance the re-usability of research data, they need to be licensed. Three Creative Commons (CC) licenses conform with the "Open Definition" for content and data *(Open Knowledge Foundation, October 2014)*: the Attribution (CC BY) license, requiring attribution of authorship, the Share-Alike (CC BY-SA) license, requiring distribution of derivative works under a license that is identical to that governing the original work, and the Public Domain Dedication (CC0) license, which waives any rights to the data. Equivalent licenses are acceptable.

With the exception of the CC0 license, CC licenses are not appropriate for software, although they can be used for software documentation. Appropriate software licenses, such as those listed as free (https://www.gnu.org/licenses/license-list#SoftwareLicenses) by the Free Software Foundation and listed as open source (https://opensource.org/licenses) by the Open Source Initiative, are strongly recommended.

Partners or researchers, as appropriate, should retain sufficient intellectual property rights to comply with the research data management requirements and to follow the research output management recommendations.

3.4 Metadata requirements

To enhance the context, findability, and potential re-use of data, machine-actionable standardised metadata frameworks are essential. The metadata should be in line with the FAIR principles and it should provide rich information on the datasets (description, date of deposit, author(s), venue and embargo); the EMPIR funding; the project acronym and number; licensing terms; persistent identifiers for the dataset, the authors involved, and, if possible, their organisations and the project.

Metadata must be open access under a Creative Commons Public Domain Dedication (CC0) license, or its equivalent, to the extent that legitimate interests are safeguarded and constraints are taken into account.

3.5 Persistent identifiers

Persistent identifiers (PIDs) are key to ensuring the findability of research outputs, including data. They are globally unique and long-lasting references to digital objects (such as data, publications and other research outputs). Importantly, they must be provided for datasets (e.g. a Digital Object Identifier (DOI) or a handle: see https://www.dpconline.org/handbook/technical-solutions-and-tools/persistent-identifiers).

3.6 Repository requirements

A repository is an online archive, where researchers can deposit digital research data and provide (open) access to them. Repositories help manage and provide access to scientific outputs including data. They also contribute to the long-term preservation of digital assets. Open access repositories assume a central role in EMPIR for the deposition of, and access to, data.

The **datasets** produced by the project should be stored in open access repositories, which can be located using OpenDOAR https://v2.sherpa.ac.uk/opendoar/ or the Registry of Research Data Repositories (http://www.re3data.org/).

There are also some general-purpose open access repositories for multidisciplinary research results including data, software and publications:

- www.zenodo.org – a general-purpose repository for data, software and publications;
- https://figshare.com/ – a repository for any research outputs of all file formats;
- open science Framework (OSF) https://arxiv.org/ – which provides an open source project management tool and repository.

Datasets can also be stored in discipline-specific open access repositories e.g. ELIXIR Deposition Databases (https://elixir-europe.org/platforms/data/elixir-deposition-databases) and ELIXIR Core Data Resources (https://elixir-europe.org/platforms/data/core-data-resources) (repositories recommended for the deposition of life sciences experimental data) or in NMI/DI open access repositories (e.g. https://oar.ptb.de).

Storing datasets on EMPIR project websites, personal websites and databases, publisher websites, and cloud storage services (e.g. Dropbox, Google drive, etc) are not acceptable alternatives to depositing datasets in a repository. Academia.edu, ResearchGate and similar platforms do not allow open access under the terms required and are also not acceptable alternatives to depositing datasets in a repository.

Further information on **Metadata standards** and Research Data Management is available from the FAIRsharing portal (https://fairsharing.org/) with information and resources on data standards, databases, and policies in the life sciences and other scientific disciplines. Data management guidelines and good practices for the Life Sciences are provided by ELIXIR (https://rdmkit.elixir-europe.org/), along with relevant data resources and repositories/databases. For more information on disciplinary metadata standards, visit the Digital Curation Centre (http://www.dcc.ac.uk/). In addition, the Research Data Alliance provides a Metadata Standards Directory (https://www.rd-alliance.org/groups/metadata-standards-directory-working-group.html) that can be searched for discipline-specific standards and associated tools.

Further **guidance on preparing DMPs** is provided by the Research Data Alliance FAIR Data Maturity Model Working Group (https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines-0). They provide a detailed annotated list of indicators to address when increasing the FAIRness of data.

Other resources include the DMPONLINE tool (https://dmponline.dcc.ac.uk/) (supports the development of project DMPs); ARGOS (https://argos.openaire.eu/) (online tool); the Data Stewardship Wizard (https://ds-wizard.org/) (a joint ELIXIR CZ and ELIXIR NL tool, which helps researchers understand what is needed for FAIR-oriented data stewardship). In addition, detailed guidance for drafting and evaluating DMPs is provided in the Science Europe Practical Guide to the International Alignment of Research Data Management (https://www.scienceeurope.org/media/4brkxxe5/se_rdm_practical_guide_extended_final.pdf).

3.7 Procedure for providing open access to research data and subsequent actions

1. **Deposit** the research data, preferably in a research data repository.
2. Provide **open access** by taking measures to enable users to access, mine, exploit, reproduce and disseminate the data free of charge:
   - either attach an appropriate creative commons licence to the data (see section 3.3).
   - or indicate that no licence is needed if the access/use is not subject to any rights.
3. Provide **information**, via the repository, about the **tools and instruments** needed for validating the results.

   Also provide these tools and instruments if possible:
   - specialised software or software code.
   - algorithms.
   - analysis protocols etc.

Once a dataset, complying with the requirements of GA Article 29.3, has been deposited in a repository, coordinators need to do 5 things:

1. First, check that the dataset has a unique and persistent identifier e.g. a DOI. Then…
2. Update the DMP.
3. Submit a link to the dataset to the EURAMET Repository link by completing the online form: https://msu.euramet.org/cgi-bin/be-submit.pl
4. Add the dataset to the output and impact (O&I) report: worksheet 12 on open research data.
5. Finally, ensure that the DMP, the EURAMET Repository Link and the O&I report include the same datasets.

Data Management Plans are only required from projects where they are included in Annex 1 of the Grant Agreement. No EMPIR projects selected in 2014-16 were required to submit DMPs.

4 Completing the Data Management Plan

4.1 Overview

It is important to address the research data management requirements described in Article 29.3 of the EMPIR Grant Agreement, for projects to deposit and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate, free of charge for any user: (1) data needed to validate the results presented in scientific publications ('underlying data'); and (2) other data as specified in the DMP.
A single DMP shall be developed for your project using Reporting Template 9. You should address the questions specified in sections 4.1.2 – 4.1.10 of these guidelines and you should provide a level of detail appropriate to the project. However, where there are specific issues for individual datasets (e.g. regarding open accessibility), you should clearly spell this out. Costs related to data management in EMPIR are eligible for reimbursement during the duration of the project.

Note that a good DMP will follow the guidance, answer all applicable questions, specify which datasets will be open, be straightforward for a non-expert to understand, and provide a plan that is easy to use.

For further guidance, see the data management presentation on the downloads page of the EMPIR participant portal: https://msu.euramet.org/downloads/documents/Data_Management_presentation_coords2.mp4

If your project will generate large numbers of datasets please group them by topic especially if the answers are the same for each group of datasets (see the example in section 4.4 of these guidelines).

4.1.1 Cover page

Please amend the header and footer as appropriate by adding the project number and short name, and the month and year of issue. Also, complete the Grant agreement number, Project short name, Project full title and the number of the data management plan.

4.1.2 Section 1.1: Data Summary

Please address the following questions:
1. What is the purpose of the data collection/generation?
2. What is its relation to the objectives of the project?
3. What types and formats of data will the project generate/collect?
4. Will you re-use any existing data and how?
5. What is the origin of the data?
6. What is the expected size of the data (if known)?
7. Outline who might find it useful ('data utility')?

4.1.3 Section 1.2: Findable, Accessible, Interoperable and Reusable (FAIR) data

Section 1.2.1: Making data findable, including provisions for metadata

Please address the following questions:
8. Are the data produced and/or used in the project discoverable with metadata?
9. Are the data identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?
10. What naming conventions will you follow?
11. Will search keywords be provided that optimise possibilities for re-use?
12. Will you provide clear version numbers?
13. What metadata will be created? If metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

4.1.4 Section 1.2.2: Making data openly accessible

Please address the following questions:
14. Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

Note that in multi-partner projects it is also possible for specific partners to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for *opting out.*

15. How will the data be made accessible (e.g. by deposition in a repository)?

16. What methods or software tools are needed to access the data?
17. Is documentation about the software required in order to access the data included?
18. Is it possible to include the relevant software (e.g. in open source code)?
19. Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories that support open access where possible.
20. Have you explored appropriate arrangements with the identified repository?
21. If there are restrictions on use, how will access be provided?
22. Is there a need for a data access committee?
23. Are there well described conditions for access (i.e. a machine readable license)?
24. How will the identity of the person accessing the data be ascertained?

4.1.5 Section 1.2.3: Making data interoperable

Please address the following questions:

25. Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, that are as far as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?
26. What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?
27. Will you be using standard vocabularies for all of the data types present in your dataset, to allow inter-disciplinary interoperability?
28. If it is essential to use uncommon, or generate project specific, ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

4.1.6 Section 1.2.4: Increase data re-use (through clarifying licences)

Please address the following questions:

29. How will the data be licensed to permit the widest re-use possible?
30. When will the data be made available for re-use? If an embargo is required to allow time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.
31. Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.
32. How long will the data remain re-usable?
33. Are data quality assurance processes described?

4.1.7 Section 1.3: Allocation of resources

Please address the following questions:

34. What are the estimated costs for making data Findable, Accessible, Interoperable and Reusable (FAIR) in your project?
35. How will these costs be covered? Note that costs related to open access to research data are eligible in EMPIR (if compliant with the Grant Agreement conditions).
36. Who will be responsible for data management in your project?
37. What are the costs and potential value of the long term preservation of the data (also state who decides on what data will be kept and for how long)?

4.1.8 Section 1.4: Data security

Please address the following questions:

38. What provisions are in place for data security (including data recovery as well as secure storage and the transfer of sensitive data)?
39. Is the data safely stored in certified repositories for long term preservation and curation?

4.1.9 Section 1.5: Ethical aspects

Please address the following questions:

40. Are there any ethical or legal issues that could impact on data sharing? You can also discuss this in the context of the outcomes of the ethics review and if relevant, include references to ethics report(s) and the ethics section in the Annex 1.

41. Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

4.1.10 Section 1.6: Other issues

Please address the following question:

42. Do you use other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

4.2 Example first data management plan

Example: Cover page

DATA MANAGEMENT PLAN

Grant Agreement number	14IND99
Project short name	MetroShine
Project full title	Metrological approaches for improving the cost efficiency of machine polishing processes in industry
Data management plan	1st ☒ 2nd ☐ 3rd ☐

TABLE OF CONTENTS

Report Status: CO Confidential, only for members of the consortium (including EURAMET and the European Commission Services)

1 Data management plan

1.1 Data summary

Questions	Answers
1 What is the purpose of the data collection/generation?	The data will originate from measurements, calibrations, comparisons and validations. It will be used in meeting the project's objectives and in conference and peer-reviewed publications.
2 What is its relation to the objectives of the project?	Experimental data will be collected by the consortium in order to meet objectives 1 - 4. Measurement and calibration data will result from objectives 1 and 3 and comparison and validation

	data from objectives 2 and 4. Data from questionnaires and market surveys will be used to support end-user uptake (objective 5).
3 What types and formats of data will the project generate/collect?	The majority of the data will be in ASCII (American Standard Code for Information Interchange) data files, eg comma-separated variable (CSV) format, which can be imported into rich-text files for word-processing or into spreadsheets. If specialised software is used, then information about free readers will be provided. Data will be generated in the following formats: • Graphics: jpeg, odg, pdf, png, pttx • Tables: odsu, opj, xlsx • Text: docx, pdf, txt • Other: nb, cpp
4 Will you re-use any existing data and how?	Some of the project's tasks will use existing data (from outside of this project) in pdf, txt and xlsx formats. These data will be used in the validation of the project's results.
5 What is the origin of the data?	The existing data will originate from several sources, which will include: partner's pre-existing data, data from the scientific literature, real-world measurement data and data from simulation experiments. The data collected from domestic properties will remain confidential and will not be included in the repository.
6 What is the expected size of the data *(if known)*?	The expected size of the data is not currently known, but it is likely to be <10 GB with individual files being ≤1 MB.
7 Outline who might find it useful *('data utility')*?	The data will be suitable for use by other research groups working on the following topics: biogas, biomethane, energy gases. It will also be useful for standards committees including ISO/TC193/SC1/WG25 Biomethane Working Group, ISO/TC 158 Analysis of Gases and regulators.

1.2 Findable, Accessible, Interoperable and Reusable (FAIR) Data

1.2.1 Making data findable, including provisions for metadata

Questions	Answers
8 Are the data produced and/or used in the project discoverable with metadata?	Yes, the data produced in the project will be discoverable with metadata. The most important search engines include https://search.datacite.org/ and https://www.base-search.net/
9 Are the data identifiable and locatable by means of a standard identification mechanism *(e.g. persistent and unique identifiers such as Digital Object Identifiers)*?	Yes, the repository (Zenodo) will assign a DOI to each of the project's deposited datasets. In addition, the project's open access peer-reviewed publications will each have a DOI.
10 What naming conventions will you follow?	The following naming conventions will be followed: Folders will be ordered hierarchically and will be clearly named. Files will be uniquely named and versioned: project name, dataset name, laboratory name, time and date.
11 Will search keywords be provided that optimise possibilities for re-use?	Yes, the following search keywords will be provided for use with the deposited datasets (Zenodo): hydrostatic weighing, oscillation-type density meter, density, calibration, traceability.
12 Will you provide clear version numbers?	Clear version numbers and dates will be provided, but it is anticipated that only one version of each dataset will be deposited.
13 What metadata will be created? *If metadata standards do not exist in your discipline, please outline what*	The metadata created for all of the project's datasets will fulfil the repository's (Zenodo) requirement for a minimum set of metadata (i.e. 1) description, 2) creator / ownership, 3) access, 4) lifecycle, 5) persistent identifiers).

type of metadata will be created and how.	

1.2.2 Making data openly accessible

Questions	Answers
14 Which data produced and/or used in the project will be made openly available as the default? *If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.*	All of the data associated with scientific publications will be made openly available as the default unless there is a specific reason not to publish the data. *Datasets which cannot be shared – voluntary restrictions* Other data may be made available on a case-by-case basis if it is relevant for third parties. The following data will not be made publicly available: • 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 partner(s) intellectual property. 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. *Datasets which cannot be shared - legal and contractual reasons* All of the data from the project will be made available, with the exception of market or customer survey data, which are commercially sensitive and cannot be shared.
15 How will the data be made accessible *(eg by deposition in a repository)*?	Once processing, quality control, organisation, analysis and publication are complete, the data will be made accessible by deposition in open access repositories (e.g. Zenodo).
16 What methods or software tools are needed to access the data?	The data will be accessible using the following software: MS Office, Matlab, Mathematica, Origin, Open Office, Adobe Reader, Image Viewer.
17 Is documentation about the software required in order to access the data included?	Standard publicly available software will be used where possible, but if specialist software tools are developed, i.e. created within Matlab, a short text file (e.g. ASCII) will be provided with the data file to explain the software required.
18 Is it possible to include the relevant software *(eg in open source code)*?	The majority of the software programmes are available as commercial products or as freeware. For the software developed in the project, the open source code will be deposited in the repository (e.g. Zenodo).
19 Where will the data and associated metadata, documentation and code be deposited? *Preference should be given to certified repositories that support open access where possible.*	The data and associated metadata, documentation and code will either be deposited in the open access repository called Zenodo (https://zenodo.org) or in PTB's Open Access Repository (https://oar.ptb.de).
20 Have you explored appropriate arrangements with the identified repository?	Yes, PTB's Open Access Repository is functional and it correctly labels datasets with a metadata scheme that is compatible with DataCite.
21 If there are restrictions on use, how will access be provided?	There are no restrictions on the use of the published data, but users will be required to acknowledge the consortium and the source of the data in any resulting publications.
22 Is there a need for a data access committee?	This 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. If necessary, some or all of a potential publication's data will be

	withheld. This will be decided in consultation with the relevant partner(s).
23 Are there well described conditions for access *(i.e. a machine readable license)*?	Yes, Zenodo provides well described conditions for access (see http://about.zenodo.org/policies/).
24 How will the identity of the person accessing the data be ascertained?	Users are required to register to use the repository.

1.2.3 Making data interoperable

Questions	Answers
25 Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (*i.e. adhering to standards for formats, that are as far as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)*?	Yes, the data produced in the project will be interoperable as the datasets will adhere to standardised formats: ASCII, txt, csv, xml, tiff. If MS Office, pdf viewer or image viewer cannot be used, a text (ASCII) file will be provided with the dataset that explains where a free reader can be obtained.
26 What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?	The datasets will be interoperable as Zenodo's basic metadata requirement (i.e. 1) description, 2) creator / ownership, 3) access, 4) lifecycle, 5) persistent identifiers) is compliant with the recommended standards used by DataCite (https://search.datacite.org/) and OpenAIRE (https://www.base-search.net/). Wording will be selected to be compatible with subject-specific vocabularies such as Scitation (American Institute of Physics) and INSPEC (Institution of Engineering and Technology).
27 Will you be using standard vocabularies for all of the data types present in your dataset, to allow inter-disciplinary interoperability?	Standard vocabularies will be used for all datasets: to ensure inter-disciplinary interoperability and re-use.
28 If it is essential to use uncommon, or generate project specific, ontologies or vocabularies, will you provide mappings to more commonly used ontologies?	The compatibility of our project-specific ontologies and vocabularies will be guaranteed through appropriate mapping to more commonly used ontologies.

1.2.4 Increase data re-use (through clarifying licenses)

Questions	Answers
29 How will the data be licensed to permit the widest re-use possible?	The data will either be licensed under a Creative Commons Attribution 4.0 (CC BY 4.0) or a Creative Commons Attribution and ShareAlike 4.0 (CC BY-SA 4.0) license. The software will be released under a GNU-GPL licence. Users will be required to acknowledge the consortium and the source of the data in any resulting publications.
30 When will the data be made available for re-use? *If an embargo is required to allow time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.*	The data used in scientific publications, posters and oral communications will be made available for re-use as soon as is reasonably possible. Some of the data is expected to be subject to an embargo period of 18 months whilst a patent application is pending.

31 Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? *If the re-use of some data is restricted, explain why.*	Any data published in open-access journals will be usable by third parties. The re-use of data that does not relate to peer-reviewed publications will be made available on a case-by-case basis.
32 How long will the data remain re-usable?	The data will remain reusable for the lifetime of the repository, which is expected to be a minimum of 20 years.
33 Are data quality assurance processes described?	Yes, data quality assurance processes are described. Data quality will be assured through repeated and comparison measurements, adherence to standards for data recording, the use of controlled vocabularies and standard terminology, through the metrological characterisation of the measurement set-ups and through the validation of the data collected. Other quality assurance processes will include the provision of test results along with the data and the peer-review of publications based on the data.

1.3 Allocation of resources

Questions	**Answers**
34 What are the estimated costs for making data Findable, Accessible, Interoperable and Reusable (FAIR) in your project?	The estimated costs for making the data Findable, Accessible, Interoperable and Reusable (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 FAIR.
35 How will these costs be covered? *Note that costs related to open access to research data are eligible in EMPIR (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?	The consortium's data access committee will also have overall responsibility for data management. 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).
37 What are the costs and potential value of the long term preservation of the data *(also state who decides on what data will be kept and for how long)*?	There are no costs associated with the long-term preservation of the data. 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 ISO/TC193/SC1/WG25 Biomethane Working Group, ISO/TC 158 Analysis of Gases and regulators. These standards bodies will need access to the data to justify the robustness of future standards. The data will also be of value as it underpins the results of published datasets. The Data Management Committee will decide on what data will be kept and for how long.

1.4 Data security

Questions	**Answers**
38 What provisions are in place for data security *(including data recovery as well as secure storage and the transfer of sensitive data)*?	*Data recovery and secure storage* All partners 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 partners 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. 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. *Transfer of sensitive data* This project will not generate sensitive data.
39 Is the data safely stored in certified repositories for long term preservation and curation?	Yes, the data will be safely stored in the Zenodo open access repository. CERN is working towards ISO certification of the organisational and technical infrastructure which Zenodo relies on for long-term preservation (https://blogs.openaire.eu/?p=1485). 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.

1.5 Ethical aspects

Questions	**Answers**
40 Are there any ethical or legal issues that could impact on data sharing? *You can also discuss this in the context of the outcomes of the ethics review and if relevant, include references to ethics report(s) and the ethics section in the Annex 1.*	There are issues that could impact on data sharing. • Data acquired from third parties, e.g. manufacturers, will not be shared without their explicit consent. • Data collected by the consortium at commercial sites will not be shared without the site owner's explicit consent. • The data from the market surveys will be made anonymous to comply with the General Data Protection Regulation (GDPR). • Ethical issues will be addressed as the project will prepare and submit a report on the Dual Use of the project's results.
41 Is informed consent for data sharing and long-term preservation 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 is collected it will be separated as soon as possible and kept secure.

1.6 Other

Question	**Answer**
42 Do you use other national/funder/sectorial/departmental procedures for data management? If yes, which ones?	Data management will be compliant with the research data policy of EMPIR and with European laws about data security and the protection of privacy (e.g. GDPR).

4.3 Example mid-term periodic data management plan

Example: Cover page

DATA MANAGEMENT PLAN

Grant Agreement number	14IND99
Project short name	MetroShine
Project full title	Metrological approaches for improving the cost efficiency of machine polishing processes in industry
Data management plan	1st ☐ 2nd ☒ 3rd ☐

TABLE OF CONTENTS

Report Status: CO Confidential, only for members of the consortium (including EURAMET and the European Commission Services)

1 Data management plan

1.1 Data summary

Questions	Answers
1 What is the purpose of the data collection/generation?	The data will originate from measurements, calibrations, comparisons and validations. It will be used in meeting the project's objectives and in conference and peer-reviewed publications.

	The project has collected/generated 3 datasets so far. The reasons for their collection/generation are specified below: **1. Illumination control for confocal microscopy – dataset** The data was collected to improve the signal-to-background ratio. **2. Materials testing – biomethane stability data** The biomethane stability data was collected to provide input to a full uncertainty budget. **3. Single photon emission data** The data was collected as it is needed in the quantification of antenna enhancement.
2 What is its relation to the objectives of the project?	Experimental data will be collected by the consortium in order to meet objectives 1 - 4. Measurement and calibration data will result from objectives 1 and 3 and comparison and validation data from objectives 2 and 4. Data from questionnaires and market surveys will be used to support end-user uptake (objective 5). The project's 3 datasets relate to the following objectives: **1. Illumination control for confocal microscopy – dataset** Contributes to meeting objective 1. **2. Materials testing – biomethane stability data** Contributes to meeting objective 3. **3. Single photon emission data** Contributes to meeting objective 4.
3 What types and formats of data will the project generate/collect?	The majority of the data has so far been in ASCII (American Standard Code for Information Interchange) data files, eg comma-separated variable (CSV) format, which can be imported into rich-text files for word-processing or into spreadsheets. Specialised software has not yet been used, but if it is information about free readers will be provided. Data has been generated in the following formats: • Graphics: jpeg, odg, pdf, png, pttx • Tables: odsu, opj, xlsx • Text: docx, pdf, txt • Other: nb, cpp
4 Will you re-use any existing data and how?	Some of the project's tasks will use existing data (from outside of this project) in pdf, txt and xlsx formats. These data will be used in the validation of the project's results. The project has re-used 3 datasets (from outside of this project) so far: **1. Ozone concentrations in ambient air – dataset (xlsx)** This dataset was re-used in A2.3.4 to test and validate a new theoretical model. **2. Selected nacelle test bench data (xlsx)** This dataset was compared with the data obtained in A1.2.3 to see if the new test bench was more effective. **3. Intra-ocular pressure measurement data (txt)**

	This dataset was re-used in A4.5.6 for comparison with simulation results.
5 What is the origin of the data?	The existing data will originate from several sources, which will include: partner's pre-existing data, data from the scientific literature, real-world measurement data and data from simulation experiments. The data collected from domestic properties will remain confidential and will not be included in the repository. The 3 datasets that the project has re-used so far originated from the following external sources (from outside of this project): **1. Ozone concentrations in ambient air – dataset** This dataset was from EMRP JRP ENV99. **2. Selected nacelle test bench data** This dataset is located at https://doi.org/10.100/s12345-123-1234-1. **3. Intra-ocular pressure measurement data** This dataset was from H2020 project 898989 EyeMet.
6 What is the expected size of the data *(if known)*?	The datasets generated so far had individual files of ≤1 MB. The expected size of the new data to be generated is not currently known, but it is likely to be <10 GB with individual files also being ≤1 MB.
7 Outline who might find it useful *('data utility')*?	The data will be suitable for use by other research groups working on the following topics: biogas, biomethane, energy gases. It will also be useful for standards committees including ISO/TC193/SC1/WG25 Biomethane Working Group, ISO/TC 158 Analysis of Gases and regulators. Data has already been provided to the ISO/TC193/SC1/WG25 Biomethane Working Group.

1.2 Findable, Accessible, Interoperable and Reusable (FAIR) Data

1.2.1 Making data findable, including provisions for metadata

Questions	**Answers**
8 Are the data produced and/or used in the project discoverable with metadata?	Yes, the data produced in the project will be discoverable with metadata. The most important search engines include https://search.datacite.org/and https://www.base-search.net/ The project's 3 datasets are all available on Zenodo. Therefore, they are discoverable with the following metadata: 1) description, 2) creator / ownership, 3) access, 4) lifestyle, 5) persistent identifiers.
9 Are the data identifiable and locatable by means of a standard identification mechanism *(eg persistent and unique identifiers such as Digital Object Identifiers)*?	Yes, Zenodo has assigned a DOI to each of the project's deposited datasets. In addition, the project's open access peer-reviewed publications will each have a DOI.
10 What naming conventions will you follow?	The following naming conventions are being followed: Folders are ordered hierarchically and are clearly named. Files are uniquely named and versioned: project name, dataset name, laboratory name, time and date.
11 Will search keywords be provided that optimise possibilities for re-use?	Yes, the following search keywords are being / will be provided for use with the deposited datasets (Zenodo): hydrostatic

	weighing, oscillation-type density meter, density, calibration, traceability.
12 Will you provide clear version numbers?	Clear version numbers and dates are being / will be provided, but it is anticipated that only one version of each dataset will be deposited.
13 What metadata will be created? *If metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.*	The metadata created for all of the project's datasets will fulfil the repository's (Zenodo) requirement for a minimum set of metadata (i.e. 1) description, 2) creator / ownership, 3) access, 4) lifecycle, 5) persistent identifiers).

1.2.2 Making data openly accessible

Questions	**Answers**
14 Which data produced and/or used in the project will be made openly available as the default? *If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.*	All of the data associated with scientific publications will be made openly available as the default unless there is a specific reason not to publish the data. *Datasets which cannot be shared – voluntary restrictions* Other data may be made available on a case-by-case basis if it is relevant for third parties. The following data will not be made publicly available: • 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 partner(s) intellectual property. 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. *Datasets which cannot be shared - legal and contractual reasons* All of the data from the project will be made available, with the exception of market or customer survey data, which are commercially sensitive and cannot be shared.
15 How will the data be made accessible *(eg by deposition in a repository)*?	Once processing, quality control, organisation, analysis and publication are complete, the data will be made accessible by deposition in open access repositories (e.g. Zenodo). The project's 3 datasets are available on Zenodo.
16 What methods or software tools are needed to access the data?	The data will be accessible using the following software: MS Office, Matlab, Mathematica, Origin, Open Office, Adobe Reader, Image Viewer. The project's 3 datasets which have been deposited in Zenodo are accessible using the following software tools: **1. Illumination control for confocal microscopy – dataset** MS office **2. Materials testing – biomethane stability data** Matlab **3. Single photon emission data** Mathmatica

17 Is documentation about the software required in order to access the data included?	Standard publicly available software has been / will be used where possible. Specialist software tools have been / will be created within Matlab and a short text file (e.g. ASCII) has been / will be provided with the data files to explain the software required. The project's 3 datasets which have been deposited in Zenodo require the following documentation about the software: **1. Illumination control for confocal microscopy – dataset** No specialist software is required. **2. Materials testing – biomethane stability data** Matlab – specialist software – an ASCII file has been provided with the dataset in Zenodo explaining the software required. **3. Single photon emission data** No specialist software is required.
18 Is it possible to include the relevant software *(eg in open source code)*?	The majority of the software programmes are available as commercial products or as freeware. For the software developed in the project, the open source code has been / will be deposited in the repository (e.g. Zenodo). The project's 3 datasets which have been deposited in Zenodo include the relevant software (where necessary): **1. Illumination control for confocal microscopy – dataset** No source codes required. **2. Materials testing – biomethane stability data** Open source codes deposited in Zenodo. **3. Single photon emission data** No source codes required.
19 Where will the data and associated metadata, documentation and code be deposited? *Preference should be given to certified repositories that support open access where possible.*	The data and associated metadata, documentation and code will either be deposited in the open access repository called Zenodo (https://zenodo.org) or in PTB's Open Access Repository (https://oar.ptb.de).
20 Have you explored appropriate arrangements with the identified repository?	Yes, PTB's Open Access Repository is functional and it correctly labels datasets with a metadata scheme that is compatible with DataCite.
21 If there are restrictions on use, how will access be provided?	There are no restrictions on the use of the published data, but users will be required to acknowledge the consortium and the source of the data in any resulting publications.
22 Is there a need for a data access committee?	This consortium has a data access committee. Their remit is 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. If necessary, some or all of a potential publication's data will be withheld. This will be decided in consultation with the relevant partner(s).
23 Are there well described conditions for access *(i.e. a machine readable license)*?	Yes, Zenodo provides well described conditions for access (see http://about.zenodo.org/policies/).
24 How will the identity of the person accessing the data be ascertained?	Users are required to register to use the repository.

1.2.3 Making data interoperable

Questions	Answers
25 Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (*i.e. adhering to standards for formats, that are as far as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)*?	Yes, the data produced in the project will be interoperable as the datasets will adhere to standardised formats: ASCII, txt, csv, xml, tiff. If MS Office, pdf viewer or image viewer cannot be used, a text (ASCII) file will be provided with the dataset that explains where a free reader can be obtained.
26 What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?	The datasets will be interoperable as Zenodo's basic metadata requirement (i.e. 1) description, 2) creator / ownership, 3) access, 4) lifecycle, 5) persistent identifiers) is compliant with the recommended standards used by DataCite (https://search.datacite.org/) and OpenAIRE (https://www.base-search.net/). Wording will be selected to be compatible with subject-specific vocabularies such as Scitation (American Institute of Physics) and INSPEC (Institution of Engineering and Technology).
27 Will you be using standard vocabularies for all of the data types present in your dataset, to allow inter-disciplinary interoperability?	Standard vocabularies will be used for all datasets: to ensure inter-disciplinary interoperability and re-use.
28 If it is essential to use uncommon, or generate project specific, ontologies or vocabularies, will you provide mappings to more commonly used ontologies?	The compatibility of our project-specific ontologies and vocabularies will be guaranteed through appropriate mapping to more commonly used ontologies.

1.2.4 Increase data re-use (through clarifying licenses)

Questions	Answers
29 How will the data be licensed to permit the widest re-use possible?	The data will either be licensed under a Creative Commons Attribution 4.0 (CC BY 4.0) or a Creative Commons Attribution and ShareAlike 4.0 (CC BY-SA 4.0) license. The software will be released under a GNU-GPL licence. Users will be required to acknowledge the consortium and the source of the data in any resulting publications. The project's 3 datasets which have been deposited in Zenodo include the following licenses (where necessary): **1. Illumination control for confocal microscopy – dataset** Creative Commons Attribution 4.0 (CC BY 4.0). **2. Materials testing – biomethane stability data** No licence is needed. **3. Single photon emission data** Creative Commons Attribution and ShareAlike 4.0 (CC BY-SA 4.0).
30 When will the data be made available for re-use? *If an embargo is*	The data used in scientific publications, posters and oral communications will be made available for re-use as soon as is

required to allow time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.	reasonably possible. Some of the data is expected to be subject to an embargo period of 18 months whilst a patent application is pending. The project has already deposited 3 datasets in Zenodo (see above). The following 3 datasets will be the next to be deposited in Zenodo: **1. A complete low-uncertainty dataset of T-T_{90}** This dataset will be made available as soon as is reasonably possible i.e. after the end of the embargo period of the following linked publication https://doi.org/10.100/s12345-123-1234-1 (expected August 2019). **2. Airborne Molecular Contamination dataset** As soon as is reasonably possible. **3. Liquid density dataset** Expected February 2020.
31 Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? *If the re-use of some data is restricted, explain why.*	Any data published in open-access journals will be usable by third parties. The re-use of data that does not relate to peer-reviewed publications will be made available on a case-by-case basis. Three of the project's datasets are subject to restrictions on re-use: **1. Dataset on the performance of a Josephson Impedance Bridge** Partner XYZ specified in the Consortium Agreement that they would keep this dataset closed as it is commercially sensitive. **2. Dataset on the 3D imaging of antibacterial agents in bacteria** This dataset does not relate to a peer-reviewed publication. **3. Dataset on the response time of an infusion pump** The consortium's data access committee decided that there would be no benefit from depositing this dataset in a repository as it was generated for comparison with a similar dataset that is already deposited in Zenodo.
32 How long will the data remain re-usable?	The data will remain reusable for the lifetime of the repository, which is expected to be a minimum of 20 years.
33 Are data quality assurance processes described?	Yes, data quality assurance processes are described. Data quality will be assured through repeated and comparison measurements, adherence to standards for data recording, the use of controlled vocabularies and standard terminology, through the metrological characterisation of the measurement set-ups and through the validation of the data collected. Other quality assurance processes will include the provision of test results along with the data and the peer-review of publications based on the data.

1.3 Allocation of resources

Questions	**Answers**
34 What are the estimated costs for making data Findable, Accessible,	The estimated costs for making the data Findable, Accessible, Interoperable and Reusable (FAIR) are 1 000 € (personnel

Interoperable and Reusable (FAIR) in your project?	costs). These costs have been kept to a minimum by using a free repository (Zenodo) and by making only relevant data FAIR.
35 How will these costs be covered? *Note that costs related to open access to research data are eligible in EMPIR (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?	The consortium's data access committee will also have overall responsibility for data management. 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).
37 What are the costs and potential value of the long term preservation of the data *(also state who decides on what data will be kept and for how long)*?	There are no costs associated with the long-term preservation of the data. 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 ISO/TC193/SC1/WG25 Biomethane Working Group, ISO/TC 158 Analysis of Gases and regulators. These standards bodies will need access to the data to justify the robustness of future standards. The data will also be of value as it underpins the results of published datasets. The Data Management Committee will decide on what data will be kept and for how long.

1.4 Data security

Questions	**Answers**
38 What provisions are in place for data security *(including data recovery as well as secure storage and the transfer of sensitive data)*?	*Data recovery and secure storage* All partners 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 partners 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. 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. *Transfer of sensitive data* This project will not generate sensitive data.
39 Is the data safely stored in certified repositories for long term preservation and curation?	Yes, the data will be safely stored in the Zenodo open access repository. CERN is working towards ISO certification of the organisational and technical infrastructure which Zenodo relies on for long-term preservation (https://blogs.openaire.eu/?p=1485). 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.

1.5 *Ethical aspects*

Questions	Answers
40 Are there any ethical or legal issues that could impact on data sharing? *You can also discuss this in the context of the outcomes of the ethics review and if relevant, include references to ethics report(s) and the ethics section in the Annex 1.*	There are issues that could impact on data sharing. • Data acquired from third parties, e.g. manufacturers, will not be shared without their explicit consent. • Data collected by the consortium at commercial sites will not be shared without the site owner's explicit consent. • The data from the market surveys will be made anonymous to comply with the General Data Protection Regulation (GDPR). • Ethical issues will be addressed as the project will prepare and submit a report on the Dual Use of the project's results.
41 Is informed consent for data sharing and long-term preservation 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 is collected it will be separated as soon as possible and kept secure.

1.6 *Other*

Question	Answer
42 Do you use other national/funder/sectorial/departmental procedures for data management? If yes, which ones?	Data management will be compliant with the research data policy of EMPIR and with European laws about data security and the protection of privacy (e.g. GDPR).

4.4 Example mid-term periodic data management plan showing how to group a large number of datasets

Example: Cover page

DATA MANAGEMENT PLAN

Grant Agreement number	14IND99
Project short name	MetroShine
Project full title	Metrological approaches for improving the cost efficiency of machine polishing processes in industry
Data management plan	1st ☐ 2nd ☒ 3rd ☐

TABLE OF CONTENTS

Report Status: CO Confidential, only for members of the consortium (including EURAMET and the European Commission Services)

1. Data management plan

1.1 Data summary

Questions	Answers
1 What is the purpose of the data collection/generation?	The data will originate from measurements, calibrations, comparisons and validations. It will be used in meeting the project's objectives and in conference and peer-reviewed publications. The project has collected/generated 38 datasets so far. These have been grouped by topic as the answers relate to multiple datasets. The reasons for their collection/generation are specified below: **GROUP A** 1. Dataset on amount fractions of n-octane in biomethane – measurement by GC-FID 2. Dataset on amount fractions of α-pinene in biomethane – measurement by GC-FID 3. Dataset on amount fractions of β-pinene in biomethane – measurement by GC-FID 4. Dataset on amount fractions of 3-carene in biomethane – measurement by GC-FID 5. Dataset on amount fractions of limonene in biomethane – measurement by GC-FID 6. Dataset on amount fractions of p-cymene in biomethane – measurement by GC-FID 7. Dataset on amount fractions of n-octane in biomethane – measurement by GC-FID 8. Dataset on amount fractions of n-octane in biomethane – measurement by GC-MS 9. Dataset on amount fractions of α-pinene in biomethane – measurement by GC-MS 10. Dataset on amount fractions of β-pinene in biomethane – measurement by GC-MS 11. Dataset on amount fractions of 3-carene in biomethane – measurement by GC-MS 12. Dataset on amount fractions of limonene in biomethane – measurement by GC-MS 13. Dataset on amount fractions of p-cymene in biomethane – measurement by GC-MS 14. Dataset on amount fractions of n-octane in biomethane – measurement by GC-MS The biomethane stability data was collected to provide input to new measurement standards for terpenes. **GROUP B** 15. Dataset on amount fractions of methyldiethanolamine in biomethane – measurement by TD-GC-MS 16. Dataset on amount fractions of diethanolamine in biomethane – measurement by TD-GC-MS 17. Dataset on amount fractions of monoethanolamine in biomethane – measurement by TD-GC-MS 18. Dataset on amount fractions of piperazine in biomethane – measurement by TD-GC-MS 19. Dataset on amount fractions of diglycolamine in biomethane – measurement by TD-GC-MS

	The biomethane stability data was collected to provide input to new measurement standards for amines. **GROUP C** 20. Dataset on amount fractions of chloromethane in biomethane – measurement by TD-GC-FID/MS 21. Dataset on amount fractions of dichloromethane in biomethane – measurement by TD-GC-FID/MS 22. Dataset on amount fractions of cis-1,2-dichloroethane in biomethane – measurement by TD-GC-FID/MS 23. Dataset on amount fractions of trichloromethane in biomethane – measurement by TD-GC-FID/MS 24. Dataset on amount fractions of trichloroethylene in biomethane – measurement by TD-GC-FID/MS 25. Dataset on amount fractions of 1,2-dichloropropane in biomethane – measurement by TD-GC-FID/MS 26. Dataset on amount fractions of 1,1,2-trichloroethane in biomethane – measurement by TD-GC-FID/MS 27. Dataset on amount fractions of tetrachloroethylene in biomethane – measurement by TD-GC-FID/MS 28. Dataset on amount fractions of trichloro trifluoroethane (freon 113) in biomethane – measurement by TD-GC-FID/MS 29. Dataset on amount fractions of vinyl chloride in biomethane – measurement by TD-GC-FID/MS The biomethane stability data was collected to provide input to new measurement standards for halogenated VOCs. **GROUP D** 30. Dataset on amount fractions of total silicon in biomethane – measurement by GC-FID/MS 31. Dataset on amount fractions of L2 (hexamethyldisiloxane) in biomethane – measurement by GC-FID/MS 32. Dataset on amount fractions of L3 (octamethyltrisiloxane)in biomethane – measurement by GC-FID/MS 33. Dataset on amount fractions of D4 (octamethyltetracyclosiloxane) in biomethane – measurement by GC-FID/MS 34. Dataset on amount fractions of D5 (decamethylpentacyclosiloxane) in biomethane – measurement by GC-FID/MS The biomethane stability data was collected to provide input to new measurement standards for total silicon and siloxanes. **GROUP E** 35. Dataset on amount fractions of ammonia in biomethane – measurement by OFCEAS 36. Dataset on amount fractions of HCl in biomethane – measurement by ion-exchange chromatography 37. Dataset on amount fractions of HF in biomethane – measurement by ion-exchange chromatography 38. Dataset on amount fractions of compressor oil in biomethane – measurement by TI-FTIR The biomethane stability data was collected to provide input to new measurement standards for corrosive impurities and compressor oil.
2 What is its relation to the objectives of the project?	Experimental data will be collected by the consortium in order to meet objectives 1 - 4. Measurement and calibration data will result from objectives 1 and 3 and comparison and validation

	data from objectives 2 and 4. Data from questionnaires and market surveys will be used to support end-user uptake (objective 5). The project's datasets have been grouped by topic as they relate to the same objectives. The project's 38 datasets relate to the following objectives. **GROUPS A,B,C,D Datasets** Contribute to meeting objective 1. **GROUP E Datasets** Contribute to meeting objective 2.
3 What types and formats of data will the project generate/collect?	The majority of the data has so far been in ASCII (American Standard Code for Information Interchange) data files, eg comma-separated variable (CSV) format, which can be imported into rich-text files for word-processing or into spreadsheets. Specialised software has not yet been used, but if it is information about free readers will be provided. Data has been generated in the following formats: • Graphics: jpeg, odg, pdf, png, pttx • Tables: odsu, opj, xlsx • Text: docx, pdf, txt • Other: nb, cpp
4 Will you re-use any existing data and how?	Some of the project's tasks will use existing data (from outside of this project) in pdf, txt and xlsx formats. These data will be used in the validation of the project's results. The project has re-used 3 datasets (from outside of this project) so far: **1. Ozone concentrations in ambient air – dataset (xlsx)** This dataset was re-used in A2.3.4 to test and validate a new theoretical model. **2. Selected nacelle test bench data (xlsx)** This dataset was compared with the data obtained in A1.2.3 to see if the new test bench was more effective. **3. Intra-ocular pressure measurement data (txt)** This dataset was re-used in A4.5.6 for comparison with simulation results.
5 What is the origin of the data?	The existing data will originate from several sources, which will include: partner's pre-existing data, data from the scientific literature, real-world measurement data and data from simulation experiments. The data collected from domestic properties will remain confidential and will not be included in the repository. The 3 datasets that the project has re-used so far originated from the following external sources (from outside of this project): **1. Ozone concentrations in ambient air – dataset** This dataset was from EMRP JRP ENV99. **2. Selected nacelle test bench data** This dataset is located at https://doi.org/10.100/s12345-123-1234-1.

	3. Intra-ocular pressure measurement data This dataset was from H2020 project 898989 EyeMet.
6 What is the expected size of the data *(if known)*?	The datasets generated so far had individual files of ≤1 MB. The expected size of the new data to be generated is not currently known, but it is likely to be <10 GB with individual files also being ≤1 MB.
7 Outline who might find it useful *('data utility')*?	The data will be suitable for use by other research groups working on the following topics: biogas, biomethane, energy gases. It will also be useful for standards committees including ISO/TC193/SC1/WG25 Biomethane Working Group, ISO/TC 158 Analysis of Gases and regulators. Data has already been provided to the ISO/TC193/SC1/WG25 Biomethane Working Group.

1.2 Findable, Accessible, Interoperable and Reusable (FAIR) Data

1.2.1 Making data findable, including provisions for metadata

Questions	Answers
8 Are the data produced and/or used in the project discoverable with metadata?	Yes, the data produced in the project will be discoverable with metadata. The most important search engines include https://search.datacite.org/and https://www.base-search.net/ The project's 38 datasets are all available on Zenodo. Therefore, they are discoverable with the following metadata: 1) description, 2) creator / ownership, 3) access, 4) lifestyle, 5) persistent identifiers.
9 Are the data identifiable and locatable by means of a standard identification mechanism *(eg persistent and unique identifiers such as Digital Object Identifiers)*?	Yes, Zenodo has assigned a DOI to each of the project's deposited datasets. In addition, the project's open access peer-reviewed publications will each have a DOI.
10 What naming conventions will you follow?	The following naming conventions are being followed: Folders are ordered hierarchically and are clearly named. Files are uniquely named and versioned: project name, dataset name, laboratory name, time and date.
11 Will search keywords be provided that optimise possibilities for re-use?	Yes, the following search keywords are being / will be provided for use with the deposited datasets (Zenodo): hydrostatic weighing, oscillation-type density meter, density, calibration, traceability.
12 Will you provide clear version numbers?	Clear version numbers and dates are being / will be provided, but it is anticipated that only one version of each dataset will be deposited.
13 What metadata will be created? *If metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.*	The metadata created for all of the project's datasets will fulfil the repository's (Zenodo) requirement for a minimum set of metadata (i.e. 1) description, 2) creator / ownership, 3) access, 4) lifecycle, 5) persistent identifiers).

1.2.2 Making data openly accessible

Questions	Answers
14 Which data produced and/or used in the project will be made openly available as the default? *If certain datasets cannot be shared (or need to*	All of the data associated with scientific publications will be made openly available as the default unless there is a specific reason not to publish the data. *Datasets which cannot be shared – voluntary restrictions*

be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.	Other data may be made available on a case-by-case basis if it is relevant for third parties. The following data will not be made publicly available: • 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 partner(s) intellectual property. 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. *Datasets which cannot be shared - legal and contractual reasons* All of the data from the project will be made available, with the exception of market or customer survey data, which are commercially sensitive and cannot be shared.
15 How will the data be made accessible *(eg by deposition in a repository)*?	Once processing, quality control, organisation, analysis and publication are complete, the data will be made accessible by deposition in open access repositories (e.g. Zenodo). The project's 38 datasets are available on Zenodo.
16 What methods or software tools are needed to access the data?	The data will be accessible using the following software: MS Office, Matlab, Mathematica, Origin, Open Office, Adobe Reader, Image Viewer. The project's datasets have been grouped by topic. The project's 38 datasets which have been deposited in Zenodo are accessible using the following software tools: **GROUPS A,B,C,D Datasets** MS office **GROUP E Datasets** Matlab
17 Is documentation about the software required in order to access the data included?	Standard publicly available software has been / will be used where possible. Specialist software tools have been / will be created within Matlab and a short text file (e.g. ASCII) has been / will be provided with the data files to explain the software required. The project's datasets have been grouped by topic. The project's 38 datasets which have been deposited in Zenodo require the following documentation about the software: **GROUPS A,B,C,D Datasets** No specialist software is required. **GROUP E Datasets** Matlab – specialist software – an ASCII file has been provided with each of the datasets in Zenodo explaining the software required.
18 Is it possible to include the relevant software *(eg in open source code)*?	The majority of the software programmes are available as commercial products or as freeware. For the software developed in the project, the open source code has been / will be deposited in the repository (e.g. Zenodo).

	The project's datasets have been grouped by topic. The project's 38 datasets which have been deposited in Zenodo include the relevant software (where necessary): **GROUPS A,B,C,D Datasets** No source codes required. **GROUP E Datasets** Open source codes deposited in Zenodo.
19 Where will the data and associated metadata, documentation and code be deposited? *Preference should be given to certified repositories that support open access where possible.*	The data and associated metadata, documentation and code will either be deposited in the open access repository called Zenodo (https://zenodo.org) or in PTB's Open Access Repository (https://oar.ptb.de).
20 Have you explored appropriate arrangements with the identified repository?	Yes, PTB's Open Access Repository is functional and it correctly labels datasets with a metadata scheme that is compatible with DataCite.
21 If there are restrictions on use, how will access be provided?	There are no restrictions on the use of the published data, but users will be required to acknowledge the consortium and the source of the data in any resulting publications.
22 Is there a need for a data access committee?	This consortium has a data access committee. Their remit is 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. If necessary, some or all of a potential publication's data will be withheld. This will be decided in consultation with the relevant partner(s).
23 Are there well described conditions for access *(i.e. a machine readable license)*?	Yes, Zenodo provides well described conditions for access (see http://about.zenodo.org/policies/).
24 How will the identity of the person accessing the data be ascertained?	Users are required to register to use the repository.
1.2.3 Making data interoperable	
Questions	**Answers**
25 Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (*i.e. adhering to standards for formats, that are as far as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)*?	Yes, the data produced in the project will be interoperable as the datasets will adhere to standardised formats: ASCII, txt, csv, xml, tiff. If MS Office, pdf viewer or image viewer cannot be used, a text (ASCII) file will be provided with the dataset that explains where a free reader can be obtained.
26 What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?	The datasets will be interoperable as Zenodo's basic metadata requirement (i.e. 1) description, 2) creator / ownership, 3) access, 4) lifecycle, 5) persistent identifiers) is compliant with the recommended standards used by DataCite (https://search.datacite.org/) and OpenAIRE (https://www.base-search.net/). Wording will be selected to be compatible with subject-specific vocabularies such as Scitation (American Institute of Physics) and INSPEC (Institution of Engineering and Technology).
27 Will you be using standard vocabularies for all of the data types	Standard vocabularies will be used for all datasets: to ensure inter-disciplinary interoperability and re-use.

present in your dataset, to allow inter-disciplinary interoperability?	
28 If it is essential to use uncommon, or generate project specific, ontologies or vocabularies, will you provide mappings to more commonly used ontologies?	The compatibility of our project-specific ontologies and vocabularies will be guaranteed through appropriate mapping to more commonly used ontologies.

1.2.4 Increase data re-use (through clarifying licenses)

Questions	Answers
29 How will the data be licensed to permit the widest re-use possible?	The data will either be licensed under a Creative Commons Attribution 4.0 (CC BY 4.0) or a Creative Commons Attribution and ShareAlike 4.0 (CC BY-SA 4.0) license. The software will be released under a GNU-GPL licence. Users will be required to acknowledge the consortium and the source of the data in any resulting publications. The project's datasets have been grouped by topic. The project's 38 datasets which have been deposited in Zenodo include the following licenses (where necessary): **GROUPS A,B,C,D Datasets** Creative Commons Attribution 4.0 (CC BY 4.0). **GROUP E Datasets** No licence is needed.
30 When will the data be made available for re-use? *If an embargo is required to allow time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.*	The data used in scientific publications, posters and oral communications will be made available for re-use as soon as is reasonably possible. Some of the data is expected to be subject to an embargo period of 18 months whilst a patent application is pending. The project has already deposited 38 datasets in Zenodo (see above). The following 3 datasets will be the next to be deposited in Zenodo: **1. A complete low-uncertainty dataset of T-T_{90}** This dataset will be made available as soon as is reasonably possible i.e. after the end of the embargo period of the following linked publication https://doi.org/10.100/s12345-123-1234-1 (expected August 2019). **2. Airborne Molecular Contamination dataset** As soon as is reasonably possible. **3. Liquid density dataset** Expected February 2020.
31 Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? *If the re-use of some data is restricted, explain why.*	Any data published in open-access journals will be usable by third parties. The re-use of data that does not relate to peer-reviewed publications will be made available on a case-by-case basis. Three of the project's datasets are subject to restrictions on re-use: **1. Dataset on the performance of a Josephson Impedance Bridge**

	Partner XYZ specified in the Consortium Agreement that they would keep this dataset closed as it is commercially sensitive. **2. Dataset on the 3D imaging of antibacterial agents in bacteria** This dataset does not relate to a peer-reviewed publication. **3. Dataset on the response time of an infusion pump** The consortium's data access committee decided that there would be no benefit from depositing this dataset in a repository as it was generated for comparison with a similar dataset that is already deposited in Zenodo.
32 How long will the data remain re-usable?	The data will remain reusable for the lifetime of the repository, which is expected to be a minimum of 20 years.
33 Are data quality assurance processes described?	Yes, data quality assurance processes are described. Data quality will be assured through repeated and comparison measurements, adherence to standards for data recording, the use of controlled vocabularies and standard terminology, through the metrological characterisation of the measurement set-ups and through the validation of the data collected. Other quality assurance processes will include the provision of test results along with the data and the peer-review of publications based on the data.

1.3 Allocation of resources

Questions	**Answers**
34 What are the estimated costs for making data Findable, Accessible, Interoperable and Reusable (FAIR) in your project?	The estimated costs for making the data Findable, Accessible, Interoperable and Reusable (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 FAIR.
35 How will these costs be covered? *Note that costs related to open access to research data are eligible in EMPIR (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?	The consortium's data access committee will also have overall responsibility for data management. 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).
37 What are the costs and potential value of the long term preservation of the data *(also state who decides on what data will be kept and for how long)*?	There are no costs associated with the long-term preservation of the data. 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 ISO/TC193/SC1/WG25 Biomethane Working Group, ISO/TC 158 Analysis of Gases and regulators. These standards bodies will need access to the data to justify the robustness of future standards. The data will also be of value as it underpins the results of published datasets. The Data Management Committee will decide on what data will be kept and for how long.

1.4 Data security

Questions	Answers
38 What provisions are in place for data security *(including data recovery as well as secure storage and the transfer of sensitive data)*?	*Data recovery and secure storage* All partners 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 partners 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. 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. *Transfer of sensitive data* This project will not generate sensitive data.
39 Is the data safely stored in certified repositories for long term preservation and curation?	Yes, the data will be safely stored in the Zenodo open access repository. CERN is working towards ISO certification of the organisational and technical infrastructure which Zenodo relies on for long-term preservation (https://blogs.openaire.eu/?p=1485). 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.

1.5 Ethical aspects

Questions	Answers
40 Are there any ethical or legal issues that could impact on data sharing? *You can also discuss this in the context of the outcomes of the ethics review and if relevant, include references to ethics report(s) and the ethics section in the Annex 1.*	There are issues that could impact on data sharing. • Data acquired from third parties, e.g. manufacturers, will not be shared without their explicit consent. • Data collected by the consortium at commercial sites will not be shared without the site owner's explicit consent. • The data from the market surveys will be made anonymous to comply with the General Data Protection Regulation (GDPR). • Ethical issues will be addressed as the project will prepare and submit a report on the Dual Use of the project's results.
41 Is informed consent for data sharing and long-term preservation 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 is collected it will be separated as soon as possible and kept secure.

1.6 Other

Question	Answer
42 Do you use other national/funder/sectorial/departmental procedures for data management? If yes, which ones?	Data management will be compliant with the research data policy of EMPIR and with European laws about data security and the protection of privacy (e.g. GDPR).

5 Submissions to the EURAMET Repository Link

Each partner must ensure open access (free of charge, online access for any user) to all peer-reviewed scientific publications relating to its results and also to all relevant data if your project has “opted in” to data management in accordance with Article 29 of the EMPIR Grant Agreement.

Coordinators should make submissions of links to their publications and data to the EURAMET Repository link by completing the online form: https://msu.euramet.org/cgi-bin/be-submit.pl

Authors can find information about this here:
http://msu.euramet.org/downloads/documents/MSU_repository_instructions.pdf