

EMPIR Contracts - Reporting Guidelines
Part 9 – Preparing data management plans

Document: P-CON-GUI-109
Approved: Programme Manager

Version: 1.1
2018-07-12



EMPIR Reporting Guidelines
Part 9 – Preparing data management plans

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.

This report addresses the requirements set out 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.

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.

2 Overview of the report

2.1 Purpose

The overall purpose of a DMP is to make research data Findable, Accessible, Interoperable and Reusable (FAIR) in order to ensure that it is soundly managed. It will describe the data management plans for all of the data sets 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).

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 submit a DMP if relevant for their planned research.

2.3 When required

The first DMP shall be provided to EURAMET in month one of the project (NB it is not necessary to provide detailed answers to all the questions in the first DMP). Intermediate DMPs shall be provided at the interim reporting times specified in Annex 1 of the EMPIR Grant Agreement (if interim reports are required). However, these intermediate DMPs only need to be provided if updates are necessary or if significant changes arise in the project such as, new data, changes in consortium policies (eg new innovation potential, or a decision to file for a patent), or changes in the consortium composition and external factors. The first periodic DMP shall be provided to EURAMET 60 days after the end of the reporting period. The final DMP shall be provided to EURAMET 60 days after the end of the project. Note that the DMP is expected to evolve during the lifetime of the project with information becoming more detailed as the project progresses and when significant changes occur.

2.4 Data management plan

A single DMP shall be developed for your project using Reporting Template 9. You should address the questions specified in sections 2.4.2 – 2.4.10 of these guidelines and you should provide a level of detail appropriate to the project. However, where there are specific issues for individual datasets (eg regarding openness), you should clearly spell this out. Costs related to data management in EMPIR are eligible for reimbursement during the duration of the project.

The data sets produced by the project should be stored in suitable repositories, which can be located using the Registry of Research Data Repositories (<http://www.re3data.org/>). Some repositories like Zenodo (<https://zenodo.org/> - an OpenAIRE and CERN collaboration), allow researchers to deposit both publications and data, while providing tools to link them. Further information on research data management is available from the Digital Curation Centre (<http://www.dcc.ac.uk/dmponline>) and from ScienceMatters (<https://www.sciencematters.io/>). In addition, the Research Data Alliance provides a Metadata Standards Directory (<http://rd-alliance.github.io/metadata-directory/>) that can be searched for discipline-specific standards and associated tools, and the EUDAT B2SHARE (<https://b2share.eudat.eu/>) tool includes a built-in license wizard that facilitates the selection of an adequate license for research data. However, please note that these services may be more suited to projects funded under Horizon 2020 rather than those funded under EMPIR.

2.4.1 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.

2.4.2 Section 1.1: Data Summary

Please address the following questions:

- What is the purpose of the data collection/generation?
- What is its relation to the objectives of the project?
- What types and formats of data will the project generate/collect?
- Will you re-use any existing data and how?
- What is the origin of the data?
- What is the expected size of the data (if known)?
- Outline who might find it useful ('data utility')?

2.4.3 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:

- Are the data produced and/or used in the project discoverable with metadata?
- Are the data identifiable and locatable by means of a standard identification mechanism (eg persistent and unique identifiers such as Digital Object Identifiers)?
- What naming conventions will you follow?
- Will search keywords be provided that optimise possibilities for re-use?
- Will you provide clear version numbers?
- 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.

2.4.4 Section 1.2.2: Making data openly accessible

Please address the following questions:

- 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*.

- How will the data be made accessible (eg by deposition in a repository)?
- What methods or software tools are needed to access the data?
- Is documentation about the software required in order to access the data included?

- Is it possible to include the relevant software (eg in open source code)?
- 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.
- Have you explored appropriate arrangements with the identified repository?
- If there are restrictions on use, how will access be provided?
- Is there a need for a data access committee?
- Are there well described conditions for access (i.e. a machine readable license)?
- How will the identity of the person accessing the data be ascertained?

2.4.5 Section 1.2.3: Making data interoperable

Please address the following questions:

- 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)?
- What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?
- Will you be using standard vocabularies for all of the data types present in your data set, to allow interdisciplinary interoperability?
- If it is essential to use uncommon, or generate project specific, ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

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

Please address the following questions:

- How will the data be licensed to permit the widest re-use possible?
- 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.
- 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.
- How long will the data remain re-usable?
- Are data quality assurance processes described?

2.4.7 Section 1.3: Allocation of resources

Please address the following questions:

- What are the estimated costs for making data Findable, Accessible, Interoperable and Reusable (FAIR) in your project?
- 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).
- Who will be responsible for data management in your project?
- 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)?

2.4.8 Section 1.4: Data security

Please address the following questions:

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

2.4.9 Section 1.5: Ethical aspects

Please address the following questions:

- 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.

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


2.4.10 Section 1.6: Other issues

Please address the following question:

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

2.5 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 |
| Technical Report (Progress) | 1 st <input checked="" type="checkbox"/> 2 nd <input type="checkbox"/> 3 rd <input type="checkbox"/> 4 th <input type="checkbox"/> 5 th <input type="checkbox"/> 6 th <input type="checkbox"/> |

TABLE OF CONTENTS

| | | |
|----------|--|----------|
| 1 | Data management plan | 6 |
| 1.1 | Data summary | 6 |
| 1.2 | Findable, Accessible, Interoperable and Reusable (FAIR) Data | 6 |
| 1.2.1 | Making data findable, including provisions for metadata | 6 |
| 1.2.2 | Making data openly accessible | 7 |
| 1.2.3 | Making data interoperable | 8 |
| 1.2.4 | Increase data re-use (through clarifying licenses) | 8 |
| 1.3 | Allocation of resources | 9 |
| 1.4 | Data security | 10 |
| 1.5 | Ethical aspects | 10 |
| 1.6 | Other | 10 |

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 |
| 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. |
| 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). |
| What types and formats of data will the project generate/collect? | <p>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.</p> <p>Data will be generated in the following formats:</p> <ul style="list-style-type: none"> • Graphics: jpeg, odg, pdf, png, pttx • Tables: odsu, opj, xlsx • Text: docx, pdf, txt • Other: nb, cpp |
| Will you re-use any existing data and how? | Some of the project's tasks will use existing data in hdf, txt and xlsx formats. These data will be used in the validation of the project's results. |
| 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. |
| 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. |
| 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 |
| 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/ |
| Are the data identifiable and locatable by means of a standard identification mechanism (eg persistent and unique identifiers such as Digital Object Identifiers)? | <p>Yes, the repository (Zenodo) will assign a DOI to each of the project's deposited datasets.</p> <p>In addition, the project's open access peer-reviewed publications will each have a DOI.</p> |

| | |
|--|---|
| 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. |
| 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. |
| 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. |
| What metadata will be created? <i>If metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.</i> | 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 |
| Which data produced and/or used in the project will be made openly available as the default? <i>If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.</i> | 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. <i>Datasets which cannot be shared – voluntary restrictions</i> 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 publically available: <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 publically 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. <i>Datasets which cannot be shared - legal and contractual reasons</i> 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. |
| 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 (eg Zenodo). |
| 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. |
| 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 (eg ASCII) will be provided with the data file to explain the software required. |
| 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 source code will deposited in the repository (eg Zenodo). |
| Where will the data and associated metadata, documentation and code | The data and associated metadata, documentation and code will either be deposited in the open access repository called Zenodo |

| | |
|--|---|
| be deposited? <i>Preference should be given to certified repositories that support open access where possible.</i> | (https://zenodo.org) or in PTB's Open Access Repository (https://oar.ptb.de). |
| 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. |
| 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. |
| 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). |
| Are there well described conditions for access (<i>i.e. a machine readable license</i>)? | Yes, Zenodo provides well described conditions for access (see http://about.zenodo.org/policies/). |
| 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 |
|--|--|
| Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (<i>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</i>)? | 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. |
| 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>i.e.</i> 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). |
| Will you be using standard vocabularies for all of the data types present in your data set, to allow inter-disciplinary interoperability? | Standard vocabularies will be used for all datasets: to ensure inter-disciplinary interoperability and re-use. |
| 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 |
| 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. |
| When will the data be made available for re-use? <i>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.</i> | 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. |
| Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? <i>If the re-use of some data is restricted, explain why.</i> | 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. |
| How long will the data remain reusable? | The data will remain reusable for the lifetime of the repository, which is expected to be a minimum of 20 years. |
| 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 |
| 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. |
| How will these costs be covered? <i>Note that costs related to open access to research data are eligible in EMPIR (if compliant with the Grant Agreement conditions).</i> | 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. |
| 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). |
| What are the costs and potential value of the long term preservation of the data <i>(also state who decides on what data will be kept and for how long)?</i> | 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 |

| | |
|---|---|
| | <p>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.</p> <p>The Data Management Committee will decide on what data will be kept and for how long.</p> |
| 1.4 Data security | |
| Questions | Answers |
| What provisions are in place for data security (including data recovery as well as secure storage and the transfer of sensitive data)? | <p><i>Data recovery and secure storage</i> 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.</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><i>Transfer of sensitive data</i> This project will not generate sensitive data.</p> |
| Is the data safely stored in certified repositories for long term preservation and curation? | <p>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).</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> |
| 1.5 Ethical aspects | |
| Questions | Answers |
| 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. | <p>There are issues that could impact on data sharing.</p> <ul style="list-style-type: none"> • Data acquired from third parties, eg 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. |
| Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data? | <p>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.</p> |

| 1.6 Other | |
|---|--|
| Question | Answer |
| 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 (eg GDPR). |