



Deliverable D8.11

Data Management Plan (DMP)



Document control sheet

Project	Fit4Micro – Clean and efficient microCHCP by microturbine based hybrid systems
Grant Agreement n°	101083536
Coordinator	MITIS
Work package n°	8
Work package title	Dissemination, Communication and Exploitation
Work package Leader	ETA Florence
Deliverable	D8.11
Title	Data Management Plan (DMP)
Version	1
Lead Beneficiary	ETA Florence
Author	Teresa Ridolfi, Stefano Capaccioli
Reference period	01/10/2022 – 31/03/2023
Due date	31/03/2023 (Month 6)
Submission date	20/03/2023
Dissemination Level	PU - Public

Funded by the European Union, Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.



History of changes

27.02.2023	Draft of the first version -> coordinator review;
01.02.2023	First update of the draft version;
20.03.2023	Partners review -> final version.

Table of contents

Executive Summary	4
Introduction	4
1. Objectives of Data Management Plan	5
2. Data Summary	5
3. Open access to scientific publications	6
4. FAIR Data	6
4.1 Making data findable, including provisions for metadata.....	6
4.2 Making data accessible.....	6
4.3 Making data interoperable.....	7
4.4 Increase data re-use	7
5. Data security	8
6. Data management responsibilities and resources	8
Conclusions	8
References	9
Definitions	9
Annex I – Data Collection Form	10
Table 1. Data collection form	10
Table 2. First project dataset description -> to regularly verify and update	11



Executive Summary

This deliverable describes the Data Management Plan (DMP) and includes all necessary processes and procedures to ensure and cover the full life cycle of data generated during the project. It describes the different types of data including the procedures how those are being collected, stored and shared.

The DMP guarantees that data follows the FAIR principle, meaning that data is findable, accessible, interoperable and re-usable.

- To ensure the data is **Findable** wherever possible existing metadata standards are followed, including multiple unique identifiers (doi, arxiv, isbn, issn).
- To ensure that data is **Accessible**, data will be stored in the official Zenodo repository. To ensure the data file itself are accessible, file types that have a high chance of remaining usable in the future, for example .txt and .csv will be used.
- To ensure the data is **Interoperable**, the data is stored syntactically parseable (grammatically broken down) to allow data exchange between researchers, institutions, organizations and countries.
- To ensure the data is **Reusable**, documentation will be provided to facilitate data re-use (e.g., readme files with information on methodology, analysis, etc). Also, open-source data types (e.g. rtf) are used wherever possible to have the least impact of restrictive licences.

Furthermore, the DMP also distinguishes and determines data that will not be open access (e.g. data linked to exploitable results/commercialisation) and the data that will be open access, such as an open data repository.

Introduction

The Data Management Plan (DMP) describes how data generated during the project will be collected and processed in both, during and after the project as well as the methodology and standard procedures applied. Basically, the DMP describes the data management life cycle for the data to be collected, processed and/or generated during the project.

This deliverable is linked to task 8.2, “Communication and Dissemination activities”. The main aim of this task is the one-off promoting and disseminating the project’s main activities and results with a series of communication activities. Within this task, two main deliverables will be implemented: The Dissemination, Communication and Exploitation plan and the Data Management Plan.

In order to follow the guidelines of the European Commission on FAIR Data Management, data has to be findable, accessible, interoperable and re-usable (FAIR).

Among others, following information should be included:

- What kind/type of data will be collected, processed and/or generated?
- Which methodology/standards will be applied?
- Will data be shared/made open access?
- How will the data be preserved?

If necessary, the DMP will be revised and updated during the project period.



1. Objectives of Data Management Plan

The Data Management Plan (DMP) is a document prepared by the project's consortium that aims at preserving and storing the data over a period of at least 5 years, in order to ensure their availability and re-usability during and after the end of the Fit4Micro project. In case the information produced has high relevance, a long term storage policy will be taken into account.

Keeping in mind that there must be a meaningful use, re-use or combination of data and that they are well documented. Datasets will be stored in file formats that have a high chance of remaining usable in the far future. In order to support the discoverability of data, a suitable data-platform like Zenodo will be selected.

The DMP will take into account the latest "Guidelines on FAIR Data Management" and the generated and gathered research data needs to be preserved in-line with the EC requirements.

This document will play a crucial role in (further) exploitation, verification of the research results and should be therefore soundly and effectively managed.

2. Data Summary

Within this project, the production of data will go in parallel with the achievement of project's results, hence activities in many project Work packages (WPs).

Another part of data will be data processed by means of statistical analysis.

Regarding the collection of data, a simple relational database will be used for raw data and processed data. Concerning data ownership, it will be documented within the database who is the provider and owner of the data and who will receive the data for further processing.

Data will be mostly collected in the form of .xls, .csv, image (.gif, .bmp, .jpeg) and textual information.

In the course of qualitative interviews, audio files may also be collected with a declaration of consent.

Possible data formats used are .pdfs, .doc, .xls, .csv, .txt or .rdt files as well as .mp3, .mpeg, .avi, .mov for video data. Together with these formats, data for simulations will be collected in their relative formats, such as DWG and/or DXF for CAD files for prototype designation and so on.

Each partner will provide a table with the potential data that they have generated (see the table in Annex I).

Personal data gained from interviews will be pseudonymized for storage, whereby the master table containing the mapping between raw identifiers and converted data will be stored separately than where the database is located. Sensitive data will be hashed to mask sensitive personal data.

Storage space required for this project will be estimated exceeding 1 TB and will be stored on a project internal volume for every partner. The shared platform chosen by the Consortium is Microsoft Sharepoint, which will be managed by the project coordinator and WP1 (Project management) leader.



3. Open access to scientific publications

The Fit4Micro openly science practices will be accessed, mined, exploited, reproduced and disseminated free of charge for the users, as explained in art. 17 of the Grant Agreement. The authors of produced results will ensure access to peer-reviewed scientific publications by depositing the publications in the project's official repository (Zenodo).

Beneficiaries can select one of the two main routes towards open access to publications, both equally valid:

1. **Green open access (self-archiving):** The published work or the final peer-reviewed manuscript that has been accepted for publication is made freely and openly accessible by the author, or a representative, in an online repository. Some publishers request that open access be granted only after an embargo period has elapsed.
2. **Gold open access (open access publishing):** The published work is made available in open access mode by the publisher immediately upon publication. The most common business model is based on one-off payments by authors (commonly called APCs – article processing charges – or BPCs – book processing charges). The costs of gold open access publications are eligible costs that can be charged against ERC grants, provided the costs are incurred during the duration of the project.

4. FAIR Data

This document, coherently with the EU guidelines, describes the data management procedures according to FAIR (findable, accessible, interoperable, and re-usable) data management.

4.1 Making data findable, including provisions for metadata

Metadata formats such as .xml (or .json) will be used in order to exchange data in between different relevant domain databases.

All raw data described as primary data and pre-processed data described as secondary data will be accompanied with a textual documentation that will be organized in a folder structured way to ensure data quality.

Documentation needed for the re-use of data will be supplied in text form. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, and so on.

Data for different WPs together with personal data will be stored and backed up at WP Leaders and project's partners servers. This storage will be managed by internal partner IT staff.

If required, a version control system for all folders and files will be used to ensure that data is not manipulated.

Acquired raw data and post-processed data will not be stored locally at all and will be synched on the volume of the relevant partner.

4.2 Making data accessible

A data repository (Zenodo) as mentioned above will be established as basis for the international community to foster further feedback on methods and acquired datasets above all.



Regarding Data protection it will be distinguished between general data (names and contact details of partners) and research data.

The open data repository will be maintained after the runtime of the project and is further planned to include the metadata standard .xml which is suitable and accepted in the community for sharing of data in the field.

In order to access data, it is planned to make it barrier-free and guarantee a high usability. To establish a user-friendly access to the acquired data sets, it is planned to use at first Microsoft Sharepoint and at a later stage a simple database with a user friendly front-end. Usage of database will help to identify a logical structure within the data, maintain data in a structured way, identify relations between datasets and query datasets to extract knowledge.

A mechanism ensuring the quality of data shall be established meaning that only data within the repository shall be processed by externals and shared which fulfils the community requirements and standards.

The interaction with the repository will be directly via upload, search and filter mechanisms established. A detailed plan of the data repository will be created within the second year of the project.

The main risk is a data breach regarding personal data, which will be taken care of by means of security measures within the coordinator infrastructure. In the event of an incident, back-up data (a copy which is not stored in the same place as the origin) can be requested by the IT support team at any time.

The data protection policy "Privacy by Default" which is required by the EU GDPR is in place.

4.3 Making data interoperable

All datasets will be described using standard descriptive metadata, such as Dublin Core and DataCite Metadata Schema, for ensuring metadata interoperability for indexing and discoverability.

All relevant documentation explaining codebooks, users' manuals, data collection procedures and analysis will be made available along with the data, in compliance with requirements of intelligibility, reproducibility and validation of project findings.

Variable names of data derived from other official sources (e.g. Eurostat, others...), will be consistent with the original source names.

4.4 Increase data re-use

In Fit4Micro, all shareable data will be shared by adopting licenses that allow re-use of the data and of datasets in their entirety by other researchers/scholars and stakeholders of policy and market sectors.

The EC Open Access policy asks researchers to make available in Open Access their peer reviewed articles.

This is easiest to comply with when the researcher retains his/her copyright and only gives the publisher of the article a licence to publish. In that case, the article can be deposited in a repository and made publicly accessible without further permission of the publisher. If the licence stipulates an embargo period, of course that must be respected.

Creative Commons licences are available in many countries for authors who wish to retain their copyright and provide their publisher with a licence.

5. Data security

When dealing with personal data the EU GDPR applies, especially for the protection of the user rights (article 12-23). In case personal information is processed, for example during expert interviews, informed consent needs to be given by the participator for preservation and/or sharing of personal data.

As mentioned, pseudonymisation of personal data and encryption is selected as a security measure (the encryption key must be stored separately from the data, for instance by a trusted third party). Based on a permission scheme (which needs to be settled in an excel list indicating who can access what) so that only authorised users can get access to personal data.

In general, all personal data will be processed within the limitations given in the EU GDPR and comply to the regulation.

The owner of the data will be the relevant partner who has acquired the corresponding raw data at hand. The ownership of the data will be noted in detail with every dataset in the header of the dataset or within a central file.

All internal partners are allowed to access and further process data which they provide and will be provided with. There will be no centrally organized ownership, meaning that each partner has the full responsibility for his data.

In case another partner is processing data and bringing in new IP to further process this data (model or processes in the context) the IP will stay at the partner who has brought in the IP.

At a later stage it will be evaluated whether data, processes or models out of this project can be integrated in a business model. Evaluation of feasibility of this will be done. The re-use of third-party data is not planned at the moment.

6. Data management responsibilities and resources

A data protection officer along with the project's coordinator will be responsible for the management of the data in regard to ownership and sovereignty, access rights to partners and non-partners, data protection (EU GDPR) and other legal restrictions.

Data within the open data repository shall be barrier-free meaning that people with disabilities can access the open data repository.

The coordinator is also responsible for sustainability issues regarding data management and data processing.

Additionally, the re-use and further maintenance of a data repository after the project has been carried out successfully will be considered.

If necessary, the DMP will be updated by partners and the coordinator during the project implementation.

Conclusions

Data Management processes, procedures and tools have been set up for the Fit4Micro project to establish an appropriate and efficient management as well as to ensure that requirements are met.



References

- https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm#A1-template
- Horizon Europe, Data Management Plan Template, Version 1.0, 05 May 2021
- Horizon Europe (HORIZON), Model Grant Agreement, Unit Grants, (HE Unit MGA — Multi & Mono), Version 1.1, 15 December 2021
- European Research Council (ERC), Guidelines on Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020, Version 1.1, 21 April 2017
- Open Research Data and Data Management Plans, Information for ERC grantees by the ERC Scientific Council, Version 4.1, 20 April 2022

Definitions

- Primary data — They are the most basic form of data. The data gathered in the first hand by a researcher is called primary data and considered to be real-time and actual data. Examples: data collected for the first time in e.g., various experiments, observations, personal interviews, questionnaires etc.
- Secondary data — Secondary Data is an advanced form of data. The data that is gathered again by a new researcher after it was originally gathered in the past is considered as secondary data. In other words, secondary data can also be referred to as known data. Examples: data is collected through already published journals, research articles, books, websites, government records and periodicals, internal and external documented records, etc. A common example includes gathering data from a research publication, documentary or previously preserved records.
- Metadata — Metadata describe other data. Metadata is "data" that provides information about other data". In other words, it is "data about data". For example, an image may include metadata that describes how large the picture is, the colour depth, the image resolution, when the image was created, and other data. A text document's metadata may contain information about how long the document is, who the author is, when the document was written, and a short summary of the document.
- FAIR principles — 'findability', 'accessibility', 'interoperability' and 'reusability'.
- Open access — Online access to research outputs provided free of charge to the end-user.
- Open science — An approach to the scientific process based on open cooperative work, tools and diffusing knowledge.
- Research outputs — Results to which access can be given in the form of scientific publications, data or other engineered results and processes such as software, algorithms, protocols, models, workflows and electronic notebooks.



Annex I – Data Collection Form

This Annex includes the template in which information about datasets will be collected for Fit4Micro project.

Fit4Micro will follow the Horizon Europe Open Research Data guidelines and FAIR (Findable, Accessible, Interoperable, and Re-usable) principles. Since good research data management is not a goal in itself, rather a key leading to knowledge discovery and innovation, Fit4Micro partners will ensure these data will be produced and archived according to the principles contained in the Data Management Plan (DMP).

DMP will be established to support the data management life cycle for all data that will be duly collected, processed or generated by the project for guaranteeing its integration and reuse. The DMP will provide the identification elements and the descriptions of the data sets, and it will include details regarding how the research data will be handled during the project and how they will be preserved after it is completed. It will specify which methodologies and standards will be used in the data creation and management and how and when the data will be shared and made open for re-use.

For each dataset the following elements will be defined:

- 1) data set name, identifier and description – including origin (if collected), scale and possible use and impact of the data;
- 2) metadata and standards;
- 3) details on data sharing, open access dissemination and licensing;
- 4) archiving and preservation. In case parts of the data cannot be openly shared, the DMP will provide justifications.

Table 1. Data collection form

In progress / Not yet available	Dataset title
Team in charge	Partner acronym
Creator/s	Name of the person/s & partner acronym who has the idea
Contributor/s	Name of the person/s & partner who is/are contributing to this dataset
Contact Person/s	Technical referee (on behalf of the creator) for the dataset
Contents	Describe with all the info at disposal what the dataset is about, its purpose etc
Data format	Please indicate the type of data (e.g. if they are raw data, derived or secondary data, etc.).
Data volume	Final volume of data is expected to be XXX MB
Accessibility	Please try to indicate (if possible) if data will be made openly available or not; if so, indicate the type of license for their access.

Table 2. First project dataset description (to regularly verify and update)

Type of data	WPs	Collection methodologies
Software	WP5	Development of specific software/programming to be used in experimental when doing measurements or integrating sub-systems.
Dissemination materials, videos to promote the project, newsletters, webpage, internal communication materials	WP1 and WP8	Information will come from partners and results in the project.
CAD models to design the prototypes	WP4,	Partners developing prototypes usually use CAD program to design the prototypes.
Experimental measurements to characterize fuels, feed stock, inorganic contaminants from residue from FPBO. Operating conditions for combustor, gas analysis feed into the combustor and analysis of exhausted gases, materials behaviour in the chamber furnace, droplet sizes in the humidification cycle, experimental measurements to confirm system functionality	WP2, WP3, WP5	Each partner will produce different experimental measurements that will be obtained for example: <ul style="list-style-type: none"> - when producing biofuels, - construction and testing of prototypes.
CFD simulations, FEM of the burners, System modelling and simulations, LCA models, socio-economic analysis results	WP3, WP5, WP6	Simulations to assess prototypes functionalities during experimental testing.