EnerMan

Energy Efficient Manufacturing System Management

D7.3 – Preliminary Data Management Plan

Date	: 30/06/2021
Deliverable No	: D7.3
Responsible Partner	: TSI
Dissemination Level	: PU

HORIZON 2020
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement **No 958478**

EnerMan

Short Description
This document is a deliverable of the EnerMan project, it provides a preliminary version of the Data Management Plan for EnerMan project. This is a preliminary version of an ongoing document that will be updated during the project to include the latest information.

Project Information	
Project Acronym:	EnerMan
Project Title:	ENERgy-efficient manufacturing system MANagement
Project Coordinator:	Dr. Ing. Giuseppe D'Angelo CRF giuseppe.dangelo@crf.it
Duration:	36 months

Document Information & Version Management			
Document Title:		Preliminary Data Management Plan	
Document Type:		ORDP: Open Research Data Pilot	
Main Author(s):		TSI	
Contributor(s):			
Reviewed by:		INTRACT	
Approved by:		CRF	
Version	**Date**	**Modified by**	**Comments**
0.1	**28/05/2021**	**Ioannis Morianos (TSI)**	**Draft ToC**
0.2	**18/06/2021**	**Ioannis Morianos (TSI)**	**TSI's Contribution**
0.3	**25/06/2021**	**Nilay Yoruk (Intract)**	**Review**
0.4	**29/06/2021**	**Ioannis Morianos (TSI)**	**Final Draft**
1.0	**30/06/2021**	**Giuseppe D'Angelo (CRF)**	**Submitted version**

Disclaimer
This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both. The publication reflects the author's views. The European Commission is not liable for any use that may be made of the information contained therein.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement **No 958478**

TABLE OF CONTENT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement **No 958478**

EXECUTIVE SUMMARY

This deliverable provides the first version of the Data Management Plan (DMP) for the EnerMan project funded by the Horizon 2020 Programme. The EnerMan DMP follows the Guidelines on FAIR Data Management in Horizon 2020 [1], and also the template suggested by the European Commission. As the project evolves, the DMP will be updated to reflect significant changes that may arise in the data sources exploited in the project and on policies and methodologies to be used. The main purpose of this deliverable is to provide an overview of the key data management policies for the datasets generated by the project. The DMP will describe in detail how the data will be collected, generated, as well as how data will be made accessible to relevant stakeholders. Moreover, the DMP details how and which datasets will meet the FAIR Data management principles [2], (i.e., Findable, Accessible, Interoperable and Re-usable), and how they will be curated and preserved during and after the lifetime of the project.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement **No 958478**

GLOSSARY OF ACRONYMS

DMP	Data Management Plan
DOI	Digital Objective Identifier
DPB	Data Protection Board
DQAP	Data Quality Assurance Process
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
H2020	Horizon 2020 Programme
OA	Open Access
ORDP	Open Research Data Pilot
SVN	Subversion

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement **No 958478**

1. INTRODUCTION

1.1. Purpose of the Data Management Plan

A Data Management Plan (DMP) is a written formal document that outlines how data will be handled during the project and after the project completion. The Guidelines on FAIR Data Management in Horizon 2020 [1] provide a set of principles (Findability, Accessibility, Interoperability and Reusability) and relevant criteria that have to be addressed. As the EnerMan project participates in the Open Research Data Pilot (ORDP) in H2020, it is required to provide a DMP, through which compliance with the aforementioned principle will be assured. The EnerMan DMP follows the template suggested by the European Commission. The EnerMan DMP will describe in detail the data that the project will collect/generate, the methodologies and standards that will be followed to make research data FAIR, the data that will be shared/made open, and how they will be curated and preserved during and after the lifetime of the project.

Open data sharing can be very beneficial to society. However, we need to balance openness on the one hand and protection of sensitive data on the other. As stated in the Guidelines on FAIR Data Management [1], data should be 'as open as possible and as close as necessary'. The partners of the consortium that are involved in personal data processing should comply with all applicable data protection or similar laws regulating the processing of any personal data.

1.2. Definitions

Dataset: An organized collection of data.

Data Management Plan (DMP): A written formal document that outlines how data will be handled during the course of the project and after the project completion.

Data Quality Assurance Process (DQAP): DQAP aims at ensuring the high quality of the data generated/collected during the lifetime of the project.

Digital Object Identifier (DOI): Digital Object Identifier is a unique permanent identifier for a published digital object, standardized by the International Organisation for Standardisation (ISO).

GitHub: GitHub [3] is a version-control online repository which supports distributed source code development and management. (https://github.com).

FAIR: Research data that is Findable, Accessible, Interoperable and Re-usable. These principles aim to provide a framework to ensure that research data can be accessed and re-used effectively.

Metadata: Metadata is a set of data that provides context or additional information about other data. The main types of metadata include descriptive metadata, structural metadata, and administrative metadata.

Open Access (OA): Open Access refers to the unrestricted access to research results including scientific peer-reviewed publications and research data. There are two complementary mechanisms for achieving open access to research. Green open access means that the authors will publish the accepted manuscript in an online repository. Gold open access means that the publication is directly available free of charge from the publisher and any related costs referred to as Article Processing Charges (APCs) shall be covered by the authors.

Open Research Data: Research data needed to validate the results of the publications that are openly available in digital form for access and re-use by anyone for any purpose.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement **No 958478**

Open Research Data Pilot (ORDP): The Open Research Data Pilot aims to improve and maximise the accessibility and re-usability of research data generated by H2020 projects, without violating the privacy of sensitive data.

Repository: A digital repository is an archive for storing and managing digital copies.

Subversion (SVN): Subversion [4] is an open-source software versioning and revision control system that tracks changes made to files, folders, and directories. It eases data recovery and provides history of changes.

Zenodo: Zenodo [5] is a research data archive/online repository created by OpenAIRE and CERN for sharing research results in a wide variety of formats for all fields of science.

1.3. Structure of the Deliverable

The remainder of this deliverable is structured as follows: Section 2 provides a summary of the datasets that will be used in the project, including the types and format, the expected size of the datasets and the data utility. Section 3 outlines how the research data will become Findable, Accessible, Interoperable and Reusable (FAIR). Section 4 describes the allocation of resources required to make data FAIR. In Section 5, the provisions regarding the curation and preservation of the data during and after the end of the project are provided. Finally, Section 6 presents the ethical principles that the partners should comply.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement **No 958478**

2. DATA SUMMARY

During the lifetime of the EnerMan project, several datasets from various industrial domains will be produced. The datasets will be divided into two categories regarding their accessibility, namely Confidential and Open Access. 'Open Access' datasets will be provided free of charge for public sharing and will be included in the Open Research Data Pilot (ORDP). Confidential datasets, on the other hand, will be shared within the consortium. The datasets indicated as Open Access will be processed prior to releasing them with anonymization techniques to guarantee the preservation of the user's personal, sensitive data. Moreover, datasets marked as Confidential will be in psedoanonymized form.

The DMP will identify the data sources to be exploited for the purposes of several pilot use cases. For this purpose, the EnerMan consortium has prepared a questionnaire addressing these datasets. With this questionnaire the consortium will also identify the devices used in data collection as well as some additional information about the datasets (format, size, origin). With this information the DMP will ensure that the consortium will make the data FAIR (findable, accessible, interoperable and re-usable) and GDPR (General Data Protection Regulation, 2016/679 of the European Parliament and of the Council of 27 April 2016) compliant.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement **No 958478**

3. FAIR DATA

FAIR data principles [2] apply to datasets that will be openly available for public use. These datasets will comply with the provisions described in this section. The rest of the datasets are considered as confidential due to internal regulations and/or legal reasons that data providers ought to comply with. Confidential datasets will be either shared within the consortium or become accessible in-house after a proper agreement is signed. The EnerMan data management policy that will be followed for all the available datasets will be described in the following sections.

3.1. Making data findable, including provisions for metadata

EnerMan will use a standardized naming convention for all the project datasets that will be constructed using the following characteristics:

1. A unique chronological number of the datasets in the project.

2. The name of the dataset.

3. The acronym of the project.

4. The number of the related work package and deliverable/task.

5. A version number for each new version of the dataset that will be incremental at each revision.

This results in a dataset's internal reference number of the following format: 01_NameOfTheDataset_EnerMan_WPX_DX.X_v1.0.xls. When a dataset is uploaded to an online data repository, a DOI will be assigned to achieve effective and persistent citation. The DOI will be further used to all the related publications so that readers will be able to link them with the underlying datasets. Search keywords describing the dataset or content of the data will be provided when a dataset is uploaded to a repository aiming at optimising possibilities of re-use.

All the project datasets will be described with metadata. Metadata is a set of data that provides context or additional information about other data. Metadata gives the ability to other research to find data in an online repository which increases the reusability of the dataset. By providing detailed and rich metadata, researchers can define easier if the corresponding dataset is relevant to their research. However, these metadata can be used not only for data ingestion and data reusability but also for a concrete view of the data being used in the project, independently from the accessibility restrictions.

Zenodo [5], a well-known data repository, will be used to provide open access to the EnerMan datasets. The datasets will be documented and uploaded with their related metadata. Zenodo follows the minimum DataCite metadata standards that EnerMan also plans to follow.

3.2. Making data openly accessible

As per Article 29.2 of the Model Grant Agreement under Horizon 2020, EnerMan consortium will ensure open access to all peer-reviewed scientific publications relating to its results. The decision on whether to publish through open access will have to account for the potential necessity for protection of sensitive data. The consortium will follow green open access in the publications. In cases, where the timely open-access dissemination is not possible by following the green access model, EnerMan will opt for gold open access. According to the Article 26 of the EnerMan Grant Agreement (GA), research data that is generated during the action of the project, is owned by the beneficiary that generates it.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement **No 958478**

The repositories that will be used for the project datasets have been decided by the consortium. Zenodo, an online repository, will be used for the generated open research data during the lifetime of the project. Zenodo can easily share datasets in various sizes and formats, provide flexible licensing, and access and re-use of research data.

Confidential datasets, on the other hand, cannot become public due to privacy restrictions that the partners ought to comply with. These datasets will be uploaded in the project's SVN repository in order to be available for the consortium.

GitHub [3] is a version control online repository which supports distributed source code development and management. EnerMan's open-source outcome will be stored in the GitHub platform to raise awareness of the project, increase its impact and ensure its long-term sustainability. GitHub repository will contain a detailed description of the EnerMan public datasets and could also be used for hosting parts of the EnerMan open-source code components that will be implemented during the project.

Research data needed for validation of results presented in scientific publications will be uploaded to Zenodo as early as possible. In case an embargo period should be applied before the publication of the results, data will be deposited in the SVN repository. Nevertheless, information about the restricted data along with the related metadata will be published in Zenodo at the same time with the publication.

Anonymization and minimization techniques will be applied to data during the project's implementation to guarantee the preservation of the user's personal, sensitive data.

3.3. Making data interoperable

The file formats of each available dataset will be included in D7.8 (Revised Data Management Plan). As described, EnerMan will follow the minimum DataCite metadata standards for datasets description. The table below presents the general overview, the content and the technical description and the access of the data.

Table 1: EnerMan Dataset Metadata

Title	Name of the dataset
Creator / Responsible Partner	Name of the partner responsible for the data created
Dataset Identifier	Dataset's internal reference number
DOI	(if applicable)
Dataset Description	A brief description of the dataset
Work Package/Deliverable	Associated work package and deliverable/task
Source	How the data have been generated
Processing	How the data have been processed

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement **No 958478**

Repository	The repository where the data will be uploaded
Language	All languages used in the dataset
Code list	Explanation of codes or abbreviations used
Type	Types of the data
Format	Formats of the data
Expected Size	An approximation of the size of the dataset
Keywords	Keywords describing the content of the data
Version	Unique identifier for each version of the dataset
Date of Repository	Release date (preferred format yyyy-mm-dd)
Submission	Necessary software needed to create, view or analyse data
Necessary software	Any rights information on the use of the data
Rights	Where and how your data can be accessed by other researchers

The aforementioned metadata contents, formats and schema may be updated as the project progresses.

3.4. Increase data re-use (through clarifying licences)

Data reusability will be achieved by uploading the datasets to the Zenodo repository. The datasets will be made available to third parties as soon as they are generated and remain public for at least one year after the completion of the project. However, further restrictions such as setting embargo periods or restrictions from editors of scientific journals and organizers of conferences are possible. These restrictions may differ and will be examined on a case-by-case basis.

A Data Quality Assurance Process (DQAP) aiming at ensuring the high quality of the data generated/collected during the lifetime of the project, will be followed. The data quality dimensions to be measured, determining the quality objectives, are the following:

1. **Validity:** Data should be sufficiently accurate/valid for the intended use.

2. **Reliability:** Methods by which the data is collected and analyzed remain stable over time. Data collection and analysis methods are properly documented; the same procedures are followed each time.

3. **Precision:** The data has sufficient detail to give a fair representation of the phenomenon of interest.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement **No 958478**

4. **Integrity:** The maintenance of, and the assurance of the accuracy and consistency of, data.

5. **Timeliness:** Data should be regularly collected, up-to-date and available when needed.

6. **Completeness:** Missing data is the minimum possible. All required data elements, records, and values are known. Reproduction of research results reported in the scientific publications is possible.

For each dataset collected/generated during the project, DQAP ensures the following:

• At least one internal review of the generated/collected data;

• Additionally, peer reviews in case of publication in Journals.

The EnerMan Management related tasks with the assistance of the Quality Assurance task will be responsible for the coordination of the data quality assurance actions of the project and will ensure that the quality assurance process will be followed for all project datasets for the whole duration of the project.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement **No 958478**

4. ALLOCATION OF RESOURCES

According to Article 6 of the Grant Agreement (GA), costs related to open access to research data in Horizon 2020 are eligible for reimbursement during the duration of the project. For this project, costs for making data FAIR are mainly related to personnel costs and include preparation of data by each project partner for publishing, updating and maintaining data, data hosting and backup, data sharing, and security, etc. These costs will be covered by the project funds.

The costs related to long-term preservation of data after the end of the project are difficult to be estimated in this version of the Data Management Plan (DMP). The data will be preserved for a minimum of 1 year. However, preserving datasets on the Zenodo repository where a single dataset file does not exceed 50 GB, is free of charge. Moreover, internal datasets of the project will be stored and preserved in the SVN repository hosted by the coordinator. These costs are planned to be free of charge. Nevertheless, the final decision regarding the costs of preserving datasets has not been taken.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement **No 958478**

5. DATA SECURITY

All original datasets are safely stored at the responsible partners' local private repositories. These repositories are secure and non-accessible to the public. All partners follow appropriate procedures for recovery, secure storage and transfers and they are GDPR compliant.

Datasets that will be shared within the consortium will be stored in the project's SVN repository. The SVN repository is a centralized password-protected repository over secure network connection. Regular backup strategy and data recovery procedures will be followed.

After the end of the project, open data will be archived and preserved in the Zenodo repository. Regarding confidential data that will be deposited to the project's repository, decisions regarding long-term preservation and curation have not been taken yet.

EnerMan will ensure the preservation of the necessary data anonymity. Proper anonymization techniques will be applied to guarantee the preservation of anonymity of the user personal, sensitive data. All personal data obtained within the project will be transmitted to partners within the consortium only after anonymization or pseudonymization techniques will be applied.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement **No 958478**

6. ETHICAL ASPECTS

Partners within EnerMan are to comply with the Ethics and research integrity as described in the Grant Agreement.

Moreover, as set out in the Description of Action, EnerMan will comply with the highest of ethical standards to ensure that there is a good balance between the objectives of the research and the means by which the project partners go about achieving these. The research will comply with ethical principles, and applicable international, EU and national law, satisfying by design all relevant compliance requirements for each specific activity. It will ensure respect for people and for human dignity, fair distribution of burden and research benefits, while at the same time it will protect the value, rights and interests of all research stakeholders.

The main ethical concern is the effective privacy and personal data protection of parties, in an accountability-driven manner. This includes both research activities and potential implementation of research results. EnerMan is fully aware of the risks for the party's privacy and shall conduct an in-debt risk analysis and assessment placing data protection by design and by default inherent in the set-up of the project. EnerMan does not involve processing of neither genetic data nor biometric or health data. EnerMan will not intentionally process any sensitive personal data, whatsoever.

The consortium will comply with all obligations and requirements of its corresponding national data protection legislation and the GDPR The beneficiaries agree that any Background, Results, Confidential Information and/or any and all data and/or information that is provided, disclosed or otherwise made available between the Parties during the implementation of the Action and/or for any Exploitation activities ("Shared Information"), may include personal data as defined by Article 4 (1), Chapter I of the GDPR, and applicable local implementing legislation (hereinafter referred to as "Personal Data"). Each Party who provides or otherwise make available to any other Party Shared Information containing Personal Data ("Contributor") represents that: (i) it has the authority and/or the authorisation to disclose the Shared Information, if any, which it provides to the Consortium; (ii) where legally required and relevant, it has obtained appropriate informed consents from all the data subjects involved, or from any applicable institution, all in compliance with applicable regulations; and (iii) there is no restriction in place that would prevent any such other Party from using the Shared Information in accordance with and for the purpose of the project execution.

The Data Protection Board (DPB) will be responsible for complying with all applicable data protection or similar laws regulating the processing of any Personal Data. The board consists of one member from each deployment site formed to attend to and discuss any privacy issues arising in the course of the EnerMan planning, implementation, use, and evaluation. The board will be aided by two external independent Ethics and Privacy advisors, who will periodically provide evaluations regarding the policies and implementation of the project.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement **No 958478**

REFERENCES

[1] Guidelines on Data Management in H2020: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

[2] FAIR data principles (FORCE11): https://www.force11.org/group/fairgroup/fairprinciples

[3] GitHub: https://github.com

[4] Subversion (SVN): http://subversion.apache.org/

[5] Zenodo: https://zenodo.org

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement **No 958478**

Energy Efficient Manufacturing System Management

enerman-H2020.eu

EnerMan

enermanh2020 | enermanh2020 | enermanh2020

HORIZON 2020
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement **No 958478**

EnerMan