Product Passport through Twinning of Circular Value Chains

Deliverable 6.2

Data Management Plan V1

WP6: Project Management

Editor:	Armend Duzha
Lead beneficiary:	MAG
Version:	1.0
Status:	Final
Delivery date:	30/06/2023
Dissemination level:	PU (Public)

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101092008

Document Factsheet

Grant Agreement No.	101092008
Project Acronym	Plooto
Project Title	Product Passport through Twinning of Circular Value Chains
Start date	01/01/2023
Duration	36 months

Deliverable Name	D6.2 Data Management Plan v1.0
Related WP	WP6 Project Management
Due Date	30/06/2023

Author	Armend Duzha (MAG)
Contributor(s)	Kostas Kalaboukas (MAG), Sonia Saracino, Linda Leonardi, Alessandro Marseglia (CETMA), Sevasti Goti, Pantelis Velanas (ACCELI), Stavros Lounis, George Zois, Yiannis Mourtos (AUEB)
Reviewer(s)	-
Approved by	All partners

Document History

Version	Date	Author(s)	Organisation	Description
0.1	05/05/2023	Armend Duzha	MAG	ToC
0.2	10/05/2023	Armend Duzha	MAG	First draft
0.3	24/05/2023	Sonia Saracino, Linda Leonardi, Alessandro Marseglia	CETMA	Contributions in Section 4.2
0.4	30/05/2023	Sevasti Goti, Pantelis Velanas	ACCELI	Contributions in Section 4.2
0.5	30/05/2023	Stavros Lounis, George Zois, Yiannis Mourtos	AUEB	Contributions in Section 3.3, 4.2, 3.3 and 4.4
0.6	29/06/2023	Margherita Forcolin	MAG	Quality check
1.0	30/06/2023	Armend Duzha	MAG	Final version ready for submission

Executive Summary

This deliverable provides the first version of the Data Management Plan (DMP) of the Plooto project. It describes what kind of data will be generated or collected during the project and how these data are then managed and published.

Such information could be the scientific publications issued, white papers published, Open-Source code generated, mock-up datasets used for supporting the development process etc. The list of research data expected during the project consists of open-source software components, original research data and anonymised statistics. These datasets are expected to be collected during the validation and evaluation phase and are therefore subject to change, considering also the definition of the Plooto business models and sustainability plans.

The publishing platforms used are the project website, Zenodo for long-term archiving (as suggested by the EC), and GitLab for open-source code. All these platforms can be accessed openly.

Table of Contents

List of Tables

Acronyms and Abbreviations

Acronym	Description
DMP	Data Management Plan
DoA	Description of Action
EC	European Commission
GDPR	General Data Protection Regulation

1 Introduction

1.1 Purpose and Scope

This document describes how research data will be handled during the project lifetime. For all data generated or collected during the Plooto project, a description will be provided including the source, the standards and metadata used for data preservation and maintenance, as well as the process of how this data will be exploited and/or shared/made accessible for verification and re-use, in accordance with the EC Guidelines on FAIR Data Management in Horizon 2020 [1].

This deliverable is a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur. This is the 1st version of the document which is going to be updated on a regular basis and the final version will be submitted in month M36.

1.2 Structure of the document

The document is structured as follows:

- **Section 2** defines the data management process that will be applied to all data collected or generated during the project, as well as the tools to be used in order to ensure that the data will be exploited and/or shared/made accessible for verification and re-use along with the data preservation and maintenance processes.
- **Section 3** provides a summary of the datasets that will be generated during the project, including information such as types and formats, expected size of the datasets and data utility.
- **Section 4** outlines how the research data will become FAIR.
- **Section 5** describes the allocation of resources required to make data FAIR.
- **Section 6** specifies the provisions regarding the curation and preservation of the data during and after the end of the project are provided.
- **Section 7** presents the ethical principles that the consortium should comply.

2 Data Management Strategy

2.1 Data Management Process

The Data Management Process in Plooto is defined as a step-wise approach for each result generated or collected during the project runtime. The following questions must be answered to classify the different datasets:

1. Does a result provide significant value to other stakeholders or is it necessary to understand a scientific conclusion?

If this question is answered with yes, then the result is classified as public (granted for open access). If this question is answered with no, the result is classified as non-public. Such a result could be code that is very specific to Plooto platform (e.g. a database initialization) which is usually of no scientific interest to anyone, nor does it add any significant contribution.

2. Does a result include personal information that is not the author's name?

If this question is answered with yes, the result is classified as non-public. Personal information beyond the name must be removed if it should be published according to the ethical principles of the project.

3. Does a result allow the identification of individuals even without the name?

If this question is answered with yes, the result is classified as non-public. As such, in order to make it publishable, the included information must be reduced to a level where single individuals cannot be identified. This can be performed by using established anonymization techniques to conceal a single user's identity, e.g., abstraction, dummy users, or non-intersecting features.

4. Can a result be abused for a purpose that is undesired by society in general or contradict with societal norms and Plooto' s ethics principles?

If this question is answered with yes, the result is classified as non-public.

5. Does a result include business or trade secrets of one or more partners of Plooto?

If this question is answered with yes, the result is classified as non-public, except if the opposite is explicitly stated by the involved partners. Business or trade secrets need to be removed in accordance to all partners' requirements before it can be published.

6. Does a result name technology that are part of an ongoing project-related patent application?

If this question is answered with yes, then the result is classified as non-public. Of course, results can be published after patent has been filed.

7. Does a result break national security interests for any project partner?

If this question is answered with yes, the result is classified as non-public.

2.2 Roles and Responsibilities

The Data Management Plan (DMP) aims to identify the project outputs to be disseminated as well as to decide on the way and means of their Open Access (if applicable). To ensure it, a dedicated time slot will be reserved at each of the project plenary meetings. The EC will be informed about related work done and publications provided in the management reports.

Individual responsibilities on data management in the consortium are:

- **Data Management Plan Leader** (MAG) – to prepare and lead related discussions at the relevant project meetings and to maintain the channels for dissemination of project outcomes.
- **Scientific Manager** (AUEB) – to identify data collected by the project and technical project outcomes eventually suitable for publication
- **Dissemination Manager** (AUEB) – to identify publications suitable for publication in the considered repositories and maintain Plooto inputs for the Open Access
- **All other partners** – to identify own project results suitable for publication.

Moreover, each partner has to respect the policies set out in this DMP. Datasets have to be created, managed and stored appropriately and in line with applicable legislation. Validation and registration of datasets and metadata is the responsibility of the partner that generates the data in the WP. Metadata constitutes an underlying definition or description of the datasets and facilitate finding and working with particular instances of data.

Backing up data for sharing through Open Access repositories is the responsibility of the partner possessing the data.

Quality control of these data is the responsibility of the relevant WP leader, supported by the Data Management Plan Manager.

If datasets are updated, the partner that possesses the data has the responsibility to manage the different versions and to make sure that the latest version is available in the case of publicly available data.

Last but not least, all partners must consult the concerned partner(s) before publishing data in the open domain that can be associated to an exploitable result.

By taking into account the methodological framework, we proceed with the early definition of data sets to be considered as part of the DMP in the following section.

3 Data Archiving and Preserving Infrastructure

The Data Archiving and Preserving infrastructure consists of web-based platforms that provide long-term open access to all generated or collected results of the project. The following list presents the platforms to be used during the project and describes their concepts for publishing, storage, and backup.

3.1 Project Website

FACTLOG has designed and setup a project webpage, which provides a general overview of the project objectives and its approach and will inform target audience on its development status. A dedicated section for resources will be integrated in order to publish deliverables as well as scientific publication. All documents will be published using the portable document format (PDF) and be enriched by using simple metadata information, such as the title and the type of the document. All information on the project website can be accessed without the need of creating an account. The webpage will be available during the project lifetime, and remain available for at least two years after the project end.

URL: https://www.plooto-project.eu/

3.2 SharePoint

SharePoint is a client-server software for file hosting services. It is very similar to the widely-used services such as Dropbox, Google Drive, iCloud, and so on. In order for desktop machines to synchronize files with their OneDrive server, desktop clients are available for PCs running Windows, OS X, or Linux. Mobile clients exist for iOS and Android devices. Files and other data (such as calendars, contacts or bookmarks) can also be accessed using a web browser without any additional software. Any updates to files are pushed between all computers or mobile devices connected to a user's account.

The SharePoint platform for Plooto is hosted by MAG and runs on a server at Microsoft premises in Italy. All the partners have been granted access, using their institutional account. The platform is securely backed in the system infrastructure and holds all project-related data, including deliverables and publications. The SharePoint platform will not duplicate any project-related information to external servers, such as issues, requirements, product code, or deployment information. The SharePoint platform will be available during the whole project duration and for at least one year after the project end.

URL: https://maggiolispa-my.sharepoint.com/PlootoHorizon

3.3 Zenodo

Zenodo is an open repository which enables sharing of research results in a wide variety of formats for all scientific disciplines. It was developed by the OpenAIRE+ project and is maintained by CERN using one of Europe's most reliably hardware infrastructures.

Zenodo not only supports the publication of scientific papers or white papers, but also the publication of any structured research data (e.g., using XML). All uploaded results are structured by using metadata, like for example the contributors' names, keywords, date, location, kind of document, license, and others. Considering the language of textual metadata items, English is preferred. All metadata is licensed under CC license (Creative Commons 'No Rights Reserved'). The property rights or ownership of a result does not change by uploading it to Zenodo.

All public results generated or collected during the Plooto project will be uploaded to Zenodo for long-term storage and open access.

URL: https://zenodo.org/communities/plooto/

3.4 GitLab

GitLab is a well-established online repository which supports distributed source code development, management, and revision control. It is primarily used for source code data. It enables world-wide collaboration between developers and provides also some facilities to work on documentation and to track issues. GitLab provides paid and free service plans. Free service plans can have any number of publics, open-access repositories with unlimited collaborators. Private, non-public repositories require a paid service plan. Many open-source projects use GitLab to share their results for free. The platform uses metadata like contributors' nicknames, keywords, time, and data file types to structure the projects and their results. The service is hosted by ESOFT in Greece.

All source-code components that are implemented during the project and decided to be public will be uploaded to an open access GitLab repository.

4 Data Summary

In this section, a list of all existing or foreseeable results for dissemination is presented, separated into public deliverables, publications and open research data. For each result and in accordance to the FAIR data management guideline [2] we provide a description, name the standards used for storage and metadata (to make data findable and interoperable), and define which open access platform is chosen.

4.1 Research Datasets

During the project lifetime several datasets form various consortium members, representing different domains, will be produced. A more detailed description of each dataset is listed below.

Name of the dataset	Owner(s)	Accessibility
Discriminating parameters (e.g., α, Tg0, η min) of prepreg waste organized in function of storage conditions.	CETMA	*open access*
Correlation between discriminating parameters of prepreg waste and physical – mechanical properties of the composites obtained by using expired prepregs in function of the storage conditions of the material.	CETMA	*open access*
Procedure to evaluate the usability of prepreg waste.	CETMA	*open access*
CAD 3D model for design	ACCELI	*confidential*
Technical specifications	ACCELI	*confidential*
Technical data of the existing parts of the drone (weight, cost, materials' process) to compare with the data provided from new materials' tests.	ACCELI	*confidential*
Optimisation results	AUEB	*confidential*
Optimization results	AUEB	*confidential*

Table 1: Datasets to be generated during the Plooto project

As the project evolves, these tables will be updated with additions or modifications. The final table will be presented in the final version of the DMP in month 36.

4.2 Public Deliverables

The following table presents the list of public deliverables of the Plooto project.

ID	Deliverable Title	Partner	Expected date
D1.1	Plooto methodological approach and business cases specifications v1	IDC	*Oct 2023*
D1.2	Plooto methodological approach and business cases specifications v2	IDC	*Dec 2024*
D1.3	Sustainability balanced scorecard framework v1	MAG	*Dec 2023*
D1.4	Sustainability balanced scorecard framework v2	MAG	*Dec 2024*
D1.5	CRIS requirements and specification v1	ESOFT	*Dec 2023*
D1.6	CRIS requirements and specification v2	ESFOT	*Dec 2024*
D2.1	RM-recovery and waste Data Space v1	TUC	*Jun 2024*
D2.2	RM-recovery and waste Data Space v2	TUC	*Apr 2025*
D2.3	Plooto complete suite of services v1	MAG	*Jun 2024*
D2.4	Plooto complete suite of services v2	MAG	*Apr 2025*
D3.1	Product passport and certification tool v1	MAG	*Jun 2024*
D3.2	Product passport and certification tool v2	MAG	*Jun 2025*
D3.3	CRIS integrated platform v1	ESOFT	*Jun 2024*
D3.4	CRIS integrated platform v2	ESFOT	*Jun 2025*
D3.5	Plooto balanced scorecard v1	MAG	*Jun 2024*
D3.6	Plooto balanced scorecard v2	MAG	*Jun 2025*
D4.1	Impact assessment methodology	AUEB	*Jun 2024*
D4.2	Report on piloting activities v1	ACCELI	*Oct 2024*
D4.3	Report on piloting activities v1	ACCELI	*Dec 2025*
D4.4	Pilot assessment report	AUEB	*Dec 2025*
D4.5	Guidelines for replication, deployment and certification	TAH	*Dec 2025*
D5.8	Feasibility study on the impact of Plooto in the EU industry	UIO	*Dec 2025*
D5.9	Educational modules and lifelong learning outcomes	JSI	*Dec 2025*

Table 2: Plooto Public deliverables

4.3 Scientific Publications

In this section, we are going to include all scientific publications produced during the project duration. There are already some publications submitted by the consortium partners in scientific journals and conferences:

1. Koukopoulos Anastasios, Lounis Stavros, Farmakis Timoleon, Vrechopoulos Adam, and Doukidis Georgios "D&C for EU Funded Projects: Towards an Integrated Omnichannel Dissemination and Communication Framework", 11th International Conference on Contemporary Marketing Issues (ICCMI) 2023, 12-14 July 2023, Corfu, Greece.
2. Psarommatis, F., May, G. and Azamfirei, V., 2023. Envisioning maintenance 5.0: Insights from a systematic literature review of Industry 4.0 and a proposed framework. Journal of Manufacturing Systems, 68, pp. 376-399.

4.4 Software Components

While Plooto project is committed to provide most of the software components as Open Source, the final decision will be made along with the IPR agreement in WP5. The analysis is performed by taking into account the list of exploitable components, defined in the DOA as open-source components. A summary is presented in the following table:

ID	Name	Owner(s)
S01	Circular and Resilient Information System	*MAG, ESOFT*
S02	Product passport and certification	*MAG, AEGIS*
S03	Balanced scorecard	*MAG, TUC*
S04	AI models and algorithms for circularity	*JSI, FRONT*
S05	Predictive maintenance algorithms	*JSI*
S06	RM-recovery and Wate Data Space	*MAG*
S07	Process modelling and simulation	*TUC*
S08	Optimization Module	*AUEB*
S09	Videolectures.NET, a platform for publishing educational videos	*JSI*
S10	X5Gon, an intelligent learning platform for collecting and indexing open educational resources	*JSI*

Table 3: Plooto open-source software components

5 FAIR Data

FAIR data principles [2] apply to available for public use datasets. 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 data management policy that will be followed for all the project datasets is described in the following sections.

5.1 Making Data Findable, including Provisions for Metadata

The naming convention to be applied for all types of datasets (*open source* or *confidential*) includes the following information:

1. A unique chronological number of the datasets.
2. The acronym of the project.
3. The name of the dataset.
4. A version number for each new version of the dataset that will be incremental at each revision.

Dataset's internal reference number:

01_Plooto_Nameofthedataset_v1.0.xls

DOIs can be used for each dataset that is uploaded to an online data repository to achieve effective and persistent citation. Each dataset's DOI should be used to all the related publications so that readers will be able to link them with the underlying datasets.

All the project datasets must 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 researches 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. An additional use of these metadata except from data ingestion and data reusability is the provision of a concrete view of the data being used in the project, independently from the accessibility restrictions.

Moreover, 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.

5.2 Making Data Openly Accessible

Plooto consortium will ensure open access to all peer-reviewed scientific publications relating to its results. For this purpose, the 'green' open access model will be used.

The repository that will be used for the project datasets is Zenodo [3], an online platform that allows easily storing of datasets in various sizes and formats, and provides 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 repository (i.e., Teams/SharePoint) in order to be available for the consortium.

GitLab [4] is a version-control online repository which supports distributed source code development and management. Plooto's produced datasets could be built on the GitLab platform to raise awareness of the project, increase its impact and ensure its long-term sustainability. GitLab will also host parts of some of the Plooto 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 soon as possible. In case an embargo period should be applied before the publication of the results, data will be deposited in the project repository. Nevertheless, information about the restricted data along with the related metadata will be published in Zenodo at the same time with the publication.

As mentioned above, some datasets cannot be shared either with the whole consortium or made publicly available due to internal business constraints. Moreover, when possible, anonymization, aggregation, minimization or sampling techniques may be applied to real data during the project to guarantee the preservation of user's personal or sensitive data.

5.3 Making Data Interoperable

To increase the interoperability of the provided data, commonly used vocabularies for the metadata within the datasets will be used. Interoperability will be ensured by using the same standards for data and metadata capture/creation to all datasets.

Table 4 Presents the general overview, the content and the technical description and the access of the data.

Name	Name of the dataset
Creator / Responsible Partner	Name of the partner responsible for the data created
Data Identifier	Dataset's internal reference number
DOI	(if applicable)
Description	A brief description of the dataset
Work Package/Deliverable	Associated work package and deliverable
Source	How the data have been generated
Processing	How the data have been processed
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 submission	Release date (preferred format dd-mm-yyyy)
Necessary Software	Necessary software needed to create, view or analyse data
Access Information	Any rights information on the use of the data

Table 4: Dataset metadata

5.4 Increase Data Re-Use

Uploading the datasets to Zenodo will help towards making data reusable. Datasets will be made available to third parties after their generation and will remain public for a specified time period after the completion of the project. However, possible restrictions such as embargo periods or restrictions from editors and organizers of conferences will be examined on a case-by-case basis. The owner of the data, is responsible for maintaining the data after the end of the project.

A Data Quality Assurance Process (DQAP) will be followed in order to ensure quality of the data that are generated during the lifetime of the project. The dimensions to be measured are: i) Validity: Data should be sufficiently accurate/valid for the intended use; ii) Reliability: Data collection methods must remain stable over time; iii) Precision: Data detail is sufficient to represent the phenomenon of interest; iv) Integrity: Accuracy and consistency of data must be maintained and assured; v) Timeliness: Data should be regularly collected, up-to-date and available when needed and vi) Completeness: All required data elements, records, and values must be known. Reproduction of research results is possible.

6 Allocation of Resources

Costs for making data FAIR are mainly related to personnel costs. Each of the partners will need to prepare the data for publication, update and maintain data, perform data hosting and backup, data sharing, and security, etc. The costs for these actions 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. The aim of the consortium is to preserve the data for a sufficient time period. We should mention that preserving datasets on Zenodo 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 SharePoint hosted by MAG.

7 Data Security

Datasets produced by consortium members will be stored at the responsible partners' local repositories, which shall be secure and non-accessible to the public. The appropriate procedures for recovery, storage, and transfers are must be followed.

Shared datasets among the consortium members will be stored in the project repository. SharePoint is the most widespread cloud-based software for collaboration that offers data security and privacy. All datasets are stored in one central location – protected from unauthorized access.

The necessary data anonymity will be ensured. Proper pseudonymisation/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.

8 Ethical Aspects

One of the main ethical concerns is privacy and personal data protection of individuals regarding research activities and potential implementation of research results. The Plooto consortium is fully aware of the risks for the individual's privacy and will conduct an in-depth risk assessment placing data protection by design and by default inherent in the set-up of the project.

Plooto is intended to ensure respect for the ethical principles and fundamental rights embedded in the regulatory framework of the European Union, including the Charter of Fundamental Rights of the European Union as well as the European Convention on Human Rights. The data processing activities within our project will be carried out in accordance with the updated privacy rules as specified in Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (GDPR) and the ePrivacy Directive 2002/58/EC.

Moreover, the corresponding national data protection legislation should be taken into consideration, and all legal documents and certifications required for compliance with such legislation will be obtained. The Parties who provide or transfer to any other Party information containing Personal Data must have: (i) the authority and/or the authorisation to disclose the aforementioned information; (ii) obtained appropriate informed consents from all the data subjects involved, or from any applicable institution, and (iii) a confirmation that there is no restriction that would prevent any other Party from using the shared information. Data protection by design and by default will be at the core of the research and development work as well as the project outputs.

The personal data will be anonymized and homomorphic encryption will be used for pseudo-anonymization. Identifiable data will be dissociated from the rest of the data in a separate database. Personal data will be processed in Plooto using a neutral code with the aim to render data non-attributable to any natural person.

Due to the way the research will be carried out, consortium partners making decisions about the collection and processing of personal data, in order to achieve the various goals and objectives of the project, will be deemed joint data controllers. As required by the GDPR, an agreement outlining the allocation of the obligations and responsibilities of the joint data controllers within Plooto, will be created. A key aspect of this arrangement is regular and effective communication between data controllers to ensure a consistent and effective approach.

Plooto implementation does not anticipate or plan any transfer of personal data to third parties. Relevant legal foundations, appropriate safeguards and compliance measures will be identified and implemented if data sharing is deemed necessary at a later stage. These safeguards will be documented in the final version of the Data Management Plan.

References

[1] Guidelines on Data Management in H2020, Version 3.0, 26 July 2016

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[2] FAIR Data Principles: https://www.force11.org/group/fairgroup/fairprinciples

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

[4] GitLab: https://about.gitlab.com/