GREEN-LOOP

Funded by the European Union

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them

GREEN-LOOP

Sustainable manufacture systems towards novel bio-based materials

WP7 – Business Model, Replication and Exploitation

D7.11 – DMP and open-sourcing approach

Version 1.0

Document information

Contractual Due date: 28.02.2023	**Delivery Date**: 10.03.2023
Author(s): RWTH AACHEN	
Lead Beneficiary of Deliverable: RWTH AACHEN	
Dissemination level: PU - Public	
Nature of the Deliverable: Report	
Internal Reviewers: AXIA INNOVATION, IDENER	

Funded by the European Union

GREEN LOOP Key Facts

Project title	Sustainable manufacture systems towards novel bio-based materials
Starting date	09/01/2022
Duration in months	36
Call (part) identifier	TWIN GREEN AND DIGITAL TRANSITION 2021 (HORIZON-CL4-2021-TWIN-TRANSITION-01)
Topic	HORIZON-CL4-2021-TWIN-TRANSITION-01-05 Manufacturing technologies for bio-based materials (Made in Europe Partnership) (RIA)
Consortium	17 organizations. 15 EU Member States + 2 non-EU state

GREEN LOOP Consortium Partners

	Partner	Acronym	Country
1	IDENER RESEARCH & DEVELOPMENT	IDE	ES
2	NATIONAL INSTITUTE OF CHEMISTRY	NIC	SI
3	SLOVENIAN NATIONAL BUILDING AND CIVIL E. I.	ZAG	SI
4	FRAUNHOFER GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V	FHF	DE
5	LABRENTA SRL	LBRT	IT
6	MIXCYCLING SRL	MYX	IT
7	NERO SU BIANCO	NSB	IT
8	GERACE MARIA CRISTINA - TERRE DI ZOE'	TDZ	IT
9	IRIS TECHNOLOGY SOLUTIONS, SOCIEDAD LIMITADA	IRIS	ES
10	GLOWNY INSTYTUT GORNICTWA	GIG	PL
11	AACHEN UNIVERISTY: PROCESS CONTROL ENGINEERING / AACHEN UNIVERISTY: INSTITUTE OF SOCIOLOGY	AAU	DE
12	AUSTRIAN STANDARDS INTERNATIONAL	ASI	AT
13	INSTITUTO DE SOLDADURA E QUALIDADE	ISQ	PT
14	AXIA INNOVATION UG	AXIA	DE
15	ASOCIACIÓN DE INVESTIGACIÓN METALÚRGICA DEL NOROESTE	AIMEN	ES
16	NATIONAL COMPOSITE CENTER	NCC	UK
17	UNIVERSITY OF BRISTOL	UBRIS	UK

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them

GREEN LOOP is a project funded by the European Commission under the Horizon Europe - HORIZON-CL4-2021-TWIN-TRANSITION-01-05- Manufacturing technologies for bio-based materials (Made in Europe Partnership) (RIA) under Grand Agreement Number 101057765.

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them.

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them

Executive Summary

This document focuses on providing the GREEN-LOOP consortium with information on how to deal with data generated during the execution of the project and how it should be managed. Specifically, this document provides information about the type of data that will be generated, the standards, and the accessibility of data for verification, re-use, curation, and preservation of data procedures. At the same time, being a public document, it will allow the easy identification of the datasets generated within the project by other researchers who may be interested in reusing or validating the information produced

The deliverable details all the data management procedures within GREEN-LOOP, in response to article 17 of the grant agreement, specifically regarding open access to research data. Accordingly, it presents information for partners in the consortium as well as for external parties about the different datasets generated and used within the project. They will be categorized and technically detailed in terms of data collection, processing, and generation in order to manage the data that is associated and generated. The monitoring of the GREEN-LOOP project with regard to this area will be the subject of continuous reviews within the context of WP7 at each consortium meeting.

Moreover, the current data management plan (DMP) is the first draft of this document. This means that the DMP will be a living document, continuously updated with information provided by all partners in the project. A final version of this document will be submitted to the European Commission by the end of the project.

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them

Table of Contents

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them

List of Figures

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them

List of Tables

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them

Abbreviations

DMP: Data management plan

GDPR: General Data Protection Regulation

LCA: Life cycle analysis

LCC: Life-cycle costing

sLCA: Social life cycle assessment

WP: Working package

RDM: Research data management

RDMO: Research data management organizer

EC: European Commission

IPR: Intellectual Property Rights

TBD: To be defined

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them

Funded by the European Union

1. Introduction

This document presents the initial version of the Data Management Plan (DMP) for the GREEN-LOOP project. This information has been prepared partially following the guidance of the RWTH IT-Center (https://www.itc.rwth-aachen.de/cms/~rbnjb/it-center/), a German-recognised centre for research data management. The center develops its own RDM tools such as the data repository Coscine and the research data planning tool RDMO.

The RWTH IT-Center provides extensive RDM courses as well as expert advice to help research organisations in storing, managing, protecting and sharing digital research data. This DMP for Green-Loop details the public datasets that the project:

- will generate.
- whether and how it will be exploited or made accessible for verification and re-use.
- how it will be curated and preserved.

The academic papers produced within the project will be made available as open access for several years (following green or gold open access approach). Similarly, the data generated as a result of the project's research activities will be partially put online and freely available in open access repository. The DMP will include the details on the datasets generated, providing detailed information in such datasets. Specifically, this DMP details the following aspects:

- metadata generation.
- data preservation.
- data storage beyond the end of the project.

In particular, the consortium partners acknowledge their responsibilities for fulfilling and updating this DMP, which includes:

- the first version of the DMP must be operative in the first 6 months of the project.
- it will be a live document continuously updated within the project timeframe.
- data identified (generated within the project) in DMP must be shared in an online repository by the producers of the data unless the data violates IPR or contain personal data, appropriate support will be provided to those involved in the project from both the DMP responsible and the project coordinator.

At a deeper level of detail, the aim of the DMP is that its execution and update will lead to:

- a better understanding of the data produced as output from the project.
- clarity on how the data is actually used within the project and outside of it.
- continuity in the work of the consortium in the event of staff leaving or entering the project during its lifecycle or equivalently staff changing roles within the project during its lifecycle. This includes such areas as:
    - avoiding duplication of effort i.e., re-collecting or re-working data.
    - enabling validation of results.

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them

- contributing to collaboration through data sharing.
- increasing visibility of output and thereby leading to greater impact. In particular, enabling other researchers to cite and use the datasets generated within the project.

Green-Loop embraces the open-access approach incentivised by the EC. This means that the project aims to improve and maximize access to and re-use of research data generated by it and considers the need to balance openness and protection of scientific information, commercialisation and Intellectual Property Rights (IPR), privacy concerns, security as well as data management and preservations. Within the GREEN-LOOP project, most of the deliverables produced are done as public documents. A similar approach will be undertaken with the generated data.

AAU has also presented a workshop on the topic "Research data management". The workshop addressed the below topics.

- Motivation/ Lifecycle of research data.
- FAIR principle.
- Publication checklist.
- Best practices.
- Possible platforms for data management.
- How to proceed concerning GREEN-LOOP.

The aim of the workshop was to provide a better overview of the fair principles and why are they important. It also addressed how can the FAIR principle be practically implemented in the GREEN-LOOP project.

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them

2. Data management plan

The DMP aims to regulate all the processes related to the data life cycle.

At this stage of the project, the consortium is working on data planning. Figure 1 shows the whole process developed in the data lifecycle, composed of mainly 4 steps:

- **Planning**: In this phase, a study of the data needed for the project, tools needed, archiving possibilities, data ownership, and timeline for execution is performed.
- **Collection**: In this phase, data is collected, annotated, and given a persistent identifier (PID).
- **Analysis**: Analysis is done as well as publication taking into consideration data context and software versioning.
- **Archiving**: Archiving is done while taking into consideration which data is important and the storage medium(long-term).
- **Access**: Access to 3rd party researchers is granted while considering access restrictions.
- **Reuse**: in this phase, data context is important as well as possibilities for data aggregation.

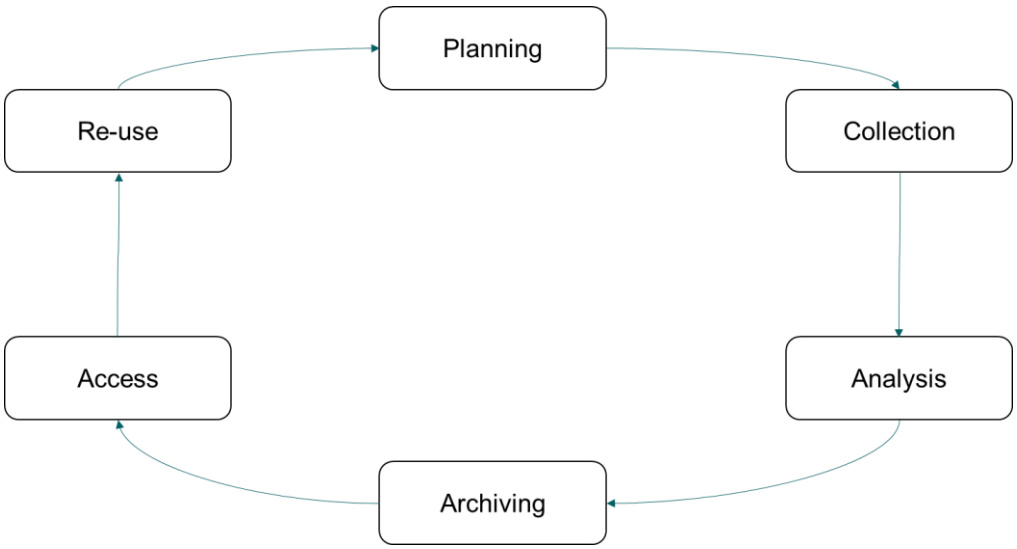

Figure 1: Lifecycle of research data

In addition, the usage of the data in the consortium will be also considered and monitored. This initial DMP provides templates to characterise the datasets generated or employed within the project (which consists mainly of created and some reused data). At the moment, personal data will only be collected to manage user accounts on the GREEN-LOOP platform within the project. Consent will be requested from users during account creation and the data won't be used for any other purpose.

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them

3. Data summary

This section includes in the initial DMP the types and formats of data generated and/or collected.

In order to define an effective DMP, the data should be categorised. In this way, the data can be categorised according to:

- The source.
- Format.
- Stability.
- Volume.

Regarding the source, the data can be categorised into 4 types: observational, experimental, simulation and derived/compiled. The GREEN-LOOP project will generate observational, experimental, and derived datasets.

- Observational data will be collected by logging manufacturing data for the three value chains during the initial studies as well as the upscaling phase.
- Experimental data will be needed to estimate the material properties of the developed sample in WP 3, 4 & 5.
- Derived data are generated by existing datasets and are reproducible. These data are characterised in the context of WP2, where the impact of the three value chains will be evaluated.

When describing the format categorisation, the data can be generated in different forms: text, numeric, audio, visual, models, computer code, discipline-specific and instrument-specific. The GREEN-LOOP project will generate data in all these forms except audio. It is important to notice that the project will be gathering mainly data in the text and numeric formats of categorizations to populate the data warehouse employed in the project.

Regarding the stability of the data, the data can be fixed or changed over the course of the project and this will affect the way it is managed. The common categories of a dataset are:

- Fixed datasets never change after being generated or collected.
- Growing datasets: new data may be added while old data are never changed or deleted.
- Revisable data: new data may be added while the old data may be changed or deleted.

In this context, partners in the GREEN-LOOP project plan to use all of the three types listed above throughout the execution of the project, requiring keeping track of data versions.

Finally, it is important to define the volume of data to be prepared. In general, the GREEN-LOOP project is expected to handle a high volume of data, including both the datasets used and the generated ones.

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them

3.1. Created datasets

At this stage of the project, some datasets that will be created are defined for different working packages. Yet more data sets might be added if the need occurs.

Created datasets			
Name	*Data owner*	*Type of data*	*Purpose of the dataset*
WP2-Circularity	AXIA & GIG	Table/ text	The data is generated to assess value chains: the analysis will study the transformation from linear value chain to circular approach, including resource optimisation, waste reduction, reuse, refurbish, recycle, circular business and Blue economy impacts.
WP2- Energy & Exergy	Idener	Table/ text	To estimate energy transformation associated with producing the newly developed products
WP2- LCA, LCC, sLCA	AIMEN	Table/ text	The data will be analysed to estimate the impact of producing the products. The analysis will focus on three main aspects: - Environmental factors such as recyclability, reuse, refurbish, and waste reduction - Financial factors such as capital, operational costs, and indirect costs - Social indicators such as new job creation, training hours needed for operators and user acceptance
WP2- Green-Loop Platform	Idener	(Protected)	The data will be used to manage user access rights on the Green-loop platform
WP2-Machine learning	Idener	Table/ text	The datasets will be used to train models. The models will be used for optimization purposes and later on for monitoring the product quality
WP3-Manufacturing properties	NCC & IRIS	Table/ text	Product manufacturing data at a lab scale setup will be collected to support the upscaling process of the bio-rubber product
WP3-Material properties	NCC	Table/ text	The material properties associated with the bio-rubber sample materials will be collected and compared
WP4-Material properties	Mixcycling	Table/ text	The material properties associated with the bio-plastic sample materials will be collected and compared.
WP5-Material properties	Labrenta	Table/ text	The material properties associated with the wood composite sample materials will be collected and compared.

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them

3.2. Reused/aggregated datasets

The project also contains reused datasets as well as aggregated datasets from potential other sources.

Reused/aggregated datasets			
Name	*Data owner*	*Type of data*	*Purpose of the dataset*
WP2-Standardisation	ASI	Table/ text	The data should describe existing standards that relate to the three developed value chains. The data will be used to estimate the standardization potential of the end products.
Test stations	Idener	Table/ text	Information on each test location will be saved in these tables

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them

4. FAIR Data

This section aims at presenting how the GREEN-LOOP consortium would save the information in order to make the data Findable, Accessible, Interoperable, and Reusable (FAIR).

4.1. File formats

The file format to be employed is a primary factor to generate FAIR data. The file format should be accessible in the future so that the selected formats should be non-proprietary, open, with documented standards, in common usage by the research community, using standard character encodings, and uncompressed.

The following table includes the acceptable formats that the GREEN-LOOP consortium is going to adapt to save the collected data:

Allowable save formats	
Name	*Formats*
Tabular data with minimal metadata (column headings, variable names)	• Widely-used formats: MS Excel(.xls/.xlsx) • MS Access (.mdb/.accdb), dBase (.dbf), • OpenDocument Spreadsheet (.ods) • Comma-separated values (.csv)
Textual data	• Rich Text Format (.rtf) • Hypertext Mark-up Language (.html) • Widely-used formats: MS Word (.doc/.docx) • Plain text, ASCII (.txt)
Documentation and scripts	• Rich Text Format (.rtf) • PDF/UA, PDF/A or PDF (.pdf) • XHTML or HTML (.xhtml, .htm) • OpenDocument Text (.odt) • XML marked-up text (.xml) according to an appropriate DTD or schema, e.g. XHMTL 1.0 • Plain text (.txt) • OpenDocument Text (.odt)
Image data	• TIFF 6.0 uncompressed (.tif) • TIFF other versions (.tif, .tiff) • RAW image format (.raw) • Photoshop files (.psd) • BMP (.bmp) • PNG (.png) • Adobe Portable Document • Format (PDF/A, PDF) (.pdf) • JPEG (.jpeg, .jpg, .jp2) if original created in this format • GIF (.gif)
Video data	• MPEG-4 (.mp4) • OGG video (.ogv, .ogg) • Motion JPEG 2000 (.mj2) • AVCHD video (.avchd)

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them

Tabular data and Textual data are going to be commonly handled in implementing the activities planned in the GREEN-LOOP project. Hence, the next recommendations given to the consortium can make datasets easier to understand and export:

- Do not put more than one table on a worksheet.
- Create charts on new sheets. Do not embed them in the worksheet with the data.
- Include a header row with a clear title for each column.
- Upon manipulating Excel worksheets, be careful when deleting the content of row or column cells. The simple deletion of the content of the cells may yield missing values when importing the data in other software packages (i.e., Stata) using selected procedures. Please take care of deleting the whole rows and/or columns.
- Be mindful of the possibility that data may not be written following the international notation employing the comma (,) as the separator of thousands and the dot (.) as the separator of decimals.
- Metadata schemas were also defined together with the partners to unify and smooth data transfer and aggregation.

4.2. Metadata: Data document

Clear and detailed documentation is essential for data to be understood, interpreted, and used. The data document describes the content, formats, and internal relationships of the data. The following template describes the documentation required within the project to document the datasets used and generated. This information will be indexed in the internal project's platform (Confluence). If during the elaboration of the data formats any field is not determined yet, the label TBD should be included. In case any field is provisional, the * should be included at the beginning of the text.

Dataset information	
Name	*Name of the dataset*
Data contained	Describing the data collected and the aggrupation of data
Dataset structure	Detailing tables structures and main indications to help understand the reading of the dataset
Dataset format	How can this data be available: format to be used
Generation process	Process followed to gather/produce the dataset
Author	Author of the data
Maintainer	Institution/company in charge of maintaining
Last actualization	Date
Additional information	Relevant information to be included for a better understanding of the data source
Publication information	Sharing policies among the following options: Green Open Access; Gold Open Access; Non-shared; Published in paid-for journals; Other (Describe)

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them

4.3. Data sharing: Publication

The Green-Loop consortium fully embraces the HE requirement for Open Access publishing, following the guidelines presented by the EC. The project will ensure 'green' and 'gold' publishing, i. e., self-archiving in one or more repositories. The associated costs of publications will be claimed as part of the HE grant.

The project will make its datasets of results publicly available through the following repositories:

- Zenodo: https://zenodo.org/

In case the data violates any confidential restrictions or IPR of any of the partners, the datasets will not be published.

The project will also be using a private repository (RWTH- Coscine) that aligns with the HE research data management principles to save all deliverables as well as context information of the reports for 10 years. The data will be structured and annotated for easy future retrieval in case data reuse is needed.

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them

5. Allocation of resources

5.1. Responsibility and roles

IDE is the project coordinator and in charge of the management of the project.

RWTH-Aachen (AAU) is responsible for maintaining and updating the DMP. Therefore, AAU will be in charge of continuously updating the DMP to produce a final version by the end of the project that provides a clear guideline on the management of personal data in compliance with the General Data Protection Regulation (GDPR).

WP leaders are also responsible for ensuring they upload all deliverables as well as supporting documents including data sets to the private repository. Data annotation is also going to be coordinated between the WP leaders and AAU. The repository is responsible for backups and archiving the data for 10 years.

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them

6. Ethical Aspects

The GREEN-LOOP project will not collect personal data for the purpose of profiling or analysis. Nevertheless, contact data will be used in task T2.5 for the purpose of role management on the GREEN-LOOP platform. The data will neither be used for analysis nor made public in any publications. During the accounts' creation, the users will be informed about the purpose of the data collection.

A separate deliverable was already submitted and approved concerning all ethics related issues as a part of WP10.

6.1. Intellectual property rights

As part of WP3, WP4, WP5 & WP6, new products will be developed and produced as part of the GREEN-LOOP- objectives. The datasets describing experimentation, manufacturing specifications, and product properties will not be made available to the public as the data constitutes intellectual properties to project partners.

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them

7. Other issues for data management

At the moment, the GREEN-LOOP project does not make use of other information requiring a special treatment different than the one already explained within this document, and no other issues have been found in this initial phase of writing the DMP.

In preparing this report, the following deliverables have been taken into consideration:

Considered deliverables					
Deliverable Number	Deliverable Title	Lead beneficiary	Type	Dissemination Level	Due date
D10.1	OEI –Requirement No.1	Idener	Report	Confidential	M01

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them

8. Conclusion

In this report, the first version of the DMP was presented. Firstly, an introduction on RDM was presented with an overview on a workshop that was done with the project consortium. Secondly, an outline on the research data lifecycle was introduced to reach common term definitions. Thirdly, the FAIR principles were presented. It was also shown how the GREEN-LOOP consortium plans to uphold and implement the principles when dealing with research data. Later on, different aspects such as resource allocation and ethical concerns were addressed. Finally, a reference to related deliverables that deals with data management was included.

The GREEN-LOOP consortium is planning to submit further iterations to the DMP as the document is regarded as a live document that will be updated periodically.

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 or HADEA. Neither the European Union nor the granting authority can be held responsible for them