HORIZON-CL6-2024-CIRCBIO-01-5

Programmed biodegradation capability of bio-based materials and products, validated in specific environments

MAGICBIOMAT

Manufacturing Guiding Tool for Circular and Programmed Biodegradable Materials Under Open Environment Conditions

Start Date: January 1st, 2025

Duration: 48 Months

Deliverable: D13.4

Deliverable Title: Initial data management plan (Period 1)

Due date of deliverable: 30/06/2025

Actual submission date: 27/06/2025

Responsible WP: WP13, Project Management and Coordination (period 1), Digiotouch

Responsible Technical Leader: Soumya Kanti Datta, Digiotouch

Revision: V1.1

Dissemination level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Funded by the European Union

Authors

Author	Institution	Contact (e-mail, phone)
Debolina Paul	**Digiotouch**	**debolina@digiotouch.com**
Soumya Kanti Datta	**Digiotouch**	**soumya@digiotouch.com**

Document History

Document version	Date DD/MM/YY	Changes
V0.1	05/06/2025	Table of content
V0.2	10/06/2025	Initial content written and sent for internal review
V1.0	27/6/2025	Updated with review comments
V1.1	27/6/2025	Released for submission

Document Approval

Reviewers		Validation date
Work Package Leader	**Yuanyuan Chen**	27/6/2025
Coordinator	**Yuanyuan Chen**	27/6/2025

Document Data

Keywords	Data management plan; Open Access; Data set
Point of Contact	Name: Soumya Kanti Datta Partner: Digiotouch Address: Narva mnt 5, 10117 Tallinn, Estonia Phone: +372 58542328 E-mail: contact@digiotouch.com
Partner:	Digiotouch

Disclaimer

Views and opinions expressed are 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

Acronyms

AI	Artificial Intelligence
ALTAI	Assessment List for Trustworthy Artificial Intelligence
BB	Bio-based
CERN	Conseil Européen pour la Recherche Nucléaire
DMA	Dynamic Mechanical Analysis
DMP	Data Management Plan
DOI	Digital Object Identifier
DSC	Differential Scanning Calorimetry
EEA	European Economic Area
EoL	End of Life
EOSC	European Open Science Cloud
EU	European Union
FAIR	Findability, Accessibility, Interoperability, Reusability
FTIR	Fourier Transform Infrared Spectroscopy
GDPR	General Data Protection Regulation
IP	Intellectual Property
KPI	Key Performance Index
LCA	Life Cycle Assessment
LCC	Life Cycle Costing
LCSA	Life Cycle Sustainability Assessment
OAIS	Open Archival Information System
TGA	Thermogravimetric Analysis
RBAC	Role Based Access Control
RP	Reporting Period
SEM	Scanning Electron Microscopy
SO	Specific Objective
TLS	Transport Layer Security

QSAR biodegradation	Quantitative Structure Activity Relationship biodegradation

Executive Summary

This deliverable (D13.4) provides an initial data management plan (DMP) for Reporting Period 1 (RP1) of the EU-funded Horizon Europe MAGICBIOMAT project. The main purpose of this deliverable is to report how research data is generated, collected, and shared throughout the RP1. In addition to that, standards adopted for common data models, interoperability, data curation, and long-term preservation are summarised. The project activities as well as the DMP follows FAIR principles. D13.4 will be revised throughout the project duration and a final version will be prepared.

Table of Contents

Acronyms 3

Executive Summary 5

List of Figures **Error! Bookmark not defined.**

List of Tables 8

1. Introduction 9

1.1. Purpose of Data Management Plan in MAGICBIOMAT 9

1.2. Ethical and legal frameworks 10

1.3. Structure of the deliverable 11

2. Purpose of Data Generation and Collection 11

2.1. Data Types, and Formats 13

2.1.1. Material characterisation data 13

2.1.2. Biodegradation study data 13

2.1.3. Life Cycle Sustainability Assessment Data 13

2.1.4. Training and validation data for the AI tool 13

2.2. Origin of Data 14

2.2.1. Internal data sources 14

2.2.2. External data sources 14

2.3. Data Storage Repository and Curation 14

3. FAIR Data Management 15

3.1. Making Data Findable 15

3.1.1. MAGICBIOMAT Community in Zenodo 15

3.1.2. Provisions for Metadata 15

3.2. Making Data Accessible 16

3.3. Making Data Interoperable 17

3.4. Increase Data Reuse 18

4. Allocation of Resources 18

5. Data and Infrastructure Security 18

6. Conclusion 19

List of Tables

Table 1 – Primary data generation goals and collection methods in MAGICBIOMAT _____11

Introduction

Uncontrolled littering of plastic waste is a global issue. Inadequate policies or the lack of enforcement of existing regulations regarding waste management and plastic use exacerbate the problem, especially in the case of marine litter[1]. While biodegradable alternatives offer a promising solution to plastic pollution in the open environment in some specific contexts, current biodegradable plastics are not a solution for inappropriate waste management or littering[2]. The use of biodegradable plastics in the open environment should be limited to specific applications for which reduction, reuse, and recycling are not feasible.

MAGICBIOMAT has deeply studied these challenges and aims to develop circular bio-based (BB) materials with programmed biodegradability in open environments. They will be demonstrated through two applications (biodegradable acrylic coated paper trays and BB agricultural mulch) with concerning rates of littering. To enable such programmed biodegradability of the BB materials, MAGICBIOMAT will develop a trustworthy AI powered tool to guide the design and manufacturing of novel materials. The process will consider requirements (properties, manufacturing) and end of life (EoL) needs and conditions, exploiting data from MAGICBIOMAT's development and biodegradation tests, complemented with open access data. Also, the proposed AI-powered tool will require access to large, diverse, and high-quality datasets to achieve robust performance and reliable outcomes. Thus, a project wide data management policy and strategy is required.

MAGICBIOMAT dedicates Task T13.5 of WP13 for DMP preparation (within RP1) and this report (D13.4) corresponds to the initial version of the DMP. It is considered a living document with the current version being delivered at M6 and will be constantly updated. The next official versions will be prepared at M36 and M48 for D14.2 and D15.2 respectively (one each corresponding to the RP2 and RP3).

Purpose of Data Management Plan in MAGICBIOMAT

The MAGICBIOMAT project participates in the Open Research Data pilot[3], a part of the open access to scientific publications and research data as well as Open Research Europe[4]. All project results (e.g., data, reports, deliverables, software tools) except those deemed to hold exploitable Intellectual Property (IP) or of sensitive nature will be released with open access to researchers, industries, policy officers, and other stakeholders. This will foster reuse of data and public results paving a way for higher impact of public investment in European research.

[1] United Nations Environment Programme, Coordinating Body on the Seas of East Asia, Marine litter and Plastic Pollution.
[2] European Commission, 2020, Biodegradability of Plastics in the Open Environment
[3] https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science/open-access_en
[4] https://open-research-europe.ec.europa.eu/

The purpose of the first version of the MAGICBIOMAT DMP is to present a consolidated view to the project partners regarding research data management. The policy on data generation, collection, and sharing throughout the project lifetime and data retention beyond are reported. Routes for reuse of the data within and outside the consortium are also governed by this deliverable. Standards for interoperability, data curation, preservation along with the legal framework are explained. The DMP ensures that the consortium adheres to the FAIR principles[5]. The European Commission's guidelines for research data quality, trusted sharing, secure storage, and privacy aware data collection are followed.

Ethical and legal frameworks

The ethical and legal framework for data collection, storage, processing, and sharing is broadly based on the GDPR[6] which is applicable to EU27 and EEA countries. GDPR regulates the personal data related to individuals within the EU by a legal and/or private person.

The stakeholder data collection (e.g., applicable in WP4 and WP9) will be performed in accordance with the GDPR and ethical standards in the respective countries of the consortium partners for the data collection (including any additional application regulations for data privacy and protection). Only some parts of the collected data are personal data (e.g., social life cycle analysis in WP9). In case of extensive personal data collection requirement arises during the project, MAGICBIOMAT partners will request personal data processing rights from the data controllers leading to formation of personal data processing agreements among the data controllers and partners.

The proposed AI tool must comply with the EU AI Act (Regulation (EU) 2024/1689)[7], which establishes the first comprehensive legal framework for AI powered solutions. The AI tool potentially falls under the EU AI Act's requirements for fostering safe and trustworthy AI systems and requires risk management protocols, continuous monitoring, and human oversight for AI systems. Key compliance requirements include:

- **Risk assessment and management**: Establishing a continuous iterative process for identifying and analyzing known and reasonably foreseeable risks throughout the AI system's lifecycle.
- **Transparency and explainability**: Ensuring users understand how the AI system makes decisions about biodegradable materials formulation and manufacturing guidance.
- **Data governance**: Implementing rigorous data quality, security, and accountability measures for the training datasets.
- **Human oversight**: Maintaining human supervision mechanisms for the AI system's operations and outputs.

[5] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18.
[6] REGULATION (EU) 2016/679, accessible at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679
[7] https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

All data collection, processing, and storage activities must comply with GDPR requirements, particularly for personal data gathered during focus groups, surveys, and stakeholder engagement workshops in WP4 as well as in WP9 (for social life cycle analysis). GDPR compliance requires:

- **Informed consent**: Obtaining freely given, specific, informed, and unambiguous consent from participants.
- **Data subject rights**: Providing processes for participants to access, correct, or request deletion of their personal data.
- **Special categories protection**: Implementing additional safeguards if collecting sensitive information about individuals.
- **Cross-border data transfer**: Ensuring appropriate safeguards when transferring data between project partners across different EU countries.

Further details on the ethical management and legal frameworks will be provided in project deliverables D16.1, D16.2, D16.3, and D16.4 which are due at M6, M18, M36, and M47, respectively.

Structure of the deliverable

This deliverable is structured as follows. Section 1 provides a general reference to the data management approach of MAGICBIOMAT with an initial consideration of the ethical and legal framework relevant to the project activities.

Section 2 provides a summary on the purpose of data generation and collection (at the pilot sites and through stakeholder survey), reuse of existing data, nature & origin of data, its utility, storage, curation, and data retention.

Section 3 dives into the FAIR data management discussing MAGICBIOMAT's approach for making data findable, accessible, interoperable, and reuseable.

Section 4 and 5 summarise allocation of resources for data storage, curation, retention, and preservation along with data security considerations.

Purpose of Data Generation and Collection

Table 1 lists the primary data generation goals (linked to the overall KPIs) and collection methods related to Specific Objectives of the project.

Table 1 – Primary data generation goals and collection methods in MAGICBIOMAT

Specific Objective	Primary data generation goals	Data collection methods/scope

SO3: AI-powered manufacturing guidance software	• **KPI 3.1**: Exploit at least 5 publicly available datasets and 2 project-generated datasets to train the trustworthy AI tool on chemical structure-biodegradation relationships. • **KPI 3.2**: Develop at least 5 new functionalities for the Paradise platform to optimize circular lifecycle of products.	• Task 3.1 focuses on data collection and AI-driven quality inspection to ensure high-quality data feeds into subsequent AI models. • Integration of experimental data from WP1-2 with biodegradation monitoring data from WP5-6
SO5: Open environment biodegradation testing	• **KPI 5.1**: Implementation of biodegradability assessments in 3 conditions (soil, fresh water, marine water) across 4 different European climate zones • **KPI 5.2**: Collection of degradation performance data showing specific percentage targets over time periods. • **KPI 5.3**: Three rounds of LCSA data collection (baseline, intermediate, final)	• 224 samples for short-term degradation studies and 224 samples for long-term studies • Environmental monitoring across diverse European climate conditions • Integration with AI tool development for predictive modelling
SO4: Societal acceptance and behaviour studies	• **KPI 4.1**: Data from at least 5 focus groups, 1 survey, and 2 stakeholder engagement events. • **KPI 4.2**: Data from 2 studies evaluating labelling system effectiveness and behaviour change strategies.	• Online focus groups with approximately 5 participants per group. • Representative surveys using platforms like Prolific. • Stakeholder engagement workshops with industry, policymakers, and environmental groups.

Data Types, and Formats

The WPs of MAGICBIOMAT will generate and collect data related to material characterisation, biodegradation study, life cycle sustainability assessment (LCSA), training & validation data for the proposed AI powered tool. While these are an initial set of data, as the project evolves, many more data types and formats will be involved. Those updates will be reported in the future versions to this deliverable.

Material characterisation data

- **Mechanical properties**: Tensile testing, 3-point bending, impact testing (numerical data)
- **Thermal properties**: DSC, DMA, TGA, rheology measurements (spectra and numerical data)
- **Physicochemical properties**: FTIR spectra, SEM images, colorimetry data
- **Morphological data**: SEM and microscope images

Biodegradation study data

- **Controlled conditions**: CO_2 emission measurements via respirometry, weight loss analysis
- **Open environment**: Mass and area changes, environmental parameter monitoring (temperature, humidity, pH, conductivity)
- **Ecotoxicity data**: Microorganism and plant impact assessments, microplastics generation measurements

Life Cycle Sustainability Assessment Data

- **Environmental data**: LCA impact assessment data, carbon footprint measurements.
- **Economic data**: LCC information and cost-benefit analyses.
- **Social data**: Social LCA data from stakeholder surveys and assessments.

Training and validation data for the AI tool

Dataset composition:

- **Public datasets**: publicly available datasets (e.g., QSAR biodegradation from UCI Machine Learning Repository)
- **Project generated data**: two comprehensive datasets from experimental work.
- **Integrated features**: Polymer structures, properties, environmental factors, biodegradability rates

Data formats:

- Structured CSV files for tabular data

- JSON formats for complex nested data structures
- Standard image formats (TIFF, PNG) for microscopy data useful for calculating area loss (for open environment degradation studies) estimation through images
- Standardised chemical notation formats (SMILES, MOL files) for molecular structures

Origin of Data

The project will bring together internal and external data sources; an estimated initial list is given below.

Internal data sources

- **TUS, UNIPD, OK**: Material characterization data from WP1-2 development activities
- **IPC**: Controlled biodegradation test results and ecotoxicity assessments
- **ISOTECH**: Open environment monitoring data from four climate zones
- **DIGI**: AI model performance data and platform integration metrics
- **USHEFFIELD**: Social and behavioural study data from consumer research
- **All Partners**: LCSA data collection coordinated by DIGI and IPC

External data sources

Public Datasets:

- UCI Machine Learning Repository for polymer biodegradation data
- Chemical databases for molecular structure information
- Environmental databases for climate and biodegradation modelling
- Literature-based datasets from systematic reviews

Collaborative Sources:

- Industry partner databases (with appropriate confidentiality agreements)
- Standardisation organisation databases for material testing protocols

Data Storage Repository and Curation

Each pilot will utilise its own existing database and infrastructure for data collection and storage. The processed data will be stored at a repository provisioned by DIGI via its cloud infrastructure (data storage location being

Europe). Open Archival Information System (OAIS), defined in ISO 14721:20120[8], will be utilised as a reference model for long term curation and preservation of processed data in this project. The post-project preservation strategy will be finalised by the partners at a consortium-wide discussion around M46. The strategy will also consider any commercial potential and IPR of the partners for future use and access of the curated data.

Through the curation process, the data will be organised, described, cleaned, enhanced, and preserved for post-project use. The consortium has taken a collective decision to archive open results (including reports, deliverables, publications, and datasets) via a 'community' on Zenodo, which is an open research data repository of CERN.

FAIR Data Management

It guides the DMP towards maximising the access to and reuse of research data collected and generated in this project. Also, datasets containing sensitive data will not be completely released as open access. The following sections detail about the FAIR practices adopted in MAGICBIOMAT.

Making Data Findable

All data produced and used in the project will be made discoverable with metadata, identifiable, and locatable by means of Digital Object Identifiers (DOI).

MAGICBIOMAT Community in Zenodo

The project has already set up the community[9] which will act as the central repository of all project results and open data (complete with metadata) and is aligned with OpenAIRE[10]. A dedicated glossary will be created to label all data sets and results, and to provide search functionality for stakeholders seeking particular data. Zenodo supports adding search keywords or tags, which optimise the possibilities for quickly finding datasets, results, and their reuse.

Provisions for Metadata

Each dataset published in MAGICBIOAT Zenodo (open data), and internal data (sensitive data) repository will be accompanied with a metadata as follows. A dataset naming convention will be prepared by DIGI covering at-least the following:

- Dataset name
- DOI
- Dataset description

[8] ISO 14721:2012 - Space data and information transfer systems — Open archival information system (OAIS) — Reference model, accessible at - https://www.iso.org/standard/57284.html
[9] https://zenodo.org/communities/magicbiomat/
[10] https://www.openaire.eu/

- Data utility
- Type
- Nature
- Keywords
- Scale
- Origin
- License
- Version number
- Personal data/anonymized data
- Standards and metadata
- Data sharing policies
- Archiving and preservation
- Project name
- Grant agreement number

Making Data Accessible

As stated previously, data will be released with open access license with exceptions such as -

- Data coming from a private source or a third party with a license restricting open sharing or open access.
- Datasets that contain personal information which must be treated in compliance with GDPR.
- Data that have commercial potential and which are therefore treated with restricted IPR licenses.

The project envisages several of its results, deliverables, and data to be necessary for data owners and other stakeholders to reproduce our results, including: (a) validation data, generated during the project's work and made available via MAGICBIOMAT community on Zenodo; (b) metadata and support documentation to enable comparable architecture (for the AI tool) and information models to be established (for the biodegradation parts); and (c) underlying data from publications. In addition to the metadata and support documentation that will accompany project data/results, tutorials and training materials will also be provided. Background context and insight will be provided on potentially sensitive data to ensure as broad as possible an understanding among stakeholders of how to replicate the work themselves.

It is therefore a central interest of MAGICBIOMAT to make the data accessible. Open results (consisting of data, technical and non-technical publications, software tools) will be centrally stored in the Zenodo community. This will facilitate reproducibility of project results where access to data for validation of the result is necessary.

For the scientific publications, project researchers will be initially directed to Sherpa-Romeo[11] to check the open access policy of any targeted Journals. This will help inform which Journal is eventually chosen. This will create a central repository, complete with metadata, of all project work in a popular database that is aligned with OpenAIRE. Green open access (self-archiving) and/or Gold open access (open access publishing) will be done for the scientific publications and public deliverables.

Making Data Interoperable

To ensure that the data collected and generated within MAGICBIOMAT can be effectively utilised by the AI tool, the project will implement a robust framework for AI-specific interoperability. This begins with the application of AI-driven data quality inspection protocols, as outlined in Task 3.1, which are designed to systematically identify, flag, and correct anomalous or inconsistent data entries prior to their inclusion in training and validation workflows. The project will standardize feature engineering processes, including normalization and scaling of numerical variables, to ensure that diverse datasets—ranging from material properties to biodegradation rates—are compatible with machine learning algorithms. Furthermore, cross-validation protocols will be harmonized across different data types, enabling the AI tool to integrate and compare results from both experimental and external datasets in a statistically rigorous manner. The development of a dedicated API layer will facilitate seamless data exchange between the various data repositories and the Paradise platform[12], supporting real-time data ingestion and model retraining. These measures collectively ensure that data from multiple sources and formats can be interoperably leveraged for AI model development, deployment, and continuous improvement, in compliance with the FAIR principles and the EU AI Act requirements.

In addition, such AI specific interoperability, semantic interoperability will also be assured in this project. It will adopt established ontologies and controlled vocabularies relevant to materials science, environmental monitoring, and biodegradation as necessary for the work of WP3. This approach ensures that data labels, descriptors, and metadata are consistently defined and interpreted across all project partners and external stakeholders. The project will utilize linked data methodologies, enabling the connection of related datasets through standardized semantic relationships, which enhances data discoverability and integration. By embedding semantic metadata and leveraging community-agreed standards, MAGICBIOMAT will facilitate not only internal data reuse but also interoperability with external datasets and platforms, supporting broader scientific collaboration and compliance with Horizon Europe open science mandates.

[11] Sherpa Romeo is an online resource that aggregates and analyses publisher open access policies from around the world and provides summaries of publisher copyright and open access archiving policies on a journal-by-journal basis, accessible at - https://v2.sherpa.ac.uk/romeo/.
[12] This is an internal R&D platform of DIGI.

Increase Data Reuse

To increase data reuse, it needs to be appropriately licensed. According to the Grant Agreement, data and results (including scientific publications and software tools) are owned by the beneficiary or beneficiaries that generate(s) them. Owner(s) of such results will be encouraged to release them with Creative Commons Attribution 4.0 (CC-BY-4.0)[13] or another open source license like Apache 2.0. The data retention policy of the project will also facilitate increased data reuse for at-least ten years beyond the project conclusion.

Allocation of Resources

The MAGICBIOMAT partners do not foresee an immediate expense for short and long term storage, curation, and preservation of the data generated in this project. The Zenodo repository also does not require allocation of any expense. In case any unforeseen cost arises for providing open access to project data, which will be covered through the responsible partners budget in this project.

Data and Infrastructure Security

MAGICBIOMAT implements a comprehensive, multi-layered approach to data and infrastructure security, in alignment with Horizon Europe requirements and relevant EU regulations. All project data is classified according to sensitivity (e.g., public, confidential, or restricted) with access strictly controlled through role-based access control (RBAC) systems. These systems enforce the principle of least privilege, ensuring that users only have access to the specific data necessary for their roles. For sensitive or restricted data, additional safeguards will be added in place, including multi-factor authentication and periodic access reviews to prevent unauthorised access.

Technical measures are central to the project's security posture. All data at rest in DIGI's Paradise platform is encrypted, while data in transit is protected by with TLS 1.3. The project infrastructure, hosted on secure cloud platforms compliant with the European Open Science Cloud (EOSC) guidelines, benefits from network segmentation, firewalls, and intrusion detection systems. Regular vulnerability assessments and security patching are conducted to address emerging threats. To ensure data integrity and availability, automated daily backups are performed, with copies stored in geographically distributed locations. These backups are regularly tested for recoverability, and disaster recovery plans are established with defined recovery time objectives.

Compliance with the General Data Protection Regulation (GDPR) is rigorously maintained throughout the project. Data minimisation principles are applied, and personal data is pseudonymized wherever possible. Data retention and deletion policies are clearly defined, and privacy impact assessments are conducted for high-risk processing

[13] http://opendefinition.org/licenses/cc-by/

activities. Continuous monitoring systems track data access and usage patterns, while anomaly detection tools help identify and respond to potential security incidents.

For AI-specific work items, MAGICBIOMAT incorporates safeguards such as Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment[14].

Conclusion

This deliverable is an initial report on the data management plan of the project following the FAIR guidelines. The developed DMP relies on standards like DOI, OpenAIRE initiative, and Zenodo open access data repository for sound and seamless implementation of the data management methodology. It will assure FAIR principles observed to the open results (e.g., data, publications, and software tools) of the project. This document will be updated throughout the project duration and a final DMP (D15.2) will be released at M48.

[14] https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment