

D6.1. Project Coordination

MONITORING SYSTEM OF THE ENVIRONMENTAL AND SOCIAL SUSTAINABILITY AND CIRCULARITY OF INDUSTRIAL BIO-BASED SYSTEMS

Grant Agreement Number 101112457

Deliverable name: Data Management Plan and IPR agreements 1
Deliverable number: D6.1
Deliverable type: DMP
Work Package: WP6: Project Coordination and Consortium
Management
Lead beneficiary: YAG
Contact person: Emad Yaghmaei
Dissemination Level: Public
Due date for deliverable: 31st October, 2023



The project is supported by the Circular Bio-based Europe Joint Undertaking and its members. 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 CBE JU. Neither the European Union nor the CBE JU can be held responsible for them.

DOCUMENT CONTROL PAGE

Author(s):	Elena Sertore, Emad Yaghmaei
Contributor(s):	All partners
Reviewer(s):	Poojan Timilsina
Version number:	4.0
Contractual delivery date:	[31-10-2023]
Actual delivery date:	[28-09-2023]
Status:	[Final]

REVISION HISTORY

Version	Date	Author/Reviewer	Notes
v.0	[31-08-2023]	Elena Sertore, Emad Yaghmaei	Creation, First draft
v.1	[04-09-2023]	Marvin Ikedo	Comments
v.2	[14-09-2023]	All partners	Input from all WP leaders
v.3	[26-09-2023]	Poojan Timilsina	Internal review
v.4	[28-09-2023]	Marvin Ikedo, Emad Yaghmaei	Final version submitted

ACKNOWLEDGEMENTS

The work described in this publication was subsidised by Horizon Europe (HORIZON) framework through the Grant Agreement Number 101112457.

DISCLAIMER

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 CBE JU. Neither the European Union nor the granting authority can be held responsible for them.

TABLE OF CONTENTS

DOCUMENT CONTROL PAGE	2
REVISION HISTORY	2
ACKNOWLEDGEMENTS	2
DISCLAIMER	2
TABLE OF CONTENTS	3
EXECUTIVE SUMMARY	4
LIST OF ACRONYMS	5
1. INTRODUCTION	6
1.1 Description of the document	6
1.2 Work Package 6 and Task 6.4	6
2. DATA SUMMARY	7
2.1 Data description	7
3. DATA MANAGEMENT STRATEGY	11
3.1 FAIR Data	11
4. ALLOCATION OF RESOURCES	13
5. DATA SECURITY	13
5.1 Provisions for data security	13
6. DATA ETHICS	13
7. CONCLUSION	14

EXECUTIVE SUMMARY

This document contains the Data Management Plan of the BIORADAR project, funded under CBE JU. BIORADAR aims at developing metrics, methods, and digital monitoring tools to be used by policy-makers and investors in order to assess the performance of industrial bio-based systems in terms of material circularity and environmental and social impact. The selected use-cases for the project belong to three industry sectors: bio-based textile, fertilizer, and packaging. The BIORADAR consortium is composed of seven partners and the project execution is structured into six work packages.

The BIORADAR DMP falls within the deliverable 6.1 of Work Package 6, namely Project Coordination and Consortium Management. In particular, it addresses the deliverable requirement of managing and maintaining an appropriate data repository. As such, the BIORADAR DMP outlines strategies, processes, and resources employed to ensure an effective management of data throughout the lifecycle of the project. The key objectives of the DMP are to promote data integrity, facilitate data sharing and reuse, and ensure compliance with relevant regulations in Europe and funding agency requirements from Horizon Europe. It is based on the template provided by the European Commission: firstly, a summary of the data expected to be relevant in each work package is provided. Secondly, a strategy is developed to follow the FAIR guidelines of data management. The DMP in its first version reflects the current status of data collection and usage of BIORADAR and, as a result, it is conceived as a living document which will be subject to changes, updates, and reviews throughout the lifecycle of the project.

LIST OF ACRONYMS

AI - Artificial Intelligence

CBE JU - Circular Bio-Based Europe Joint Undertaking

DMP - Data Management Plan

EBP - Exploitation and Business Plan

FAIR - Findable, Accessible, Interoperable and Reusable

MOOC - Massive Open Online Course

WP - Work Package

YAG - YAGHMA

1. INTRODUCTION

The Horizon Europe Model Grant Agreement requires the establishment of a regularly updated DMP which complies with the relevant regulations in Europe and funding agency criteria. This DMP is therefore based on the template recommended for Horizon Europe beneficiaries. By completing the sections of the template, the requirements for research data management of Horizon Europe as described in article 17 and analysed in the Annotated Grant Agreement, are all addressed.

1.1 Description of the document

This report describes the strategy adopted by the BIORADAR consortium to ensure an effective and responsible management of the data collected and used throughout the lifecycle of the project. It therefore contains preliminary information on the type of data used and/or expected to be collected and used within each WP, and the FAIR principles regulating their management.

Furthermore, the report includes information on the infrastructure in place to support all BIORADAR partners in managing non-commercially sensitive data. It consists of three main building blocks: the BIORADAR Research Service Center, the repository media BIORADAR, and the software YAG Workbench.

1.2 Work Package 6 and Task 6.4

This deliverable refers to Task 6.4 included in WP6: Project Coordination and Consortium Management. The Task establishes that within four months from the start of the project, a DMP is drafted and kept up-to-date throughout the implementation of the project. The DMP shall be available to all consortium partners at any time and revised after the main data collection phase and at the end of the project. Additionally, the DMP shall evaluate the publication and dissemination of the data collected throughout the project. It follows that the exploitation plan shall be developed by observing agreements and regulations of the DMP. If necessary, the DMP can be revised and necessary updates will be indicated in the Periodic report so that future efforts with BIORADAR remain compliant with EU regulations.

2. DATA SUMMARY

The main types of data currently envisaged to be collected in the project include:

- (1) Specifications of the use-cases by use case partners (including potentially sensitive commercial information);
- (2) Meetings, workshops and event summaries;
- (3) Results of surveys among stakeholders;
- (4) Data on AI-driven benchmarking;
- (5) Reports (All deliverables);
- (6) Scientific publications;
- (7) Qualitative and quantitative data generated in the self-assessment tool;
- (8) Data generated in the global bio-based systems regulatory tracker tool.

2.1 Data description

The table below provides an overview of the data collected and used in each work package of the BIORADAR project. It is currently based on preliminary plans of each WP, thus it might be subject to changes during the execution and implementation of the project.

Work Package	Data	Data Format	Description
WP1 <i>Identifying and Assessing Sustainability aspects (Environmental, Economic, Social) of Industrial Bio-based Systems and embedding them into BTI Framework</i>	Data on bio-based value systems and products pertaining to the sectors to which BIORADAR applies, namely: packaging, textile, and fertiliser. Source of data may include literature and other projects on the topic; guides and reports from public and private institutions on bio-based products and value chains	Structured data / Real data / Text data	Data characterising bio-products; e.g., TRL, bio-based carbon content, chemical classification, etc.
		Structured data / Real data / Text data	Data characterising the boundaries of EU supply chains, associated countries and across borders
		Structured data / Real data / Text data	Data on existing methods/metrics assessing environmental impact of the selected bio-based systems, e.g., greenhouse emissions, carbon removal potential of bio-based systems, land use indicators
		Real data / Text data	Data on standards of industrial bio-based systems
		Structured data / Real data / Text data	Data on capital, operational costs and cost variations deriving from the application of the proposed technologies and from the results of cost

			identification and cost-benefit analysis
		Real data / Text data	Data on social and socio-economic benefits of the implementation of the proposed products; e.g., local employment and number of jobs created
WP2 <i>Identifying and Assessing Circularity aspects of Industrial Bio-based Systems and embedding them into BTI Framework</i>	<p>Data on circularity indicators methodologies for industrial bio-based systems. Source of data may include literature and other projects on the topic, as well as the study of existing methodologies.</p>	Real data / Structured data / Text data	Data on existing metrics on circularity for industrial bio-based systems and use-cases; data on resources efficiency, effectiveness, energy, and a life-cycle perspective.
		Real data / Structured data / Text data	Data on new indicators of circularity for industrial bio-based systems; e.g., health and safety.
		Real data / Structured data / Text data	Data on end-of-life issues; e.g., energy recovery processes, secondary materials, proportions of recycled or reused.
		Real data / Structured data / Text data	Economic data on circularity of industrial bio-based systems; e.g., surveys, bibliography, LCA databases.
WP3 <i>Developing and Validating Digital Monitoring Dashboard/Tools: Self-Assessment, AI-driven Analytics Platform and Regulatory Tracker Tool</i>	<p>A synthetic dataset used to train the AI-driven benchmarking and analytics platform based on discussion with the use cases and indicators received from WP1 and WP2.</p>	Structured data / Real data / Text data	Data to design and develop a deterministic model to reconfigure the bio-based product
		Structured data / Real data / Text data	Data on possible changes in production requirements
	<p>Data collected to build and test the Self-Assessment Tool, composed of different layers, e.g. data acquisition layer, data</p>	Structured data / Real data / Visual data (images/video)	Data visualization using software app, for example Tableau, based dashboard and charting engine
		Structured data / Real data	Dataset of bio-based project sustainability and

	treatment, and data storage layers	data / Text data	circularity data for AI-driven Benchmark
		Structured data / Real Data / Text data	Data from streamlined Reporting
		Structured data / Real Data / Text data	Organisation data
	Data to build the Regulatory Tracker tool	Structured data / Real Data / Text data	Data on hard laws, soft norms, and regulations in the bio-based field
WP 4 <i>Upscaling and Replicating the project results</i>	Data from the “BIORADAR Replication Facility” and the BIORADAR implementation scorecard	Structured data / Numerical data	Economic, environmental, social, and circularity data of the use-cases used by the BIORADAR implementation scorecard.
		Unstructured data / Text reports	Data on scientific results, novel business models, inputs from stakeholders.
	Data on EBP	Unstructured data / Text report	Data on economic expected quality, marketing support services, expected levels of investment, financial structures, and environmental and social indicators
	Data in the MOOC	Unstructured data / Text / Images / Video	Data on bio-transition framework, its methodologies and metrics
		Unstructured data / Text / Images / Video	Training components data from UNI on standardization and bio-based products

WP5 <i>Communication, Dissemination and Exploitation activities</i>	Data on project Dissemination, Exploitation, and Communication Plan	Text data	Data on project results is delivered in the format of project leaflets, brochures, posters, project video, newsletters, promotional materials, social media posts, online content on project website; scientific articles and papers, sessions, scientific workshops, course materials; events part of the BIORADAR agenda; stakeholder survey
	Data to set up stakeholder analysis and stakeholder engagement strategies	Contact data / Text data	Data on ecosystem and stakeholders analysis, e.g. stakeholders' interest, attitude, influence, knowledge, and expectations
	Data to set up BIORADAR Clustering	Contact data / Text data	Data on BIORADAR's network with other projects and initiatives; Data on other EU/national projects
	Data from pre-standardization activities	Text data	Technical experts' feedback on standardization of project findings
WP6 <i>Project Coordination and Consortium Management</i>	Data pertaining internal communication and project management	Text data	Data related to financial administration, risk management, quality assurance plan, and ethical impact assessment are collected and shared with project partners; Minutes of meetings, summary of events, and notes of workshop

3. DATA MANAGEMENT STRATEGY

All data gathered in the project are managed according to Open Research Europe's Guidelines for Handling Research Data. The Data Management Plan for BIORADAR follows YAG's internal research data handling guidelines, Research Code of Conduct and the Open Access Policy. The plan includes detailed roadmaps on the planning, collection, pro-cessing, and publication of research data.

A well-established infrastructure for data management is used to support all BIORADAR partners in managing non-commercially sensitive data. It consists of three main building blocks:

- the BIORADAR Research Service Center, which supports BIORADAR partners (and YAG as the coordinator) on aspects related to the infrastructure for data management, individual data management, questions, and publication.
- All project data will be stored in a password-protected OneDrive & Share folder which ensures frequent back-ups and contains the possibility to restore data. The data will be stored during and up to two years after the project finalisation.
- All data is treated according to the FAIR principles¹, including research papers as well as open datasets.
- BIORADAR publish data with public dissemination level in a truly open science approach, i.e., making data not only available for project partners, but enabling proactive collaboration with stakeholders outside of the consortium to exploit the project results. BIORADAR allows for easy data sharing according to acknowledged standards and enables collaborative activities (e.g., peer review, creation of digital lab journals). The largest amount of data generated in the project can be expected in the area of process data. A long-term storage the DSS (Data Science Storage) building block will be available for BIORADAR. Furthermore, BIORADAR allows the link of metadata to primary data. Metadata can, for example, be the indicators for environmental and social sustainability and circularity.

3.1 FAIR Data

All data will be handled according to the FAIR principles described below.

Making Data Findable

The following measures will be applied to ensure the data generated during the project will be findable: firstly, globally unique persistent identifiers will be issued and supported for publications. Furthermore, discovery metadata schemes such as Dublin Core, DDI and Schema.org will be applied. Lastly, the consortium will ensure that all published data will be registered or indexed in searchable

¹ https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

resources (e.g., DataCite, Schema.org), which will be supported by publishing and managing research data open source. Keywords will be used whenever it fits the purpose and features of the publication or document. The list of keywords will be updated continuously throughout the project.

Making Data Openly Accessible

BIORADAR follows an open science strategy, which ensures that items are retrievable by their identifier using a standardised communications protocol, e.g., Hypertext Transfer Protocol, Rsync over Secure Shell, Representational State Transfer, etc.; furthermore, it supports both session and API key base authentication and authorisation protocols and facilitates appropriately robust deaccession procedures should changes need to be made to the items, or should items need to be removed completely.

The choice of trusted data repository will be based on the sensitivity of the data used in the project. The project will deposit the generated synthetic data that is evaluated as “non-personal data” in an appropriate repository. The real-world data will be deposited in a trusted repository after being anonymized and generalized to ensure the processed data is not individually identifiable. The sensitive real-world data and synthetic data that are still considered as “personal data” will be managed and maintained at their source site in controlled environments. The metadata of these datasets will be published in a trusted repository as well as the requirements and procedure of requesting the access to the data.

Making Data Interoperable

BIORADAR uses formal accessible, shared and broadly applicable language for knowledge representation via a Linked Data approach supported with JSOB-LD for Schema.org. The use of FAIR controlled vocabularies and data models and the establishment of qualified references to other relevant data or research outputs via the DDI schema are further important guidelines that will be followed.

The project will re-use existing datasets from data repositories and may link with newly generated or collected data from the project. The project will ensure the data can be linked using common/shared ontologies and vocabularies.

Increase Data Re-Use

The consortium will ensure that results from the project will be released with a clear and accessible usage license, i.e., published under the default license CC0. Moreover, the consortium will ensure that items are associated with detailed provenance, including information about authors, providers, distributors, and other related data, and supported by W3C PROV and that they meet domain-relevant community standards.

4. ALLOCATION OF RESOURCES

The costs for making data FAIR in the project are estimated to 0,5 to 1 person months per partner in the form of personnel costs. The owner of the data will take responsibility for its data in collaboration with the Data Protection Manager. The resources for long-term preservation (costs and potential value, time to keep the data) are decided on a per-user/per-partner basis.

5. DATA SECURITY

5.1 Provisions for data security

All project data will be stored in a password-protected OneDrive & Share folder which ensures frequent back-ups and contains the possibility to restore data. Data transfer between partners should always be protected by a password. All partners are responsible for ensuring data security by appropriate measures (e.g., encryption) and regular back-ups.

The non-sensitive synthetic data will be stored in trusted/certified research data repositories. The security standards of the repository will apply in this case. The real-world data, if any, in the project will be pseudonymized and stored securely in encrypted servers with access controlled by strict authentication mechanisms such as university servers or private cloud infrastructure. The data will be backed up periodically to prevent data loss and ensure data recovery if needed. The data transfer BIORADAR D6.1 may happen, if necessary, through encrypted channels. The minimum necessary amount of data can be transferred after anonymisation or pseudonymisation.

6. DATA ETHICS

The project will combine qualitative and quantitative data collection to gain a deep understanding of industrial bio-based systems. Specifically, the project will conduct a quantitative survey; may conduct qualitative interviews and/or focus groups with representatives of bio-based transition initiatives, NGOs/CSOs; policymakers; local businesses and bio-based industry representatives; and scientists/researchers. Coordination and support action will be carried out in use-cases in bio-textile, bio-fertiliser, and biopackaging. Any cross-border transfers of identifiable information will be subject to appropriate safeguards per GDPR Article 46. The BIORADAR' Ethics Policy sets out in detail the ethics principles that govern universal research involving bio-based industrial stakeholders, their data namely principles of safety and well-being, consent, anonymity, confidentiality, and data protection.

In this project the partner IRIS aims to develop an AI-matchmaker engine for AI-driven benchmarking analytics platform with data in sustainability and circularity of industrial bio-based systems collected and available in the BIORADAR WP1 and WP2. The development of the AI-engine will include the generation of synthetic datasets to train the engine and the demonstration of the engine. Besides, we will validate the tool with anonymised real data coming from the use-cases. The partner YAGHMA will be responsible for the ethical assessment during the development of the AI tool.

7. CONCLUSION

This report describes the BIORADAR project DMP, which will serve as a guide for all consortium partners to collect, use, and manage the data during the project lifecycle, as established by Deliverable 6.1.

In its current version, the DMP reflects the initial phase of the BIORADAR project in terms of data usage and collection. It provides an overview of the data used or expected to be used during the duration of the project, as well as the strategies, processes, resources, and infrastructure in place to ensure an effective and responsible management of data that complies with data security and ethics standards. The DMP is therefore expected to be subject to changes and updates throughout the implementation of BIORADAR.