Data Management Plan

Deliverable 7.2

Author: **Helena Vilaça**

Date: 30 November 2022

Funded by the European Union

Project	
Action Number	GA 101058654
Action Acronym	Waste2BioComp
Action Title	Converting organic waste into sustainable bio-based components

Document	
Document Identifier	D7.2
Due date of delivery to EC	30 November 2022
Actual date of delivery to EC	30 November 2022
Dissemination level	Public
Work package	7

Contributors	
Main Contributor	Helena Vilaça (CITEVE)
Contributor	Carla Silva (CITEVE)

Control Sheet			
Version	**Date**	**Editor**	**Summary of Modifications**
1.0	30/11/2022	Helena Vilaça	Initial version

Table of Contents

List of Abbreviations

Acronyms	Description
DMP	Data Management Plan
DoA	Description of the Action
DOI	Digital Object Identifier
FAIR	Findable, Accessible, Interoperable, Reusable
IPR	Intellectual Property Rights
ISBN	International Standard Book Number
MS	Microsoft
n.a.	Not applicable
R&I	Research and Innovation

1. Data summary

This Data Management Plan (DMP), issued by M06 and updated during the project, will contain the rules for data and results privacy and protection (i.e., collection, storage, and handling), ensuring that research data and other outputs will be managed in line with the FAIR principles (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable). The consortium partners will collaborate to ensure that quality procedures, based on a "sustainable and innovative scholarly communications ecosystem" are adopted, and they will promote bidirectional cooperative work and systematic sharing of knowledge allowing the use and re-use of data and protocols by unforeseen collaborators, including citizens, to increase the social impact of research. The workflow will go through different stages of the research process, including:

- discovery or compilation of information;
- analysis of the generated data;
- writing;
- publication;
- outreach;
- results' assessment.

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

Different existing data will be re-used, namely the knowledge already published (in scientific papers or patents) by RTDs and HES partners, but also some from the companies, so that this knowledge can be used and adapted for the new raw materials and processes being used in the W2BC project (e.g., fermentation processes and use of cyanobacteria by HSKL, encapsulation technology on PHBs by UDC, chemistry approaches to produce PHAs by HSKL, robotic developments by INESC TEC, inkjet printing technologies by MTEX NS, production of bio-based pigments by PILI).

What types and formats of data will the project generate or re-use?

The research outputs (types and formats of data) expected from the project are data regarding the different developed bio-based components; adaptation of smart manufacturing processes for bio-based materials; testing of all components & materials; production & evaluation of the demonstrators in a relevant scenario; different EoL alternatives of the bio-based materials; toxicology, impact & regulatory assessments; training activities. This data will be in the form of e.g., written synthetic methods/protocols, images, charts, properties reports, schemes (which will all be published in scientific journals with open access, and/ or in the project deliverables), physical samples, reference architectures, and will be organized in dedicated reports and deliverables (see work plan).

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

The generated data will be of huge value to the project, as this will contain all the information about project outputs, exploitable results, and increase its impact in industry, academia, and general public. Without the generation of this data, it would not be possible to achieve the project objectives, expected results, outcomes, and quantified dissemination, exploitation, and communication measures.

What is the expected size of the data that you intend to generate or re-use?

In a phase so initial of the project it is difficult to predict the size of the data to be generated. This question will be answered in the next version of the deliverable (to be submitted on M24).

What is the origin/provenance of the data, either generated or re-used?

The data came from scientific publications (available online), patents and the in-house generated knowledge of each partner.

To whom might your data be useful ('data utility'), outside your project?

The data generated by W2BC is expected to be of great use to different targeted audiences, both on the academic side and on the industry side, namely:

- Manufacturing industries in the three VCs (textiles, packaging, and footwear, namely, textile finishing & printing industries, mask producers, sole & insole and shoes producers, rigid plastic packaging & plastic films producers);

- Players across the textile, packaging, and footwear VCs, e.g., designers, brands working with personalized collections;
- Players from other VCs, e.g., automotive and building & construction industries (PHA-basedfoams, composites, etc.), producers of plastic furniture and/or household articles, pharma and cosmetic companies (e.g. micro- and /nanoparticles, biodegradable packaging);
- Scientific community (fields of biopolymers, biofoams, bio-composites / bio-films, bio-based micro- and NCs, bio-inks);
- Workers & Students;
- Standardisation bodies on bio-based materials;
- Consumers in general (textiles, plastic packaging and footwear consumers).

2. FAIR Data

2.1. Making data findable, including provisions for metadata

Will data be identified by a persistent identifier?

Persistent identifiers, e.g., digital object identifier (DOI) and international standard book number (ISBN) will be used and make the generated outputs more findable and citable without compromising the Intellectual Property Rights (IPR).

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Ain this early stage of the project, the consortium still does not know if metadata will be used/generated. In case it is created, formal, accessible, shared, and broadly applicable formats and vocabularies such as Resource Description Framework, comma-separated values, and JavaScript Object Notation for Linked Data will be used, to ensure its Findability and Interoperability.

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

n.a. to this version of the deliverable.

Will metadata be offered in such a way that it can be harvested and indexed?

n.a. to this version of the deliverable.

2.2. Making data accessible

2.2.1. Repository

Will the data be deposited in a trusted repository?

W2BC impact will not rely only on traditional citation metrics. Diverse online research outputs, such as social media, online news media or online reference managers, will also be considered. A open collaboration platform (MS 365 SharePoint) is being used to share data between partners, to facilitate knowledge sharing during the project timeline, serving also as a trusted repository.

Have you explored appropriate arrangements with the identified repository where your data will be deposited?

CITEVE has a contract with Microsoft (MS), and MS 365 SharePoint is the standard trusted repository for all of the institute information and that of the R&I projects it is involved with.

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

This will be analysed together with the MS 365 team of CITEVE, and, if needed, inquired to the Microsoft company.

2.2.2. Data

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

The **Accessibility** of the data will follow the principle "*As open as possible, as closed as necessary*", with sensitive data being made publicly available when IP protection is enabled and when it is deemed commercially appropriate, via conference presentations, peer reviewed open-access publications, and regular updates on the W2BC website. The access to restricted data will be considered in instances where the consortium decides to share data confidentially with a select group outside the consortium (e.g., textile, packaging, shoe soles/insoles or PES recycling professionals, renowned researchers & potential customers, EC members for verification purposes), for which an NDA will be signed before data sharing.

One of the main advantages of an Open Science policy is the facilitation of scientific knowledge circulation and its **Reusability**, avoiding duplication of data, following the principles of European initiatives like GAIA-X and International Data Spaces.

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g., patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

The dates for each partner to communicate to the consortium that it intends to make a publication, and the possibility of a partner to embargo that publication is set in the CA, as follows:

"Prior notice of any planned publication shall be given to the other Parties at least 45 calendar days before the publication. Any objection to the planned publication shall be made in accordance with the Grant Agreement by written notice to the Coordinator and to the Party or Parties proposing the dissemination within 30 calendar days after receipt of the notice. If no objection is made within the time limit stated above, the publication is permitted.

An objection is justified if:

a) the protection of the objecting Party's Results or Background would be adversely affected, or

b) the objecting Party's legitimate interests in relation to its Results or Background would be significantly harmed, or

c) the proposed publication includes Confidential Information of the objecting Party.

The objection has to include a precise request for necessary modifications.

If an objection has been raised the involved Parties shall discuss how to overcome the justified grounds for the objection on a timely basis (for example by amendment to the planned publication and/or by protecting information before publication) and the objecting Party shall not unreasonably continue the opposition if appropriate measures are taken following the discussion.

The objecting Party can request a publication delay of not more than 60 calendar days from the time it raises such an objection. After 60 calendar days the publication is permitted, provided that the objections of the objecting Party have been addressed."

Will the data be accessible through a free and standardized access protocol?

For the public, data without IP issues will be made accessible through the W2BC website and the publication of periodic newsletters.

For the consortium, the access to all generated data is made accessible by giving access to a shared MS 365 SharePoint folder.

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

n.a. to this version of the deliverable

How will the identity of the person accessing the data be ascertained?

MS 365 SharePoint allows to follow the history of access to the documents on the shared folder, so it is simple to follow who has accessed what document, and if that person has made changes to the document. The SharePoint stores the different versions of the same document, which can be transferred or restored, ensuring a safe recovery of any information that might be lost during documents editing.

Is there a need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?

No.

2.2.3. Metadata

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

n.a. to this version of the deliverable

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

n.a. to this version of the deliverable

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

n.a. to this version of the deliverable

2.3. Making data interoperable

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

Formal, accessible, shared, and broadly applicable formats and vocabularies such as Resource Description Framework, comma-separated values, and JavaScript Object Notation for Linked Data will be used. The principles of European initiatives like GAIA-X and International Data Spaces will be followed.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

Yes.

Will your data include qualified references[1] to other data (e.g., other data from your project, or datasets from previous research)?

[1] [1] *A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to*

Yes.

2.4. Increase data re-use

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g., readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

Via conference presentations, peer reviewed open-access publications, regular updates on the W2BC website, and publication on time of the deliverables of the project (some of which have a public level of dissemination). Each partner will also be made to save (at least) for the duration of the project the generated files from characterization equipment's, readings, etc., in the original format.

Will your data be made freely available in the public domain to permit the widest re-use possible?

Yes. In the case of sensitive data, this will be made publicly available when IP protection is enabled and when it is deemed commercially appropriate.

Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

n.a. to this version of the deliverable.

Will the data produced in the project be useable by third parties, in particular after the end of the project?

Yes.

Will the provenance of the data be thoroughly documented using the appropriate standards?

Yes.

Describe all relevant data quality assurance processes.

n.a. to this version of the deliverable.

Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

n.a. to this version of the deliverable.

3. Other research outputs

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g., software, workflows, protocols, models, etc.) or physical (e.g., new materials, antibodies, reagents, samples, etc.).

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

Research outputs expected from W2BC, such as bio-based materials, dyes and pigments, microparticles, functional capsules, foams, nonwovens, films, composites, etc., will be send from the partner generating the output to the partner requiring that output to continue that work describe in the DoA (develop those materials into other materials, test toxicity, test end-of-life alternatives, etc.). This is already agreed among the partners, and the monthly meeting of each WP, where the

enrich the contextual knowledge about the data. (Source: https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/)

coordinator will always be present will assure a proper flow of research outputs for the achievement of the project goals.

4. Allocation of resources

What will the costs be for making data or other research outputs FAIR in your project (e.g., direct and indirect costs related to storage, archiving, re-use, security, etc.)?

No costs are foreseen, but this will be analysed through the project, as the data is being generated.

How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions).

These will be covered by the coordinator.

Who will be responsible for data management in your project?

The coordinator – CITEVE.

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

Long term preservation of the generated data, including organization, selection of information, integration, maintenance, reuse, and recovery of the data collected during the project, will be done by the coordinator, through the trusted repository above mentioned – MS 365 SharePoint – which is already being used by the coordinator for all of the entity data, so no further cost are predicted for preserving the data generated in the Waste2BioComp project.

5. Data security

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

The coordinator already has in place a contract with MS, ensuring that licenses are provided for the use of MS 365 SharePoint in such a way that all the data generated in the project can be safely stored in this cloud service, and shared in different modes (read-only mode, editing, etc.). Also, SharePoint also has the advantage of being intuitive and having an automatic history track record of the data versions, so any data lost can be recovered, and eventually transferred to other services, if required.

Will the data be safely stored in trusted repositories for long term preservation and curation?

Yes, at least for 5 (five) years after the project ends.

6. Ethics

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

Some ethics issues in relation with the data collected for the training actions, and allergy test with volunteers. These have already been approached in the DoA (chapter 4 of the DoA in the GA), but will be further discussed with the appointed Ethics Advisor, before the collection of any personal data.

Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

Yes.

7. Other issues

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

n.a. to this version of the deliverable.

waste2biocomp

www.**waste2biocomp**.eu

Waste2biocomp @waste2biocomp Waste2biocomp Project

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency (HADEA) Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union