Data Management Plan - intermediate

Deliverable 7.8

Author: **Helena Vilaça**

Date: 27 May 2024

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	D7.8
Due date of delivery to EC	31 May 2024
Actual date of delivery to EC	29 May 2024
Dissemination level	Public
Work package	7

Contributors	
Main Contributor	Helena Vilaça (CITEVE)

Control Sheet			
Version	**Date**	**Editor**	**Summary of Modifications**
1.0	27/05/2024	Helena Vilaça	Initial version of deliverable 7.8 (update of v1.0 of deliverable D7.2)

Table of Contents

List of Abbreviations

Acronyms	Description
CA	Consortium agreement
D	Deliverable
DCAT	Data Catalogue Vocabulary
DMP	Data Management Plan
DoA	Description of the Action
DOI	Digital Object Identifier
EC	European Commission
EU	European union
FAIR	Findable, Accessible, Interoperable, Reusable
GA	Grant agreement
HES	Higher or Secondary Education
IP	Intellectual property
IPR	Intellectual Property Rights
ISBN	International Standard Book Number
KER	Key exploitable result
LCSA	Life cycle sustainability assessment
M	Month
MS	Microsoft
n.a.	Not applicable
NC	Nanocapsule
NDA	Non-disclosure agreement
ORCID	Open Researcher and Contributor Identifier
PHA	Polyhydroxyalkanoate
PHB	Polyhydroxybutyrate
RDA	Resource Description Access
R&I	Research and Innovation
ROR	Research Organisation Registry
RTD	Research and technology developer
URL	Uniform resource locator
VC	Value-chain
W2BC	Waste2BioComp
WP	Work package

1. Introduction

The **W2BC** project aims to demonstrate relevant scale production of bio-based products and materials, as alternatives to traditional materials with high environmental footprint, using innovative manufacturing technologies. The project integrates all stages in the bio-based products' life cycle, starting from R&I activities regarding the sourcing of feedstocks for the development of bio-based precursors and intermediate materials, smart inkjet printing techniques, and smart manufacturing technologies for final products, and the final demonstrators, which will entail the production on a relevant scale of the following bio-based products:

- shoe sole materials with different hardness;
- three-layered shoe insoles;
- plastic films/packaging with different flexibilities;
- technical textiles for sportswear;
- fashion garments printed with bio-based inks;
- leather and textile shoes printed with bio-based inks;
- paper for packaging printed with bio-based inks.

W2BC is also developing sustainability and toxicity assessments to the developed materials and products, as well as re-manufacturing and recycling approaches to ensure circularity by closing the material loop. Furthermore, the project is developing training activities to support the creation of a skilled workforce in the biomaterial-based manufacturing sectors, particularly for the textile, footwear, and packaging activities. Therefore, it is expected that **W2BC** will have a significative impact on the reduction of the use of fossil-based materials, not only in the approached three value chains (textiles, packaging, and footwear), which are highly resource and polluting intensive sectors, but also with potential for several other sectors and applications.

The project runs for 36 months, and has a consortium of 13 partners, from France, Germany, Italy, Portugal, Spain, and Switzerland.

This Data Management Plan (DMP), issued by M06 (D7.2), updated by M24 (D7.8), and with another predicted update by the end of the **W2BC** project on M36 (D7.11), contains 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 are collaborating to ensure that quality procedures, based on a "sustainable and innovative scholarly communications ecosystem" are adopted, and they are promoting 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 the research being carried out in the project. It is encouraged to make data generated during the project available not only between the consortium partners for research purposes, but also to the broader research community, trainees, stakeholders, and policy makers, while respecting the ethic-legal framework.

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.

The DMP describes the data management life cycle for all datasets to be collected, processed and/or generated by the **W2BC** project. It describes, among others:

- the handling of research data during and after the project;
- the type of data that will be collected, processed, or gathered;
- what methodology and standards will be applied;
- whether and how the data will be made (openly) accessible;
- how the data is stored.

2. Data summary

This section provides a summary of the data addressing the following issues:

- The purpose of the data collection or re-use and its relation to the objectives of the project;
- The types and formats of data generated or re-used;
- If existing data is being re-used (if any),
- The origin of the data;
- The expected size of the data (if known);
- The data utility: to whom will it be useful.

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

The main purpose of the data generated and re-used in the project is to develop new bio-based materials for three value-chains (textile, packaging, and footwear), characterize them in a relevant environment, and assess their life cycle and sustainability, as also create a skilled workforce for these materials. The creation (and re-use) of the data is necessary to achieve the project objectives, namely:

- WP1: development of new bio-based "raw" materials (PHAs, microparticles and nanocapsules, pigments, inkjet inks) and their characterization – generation of experimental protocols, characterization (of the materials developed) raw data, methodologies and results, and materials developed (physical samples of polymers, capsules, microparticles, pigments and inks), internal deliverables of the project that will not be made public for IPR reasons;
- WP2: development of new inkjet systems for both 2D and 3D substrates adapted for bio-based inks – generation of specifications sheet for the inks, designs and diagrams of the print engines and printing systems, physical printed samples and equipments developed, trajectories for the 3D model object, internal deliverables of the project that will not be made public for IPR reasons;
- WP3 & WP4: application of the "raw" materials into new materials (shoe sole materials with different hardness, three-layered shoe insoles, plastic films/packaging with different flexibilities, textiles for sportswear, different textile, plastic, leather and paper-based materials printed with bio-based inks), using innovative technologies or adapting the existing ones, and the characterization and validation of these new materials in relevant environments – generation of experimental protocols, characterization (of the materials developed) raw data, methodologies and results, and materials developed (physical samples), internal deliverables of the project (only those from WP4 will be made public for IPR reasons);
- WP5: colour removal and reprinting, and recycling of the materials developed in the project – generation of experimental protocols, characterization raw data, methodologies and results, and new materials developed (physical samples), internal deliverables of the project that will not be made public for IPR reasons;
- WP6: assessment of the toxic and sustainability profiles of the materials and processes developed in the project, as also their compliance with regulations – generation of toxicity tests raw data and reports, LCSA calculations/outputs, and regulation compliance reports, internal deliverables of the project which will be made public;
- WP7: besides the dissemination and communication activities, it is also goal of this WP to create a skilled workforce for the new materials/technologies, and develop exploitation strategies and business plans for the KERs of the project – generation of reports with dissemination, communication and exploitation plan, business plans, different training materials (ppt presentations, training plans, participation sheets, etc.), dissemination and communication materials (news published on the social media, posters, flyers, etc.), personal data of external users for providing **W2BC** community functionality (newsletters, event invitations, etc.);
- WP8: smooth and efficient project management – generation of internal consortium information (encompasses mainly personal data of **W2BC** partner personnel for operating the project's mailing lists, collaboration platform MS 365 Sharepoint, but also the internal deliverables of the project that will not be made public for IPR reasons.

Thus, the data generated in each WP will be re-used in other WPs ("feed" that WP), namely physical samples from WP1 will feed WPs2-5, and physical samples and protocols from WPs2 and 3 will feed WP4 for the final prototype's development. Both physical samples and processes information (steps taken, entries and outputs in each step, etc.) will be used in WP6 for the toxicity and sustainability (LCSA) assessments.

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.

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

The types of data expected from the project are mainly research outputs, i.e., data related with:

- 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 end-of-life alternatives of the bio-based materials;
- toxicology, impact & regulatory assessments;
- training activities.

This data will be in the form of written protocols (MS word compatible format - .docx), images (preferably in png format, but .ai, .svg, .tiff and .jpeg will also be acceptable, depending on each case), numerical spreadsheets and respective charts (MS excel compatible format - .xlsx, but other formats will also be acceptable), properties reports (MS word - .docx, MS powerpoint - .pptx, MS excel - .xlsx compatible formats and in some cases pdf documents), schemes (can be in different formats, such as MS powerpoint compatible - .pptx, png, or others), physical samples (photographs will be used to digitise, as much as possible, the samples, generating images in the formats already above-mentioned), video (preferably in .mp4 format). The main outputs from each task will be organized in dedicated reports and deliverables (.docx and pdf formats).
In some specific cases, files in other formats may be generated, for example for the characterization and properties assessment of the materials (file formats depending on the software of the equipment used), but in these cases, the main information will be extracted to the more common used formats above mentioned (.docx, .xlsx, .png), to facilitate its sharing and longer-term re-use.

2.3. Will you re-use any existing data and what will you re-use it for?

Different existing data will be re-used, namely the knowledge already published (in scientific papers or patents) by the RTDs and HES partners, but also from some of the companies, so that the knowledge associated with this data 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, synthetic protocols for biopolymers modification and encapsulation by IVW, 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, chemical protocols for recycling of polyesters by GR3N).

2.4. What is the origin of the data?

The re-used data came mainly from scientific publications (available online) and patents from the consortium partners and other researchers, but also from the in-house generated knowledge of each partner.

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

This aspect will be evaluated during the project, as it is quite a difficult parameter to estimate. The expected size depends on the extend and the nature of the data that are made available.

From the datasets re-used and generated so far, the expected size of individual datasets ranges between a few KB and 300 MB. The overall size of the generated data is expected to overcome 20 GB, stored in an internal restricted repository under the regime of data protection and privacy and will not be made openly available. From this, only a small part will be made publicly available, after assessing IPR.

2.6. 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, producers of technical textiles for sportswear, sole & insole and shoe producers, rigid plastic packaging & plastic films producers);
- Players across the textile, packaging, and footwear VCs, e.g., designers, brands working with personalized collections, inks producers looking for bio-based alternatives;
- Biopolymer producers, namely PHA producers;
- Players from other VCs, e.g., automotive and building & construction industries (PHA-based foams, 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, bio-foams, 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).

3. FAIR Data

This section addresses how the data generated will follow the FAIR principles: findable, accessible, interoperable, and re-usable.

3.1. Making data findable, including provisions for metadata

This section addresses the following points, required to make the data findable:

- Metadata provision (discoverability of data);
- Specify standards for metadata creation (if any);
- Approach towards search keywords,
- Use of persistent identifier(s) (identifiability of data);
- Naming conventions used;
- Approach for clear versioning.

3.1.1. Metadata provision

Metadata to describe the data that is collected and generated by **W2BC** is needed to facilitate open access to the data. Thus, a pragmatic and feasible approach is to agree on a common and minimal catalogue metadata schema for those datasets that are published in public catalogues and data repositories.

In general, the following minimum and recommended terms will be used for open data generated by the project and deposited in an appropriate repository (according to the publisher), according with the FAIR principles:

- Publication
  - Author(s)
  - Title
  - Date of publication
  - Publication venue
  - A short description (abstract)
- Keywords
- Horizon Europe or Euratom funding
- Grant project name, acronym and number
- Licensing terms
- Persistent identifiers for the publication (DOI), the authors involved in the action (Open Researcher and Contributor Identifier - ORCID) and, if possible, for their organisations (Research Organisation Registry - ROR) and the grant
- Persistent identifiers of related publications and datasets.

The data generated in the project, including the one that will not be made public, will be organized in a shared folder on the MS 365 SharePoint, created and managed by the coordinator, CITEVE. This folder is organized in several sub-folders, for each WP - to which only the consortium members of that WP have access, and broader folders to which all consortium members have access: consortium meetings, GA & CA information, deliverables, milestones, timelines, and consortium names and contacts. The datasets contained in those folders will have dates associated and, when is pertinent, a version control will be made (mainly for deliverables and reports from the different tasks).

The template for the project data (.docx and .ppt files) already have included a document identification and control sheet tables, which include:

- Document Identifier
- Due date of delivery to EC
- Actual date of delivery to EC
- Dissemination level
- Work package
- Main Contributor(s)
- Contributor(s)
- Version
- Date
- Editor
- Summary of Modifications

The relation between different datasets will be made clear either by referencing other datasets in a dataset, or by creating a .txt file that explains the relation between the different datasets generated, so that a user can easily know what datasets need to consult.

As for the information generated be each partner, and that is related with IPR only from that partner, it will be organized and stored by that partner, following its internal rules.

3.1.2. Standards for metadata provision

Ensuring that metadata records are formatted by a common standard facilitates the readability of the metadata by both humans and machines. Generic metadata standards are widely and easily

used, while domain-specific metadata standards are richer in vocabulary understood by researchers in a specific discipline. The Resource Description Access (RDA), Data Catalogue Vocabulary (DCAT) and Dublin Core are examples of metadata standards applicable, but not limited, to research data.

A common and minimal catalogue metadata schema will be adopted, based on DataCite's metadata schema.[1] This minimal metadata schema can be extended, if required. It is compatible with the Dublin Core metadata standard and thus can be interpreted by catalogue harvesters used by e.g., OpenAIRE.

3.1.3. Approach towards search keywords

Keywords will be provided to optimize possibilities for re-use.

3.1.4. Will data be identified by a persistent identifier?

Open **W2BC** results that are deposited in institutional repositories, the project website, CORDIS, repositories of scientific publishers or other data and research repositories will be indefinable at least by a persistent URL. If the institution is a DOI registrant that has an agreement with a DOI registration agency, a DOI will be assigned, too, thus making the generated outputs more findable and citable without compromising the IPR. If results are published in a book, then an International Standard Serial Number (ISSN) will be attributed.

Zenodo (https://zenodo.org/) is an example of an open data repository that can generate DOIs for research results.

3.1.5. Naming conventions

For deliverables, the following naming convention will be adopted:

- [number of deliverable]_[month of delivery]_[task related with the deliverable OR deliverable short name]_[version]_[status]
  - Examples: D1.1_M24_Task1.1_v1

    D7.2_M06_DMP_v1.0_submitted

All files made publicly available reference **W2BC** in their name, with the recommendation that the convention **waste2biocomp**_[name]_[date], where [name] is a meaningful short description of the file.

For metadata, dataset and template names deposited in the MS 365 SharePoint repository, all will consist at least of a short and meaningful name of the dataset/template, and stored under the respective folder (WP -> event OR photos OR test OR etc.). Each file automatically indicates the date and author of deposition in the repository.

For .ppt files prepared for the meetings, their name will indicate at least the type of meeting (consortium, WP or other), date, and partner if applicable. Examples:

WP1 monthly meeting_Apr-2023_HSKL

Presentation_Consortium Meeting+Minute_17-01-2024

W2BC_Contingency Plans presentation_9-10-2023

[1] https://schema.datacite.org/archive/kernel-2.1/doc/DataCite-MetadataKernel_v2.1.pdf

3.1.6. Clear versioning

The versioning management of the data, and in general the files stored into the Repository will be applied at three levels:

- Via the naming convention and the use of the date as suffix, indicating the last version of the file uploaded into the Repository;
- By filling a "Control sheet" table in each document prepared;
- By checking the Version History for each dataset/file, which is possible when using MS 365 SharePoint as the repository.

Additionally, all open publications deposited in the publisher repositories will use DOI versioning. DOI versioning allows for updating a dataset after it has been published and to cite either a specific version of a dataset or all versions of a dataset.[2]

3.2. Making data accessible

This section addresses the following points, required to make the data accessible:

- If the data will be deposited in a trusted repository and if identifiers will be assigned;
- Which data will be made openly available;
- How access will be provided in case there are any restrictions;
- How the data will be made available;
- What methods or software tools are needed to access the data.

3.2.1. Repository

Will the data be deposited in a trusted repository?

An 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. This repository is the one chosen by the coordinator, and has an automatic backup, therefore safeguarding the information there deposited.

The publication in open access repositories, such as project website, Zenodo, and others will be considered for open access publication unless restricted by third-party copyrights or confidentiality agreements.

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

The coordinator, CITEVE, has a contract with Microsoft (MS), and MS 365 SharePoint is the standard trusted repository for all the institution 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?

Each data deposited in the repository will have an URL as identifier, allowing the easy share of information between partners.

3.2.2. 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

[2] https://blogs.OpenAIRE.eu/?p=2010

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 polyester recycling professionals, renowned researchers & potential customers, EC members for verification purposes), for which an NDA will be signed before data sharing. For example, the partners might enter into legal/contractual restrictions imposed by a company that they enter into contact with that are interested in the developments (possible exploiters of the results of the project).

As set up in the GA, each partner will be the owner of the data generated by its own work in the project, and is up to each partner to decide what data can be shared with the consortium (the coordinator will ensure that enough data is shared to allow the follow up of the state of the work, and a smooth flow of the project), and what data can be made publicly available. For that, before each partner shares any data that is not only from its own work, it has to communicate what is intended to divulge to the whole consortium, as set up in the CA.

For most of the research carried out and resulting outputs, the possible restrictions for data sharing are intentional, as these are related with the need for IPR protection before divulging data outside the consortium.

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 and overall scientific community, data without IP issues will be made accessible through the **W2BC** website, social media, the publication of periodic newsletters, and the publication of open-access papers.

For the consortium, the access to all generated data that each partners sees feasible to share is made accessible by giving access to a shared MS 365 SharePoint folder, to which the partners have free access.

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

As above mentioned, the data will have three levels of sharing:

- Full public – shared in the free access platforms above mentioned;
- Sensitive – shared only with the consortium – this corresponds to most of the data generated in the project – shared in the MS 365 SharePoint folder of the project between consortium members;
- Confidential – not shared with the consortium – data "stays" only with the partner that generates it.

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

By using MS 365 SharePoint as a trusted repository of the data shared among the consortium, it is possible to follow the history of access to the documents, so it is simple to follow who has accessed what document, and if that person has made changes to the document.

As for the data publicly shared, there is no need to identify who has accessed the data, only the number of accesses for statistical purposes.

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

No.

3.2.3. 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?

Open access data will be published as scientific articles under an open access Creative Common license, mainly CC BY license, under which any part of the article may be reused without permission provided that the original article is clearly cited. However, the metadata of the publications will be published under a Creative Common Public Domain Dedication (CC 0) license, in line with the FAIR principles.

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

The data published on the project website will remain available for a period of at least 5 years after the conclusion of the project, while the data on the MS 365 SharePoint folder will remain available for the consortium partners for 1 month after the complete closure of the project process (final payment by the EC to the consortium), and after that it will stay accessible for the coordinator members, although with different URLs, but with the same organization/structure.

As for the data published in the repositories of scientific publishers (which will also be made available on the project website), it will remain available for as long as the publisher keeps the repository accessible.

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)?

Not applicable.

3.3. Making data interoperable

This section addresses the following points, required to make the data interoperable:

- What data and metadata vocabularies, standards or methodologies you be followed to facilitate interoperability;

- If standard vocabulary will be used for all data types present in the data set, to allow inter-disciplinary interoperability.

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 Linked Open Data will be used. The principles of European initiatives like GAIA-X and International Data Spaces will be followed.

Proprietary formats (created by scientific instruments and/or specific software applications) will be converted into open and standard formats, such as xlsx or csv formats, to allow their interoperability and re-usability.

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?

W2BC does not intend to introduce new project specific ontologies or vocabularies.

Will your data include qualified references[3] to other data (e.g., other data from your project, or datasets from previous research)?

Yes.

3.4. Increase data re-use

This section addresses the following points, required to make the data re-usable:

- How documentation needed to validate data analysis and facilitate data re-use will be provided;
- If data will be made freely available in the public;
- How the data will be licenced to permit the widest reuse possible;
- If the data produced in the project be useable by third parties;
- If the provenance of the data will be thoroughly documented using the appropriate standards;
- The relevant data quality assurance processes.

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.)?

Any data that is sharable among the consortium will be deposited in a trustworthy repository – a folder in MS 365 SharePoint that all the consortium has access, but which has an organization by WPs and the access to each sub-folder is dependable on the participation of the partner in that WP. That way, any partner can validate the data generated by another partner. Data will be deposited in formats that are understandable and allow analysis from all participating partners.

As for the generated files from characterization equipment's, readings, etc., in the original format, these will be saved by each partner that generates them (at least) for the duration of the project. The treated data will be shared among the consortium by regular presentations (.ppt files) and reports (.docx files). Is case of any doubt, raw data from a partner may be shared upon request by another partner by email.

The data that is feasible to publish will then be submitted for peer reviewed open-access publications, which also ensure a further analysis of the data, its quality, and facilitates data re-use.

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

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

Yes, some of it, through publication (in CORDIS and the project website) of public project deliverables, open access papers, communications in conferences and fairs, project public events, training activities, etc. In the case of sensitive data, this will be made publicly available only 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?

Yes. Authors of scientific publications arising from the project are encouraged to seek an agreement with the scientific publisher of the publication that allows the authors to:

- retain the ownership of the copyright for their work;
- deposit the publication in an Open Access repository.

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

Yes, the data related with results that show greater exploitation potential.

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

Yes.

Describe all relevant data quality assurance processes.

In research is a common practice to make several replicas for each trial/protocol, to demonstrate its replicability, to use equipment's regularly calibrated, and to establish replicable protocols. The partners involved in these processes have these practices implemented, therefore data consistency and quality is ensured. Also, the characterization of the materials will be done using standards, when these exist, or using protocols adapted from standards with a focus on developing new standard protocols.

4. Other research outputs

Other research outputs expected from **W2BC**, namely the bio-based materials, dyes and pigments, microparticles, functional capsules, foams, fibres & nonwovens, films, composites, etc., will be managed mainly by the partner generating that output, which is responsible for ensuring that the physical samples are stored for at least the timeline of the project, but preferably also after its conclusion. As for detailed protocols and models, these will also be stored by the partner generating them. Any pertinent data generated in handwritten format, will be or digitised, or passed to digital format (reports, ppt…) to safeguard the information and its accessibility and re-usability.

During the project there will be a need for materials exchange, which is managed by an excel file shared among the partners showing which samples are needed by which partner for which task, amounts and expected dates of delivery. The partner producing the requested sample/material will send it to the requesting partner, therefore allowing the development of the project as described in the DoA (develop those materials into other materials, test toxicity, test end-of-life alternatives, etc.). This is already agreed among the partners.

5. Allocation of resources

This section addresses the allocation of resources, addressing the following issues:

- Estimated costs for making the data FAIR, and how these costs will be covered;
- Responsibilities for data management in the project;
- Costs and potential value of long-term preservation.

5.1. Estimation of costs

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.)?

For the trusted repository in MS 365 SharePoint no direct costs are foreseen, as this is the repository used by the coordinator, regardless of the project.

As for the deposit of data in the project website (data made public) the costs predicted are those of maintaining the website.

Also, there are estimated costs for the publication of open access peer-review publications, which will have a medium cost of 2500€ each.

Data preparation for sharing and deposit requires time from the consortium partners generating the data, and the ones that gather the data (mainly the coordinator), but this time was also estimated in the PM required by each partner, as they new beforehand that any data generated needed to be prepared for presentations, reports and eventually for publication.

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).

The costs related with the website will be covered by the partners MAGELLAN ASSOCIATION (M01-M18) and MAGELLAN CIRCLE (M19-M36) and were already estimated in the budget of the partner for the project and are covered by the financial budget of the project.

The costs for open-access publications will be covered by the partner submitting the publication, and these were also estimated in the budget of each partner intending to publish, and will be covered by the financial budget of the project.

5.2. Responsibilities for data management

Who will be responsible for data management in your project?

The main responsible for data management in **W2BC** is 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 **W2BC** project.

6. Data security

This section addresses the data security, addressing the following issues:

- Data confidentiality and integrity;
- Data availability.

6.1. Data confidentiality and integrity

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.), and only emails that the folder has been shared with have access to the information therein contained (not possible to forward access invitations to other persons). Also, SharePoint 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. As above explained, the shared folder among the consortium is organized by WP, and access to each WP sub-folder is limited to the partners involved in that WP.

6.2. Data availability

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

All the technical and financial data of the project, either deposited in the MS 365 SharePoint folder, or stored by each partner, will be kept for at least for 5 (five) years after the project ends.

7. 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).

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

With regards to personal data, the Consortium will ensure that data on individuals are transmitted and used in a secure environment, that the use of the data complies with ethical and legal requirements (including signed informed consent and applicable data protection laws and EU regulation), and that the use of both existing and new data is agreed with the data owner/data provider. Data records containing personal data are stored and processed in compliance with the General Data Protection Regulation of the EU (2016/679).

W2BC has a task dedicated to address Responsible R&I Ethics, and Gender Issues, to ensure that the project complies with the ethical requirements such as data protection and privacy. This implies that data management will follow the highest ethical standards, and apply international, EU and national law (in particular, the General Data Protection Regulation of the EU (2016/679), national data protection laws and other relevant legislation). Data will be made anonymised before any public share, and even among the consortium members only data about themselves (names, emails and institutions) will be shared.

The data collected during the allergy tests will assure anonymity, thus not raising a Personal Data ethical issue.

As for the workshops and training activities, collection of certain personal data of the participants (e.g., name, e-mail address, age, occupation, schooling), and its processing (e.g., creation and management of a database of the participants, creating a mailing list of participants) will be required, in order to identify the persons interested in the training activities, if their professional profile is suitable for that specific activity, allow to contact those persons to participate in the activity, and after the activity, to process their participation certificates. The overall data collected will be set to the minimum necessary for the training activities, i.e., follow the "data minimisation principle", and the participants will be clearly informed of what, how and what for the data is being collected, so that they can give informed consent for data processing. In case the participants agree, their e-mail addresses may be added to pre-existing mailing lists (with the option of being removed from this list anytime, without any consequence) of persons interested in training activities for a certain sector, in order to be informed of any subsequent training activities to be carried out during the project.

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

Yes.

8. 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)?

No.

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