



Funded by  
the European Union



UK Research  
and Innovation



high-value products from agricultural  
residues through sustainable chains

## Deliverable D6.4

### Data management plan

This report only reflects the views of the author(s).

The Commission is not liable for any use that may be made of the information contained therein.

Project funded under the Horizon Europe Research and Innovation Programme.

#### Dissemination Level

<b>P</b>	<i>Public</i>	<input checked="" type="checkbox"/>
<b>CO</b>	<i>Confidential, only for members of the consortium (including the Commission Services)</i>	<input type="checkbox"/>
<b>CI</b>	<i>Classified, as referred to Commission Decision 2001/844/EC</i>	<input type="checkbox"/>
<b>SEN</b>	<i>Sensitive</i>	<input type="checkbox"/>

#### Type

<b>R</b>	Document, Report	<input checked="" type="checkbox"/>
<b>DEM</b>	Demonstrator, Pilot, Prototype	<input type="checkbox"/>
<b>DEC</b>	Websites, Patent Fillings, Videos, etc	<input type="checkbox"/>
<b>Other</b>	(Please describe the type)	<input type="checkbox"/>

## Summary

**Project Number:** 10108176

**Project:** AgriLoop : pushing the frontier of circular agriculture by converting residues into novel economic, social and environmental opportunities

**Topic:** HORIZON-CL6-2022-CIRCBIO-01-05

**Duration:** 48 Months

**Start date of Project:** 1st December 2022

**End date of the Project:** 30 November 2026

**Coordinator:** INRAE

**Deliverable:** D6.4

**Due date of deliverable:** 31/05/2023

**Actual submission date:** 31/05/2023

**Work package number:** WP6

**WP Leader (name and organisation):** Nathalie Gontard INRAE

**Person in charge of the deliverable (name and organisation):** Patrice Buche INRAE

**Author(s):** Patrice Buche

**Contributor(s):** all partners

**Funding acknowledgement**

This project has received funding from the European Union's Horizon Europe research and innovation programme and the UK Research and Innovation fund under the UK government's Horizon Europe funding guarantee, grant agreement No. 101081776. Views and opinions expressed are however 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.

**Version:** 1.0

## Table of content

1. Executive Summary .....	5
2. Introduction.....	6
3. Data collected or produced by WP.....	8
3.1. WP 1 Foundational circular & strategic flows .....	8
ECOZEPT .....	8
INRAE.....	8
SDU .....	10
DSS .....	11
UNIROMA .....	11
USC .....	12
WR.....	14
3.2. WP 2 Upstream Proteins & Chemicals recovery .....	14
FHNW .....	14
ITQB .....	15
INRAE.....	16
UNIBO .....	17
TOMA .....	18
3.3. WP3 Microbial fermentations for proteins and PHA production .....	19
CSIC .....	19
UGENT .....	20
NID .....	21
ENTO.....	22
3.4. WP4 End-products and value chains .....	22
BIO-MI .....	22
UNIVR .....	23
AVECOM.....	24
UM .....	24
3.5. WP5 Communication, dissemination and exploitation .....	26
BDC .....	26
FCAC.....	27
4. Personal data collection and/or processing by WP .....	28
4.1. WP 1 Foundational circular & strategic flows .....	28
SDU .....	28
USC .....	28
4.2. WP5 Communication, dissemination and exploitation .....	30
FCAC.....	30
5. Conclusion.....	31

Annex I .....	32
List of partners .....	32

## 1. Executive Summary

<b>Background</b>	<p>This document D6.4 is the Data Management Plan (DMP) deliverable of the AGRILLOOP project, which is funded by the European Union's Programme under Grant Agreement #101081776.</p>
<b>Objectives</b>	<p>The main objective of this deliverable is- to create the first version of the AgriLoop data management plan (WP6). Open Access (OA) to research data via trusted repositories under the principle "as open as possible, as closed as necessary" following the FAIR (Findable, Accessible, Interoperable and Reusable) principles that must be ensured during and after the AgriLoop project. To reach this aim, the implementation steps of the Data Management Plan (DMP) is based on one hand on three main deliveries, DMP initialization version (M6), intermediate version (M24) and final version (M48) and on the other hand on a continuous effort of data "FAIRification" all along the project synchronized with AgriLoop project deliverable production.</p> <p>The purpose of the Data Management Plan (DMP) is to provide a global view of data collected/produced in the project. Detailed information is provided on the procedures that have been implemented for data collection, storage, protection, retention and destruction and confirmation that they comply with national and EU legislation.</p>
<b>Methods</b>	<p>A web survey has been designed by INRAE. It has been fulfilled by AGRILLOOP partners to collect the required information at the beginning of the project. A synthesis of information is provided in this document. This synthesis will be updated twice by the partners at the middle and at the end of the project.</p>
<b>Results &amp; implications</b>	<p>The DMP provides a global picture of types of data which have been collected/produced in all the WPs. The list of partners who are concerned by ethic aspects (mainly about personal data management) has also been achieved. Those partners have provided information about their personal data management policy. INRAE has implemented a route implementing FAIR principles to manage data collected/produced during the project based on INRAE dataverse.</p>

## 2. Introduction

AgriLoop will manage research data and other research outputs responsibly, following the FAIR (Findable, Accessible, Interoperable and Reusable) principles, using a Data Management Plan and providing OA to research data via trusted repositories under the principle “as open as possible, as closed as necessary”. To ensure AgriLoop results’ reproducibility, scientific process and data will be documented at the early stage of research in a transparent way: successive raw data saved; lab notebooks to track reasoning, experimental, final protocol, deviations, linked to data files; operating protocols in tracked versions; protocols shared in supplementary files; statistics described and pilot processes quality-controlled.

To ensure a smooth management of AgriLoop’s data, T6.4 in WP6 will be dedicated to specify how AgriLoop collected, processed or generated data will be handled by the consortium by following FAIR principles during the project in a Data Management Plan (D6.4, D6.5 and D6.6) and how to ensure their preservation beyond the project lifetime. The DMP will be generated at the start of AgriLoop and will evolve over the lifespan of the project. AgriLoop will include the following data types: Stakeholders’ preferences about potential residues conversion pathways (WP1) - Data collection on volumes and compositions of primary and secondary residues - data on physical-chemical and functional properties of extracted/fermented compounds (from WP2 and WP3) and of end-products (WP4) from the agri-residues with identification of potential end-uses, and data on identified hazards (contaminants) in original residues and end-products - Estimated sustainability indicators of the circular solutions (climate, ILUC, effect on soil health - Experimental methods and processes set up for the residue valorisation pathways - Data on end-product purity and quality - Techno-economic data on the upscaling of the AgriLoop.

WP6 will concentrate data collection that is needed for the rest of the project. Data generated by WP1, WP2, WP3, WP4 and WP5 will be examined by the involved partners to decide which data should be made available on the data management platform following the principle ‘as early as possible, as open as possible’. Specifically, they will decide of an action plan during deliverable production: which metadata and license associated with the data, what are the most appropriate open repositories / channels for exploiting those data during and after the end of the project in addition to INRAE dataverse. OpenAIRE+ and Zenodo platform. Then, the compiled and curated data will be uploaded by partners for preservation guaranteed at least for 10 years in a dataverse on the INRAE Data Portal (INRAE dataverse), and metadata replication on OpenAIRE+. This will allow and optimise the exploitation of the collected/generated data.

In Task 6.4, INRAE proposes a FAIR data management implementation to help AgriLoop partners to make their data durable. The solution is based on the notion of Digital Object Identifier - DOI, which is a persistent identifier used to uniquely identify objects, standardised by the International Organisation for Standardisation (ISO). DOIs are in wide use mainly to identify academic, professional, and government information, such as journal articles, research reports and data sets. AgriLoop partners will have the possibility to upload in the INRAE dataverse their data files stored in a dataset, then automatically associated with a DOI. The stored datasets are automatically harvested to replicate their metadata to OpenAIR+. FAIR principles implementation in INRAE dataverse are the following:

- Findable: Digital Object Identifier (DOI) including a unique identifier and metadata is associated with a given dataset. DOI creation allows metadata indexation by search engines (e.g. DataCite, INRAE dataverse);
- Accessible: INRAE guarantees at least 10 years of persistent storage for datasets;
- Interoperable: data files should be registered in open formats like CSV or common formats (e.g. Excel, Open Office); all datasets will be described using standard descriptive metadata, such as Dublin Core and DataCite

- Metadata Schema to ensure metadata interoperability for indexing and discoverability;
- Reusable: License terms for data sharing and reuse are associated with datasets (by example CC-BY 4.0).

INRAE sent a web survey to AgriLoop partners to build the first version of the DMP, which INRAE will deliver in M6 (D6.4). Before each period of deliverable production, INRAE will recall the procedure to implement the data management plan to partners during monthly online meetings. INRAE will create a dataverse dedicated to the Agriloop project on the INRAE dataverse platform. Partners in charge of deliverable writing will be invited to add a specific section at the end of their deliverable in order to explain how the DMP has been implemented (mainly the DOIs of datasets created on INRAE dataverse or other institutional dataverses). INRAE will send a short INRAE dataverse user guide to partners. It will explain the procedure to create datasets, add metadata to them, upload data files in the datasets.

The first data management plan delivered in M6 will be revised by the partners twice during the project at month M24 (D6.5) for an intermediate version and M48 (D6.6) for the final version. INRAE will use the collaborative platform to realize those revisions with partners.

The types of research data that collected or generated along the project lie in the WPs listed in the following table:

WP Number	WP Title
WP1	Foundational circular & strategic flows
WP2	Upstream Proteins & Chemicals recovery
WP3	Microbial fermentations for proteins and PHA production
WP4	End-products and value chains
WP5	Communication, dissemination and exploitation
WP6	Coordination and project management

In section 3, data which are intended to be collected/produced in each of these WPs by partner, WP and task are described. Section 4 provides information on the partners who have been involved in personal data management. Section 5 concludes the report and provides perspectives.

### 3. Data collected or produced by WP

Nota bene: To avoid redundancy, responses provided by partners have been grouped in the first WP in which they participate; except in case of several respondents with the same partner in different WPs.

#### 3.1.WP 1 Foundational circular & strategic flows

##### ECOZEPT

Contact: Angela Funk (funk@ecozept.de)

*Task and deliverable associated the collected/produced data*

Task 1.1 and 4.4

*Type of data*

We will conduct interviews with different types of stakeholders, so data of experts' assessment will be collected.

*Metadata description*

Data will be collected by interviews conducted via phone. The answers of participants to our questionnaire will be documented in written form. The data will be specifically collected for this project and is not being re-used.

*Legal issues management*

No claims will be made for intellectual property rights, no inventions will be made by conducting expert interviews.

*Data storage during the project*

The storage server is located in a secure server room with key and access control. Data are hosted on a secured server, operating on Windows storage server 2012 system. Access is secured and accessible through password-based user authentication.

*Data sharing and long term preservation*

Infrastructure INRAE dataverse.

Suggested key words: interviews, Delphi, market study.

*Resources dedicated to data management*

ND

*Person in charge of data management*

ND

##### INRAE

Contacts: Pierre Bisquert ([pierre.bisquert@inrae.fr](mailto:pierre.bisquert@inrae.fr)) for Task 1.1. and Task 1.4.  
Thierry Gauthier ([thierry.gauthier@inrae.fr](mailto:thierry.gauthier@inrae.fr)) for Task1.3.



#### *Task and deliverable associated the collected/produced data*

Task 1.1. Multi-stakeholders knowledge exchange, market pre-assessment and decision-making guidance.

Task1.3 : Chemical & biological contaminants fate, circular safety.

Task 1.4. Environmental & circularity performance early guidance.

#### *Type of data*

Task1.1:

We will produce data in the context of T1.1. This data will be a collection of preferences (in particular about valorization pathways) and their associated justifications. They will be gathered together in a file format such as CSV or JSON that will allow to easily accessing them. The volume will be highly dependent on the number of persons interviewed, but we expect a reasonably small number of data.

Task1.3:

- Bibliographic studies in order to identify potential toxic contaminants possibly present in the food residues used in WP2 and WP3.
- Cytotoxicity and genotoxicity data from in vitro experiments.

#### *Metadata description*

Task1.1:

New data will be collected according the following general protocol: some stakeholders will be interviewed through the course of the project; their answers (preferences and justifications) will be collected in text format and translated in some formal language (through some ontology/logical language). Both the original answers, their formalization, and meta-data describing the produced data, will be stored (after thorough anonymization) in a data-warehouse allowing the easily access them.

Task1.3:

Data collected from scientific papers registered in the PubMed collection. Key words will be used to focus on natural contaminants that could end up in food residues.

#### *Legal issues management*

Task1.1:

Given the nature of the data we will handle, intellectual property rights and ownership should not be a concern. We will however ensure that no cross-identification of interviewees might happen through appropriate anonymization.

Task1.3:

New data will be presented to the consortium before being published. INRAE Transfer will be responsible for managing the intellectual property of the results issued from INRAE laboratories.

#### *Data storage during the project*

Task1.1:

Data sets under production will be locally stored as to ensure anonymity.  
Finalized data sets will be stored on the INRAE Dataverse.

Task1.3:

The data will be saved on the INRAE server.

#### *Data sharing and long term preservation*

Infrastructure INRAE dataverse.

Task1.1:

Suggested key words: Agriloop; preferences; justifications; arguments; valorization pathways.

Task1.3:

Suggested key words: Toxic compounds, microbial hazards, mycotoxins, cytotoxicity, genotoxicity.

#### *Resources dedicated to data management*

ND

SDU

Contact: Benyamin Khoshnevisan ([bekh@igt.sdu.dk](mailto:bekh@igt.sdu.dk))

#### *Task and deliverable associated the collected/produced data*

Task 1.4. Environmental & circularity performance early guidance.

Task 4.3. Environmental, safety, economic and social assessment.

#### *Type of data*

Data collected from other activities and work packages:

- 1 - Feedstock characteristics (physicochemical characteristics)
- 2 - Conversion pathways and conversation rate
- 3 - Operating conditions including temperature, pressure, retention time, flow rate
- 4 - Yield
- 5 - Product characteristics
- 6 - Characteristics of residuals and waste streams
- 7 - Materials and chemical use
- 8 - Mass flow and fractionation
- 9 - Potential recovery rate of products

Please consider that this is a non-exhaustive list

Data generated by SDU:

- 1 - System modeling including mass and energy flows
- 2 - Environmental profile of various scenarios
- 3 - Social indicators of scenarios

#### *Metadata description*

To collect data from other work packages and other activities, questionnaires will be provided in word or excel format. A detailed list of life cycle inventory data will be provided and shared with other partners. If needed, online interviews will be done as complementary to the life cycle inventory data. The results generated will be documented in word and excel format for further use and referencing.

#### *Legal issues management*

This will be done with respect to the internal regulations and agreements with AgriLoop partners.

#### *Data storage during the project*

All data will be stored in SDU OneDrive. Moreover, periodic report along with the supplementary data will be shared and stored in AgriLoop shared folder

#### *Data sharing and long term preservation*

INRAE dataverse.

Key words: Life cycle inventory; system models; material flow analysis; life cycle impact assessment.

*Resources dedicated to data management*

ND

*Person in charge of data management*

ND

DSS

Contact: ANNE VERNIQUET (anne.verniquet@consultdss.com)

*Task and deliverable associated the collected/produced data*

1.1, 1.4, 4.4, 5.1, 5.2, 5.3, 5.4

*Type of data*

WP1 : quantities of feedstock, typology of stakeholders to valorise these feedstock, stakeholders' preferences (lead Ecozept).

WP4 : valorisation of the data gathered via the various WPs and Tasks, key economic data.

*Metadata description*

New data will be collected in WP1 via interviews or online research. Data produced in other WPs are going to be valorised in WP4.

*Legal issues management*

No IP rights foreseen in our tasks, valorisation of existing data from different perspectives (Strategic thinking).

*Data storage during the project*

INRAE dataset I suppose. So far in our Drive.

*Data sharing and long term preservation*

Infrastructure INRAE dataverse.

Suggested key words: Feedstock ; stakeholders' typology.

*Resources dedicated to data management*

ND

*Person in charge of data management*

ND

UNIROMA

Contact: Marianna Villano (marianna.villano@uniroma1.it)

*Task and deliverable associated the collected/produced data*

WP1: Task 1.2 (Integrate feedstock & valorization pathways potential: early assessment); Task 1.3: Chemical & biological contaminants fate, circular safety).

WP3: Task 3.1 (Upstreams steps for precursor production); Task 3.3 (Microbial PHA production); Task 3.4 (PHA extraction).

WP4: Task 4.1 (End-users driven products); Task 4.3 (Environmental, safety, economic and social assessment).

WP5: All tasks.

WP6: Task 6.5 (Responsible Research and Innovation).

*Type of data*

Data will be mainly collected in WP1 (task 1.3 with reference to Chemical and Biological contaminants fate) and in WP3 (Task 3.3 for PHA production at lab scale, approximately 150 g PHA; and Task 3.4 for PHA extraction with different approaches).

*Metadata description*

Procedures to collect data are still in progress. New protocols are being established for PHA production and extraction.

*Legal issues management*

ND

*Data storage during the project*

Probably by using google drive.

*Data sharing and long term preservation*

Infrastructure INRAE dataverse.

Suggested key words: PHA (polyhydroxyalkanoates), Waste feedstock, Contaminants fate.

*Resources dedicated to data management*

ND

Person in charge of data management.

ND

USC

Contact: Lucas Van der Hauwaert ([lucas.vanderhauwaert@usc.es](mailto:lucas.vanderhauwaert@usc.es))

*Task and deliverable associated the collected/produced data*

Task 1.2: Integrate feedstock and valorisation pathways potential: early assessment.

Task 2.3: Decision criteria on extraction techniques.

Task 3.5: Library of models of microbial fermentations and PHA production,.

Task 4.4: Business concept of valuable sustainable value chains.

*Type of data*

Task 1.2, Task 2.3 and Task 3.5:

The models and the superstructure designs will be coded in python: The output will be scripts with a .py extension. The entire project will also be available on a Github repository, which will be accessible to the general public. The results of the flow sheet diagrams and created models will be reported in a word document.

Task 4.4:

A report as a word document will be made to associate projections and risks considered by the robust optimisation analysis.

*Metadata description*

We will use Python scripts to create models and flow charts using superstructure optimization. To create these models and solve superstructure optimization problems, data from other work packages (1, 2, and 3) will be required.

To obtain data from other partners, we will conduct online face-to-face meetings to discuss the data and its intended use. We will seek their consent to use their data. Partners will also receive an Excel template in which they can fill in their data.

#### *Legal issues management*

The code will eventually be made public through repositories like GitHub under the MIT license. Respecting the clauses of the Grant Agreement and the Consortium Agreement. Internally (and always respecting these two agreements) taking into account national laws and the corresponding internal regulations at USC.

#### *Data storage during the project*

The code is backed up on a repository using version control.  
Using repositories like GitHub and cloud services (one drive) provided by the USC.

#### *Data sharing and long term preservation*

Infrastructure INRAE dataverse.

Code will also live in the repository on GitHub.  
The word reports will certainly be placed in the INRAE dataverse.

Suggested key words: Superstructure optimization, Biorefinery design, surrogate modeling, Life Cycle assessment.

#### *Resources dedicated to data management*

ND

#### *Person in charge of data management*

ND

WR

Contact: Jan Broeze (jan.broeze@wur.nl)

*Task and deliverable associated the collected/produced data*

Task 1.2. Integrate feedstock & valorisation pathways potential: early assessment

Task 4.4. Business concept of valuable sustainable value chains

*Type of data*

T1.2:

- Estimates of volumes of agricultural residues (per type) per country.
- Compositional information of agricultural residues.

T4.4: Cost-benefit estimates of circular value chains.

*Metadata description*

Data collection (T1.2): combining data from public data sources (like FAOSTAT, other EU projects, scientific literature, etc.)

Data generation (T4.4): through calculations with our own tools, based on parameters derived from AgriLoop datasets.

*Legal issues management*

Data sources will be referenced, including specific referencing of individual data from scientific publications.

*Data storage during the project*

They will be stored in protected environment (which is continuously backedup) (OneDrive for business/teams).

*Data sharing and long term preservation*

Infrastructure INRAE dataverse.

Suggested key words: Agricultural residues, Residues composition

*Resources dedicated to data management*

ND

*Person in charge of data management*

ND

## 3.2.WP 2 Upstream Proteins & Chemicals recovery

FHNW

Contact: Philippe Corvini (philippe.corvini@fhnw.ch)

*Task and deliverable associated the collected/produced data*

Task 2.1 Raw material pre-treatment & extraction.

Task 2.2 Characterisation of the structural and biological properties of the extracts.

Task 4.2 Pilot-scale production, biorefineries schemes and end-usages in close to real conditions.

Task 5.1. Communication, dissemination and exploitation (C&D&E) strategy.

#### Type of data

Enzymatic treatment protocols; text; few Mb  
Mass spectra

#### Metadata description

ND

#### Legal issues management

ND

#### Data storage during the project

ND

#### Data sharing and long term preservation

ND

Suggested key words: ND

#### Long term preservation implementation out of INRAE dataverse

ND

#### Resources dedicated to data management

ND

#### Person in charge of data management

ND

ITQB

Contact: Cristina Silva Pereira (spereira@itqb.unl.pt)

#### Task and deliverable associated the collected/produced data

- T2.1. Raw material pre-treatment & extraction.
- T2.2. Characterization of the structural and biological properties of the extracts.
- T2.3. Decision criteria on extraction techniques/products.
- T4.1. End-users driven products: frugal design of bio-based alternatives to plastics materials.  
(also WP5 for dissemination and WP6 as member of the ExCom, co-leader of WP2).

#### Type of data

T2.1. Raw material pre-treatment & Extraction & T2.2. Characterization of the structural and biological properties of the extracts.

Chemical da NMR, GC-MS and elemental analysis data (CSV files); extraction yields for each raw material (excel files); microscopy data (SEM, TEM) (TIFF files); antimicrobial data (CSV or Excel files and TIFFs files); specific (target) structural identification of chemicals/polymeric structures (MS-based CSV data); thermal properties of plant polyesters (CSV data); additional biological testing (CSV and TIFF files).

T2.3. Decision criteria on extraction techniques/products.

Protocols for integration and cascading of the different extraction (doc/ppt files).  
Quantitative structural properties correlations (Excel or similar).

T4.1. End-users driven products: frugal design of bio-based alternatives to plastics materials.

Same data types as in T2.1 and 2.2.

(also WP5 for dissemination and WP6 as member of the ExCom, co-leader of WP2).

#### *Metadata description*

Data (raw data) is deposited (and preserved) in the team server; agriloop team members have full access; data analysis generates other types of files that are also stored in our server; when thesis/papers are ready they are deposited in the Run depository of the university.

#### *Legal issues management*

ITQB NOVA has an innovation office (the coordinator will join the Agriloop IP team) that provides support to all ITQB NOVA researchers and have been involved in the GA preparation.

#### *Data storage during the project*

The team server and in the INRAE dataverse; for open access datafiles (eg papers) also in the RUN depository of NOVA University.

#### *Data sharing and long term preservation*

Infrastructure INRAE dataverse.

Suggested key words: Plant-polyesters, Cutin, Suberin, Antimicrobial, Raw materials.

#### *Resources dedicated to data management*

ND

#### *Person in charge of data management*

ND

INRAE

Contact: Chahinez Aouf (chahinaz.aouf@inrae.fr)

#### *Task and deliverable associated the collected/produced data*

T2.1: raw material pre-treatment and extraction.

T2.2: characterization of the structural and biological properties of the extracts.

T2.3: Decision criteria on extraction techniques/products.

T4.1: frugal design of biobased alternatives to plastic materials.

#### *Type of data*

- scientific reports
- milestone and deliverable reports
- scientific articles
- excel sheets

#### *Metadata description*

Data will be produced from experimental results. They will be collected in laboratory books and scientific reports. They will be organized and stored in internal folders. Part of them will be valorized in scientific articles.



*Legal issues management*

In the case of obtained results that require IP, the IP service on INRAE will be contacted, in order to manage the rights and ownership.

*Data storage during the project*

Data will be stored in internal folders and required data shared in the sharepoint. Use of INRAE dataverse.

*Data sharing and long term preservation*

Infrastructure INRAE dataverse.

Suggested key words: agri-residues valorisation, cascading bioconversion, added-value bioproducts, biorefinery.

*Resources dedicated to data management*

ND

*Person in charge of data management*

ND

UNIBO

Contact: Annalisa Tassoni (annalisa.tassoni2@unibo.it)

*Task and deliverable associated the collected/produced data*

Task 2.1. Raw material pre-treatment & extraction.

Task 2.2. Characterisation of the structural and biological properties of the extracts.

*Type of data*

Task 2.1. Raw material pre-treatment & extraction.

Data on amounts of extracted proteins/peptides (g/L or g/gDW), peptide amino-acid sequences - Excel Tables.

Task 2.2. Characterisation of the structural and biological properties of the extracts.

Data on biological activities of extracts obtained in Task 2.1 - Excel tables.

*Metadata description*

Task 2.1 - Data will be collected after enzymatic digestion of biomass and after quantifications of yields of proteins/peptides by spectrophotometrical analysis, and after LC-MS sequencing. All protocols are already published and will be available together with dataset. Data will be organised in excel Tables.

Task 2.2 - Data will be collected after in vitro, enzymatic or colorimetric assays by spectrophotometrical analysis. All protocols are already published and will be available together with dataset. Data will be organised in excel Tables.

*Legal issues management*

UNIBO will consult the Knowledge Transfer Office for specific issues related to IPR ownership.

*Data storage during the project*

Personal computers of researchers and replication on the AgriLoop collaborative platform.

*Data sharing and long term preservation*

Infrastructure INRAE dataverse.

Suggested key words: Proteins. peptides, biological activity.

*Resources dedicated to data management*

ND

*Person in charge of data management*

ND

TOMA

Contact: Tommaso Barbieri ([tommaso.barbieri@tomapaint.com](mailto:tommaso.barbieri@tomapaint.com))

*Task and deliverable associated the collected/produced data*

WP2: task 2.1 Raw material pre-treatment & extraction.

WP4: task 4.2 Pilot scale production, biorefinery schemes and end-usage in close to real conditions.

*Type of data*

We supply the partners with tomato pomace, tomato peels and tomato seeds (from 1kg to 40kg each).

We produce data on humidity %, microbiological charge, lactic acid bacteria, dry matter, proteins, total fat, total fibre, total sugars, ash, lycopene.

*Metadata description*

We measure the humidity % in double for each sample we produce using the thermogravimetric method.

Regarding the others parameters our external laboratory apply standard method ISO.

*Legal issues management*

Data are owned by Tomapaint that share them with the partners.

*Data storage during the project*

We have an internal cloud in which we save all the data.

*Data sharing and long term preservation*

Infrastructure INRAE dataverse.

Suggested key words: tomato, peels, pomace, seeds, humidity, microbial.

*Resources dedicated to data management*

ND

*Person in charge of data management*

ND

### 3.3.WP3 Microbial fermentations for proteins and PHA production

CSIC

Contact: Natalia Hernández (nhernandezh@cib.csic.es)

*Task and deliverable associated the collected/produced data*

#### Task 3.4 PHA extraction

##### *Type of data*

Microbial physiology from mixed cultivations to accumulate biomass containing PHA. Protocols, experimental measurements, tabular data, graphical data. Experimental measurements: plain text (.txt), comma-separated values (.csv), MS Excel (.xls/.xlsx), MS Word (.doc/.docx), tab-delimited file (.tab). Tabular data: plain text (.txt), comma-separated values (.csv), MS Excel (.xls/.xlsx), tab-delimited file (.tab). Graphical data: MS Excel (.xls/.xlsx), images (JPG, TIFF, PNG). Synthetic biology such as genetic constructs and novel strains to improve the proposed lytic system. DNA sequences, graphical data, genetic models (scripts and tabular data), and experimental measurements. DNA sequences: plain text (.txt), MS Word (.doc/.docx), web links. Graphical data: MS Excel (.xls/.xlsx), images (JPG, TIFF, PNG). Genetic models (scripts and tabular data): plain text (.txt), comma-separated values (.csv), MS Excel (.xls/.xlsx), XML marked-up text (.xml), tab-delimited file (.tab). Experimental measurements: plain text (.txt), comma-separated values (.csv), MS Excel (.xls/.xlsx), MS Word (.doc/.docx), tab-delimited file (.tab).

##### *Metadata description*

Usage of the group's own bacterial strain and plasmid collections for establishing initial experiments. This also includes genome sequences, models and other OMICS data from existing public databases (e.g., GenBank from NCBI, BioModels Database).

Files will be stored locally in a folder "AgriLoop\_CSIC" organized in subfolders according to the type of data. File names should contain the date in YYYYMMDD format, followed by "AgriLoop", followed by the short partner name "CSIC", followed by a short description of the document/experiment/strain/construct, followed by the version number (if applicable). E.g. 20230328\_AgriLoop\_CSIC\_DMP.D6.4.docx.

These formats are provisional and depend on the preferences of the project coordinator.

##### *Legal issues management*

IPR issues will be managed with the collaboration of the institute's (CIB's) own Transfer Unit (<https://www.cib.csic.es/ip-portfolio>) and the CSIC's Deputy Vice-Presidency for Knowledge Transfer (VATC; <https://www.csic.es/en/innovation-and-transfer>). The colleagues in these departments have profound experience in these fields.

##### *Data storage during the project*

Periodic back-ups of locally stored data will be made to an external hard drive. Also, all data will be saved on SACO (SACO is a CSIC's shared storage tool for large data volumes; "Servicio de Almacenamiento Compartido de la Organización") and in parallel on the INRAE collaborative platform.

##### *Data sharing and long term preservation*

Infrastructure: We may intend to use the INRAE dataverse system for long term preservation.

##### *Responsible of data management*

Postdoctoral researcher at the Polymer Biotechnology laboratory (CIB-CSIC).

#### *Resources dedicated to FAIR data management*

We expect to apply 0.5 PM for the data management activities during the whole project running time. The fundamental FAIR approaches will be implemented which we gained in the last years within other European projects.

#### UGENT

Contact: Estevez Alonso (angel.estevez.alonso@ugent.be)

#### *Task and deliverable associated the collected/produced data*

Task 3.1. Upstream steps for precursor production (M1-36).

#### *Type of data*

In Task 3.1., the research will generate data such as IC and GC chromatograms (.gcd) and rRNA sequences (.fastq) (expected volume of 5 GB). We will create spreadsheets (.xls) to analyze the generated data and different manuscripts (.doc and .ppt) such as documentation, protocols, scientific papers, posters, presentations, internal reports, etc. (expected volume of 1 GB). Expected total volume of data is approximately 6-7 GB.

#### *Metadata description*

Raw measurement data will be collected per measurement test, including a text file with a clear description of what the data represent and how they were generated (e.g., and experimental plan, an analytical method, an SOP, etc.). The input-files used for the measurements will be kept inside the same folder. The name of the folder will contain the experimental description and measurement conditions. A text file explaining the naming will be maintained.

#### *Legal issues management*

The work in Task 3.1. will result in ways to hydrolyze complex organic wastes and reactor configurations to produce carboxylic acids from these hydrolyzed wastes. Valorisation will most likely be implemented by means of follow-up research projects with industrial partners. Reactor designs and operational conditions can be protected as trade secrets. Patents are difficult to obtain in this field.

#### *Data storage during the project*

In Task 3.1. data files (both raw and treated), manuscripts, reports, ... will be stored in OneDrive for Business and Sharepoint provided by UGent. Large (> 10 Mb), raw measurements data (spc, fcs, fasta, tiff, etc.) will be stored on the additional shared disk space of UGent called "shares". Code scripts for data analysis will be stored at UGent GitHub Enterprise. After publishing, data will be stored in the public available repositories dedicated for each specific type of data.

There is automated backup system provided for each type of data storage. OneDrive and Sharepoint are backed up in the cloud and SharePoint enables version control that is indefinite in time. In the shares, the "snapshots" of the data are made, that allow to retrieve (older versions of) files. Daily snapshots are made for 5 weeks and weekly snapshots are made for 15 weeks. This enables to retrieve previous versions 15 weeks back in total. The shares also has the off-site backup system – the disk with data is copied and transferred to another location.

#### *Data sharing and long term preservation*

Infrastructure INRAE dataverse

Suggested key words: agro-waste, hydrolysis, thermophilic fungi, thermophilic fermentation, carboxylic acids, volatile fatty acids, medium chain fatty acids, EGSB reactor

*Resources dedicated to data management*

ND

NID

Contact: Cristiana Torres (c.torres@fct.unl.pt)

*Task and deliverable associated the collected/produced data*

Task 3.1 - Upstream steps for precursor production.

Task 3.3 - Microbial PHA production.

Task 4.2 - Pilot scale production, biorefineries schemes and end-usages in close to real conditions.

*Type of data*

Qualitative and quantitative data from experimental measurements, as raw or analysed data; numerical data (datasets, spreadsheets); textual data (protocols, methodological descriptions, reports); mixed media data (image/ video) and presentations.

Formats: .docx, .pdf, .xlsx, .jpg, .png, .bmp, .pptx, .zip, PDF, etc.

*Metadata description*

Data will be collected and/or produced by sampling, experimental conditions, variables, units of measurement, as well as procedures for processing and analysing experimental data, which enables validation of the conducted analysis and displayed results, as well as re-use of the data.

*Legal issues management*

Regardless of the type of research output, Agriloop partners will seek to ensure which is or not confidential due to legal or ethical requirements, IP protection or commercial exploitation limitations.

*Data storage during the project*

For data storage, backup and archiving, researchers will be assisted by institutional IT team, in charge of maintaining storage capacity. Backups will be performed regularly in the Agriloop INRAE dataverse.

*Data sharing and long term preservation*

Infrastructure INRAE dataverse.

Suggested key words: Wastes pre-treatments, PHA production, PHA extraction. procedures, Characterization PHA, Pilot scale PHA production.

*Resources dedicated to data management*

ND

## ENTO

Contact: Juan Antonio Cortes (jcortes@entomotech.es)

*Task and deliverable associated the collected/produced data*

Task 3.4. PHA extraction

*Type of data*

Insect growth performance, food conversion ratio, mass gain, conversion ratio and insect biomass composition as proteins, oils and chitin

*Metadata description*

Data will be collected by simple observation with necessary repetitions and registered on Excel for the later use and statistics.

Report commenting and highlighting the results will be provided along with the raw data (Excel)

*Legal issues management*

Any result will be keep in secret following Entomotech IP protection protocol, and if any publication, will be first share and informed to the Agriloop partners and if agree could be disclosed

*Data storage during the project*

Copies are done in the working computers/laptops and on the cloud

*Data sharing and long term preservation*

Supported by INRAE at due time

*Person in charge of data management*

The R&D manager

*Resources dedicated to data management*

1 person month

### 3.4.WP4 End-products and value chains

## BIO-MI

Contact: Valentina Siuc (valentina.siuc@bio-mi.eu)

*Task and deliverable associated the collected/produced data*

T4.2. Pilot-scale production, biorefineries schemes and end-usages in close to real conditions.

*Type of data*

Collected data will be numeric, text and visual in formats .xlsx, .doc, .jpeg and .ppt. with expected size in megabytes.

*Metadata description*

Data will be collected from experimental procedures, process optimization and laboratory testing (observational and experimental data). All data is organized according to the company's internal policy (know-how).

*Legal issues management*

Know-how, EU legislation

*Data storage during the project*

Office computer, backup external Hard drive.

*Data sharing and long term preservation*

Infrastructure INRAE dataverse.

Suggested key words for dataset indexation: flexible films, rigid materials.

## UNIVR

Contact: David Bolzonella ([david.bolzonella@univr.it](mailto:david.bolzonella@univr.it)), Giovanna Pesante ([giovanna.pesante@univr.it](mailto:giovanna.pesante@univr.it))

*Task and deliverable associated the collected/produced data*

Task 1.2. Integrate feedstock & valorisation pathways potential: early assessment.

Task 4.2. Pilot-scale production, biorefineries schemes and end-usages in close to real conditions.

Task 4.3. Environmental, safety, economic and social assessment.

*Type of data*

Task 1.2 Numeric data for mass balances, produced in one excel file.

## Task 4.2

- 15-20 kilos of purified PHA will be produced in a pilot plant in 100 batches. The data relative to this production will be numeric and will be collected in a multiple-tab excel file, which we will share with the consortium.
- 1-2 kg of concentrated nutrients (N, P) will be recovered in 5 batch tests. The data relative to this production will be numeric and will be collected in a multiple-tab excel file, which we will share with the consortium.

Task 4.3 The framework based on multi-criteria decision analysis developed in T1.4 is applied to derive meaningful and transparent decision from the vast number of environmental, economic and social performance indicators derived in each scenario (UNIVR, UNIROMA, FCAC).

*Metadata description*

The data will be produced by chemical -physical characterisation of the samples (analytical data) and by elaboration of the obtained values (mass balances and yields).

All this data will be collected in dedicated excel files, whose template is available on request.

*Legal issues management*

Intellectual property rights will be managed according to the decisions taken by the Agriloop consortium as stated in Task 5.4 of the grant agreement signed by all the partners.

*Data storage during the project*

Data and metadata will be stored on OneDrive Storage system during data collection, and then periodically uploaded into the storage system suggested by the Agriloop consortium (INRAE dataverse).

*Data sharing and long term preservation*

Infrastructure INRAE dataverse.

Key words: Mass balance, PHA production, nutrient recovery.

## AVECOM

Contact: Michael Pil (michael.pil@avecom.be)

*Task and deliverable associated the collected/produced data*  
task 4.2

### *Type of data*

We will perform the upscaling of the production of MP, which means we can provide tables with parameters over time. These parameters are CODsoluble, N, P, protein %, TSS, influent compositions, productivities, yields, COD conversion,..will be provided in an excel format

### *Metadata description*

Data will be collected by performing chemical and physical analyses on the influent and effluent of the process, will be put in tables and graphs.

Data will be collected by analysis and calculations of the parameters, no or little data will be re-used for our process. The data production is following the general protocols for certain analysis (hach large test kits or standard protocols)

### *Legal issues management*

Avecom has several patents concerning the production of microbial protein. These IP rights and ownerships should be discussed with the IP responsible: Stijn Boeren (stijn.boeren@avecom.be).

### *Data storage during the project*

Data will be stored on a private google drive from AVECOM.

### *Data sharing and long term preservation*

INRAE dataverse.

Suggestion of key words: upscaling of the microbial protein production process.

### *Resources dedicated to data management*

WP5 person months

### *Person in charge of data management*

Michael Pil, project engineer at Avecom.

## UM

Contact: Angellier-Coussy (helene.coussy@umontpellier.fr)

*Task and deliverable associated the collected/produced data*  
Task 4.1.

### *Type of data*

Excel files, SEM pictures, DSC thermograms

### *Metadata description*

Characterization of PHA-based materials. Characterization in terms of thermal properties assessed by differential scanning calorimetry analysis, thermal stability assessed by thermogravimetric analysis, mechanical properties assessed by tensile tests, microstructure assessed by SEM observations and WAXS analysis, gas barrier properties.



*Legal issues management*

ND

*Data storage during the project*

Data will be stored on my computer, as well as on an external disk.  
The collaborative workspace can also be a place of storage.

*Data sharing and long term preservation*

Infrastructure INRAE dataverse

Suggested key words: bioeconomy, PHAs, polyhydroxyalcanoates, agricultural residues, biosourced materials.

*Resources dedicated to data management*

ND

*Person in charge of data management*

ND

### 3.5.WP5 Communication, dissemination and exploitation

BDC

Contact: Emma Needham (emma.needham@york.ac.uk)

*Task and deliverable associated the collected/produced data*

Task 5.1 , 5.2, 5.3

*Type of data*

Task 5.1

Internal data on communication audiences (targeted) and which events to attend.

Which other EU projects to collaborate with.

Communication metrics e.g. metrics: media coverage, social media following, mentions on websites or reports, website visitor numbers, webinar attendance, downloads of reports or resources, numbers of stakeholders directly engaged with at events.

Task 5.3

Training materials include technology case studies, best practice guidance, product samples and product performance data. Training events and other project webinars are recorded and made available as further training content on the website (IT) and YouTube channel (BDC).

*Metadata description*

Data will be recorded from other software platforms such as website analytics and social media metrics.

*Legal issues management*

ND

*Data storage during the project*

ND

*Data sharing and long term preservation*

Infrastructure INRAE dataverse.

Suggested key words: social media, website analytics, event attendees, media coverage, mentions on websites or reports, website visitor numbers, webinar attendance, downloads of reports or resources, numbers of stakeholders directly engaged with at events.

*Resources dedicated to data management*

ND

*Person in charge of data management*

ND

## FCAC

Contact: Màrius Simon Monrós (marius.simon@fcac.coop)

*Task and deliverable associated the collected/produced data*

WP1= Tasks 1.1, 1.2

WP4= all

WP5= all

*Type of data*

Residues = volumes, nature, location, process details.

Information about Know how existing business value chains , including prices, etc

Know how related to Farmer needs.

Training activities: attendants' information (name, institution, etc...).

*Metadata description*

-FCAC own data/Methodolgy= annual surveys and regular contacts with the cooperatives and policy makers.

-Information from official sources (National or UE) about Agriculture, rural areas and bioeconomy, etc / Methodology= public availability.

-Data from affiliate cooperatives . surveys or special data requirement depending the task to be achieved.

-Other institutions (through regular agreements with FCAC). Data requiriment depending the task to be achieved.

*Legal issues management*

Catalan, Spanish and European legislation.

In some case we have our own legal department to consult from some issues.

*Data storage during the project*

FCAC own server. Backup every day.

*Data sharing and long term preservation*

Infrastructure INRAE dataverse

Suggested key words: volume wastes; crops location; current valorization;

*Resources dedicated to data management*

ND

*Person in charge of data management*

ND

## 4. Personal data collection and/or processing by WP

### 4.1.WP 1 Foundational circular & strategic flows

SDU

Contact: Benjamin Khoshnevisan ([bekh@igt.sdu.dk](mailto:bekh@igt.sdu.dk))

*Task and deliverable associated the collected/produced data*

Task 1.4. Environmental & circularity performance early guidance.

Task 4.3. Environmental, safety, economic and social assessment.

*Acquisition of an informed consent providing details of the procedures for data collection, retention, destruction or re-use.*

Data collection involves survey through questionnaires in word or excel format. It can also include online interviews. The consent will be obtained to use/reuse the data.

*Methods of storage and exchange*

SDU OneDrive.

Agriloop collaborative platform.

*Data safety procedures used*

SDU OneDrive is protected by SDU IT.

*Data flow between partners and WP*

ND

*Tracking or observation of participants*

no

*Collection or processing of sensitive personal data*

no

*Transfer personal data to non EU partners*

no

*Internal process to register personal data processing/project in the register of the establishment*

SDU does not use personal data to deliver its tasks.

*Name of the Data Protection Officer (DPO)*

ND

USC

Contact: Lucas Van der Hauwaert ([lucas.vanderhauwaert@usc.es](mailto:lucas.vanderhauwaert@usc.es))

*Task and deliverable associated the collected/produced data*

Task 1.2: Integrate feedstock and valorisation pathways potential: early assesment.

Task 2.3: Decision criteria on extraction techniques.

Task 3.5: Library of models of microbial fermentations and PHA production.

Task 4.4: Business concept of valuable sustainable value chains.

*Acquisition of an informed consent providing details of the procedures for data collection, retention, destruction or re-use.*

To collect data from our partners, we will conduct virtual face-to-face meetings to discuss their data and its intended use. We will obtain their consent before using their data. Our partners will receive an Excel template to enter their data.

Depending on the consent the used data will be made available to the general public.

The generated code can be reused by downloading it from a GitHub repository.

*Methods of storage and exchange*

One drive cloud services.

GitHub for version control of the code.

*data safety procedures used*

We use a reinforced security credential system with two-factor authentication for system access at the university. The official data storage is OneDrive, which offers data encryption and secure connections. However, data security depends on the research group's practices, and USC cannot guarantee protection if external devices are used or personal data isn't adequately anonymized.

*data flow between partners and WP*

Data from WP 1, 2 and 3 need to flow to the USC.

*tracking or observation of participants*

no

*collection or processing of sensitive personal data*

no

*transfer personal data to non EU partners*

no

*internal process to register personal data processing/project in the register of the establishment*

USC has a comprehensive privacy and data protection policy that can be found on the page: <https://www.usc.gal/gl/normativa/protecciondatos/index.html>, along with a data protection protocol.

*name of the Data Protection Officer (DPO)*

Mr. Marcos Almeida Cerrada

## 4.2.WP5 Communication, dissemination and exploitation

FCAC

Contact: Màrius Simon Monrós (marius.simon@fcac.coop)

*Task and deliverable associated the collected/produced data*

WP1= Tasks 1.1, 1.2

WP4= all

WP5= all

*Acquisition of an informed consent providing details of the procedures for data collection, retention, destruction or re-use.*

- informed consent will be obtained and provide details for our procedures for data collection (face to face, online survey, (ZOOM WEBINARS).
- retention (duration, 10 years...).
- destruction: all the records are digital.

*Methods of storage and exchange*

- FCAC network server for FCAC team.
- Within the consortium Agriloop collaborative platform.
- with cooperatives, other companies and institutions: shared files through <https://owncloud.com/>

*Data safety procedures used*

- FCAC staff internal rules.
- Storage only in secure medias (FCAC local area network, owncloud.com)

*Data flow between partners and WP*

Data collected by FCAC from different sources will be shared with the other partners through medias described above.

*Tracking or observation of participants*

no

*Collection or processing of sensitive personal data*

no

*Transfer personal data to non EU partners*

no

*Internal process to register personal data processing/project in the register of the establishment*

ND

*Name of the Data Protection Officer (DPO)*

Jordi Vives Xiol (CEO)

## 5. Conclusion

Almost all AgriLoop partners have fulfilled the web survey, which has been used to write this first version of the DMP. It must be noticed that all respondent expressed the wish to use INRAE dataverse platform to manage their data for long term preservation. They produced a list of key words that will be used to index their datasets. Consequently, a dataverse dedicated to AgriLoop project has been created (<https://entrepot.recherche.data.gouv.fr/dataverse/agriloop/>) and a short guide explaining how to use it will be sent to partners soon.

A short term perspective is now to check if partners' responses in terms of personal data collection conform General Data Protection Regulation (GDPR). Furthermore, a list of partners who are concerned by personal data transfer to non-EU partners will be elaborated. They will have to sign data transfer agreement with non-EU partners.

## Annex I

### List of partners

Short name	Full name	Type	Country
<b>INRAE</b>	Institut National de Recherche pour l'Agriculture, l'alimentation et l'Environnement	RE	FR
<b>WR</b>	Wageningen Research	RE	NL
<b>SDU</b>	University of Southern Denmark	RE	DN
<b>ECOZEPT</b>	Ecozept	SME	FR
<b>ECOZEPT DE</b>			DE
<b>USC</b>	University of Santiago de Compostela	RE	ES
<b>FCAC</b>	Federació de Cooperatives Agràries de Catalunya	Other	ES
<b>UNIBO</b>	University of Bologna	RE	IT
<b>ITQB</b>	Institute of Chemical and Biological Technology, Universidade NOVA de Lisboa	RE	PT
<b>TOMA</b>	TomaPaint SRL	SME	IT
<b>NID</b>	Innovation and Development Association, NOVA School of Science and Technology	RE	PT
<b>UNIROMA</b>	University of Roma	RE	IT
<b>ENTO</b>	ENTOMOTECH S.L.	SME	ES
<b>CSIC</b>	Consejo Superior de Investigaciones Biológicas	RE	ES
<b>UGENT</b>	University of Gent	RE	BE
<b>UNIVR</b>	University of Verona	RE	IT
<b>BIO-MI</b>	Bio-Mi	SME	HR
<b>UM</b>	University of Montpellier	RE	FR
<b>AVECOM</b>	Avecom	SME	BE
<b>IT</b>	INRAE Transfert	Other	FR
<b>FHNW</b>	University of Applied Sciences Northwestern Switzerland	RE	CH
<b>BDC</b>	Biorenewables Development Centre	Other	UK



