Deliverable 1.4
Data Management Plan

primesoft

Development of innovative priming technologies safeguarding yield security in soft fruit crops through a cutting-edge technological approach

Funded by the European Union

The project has received funding from the European Union's Horizon Europe programme under Grant Agreement 101079119

Table of Contents

Funded by the European Union

The project has received funding from the European Union's Horizon Europe programme under Grant Agreement 101079119

Document Summary

Deliverable number & title: **D1.4 – Data Management Plan**

Version & submission date: **v1, 28 April 2023**

Lead Beneficiary: **CSIC**

Related Work package: **WP1**

Authors: **Francisco Tomas-Barberan (CSIC), Carlos Garcia (CSIC)**

Contributor(s) to deliverable: UP, KUL, CUT, CSIC, NTUA

Reviewers: **Maarten Hertog (KUL), Vassilis Fotopoulos (CUT), George Manganaris (CUT), Magda Krokida (NTUA), Bernd Müller-Röber (UP), Katrin Czempinski (UP),**

Communication level:
☒ **PU** Public
☐ **CO** Confidential, only for members of the consortium (including the Commission Services)

Approved by:
Final version of the deliverable to be approved by the respective GA members (before submission to the EU).
☒ CUT (Coordinator) ☒ KUL ☒ CSIC ☒ UP ☒ NTUA

Grant Agreement Number: 101079119

Call: HORIZON-WIDERA-2021-ACCESS-03

Type of action: HORIZON Coordination and Support Actions

Granting authority: European Research Executive Agency

Acronym: PRIMESOFT (priming in soft fruits)

Start date of Project: 1st November, 2022

Duration: 3 years

Project coordinator: **CUT**

1. Introduction

The PRIMESOFT data management plan (DMP) is part of Work Package 1 (WP1) and is based on FAIR concepts making the data findable, accessible, interoperable and reusable. This document is the first edition of the DMP, delivered in Month 6 of the project. The purpose of the DMP is to provide an overview of the datasets to be generated by the PRIMESOFT consortium as well as guidelines for the collection, analysis, storage, dissemination, and the overall management policy. The DMP will be a living document, constantly evolving during the lifespan of the project to cover new datasets, methodologies and policies.

2. Data summary

The main PRIMESOFT objective is to explore innovations in the application of priming agents in value-added soft fruit crops from a range of perspectives and strengthen educational, research and innovation activities among the Widening Institution (WI) and the Advanced Partners (APs). Due to the multidisciplinary nature of the PRIMESOFT project, different data types such as transcriptomics and metabolomics will be stored. To this end, the data summary of the DMP has to necessarily contemplate all required data types.

Dataset	**1. Crop growth data**
Description and type of data	Data regarding agronomic/morphological parameters as well as yield parameters following priming treatments will be collected in comparison with appropriate controls.
Format & size of the data	Format: Word (Docx), Excel (xlsx), jpeg Size: ~500 MB.

Funded by the European Union

The project has received funding from the European Union's Horizon Europe programme under Grant Agreement 101079119

Storage of the data	The dataset will be preserved in the PRIMESOFT online repository, during the lifetime of the project.
Re-use of the data	It is not foreseen to be used for other projects but published data will be made available under the CC0-1.0 license to stimulate re-use.
Utility	The results could be used by researchers, farmers, agronomists, etc.
Origin	Primary data
Related WP and Task(s)	WP3: Research & Innovation activities Task 3.1: Experimental approach
Partners involved	CUT
Data sharing, publication	Any relevant data will be published through open access journals being made available either as supplementary materials or through dedicated online repositories.
Ethical issues	No
Dataset	**2. Encapsulation of priming agents**
Description and type of data	Data regarding the specifications of the final products (size of the nanoparticles, concentration, type of encapsulating agents, etc.) as well as the efficacy of the encapsulated priming agents will be collected.
Format & size of the data	Format: Word (Docx), Excel (xlsx), jpeg Size: ~500 MB.
Storage of the data	The dataset will be preserved in the PRIMESOFT online repository, during the lifetime of the project.
Re-use of the data	It is not foreseen to be used for other projects but published data will be made available under the CC0-1.0 license to stimulate re-use.
Utility	The results could be used by researchers, farmers, agronomists, etc.

Funded by the European Union

The project has received funding from the European Union's Horizon Europe programme under Grant Agreement 101079119

Origin	Primary data
Related WP and Task(s)	WP3: Research & Innovation activities Task 3.2: Encapsulation of priming agents
Partners involved	NTUA
Data sharing, publication	Any relevant data will be published through open access journals with the corresponding datasets being made available either as supplementary materials or through dedicated online repositories.
Ethical issues	No
Dataset	**3. Transcriptomic data**
Description and type of data	mRNA sequencing data; RT-qPCR data; of strawberry (*Fragaria* x *ananassa*) and raspberry (*Rubus idaeus* L.) will be obtained from fruits (different developmental stages) of primed plants.
Format & size of the data	RNA-seq will be performed using e.g. Illumina or the DNBseq technology platforms which provide sequence reads of 100 or 150 bases length. Preferentially, paired-end (PE) sequencing will be employed. It is envisaged that approximately 20-25 Mio. sequence reads will be obtained for each experimental condition (type of priming agent, treatment timepoint, plant species, replicate number). Data volume will be in the Tb range. Raw data format will typically be FASTQ. For RT-qPCR analyses, REST-XL calculated deltaCP values subsequently tested for significance by a Pair Wise Fixed Reallocation Randomisation Test will be calculated and included in the data repository.
Storage of the data	The dataset will be preserved in the PRIMESOFT online repository, during the lifetime of the project.
Re-use of the data	It is not foreseen to be used for other projects but published data will be made available under the CC0-1.0 license to stimulate re-use.

Funded by the European Union

The project has received funding from the European Union's Horizon Europe programme under Grant Agreement 101079119

Utility	The results could be used by researchers.
Origin	Primary data
Related WP and Task(s)	T3.4 Global transcriptomic and SIGS analyses of plants primed with the prototypes
Partners involved	UP, CUT
Data sharing, publication	Any relevant data will be published through open access journals with the corresponding datasets being made available either as supplementary materials or through dedicated online repositories.
Ethical issues	No
Dataset	**4. Metabolomics data**
Description and type of data	The metabolomics data will be collected by UPLC-MS and used for phytochemicals analysis.
Format & size of the data	Format: .d; MZml; csv. 100MB-200MB
Storage of the data	The dataset will be preserved in the PRIMESOFT online repository, during the lifetime of the project.
Re-use of the data	It is not foreseen to be used for other projects but published data will be made available under the CC0-1.0 license to stimulate re-use.
Utility	The results could be used by researchers.
Origin	Primary data
Related WP and Task(s)	WP3: Research & Innovation activities Task 3.5 Metabolomic analysis of primary and secondary metabolism
Partners involved	CSIC

Funded by the European Union

The project has received funding from the European Union's Horizon Europe programme under Grant Agreement 101079119

Data sharing, publication	Any relevant data will be published through open access journals with the corresponding datasets being made available either as supplementary materials or through dedicated online repositories.
Ethical issues	No
Dataset	**5. Extraction of bioactive compounds**
Description and type of data	The extraction will be optimized according to previous publications and the extracts will be used for their total phytochemical content analysis.
Format & size of the data	csv; values mg/100 g fresh weight
Storage of the data	The dataset will be preserved in the PRIMESOFT online repository, during the lifetime of the project.
Re-use of the data	The data can be used for future research proposals. Published data will be made available under the CC0-1.0 license to stimulate re-use.
Utility	The data could be used by researchers to improve extraction methodologies, and potentially by the private sector for the development of new products.
Origin	Primary data
Related WP and Task(s)	WP3: Research & Innovation activities Task 3.5: Metabolomic analysis of primary and secondary metabolism
Partners involved	CSIC, CUT
Data sharing, publication	Any relevant data will be published through open access journals with the corresponding datasets being made available either as supplementary materials or through dedicated online repositories.
Ethical issues	No
Dataset	**6. Life Cycle Analysis**

Funded by the European Union

The project has received funding from the European Union's Horizon Europe programme under Grant Agreement 101079119

Description and type of data	The life cycle analysis will be applied on the whole process chain in order to evaluate the sustainability of (i) cultivation of berries, (ii) using of priming agents, and compare them with the existing conventional cultivation routes. It will include quantitative information about the flow of materials and energy and the emissions (air, water, soil).
Format & size of the data	Format: word (Docx), excel (xlsx), jpeg Size: up to 1 Gb.
Storage of the data	The dataset will be preserved in the PRIMESOFT online repository, during the lifetime of the project.
Re-use of the data	It is not foreseen to be used for other projects but published data will be made available under the CC0-1.0 license to stimulate re-use.
Utility	The data could be used by researchers and by the private sector for the development of new products.
Origin	Primary data
Related WP and Task(s)	WP3: Research & Innovation activities Task 3.6: Life Cycle Assessment
Partners involved	NTUA
Data sharing, publication	Any relevant data will be published through open access journals with the corresponding datasets being made available either as supplementary materials or through dedicated online repositories.
Ethical issues	No
Dataset	**7. Volatilomics data**
Description and type of data	The volatilomics data will be collected by SITF-MS and GC-MS

Funded by the European Union

The project has received funding from the European Union's Horizon Europe programme under Grant Agreement 101079119

Format & size of the data	Format: .csv; .xlsx 100MB-200MB
Storage of the data	The dataset will be preserved in the PRIMESOFT online repository, during the lifetime of the project.
Re-use of the data	It is not foreseen to be used for other projects but published data will be made available under the CC0-1.0 license to stimulate re-use.
Utility	The results could be used by researchers, farmers, agronomists, etc.
Origin	Primary data
Related WP and Task(s)	WP3: Research & Innovation activities T3.5 Metabolomic analysis of primary and secondary metabolism
Partners involved	KU Leuven
Data sharing, publication	Any relevant data will be published through open access journals with the corresponding datasets being made available either as supplementary materials or through dedicated online repositories.
Ethical issues	No

3. FAIR Data - Making PRIMESOFT data Findable, Accessible, Interoperable and Reusable

In accordance with EU guidelines and following the official Horizon Europe instructions, the data by PRIMESOFT will be 'FAIR', that is Findable, Accessible, Interoperable and Re-usable. To facilitate the flow of information, a project-dedicated data repository, in a secured PRIMESOFT drive was developed to allow data exchange among project partners of raw and analyzed experimental data derived from spectra, chemical and physical properties, time-dependent chemical and physical

Funded by the European Union
The project has received funding from the European Union's Horizon Europe programme under Grant Agreement 101079119

processes and computer simulations, as well as experimental protocols (wet lab and fieldwork), algorithms for computational calculations, educational resources, users' information, tutorials and workshops material.

3.1. Making data findable

The Steering Committee developed a set of guidelines for the implementation of the data that will follow a common framework for dataset description, reference and name, standards and metadata, data sharing, archiving and preservation.

To make data findable to the consortium partners, a meta-datafile will be created and stored in a central folder on the PRIMESOFT intranet describing, per experiment, in short the experimental setup, the main experimental factors involved and the outputs registered. This meta-datafile will refer to the various datasets, their location on the PRIMESOFT intranet and a short description of the content of each dataset file. Each dataset file will be labeled with the project code (PRIMESOFT), followed by an acronym referring to the analytical method (met=metabolomics; vol= volatilomics; trans=transcriptomics; tab=generic table of result; prot=protocol etc.) and the date (day, month and year not separate). i.e. PRIMESOFT_MET45_281022. Raw experimental data will be collected in their native form and exported in ASCII format (equivalent to International Standard ISO/IEC 646) as easily manipulatable, comma- or tab-delimited value files, and will be incorporated in Microsoft Excel with the final data file output in .xlsx format.

To make data findable to the wider audience, the PRIMESOFT website (https://www.prime-soft.eu/) and the dedicated CORDIS page (https://cordis.europa.eu/project/id/101079119) will be the main starting points. All (open access) publications resulting from the project will be listed on these websites using unique identifiers to refer to the publisher's websites hosting the original publication and their

Funded by the European Union

The project has received funding from the European Union's Horizon Europe programme under Grant Agreement 101079119

supplementary data files. Additional open access data deposited elsewhere will be listed as well providing their unique identifiers referring to the repositories hosting the data.

3.2. Making data openly accessible

Initially, PRIMESOFT project data in its raw and pre-processed form will only be directly accessible to the project partners. However, all relevant data will be published through open access resources. Once published, the curated data will either be made available as supplementary data on the publisher's website or through dedicated open access repositories.

Depending on the nature and volume of the data, curated datasets will be deposited in open access repositories such as Metabolights (https://www.ebi.ac.uk/metabolights/) for metabolomics data, ArrayExpress (https://www.ebi.ac.uk/biostudies/arrayexpress) for transcriptomics data. By submitting data to these repositories, unique identifiers will be generated that will be made available through the published manuscripts referring to the datasets and will be communicated through the project website as indicated above.

Additionally, the project's website (www.prime-soft.eu) will present PRIMESOFT's major achievements in a language easily understandable by agricultural practitioners and the general public, highlighting the main benefits for the farmers, consumers, and other relevant stakeholders.

3.3. Making data interoperable

The Steering Committee of the PRIMESOFT project, due to their high expertise on the fields implicated, will guarantee the usage of the correct standards and methodologies for the PRIMESOFT results following the requirements of the repositories mentioned, therefore facilitating their interoperability. Additionally, we will use standard vocabulary for all data and metadata types in order to ensure interdisciplinary interoperability, using open software

Funded by the European Union

The project has received funding from the European Union's Horizon Europe programme under Grant Agreement 101079119

environments whenever possible, supporting file formats that will be easily downloaded and read by the majority of the relevant stakeholders (e.g., pdf., docx., xlsx.).

3.4. Increase data re-use (through clarifying licenses)

All published data resulting from the PRIMESOFT project will be made available to stimulate future re-use of the data. Depending on the publisher's website (in case of supplementary datafiles) or the various open access repositories (in case of additional open access data files), the exact user license might vary. However, to be 100 % FAIR compliant the PRIMESOFT project will, by default, apply a CC0-1.0 license with an additional request to give credit to the project to acknowledge the provenance of the data.

4. Allocation of resources

The cost of ensuring that the data generated in the PRIMESOFT project meets the FAIR conditions is included in the budget of the project as well as the cost related to the open access publication. The PRIMESOFT project will preferably make use of free repositories as indicated above. Each partner of the project will be responsible to manage their data with the dedicated repositories.

5. Data security

The data generated by the PRIMESOFT will be shared between the partners through online server (Dropbox) to guarantee the protection and minimizing the risk of information leak and destruction. All data generated will be backed up routinely and appropriately each month in an external solid disk with no internet connection.

Funded by the European Union

The project has received funding from the European Union's Horizon Europe programme under Grant Agreement 101079119

6. Ethical aspects

The PRIMESOFT project will produce and process non-sensitive data [Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (GDPR)], therefore, it does not raise significant ethical issues. Furthermore, the PRIMESOFT website will include a Data Policy, Privacy Statement, and cookies bar, informing website visitors about what the project does with their personal data.

Funded by the European Union

The project has received funding from the European Union's Horizon Europe programme under Grant Agreement 101079119

The project has received funding from the European Union's Horizon Europe programme under Grant Agreement 101079119