



## D1.3

# Data Management Plan

Date: 25.03.2025

Version: 2.0

## Deliverable 1.3 Data Management Plan

### Document Information

<b>Project title</b>	SOIL BIODIVERSITY AND FUNCTIONALITY OF MEDITERRANEAN OLIVE GROVES
<b>Project acronym</b>	SOIL O-LIVE
<b>Grant Agreement</b>	101091255
<b>Work package</b>	WP1 – Coordination and management
<b>Deliverable number and title</b>	D1.3 Data Management Plan
<b>Deliverable description</b>	Elaboration of Data management plan for the project. To ensure an effective contribution to the EUSO, the consortium will identify in the DMPs which knowledge and specific datasets might be of relevance to the EUSO and appoint a contact person to participate in discussions on data management with the JRC's EUSO
<b>Deliverable type</b>	DMP — Data Management Plan
<b>Dissemination level</b>	PU – Public
<b>Lead beneficiary</b>	UJA
<b>Due submission date</b>	30 June 2023 (M6)
<b>Submission date</b>	27.04.2025

### Document Author(s) and Reviewer(s)

<b>Name</b>	<b>Entity</b>	<b>Role (author, reviewer)</b>
F.J. Ariza-López	UJA	Author / reviewer
P. Belart Rodríguez	UJA	Author / reviewer
V. Valenzuela Polo	UJA	Author / reviewer
A.J. Manzaneda Ávila	UJA	Author / reviewer

### Document history

<b>Version</b>	<b>Date</b>	<b>Made by</b>	<b>Short Description of Changes</b>
1.0	30.06.2023	A.J. Manzaneda Ávila	Initial version
2.0	25.03.2025	A.J. Manzaneda Ávila	Updated version after RP1

## Table of Contents

Document Information.....	2
Table of Contents .....	3
Executive Summary .....	4
1. Introduction and purpose .....	5
2. Type of data .....	6
3. FAIR Data.....	11
3.1. Making data findable, including provisions for metadata .....	11
3.2. Making data accessible .....	12
3.3. Making data interoperable .....	13
3.4. Increase data reuse.....	14
4. Other research outputs .....	15
5. Allocation of resources .....	15
6. Data security.....	16
7. Ethics.....	16
8. Other issues.....	16
ANNEX I. Data collection template.....	17

## **Executive Summary**

This deliverable describes the preliminary version of the **Data Management Plan (DMP) of the Soil O-live project**. The aim of the DMP is to provide an analysis of the main elements of the data management policy that will be used throughout the project, with regard to all the datasets that will be generated and used. DMP is a living document that will be updated as the implementation of the project progresses and when significant changes occur. The format of DMP follows the guidelines suggested for Horizon Europe calls.

## 1. Introduction and purpose

This Data Management Plan (DMP) aims to outline how the project's datasets will be managed throughout its duration and after the end of the project. It will cover data standards and metadata, sharing, archiving preservation, and security to ensure proper management principles are followed. This DMP describes how research data was collected, generated, shared, and preserved in the project context and after the end of the project, following FAIR data principles. The data generated by the project through the different work packages (WPs) will primarily be utilized to accomplish the objectives outlined in the project proposal. The project will collect data from innovative observations obtained from laboratory and/or field experiments, surveys, and assays specified in the project. Additionally, if needed, existing data from public repositories, digital archives, and databases, such as geographical, climatic, economic, etc., will be incorporated to complement the project's data analysis. The nature of the SOIL O-LIVE data set is diverse in line with its multidisciplinary and interdisciplinary approach. It comprises ecological data (e.g., soil biodiversity, climate, soil physical and chemical properties, land degradation estimates, restoration methodologies, and effectiveness, etc.), agronomic data (e.g., olive yield and quality, land management, tree physiological status, etc.), geographical data (latitude, longitude, elevation, slope, etc.), genomic data (e.g., genetic polymorphisms, genome annotations, genome size, chromosome number, etc.), chemical data (e.g., soil pesticide-toxicity data, soil copper concentration, antibiotic, and microplastic pollution). Lastly, it encompasses socio-economic data such as surveys on sustainability and market trends related to olive oil consumption and health.

The DMP is a public deliverable of the SOIL O-LIVE project. It includes initial, mid-term, and final reports (D1.3, D1.4, and D1.5). Such deliverables will be available (i) at the internal SOIL O-LIVE repository, which constitutes the internal Consortium document and data repository for the project (Google® Drive Tool); and (ii) at the ARGOS platform (<https://argos.openaire.eu/home>). In addition, a permanent link to deliverables is available on the SOIL O-LIVE webpages (<https://soilolive.eu>).

The Joint Research Centre of the European Commission (JRC) and the Soil O-live Consortium have a collaboration agreement. As per this agreement, after the SOIL O-LIVE project ends, they will transfer relevant data, knowledge except, if any of the information has potential value for exploitation, and indicators to the EU Soil Observatory (EUSO) and the European Soil Data Centre (ESDAC), except for any personal data collected during the project. For this reason, relevant soil data and information will be available also in the ESAC repository (<https://esdac.jrc.ec.europa.eu/resource-type/datasets>).

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

SOIL O-LIVE will produce various datasets of different types, including both quantitative and qualitative data. The management of this data will aid in achieving the project's scientific goals and disseminating its results. Two main categories of data management are:

- 1) **Research objectives.** The datasets in this category provide all the necessary data for users to reproduce the scientific results of the project. This encompasses experimental data

(including standardized analysis methods and agreed protocols), observations, genome annotations, computer simulations, and data production and analysis codes. Likewise, it includes all relevant metadata necessary for EUSO development and validation of indicators for “Soil health” as listed in the implementation plan of the EU-Soil Mission.

2) **Outreach, dissemination, and communications data.** This collection includes preprints, technical reports, conference presentations or abstract, educational resources, and data related to outreach, dissemination, and communications. However, if any of the information has potential value for exploitation, particularly for the new prototype developed by UCLM, it will be protected by patents, copyrights, or other appropriate means.

## 2. Type of data

### **What types and formats of data will the project generate or reuse?**

The SOIL O-LIVE project will produce different types of data in various formats. Examples of these types of data, as well as their descriptions, are listed in the table below (Table 1). The sources, formats, and levels of data sharing will be determined based on a survey conducted among the consortium members (see Annex I).

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

Elevated (Terabyte order) in all the cases because of the extent and the temporal frame of the study and the type of data collected.

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

The EU Green Deal for 2030 and related programs require agriculture production to meet certain standards. This makes the data collection particularly valuable for stakeholders in the olive oil sector, such as producers, cooperative managers, regional and national representatives, politicians with agro-environmental responsibilities, and distribution companies. The data will also benefit olive oil producers from other regions who want to implement sustainable practices in their production. Additionally, the data will aid in cultivating other Mediterranean woody crops, like almonds or pistachio, that grow in similar ecological conditions. Finally, data will be very important for the scientific community in several fields of knowledge because the project will provide quality data to answer unsolved questions on agroecology and the sustainability of agriculture.

**Table 1.**

Type of data	Description	Formats <sup>1</sup>	Source	Dissemination Level/ Data Sharing	Owner or responsible Partner	Licence for the dataset	Archiving and preservation
<b>Geographical (WP1)</b>	Data on the longitude, latitude, elevation, and slope of the 53 orchards selected in the project. It includes metadata with the primary descriptors of the orchards (phytochemical input/hectare/year, orchard size, the density of plantation, age, cultivar type, owner gender, and mode of governance). Land cover and land use data for each orchard is also generated	csv, xml, kml, shp, shx, jpg, pdf, Geotiff	Gathered by the Consortium:  In situ collection using a device GPS from the orchard centroid.  Gathered by the Consortium:  Laboratory validation for polygons and perimeter by using GIS analyses.  Digital Elevation Model	Public	UJA	The license used for this dataset:  <input type="checkbox"/> CC0  PDDL CC-BY-4.0  ODbL <input type="checkbox"/> Other, please specify:  <input checked="" type="checkbox"/>	The portion of the data that will be relevant for a publication will be available through the following platform(s) and/or repositories: Zenodo  <i>The duration of the preservation will be: N/A</i>  <i>Foreseen costs of the preservation: N/A</i>  <i>Means to cover preservation costs: N/A</i>
<b>Ecological (WP2-WP3-WP4-WP5-WP6)</b>	<u>Soil Biodiversity</u> : This database contains geo-referenced information about the diversity of soil-living organisms, including bacteria (symbiotic or not), fungi, nematodes, ants, and plant species. It includes taxonomic, functional, and genomic estimates of biodiversity, beta-diversity estimates for different cultivation methods and comparisons of diversity over time and between regions. Additionally, it provides estimates of gamma diversity variation for each region.  <u>Soil properties</u> : Geo-referenced data on chemical and physical parameters that it	csv, xml, HTML GeoTIFF, txt  Optical microscopy data (TAR.GZ )	Gathered by the Consortium:  In situ collection from soil sampling at study orchards.  Metabarcoding and metagenomic data. (BLAST)  OTU clustering  Functional biodiversity in existing databases				

<sup>1</sup> Putative type of formats.

**Deliverable 1.3 Data Management Plan**  
**Project SOIL O-LIVE (ID: 101091255)**

	<p>includes all soil components of the EU Statistical Office's Land Use and Coverage Area Frame Survey (LUCAS)<sup>2</sup>. Information on two soil horizons, specifically the 0-10 cm and 10-20 cm layers, will be provided separately.</p> <p><u>Climate</u>: Geo-referenced data on precipitation, temperature, and aridity at the scale of 1km</p> <p><u>Soil Restoration</u>: Geo-referenced multivariate data on variation in degraded/polluted soils after restoration amendments</p> <p><u>Ecosystem services</u>: Geo-referenced data on pest abundance (<i>Prays oleae</i> and <i>Bactrocera oleae</i>) and fruit damage. Geo-referenced data on soil gas exchange and water flux. Carbon fixation potential</p>		<p>(e.g., TRY: Plant Trait database, BacDive)</p> <p>Gathered by the Consortium:</p> <p>In situ collection from soil sampling at study orchards.</p> <p>WorldClim, Chelsa Climate, CGIAR</p> <p>Gathered by the Consortium:</p> <p>In situ collection from experimental plots in selected orchards</p> <p>Gathered by the Consortium</p> <p>In situ periodic surveys and experiments</p>				
<b>Pollution<sup>3</sup></b> <b>(WP2-WP6)</b>	<p><u>Pesticides</u>: The data provided is geo-referenced and contains information on the quantity and presence of herbicides, fungicides, and insecticides found in soils and olive oils. It also includes details on the concentration of copper in the soil. Special attention to glyphosate and</p>	<p>csv, xml, GeoTIFF,</p>	<p>Gathered by the Consortium:</p> <p>In situ periodic surveys and experiments</p>				

<sup>2</sup> Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., & Fernández-Ugalde, O. (2018). LUCAS soil, the largest expandable soil dataset for Europe: A review. European Journal of Soil Science, 69(1), 140– 153.

<sup>3</sup> Information on two soil horizons, specifically the 0-10 cm and 10-20 cm layers, will be provided separately



**Deliverable 1.3 Data Management Plan**  
**Project SOIL O-LIVE (ID: 101091255)**

	<p>oxyfluorfen.</p> <p><u>Microplastic</u>: geo-referenced data on the presence and abundance of soil microplastics.</p> <p><u>Nitrogen excess</u>: geo-referenced data on soil Nitrogen concentration.</p> <p><u>Antibiotics</u>: geo-referenced data for the presence of antibiotics and other veterinarian residues in olive orchard soils</p>	<p>csv, xml, GeoTiff</p> <p>csv, xml, GeoTiff</p> <p>csv, xml, GeoTiff</p>	<p>Gathered by the Consortium:</p> <p>In situ periodic surveys and experiments</p> <p>Gathered by the Consortium:</p> <p>In situ periodic surveys and experiments</p>				
<b>Agronomical (WP6)</b>	<p><u>Olive oil production and quality</u>: geo-referenced data on the number of kilograms of oil/ha for each orchard. Data on olive oil quality from standardized procedures provided by the International Olive oil Council based on the physico-chemical profile. Qualitative evaluation of olive oil by organoleptic tasting (IOC criteria)</p> <p><u>Tree-Physiology</u>: geo-referenced evaluation of post-summer drought tree health in terms of foliar analyses (C/N) and water use efficiency (WUE) from carbon isotopic signatures.</p> <p><u>Digital twin</u></p>	<p>csv, xml, GeoTiff</p> <p>csv, xml, GeoTiff</p>	<p>Gathered by the Consortium:</p> <p>In situ sample collection of olive and oils</p> <p>International Tasting Panel</p> <p>In situ sample collection of leaves</p> <p>Remote sensing data</p>				
<b>Land degradation (WP3)</b>	<p><u>Soil erosion indicators.</u></p> <p><u>Soil erosion and degradation modelling</u></p>	<p>csv, xml, GeoTiff</p>	<p>Gathered by the Consortium:</p> <p>In situ data collection</p> <p>Remote mapping (Copernicus monitoring data and infrastructures)</p>		<p>Università RomaTre &amp; University of Basel</p>	<p>CC-BY-4.0</p>	<p>All data supporting the findings of the WP3 research activity will be made available in ZENODO and, if applicable, at the European Soil Data Centre (ESDAC), the institutional soil data repository of the European Commission Joint Research Centre.</p>

**Deliverable 1.3 Data Management Plan**  
**Project SOIL O-LIVE (ID: 101091255)**

			Modelling outcomes Isotopic signatures				
<b>Genomic and genetic (WP5)</b>	Plant DNA C-values and chromosome number <u>data</u>  <u>Genome assemblies, gene annotation based on RNAseq, genomic diversity (GBS/ddRAD) data.</u>	csv, xml, GeoTiff, jpg (Fluorescence intensity graphs)  FASTQ, FASTA, GFF, VCF	Gathered by the Consortium:  in situ sampling collection and laboratory flow cytometry  in situ sampling collection and nucleotide sequencing				
<b>Social Surveys and dissemination &amp; communication (WP7)</b>	<u>Survey data:</u> data from surveys on sustainability practices conducted on farmer populations at regional level. Gender studies in olive farming  <u>Dissemination and Communication data:</u> Presentations or abstract in international/national conferences for expert audiences, outreach talks, teaching materials (including scientific papers, talk presentations, videos, and posters)	csv, xml  PDF, Keynote, PPT, JPG, mp4	in situ sampling collection  Consortium elaboration				

### 3. FAIR data

#### 3.1. Making data findable, including provisions for metadata

##### Will data be identified by a persistent identifier?

The data will be identified through Digital Object Identifiers (DOIs) and INSDC accessions provided by Zenodo repository where data will be stored (see below).

##### PROVISIONS FOR METADATA

Each data set will have specific metadata standards; examples of metadata are summarized in the table below. Procedures to generate metadata are given in the **Deliverable 7.1 (D7.1)**.

**Table 2.**

Type of data	Metadata format	Elements	License
<b>Biodiversity and Taxonomy</b>	GBIF <sup>4</sup>	Title, description, creator, contact information, license, dataset ID, geographic coverage, taxonomic coverage, temporal coverage, data quality, data format, data access details, and citation.	Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International
<b>Ecological including land degradation</b>	Ecological Metadata Language (EML)	Title, abstract, creator, temporal and spatial coverage, keywords, methods, variables, data quality, publications, data access, and usage.	Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International
<b>Pollution</b>	Ecological Metadata Language (EML)	Title, abstract, creator, temporal and spatial coverage, keywords, methods, variables, data quality, publications, data access, and usage.	Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International  CC-BY-NC-4.0  ODbL
<b>Agriculture</b>	AGRI <sup>5</sup>	Title, alternative title, creator: personal, corporate and conference, publisher, place of publication, date of publication, subject: classification and thesaurus, description: notes, edition, abstract, identifier, type, format, language relation, availability source coverage: spatial, temporal rights: statement, terms of use citation: title, identifier, number, chronology	Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International
<b>Genomic and genetic</b>	Sequence Read Archive (SRA) <sup>6</sup> , BioSamples <sup>7</sup> , European Nucleotide Archive	Study (title, description, principal investigators, affiliations, and funding sources), Experiment (experimental design, sample attributes, library preparation methods, sequencing platforms used, and any relevant protocols followed), Sample (sample name or ID, organism name, strain, tissue type, disease status, and other relevant	EMBL-EBI will minimise barriers to reuse of data in EMBL-EBI resources by adopting the Creative Commons (CC) license framework across all its data resources in the next 5 years.

<sup>4</sup> Global Biodiversity Information Facility (<https://www.gbif.org/>).

<sup>5</sup> <https://www.fao.org/3/ae909e/ae909e00.htm>

<sup>6</sup> <https://www.ncbi.nlm.nih.gov/sra>

<sup>7</sup> <https://www.ebi.ac.uk/biosamples>

	(ENA) <sup>8</sup> , European Variation Archive (EVA) <sup>9</sup>	annotations). Run: (sequencing platform used, sequencing file format, read length, number of reads, and any associated quality control measures), Data Processing (data processing steps and bioinformatics pipelines applied to the raw sequencing data, alignment, variant calling, quality control), Data Availability, Experimental Factors, Data Quality.	[2023, <a href="https://www.ebi.ac.uk/licencing">https://www.ebi.ac.uk/licencing</a> ]
<b>Flow cytometry</b>	MIFlowCyt <sup>10</sup>	Experimental Overview (experimental design, sample types, and any controls or standards used), Instrument Description (instrument manufacturer, model, configuration, and settings such as laser excitation wavelengths and detection channels), Sample Preparation (sample handling, preparation, and staining protocols, sample source, cell types, labelling reagents, staining procedures, and any treatments or manipulations performed), Data Acquisition (acquisition parameters, settings, gating strategies, compensation matrices, and any data pre-processing steps applied during acquisition). Data Analysis (gating strategies, data transformation, statistical analysis, and any specific algorithms or software used for data processing and visualization. Results, Standardization and Quality Control	Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International
<b>Social surveys</b>	DDI <sup>11</sup>	Study Description, Documentation (protocols, data collection manuals, codebooks, questionnaires, and any other relevant documents that provide additional information about the study design and data collection procedures), data collection (sample size, data collection mode, data collection timeline, and any training provided to interviewers or survey administrators), data variables, data structure, data quality, data access and use conditions, and publications	Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International

### 3.2. Making data accessible

#### Will the data be deposited in a trusted repository?

The official repository for data and software codes generated by the Soil O-live project will be located on ZENODO, stored in an open file format and with a persistent identifier (DOI) under an open license. Final data, knowledge except, if any of the information has potential value for exploitation, and indicators will also be available on the ESAC repository through a direct link from the Soil O-live web pages. Samples, raw reads, assembled transcripts, and genome

<sup>8</sup> <https://www.ebi.ac.uk/ena>

<sup>9</sup> <https://www.ebi.ac.uk/eva>

<sup>10</sup> <http://flowrepository.org/>

<sup>11</sup> <https://ddialliance.org/>

assemblies will be submitted to BioSamples and the European Nucleotide Archive. VCF files containing variation data mapped to reference sequences will be submitted to the European Variation Archive. In collaboration with Ensembl, the genome data and the mapped variants will be available, together with regularly updated comparative genomics data, in the Ensembl genome browsers. The Soil O-live proposal for collaboration in genomic bioinformatic analyses and the development of open genomic data resources has been explicitly endorsed by EMBL's European Bioinformatics Institute.

**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.**

If deemed to have potential exploitation value (especially for the novel, cutting-edge prototype built by UCLM), the knowledge generated will be protected by patents, copyright, or other means wherever appropriate.

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

The data access will be free of restrictions, but respecting at all times the specific licenses of each dataset.

**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)?**

Source code and documentation for data analyses (in languages such as R, Python, etc.) will be available in Zenodo and other specific repositories such as Github, CRAN, etc.

### **3.3. Making data interoperable**

**What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines?**

See table 2.

**Will you follow community-endorsed interoperability best practices? Which ones?**

Here we provide five ways to improve the quality and interoperability of Soil O-live data:

1. Use standardized data formats (refer to Table 1).
2. Comply with metadata standards (refer to Table 2).
3. Use standardized identifiers, such as DOI for scholarly publications and datasets or ORCID for researcher identification. This enables the unique and persistent identification of resources and individuals.
4. Adopt linked data principles based on Uniform Resource Identifier (URI). This allows for the interconnection and integration of diverse datasets across the web, enabling data discovery, exploration, and aggregation.

5. Open Standards and Open Data: Embracing open standards and open data practices promotes interoperability by removing barriers to data access, encouraging data sharing, and enabling collaboration among different stakeholders.

**Will your data include qualified references to other data (e.g., other data from your project)?**

Yes, it is absolutely necessary to accomplish the SEM model stated in the proposal. Please refer to Figures 4 and 5 in the GA for more information.

### **3.4. Increase data re-use**

**How will you provide the 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.)?**

The Soil O-live project's methods, procedures, and protocols are available to the public, and they will be accessible through the project's webpage, which will link to permanent repositories (e.g., Zenodo).

**Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?**

The data will be freely available for the public by using a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International license.

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

Yes, if open dataset-specific licences are respected at all times.

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

Yes (explained in Table 2).

**Describe all relevant data quality assurance processes.**

Key data quality assurance processes:

#### **1. Data Collection Process:**

- The data collection process at Soil O-live follows reliable and standardized methods, protocols, and guidelines that have undergone evaluation and agreement. Any consortium member can access these standardized protocols and are open for discussion should any issues arise during sampling. Protocols are revised in each semestral in the ordinary consortium meetings.

- The chosen teams of the consortium have the necessary skills to carry out all the proposed samplings and surveys using common protocols. However, training sessions will be provided for certain tasks, such as the ant taxonomy workshops in September 2023. These workshops aim to ensure accurate taxa classification.

## 2. Data Validation and Cleaning:

- Before any analysis, the person in charge of gathering data, under the supervision of task leaders, thoroughly reviews all data sets to check for errors, inconsistencies, and outliers. The consortium member can access and review their own data anytime to detect inconsistencies or errors. To clean and correct data, we use procedures such as removing duplicates, handling missing values, and resolving discrepancies. A general supervision will be performed by the responsible for task 7.1 (Data communication and interoperability assurance) and task 7.2 (Standardization activities)

## 3. Data Documentation and Metadata:

- The task leader's responsibilities include validating the dataset, creating and maintaining metadata, and preparing all necessary documentation according to the standards in Table 2. Taskleaders will oversee documentation and metadata quality control, ensuring data integrity through range checks, consistency checks, and logical validations. WP leaders will communicate with task leaders to resolve inconsistencies in case of discrepancies or errors. A general supervision will be performed by the responsible for tasks 7.1 (Data communication and interoperability assurance) and 7.2 (Standardization activities).

## 4. Regular Audits and Reviews:

- Regular audits or reviews of the data quality assurance procedures will be performed by tasks leaders to monitor and evaluate data quality to identify areas for improvement and maintain data integrity. A general supervision will be performed by the responsible for task 7.1 (Data communication and interoperability assurance) and task 7.2 (Standardization activities)

## **4. Other research outputs**

Other research outputs expected throughout Soil O-live project are:

- Agreed Protocols, guidelines, and procedures will be managed similarly – in line with the FAIR principles – as the data detailed in this document.
- Prototype for the chemical restoration of polluted soils in olive groves.

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

## **5. Allocation of resources**

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

Soil O-live will use OpenAIRE's permanent open scholarly communication infrastructure, including storage and archiving (e.g, Zenodo). This implies that data storage and archiving will

be free of charge. To ensure security, we will use external hard drives as a backup. We will purchase these devices using funds allocated for small equipment items. Additional virtual space may be acquired if necessary for data backup.

**Who will be responsible for data management in your project?**

Soil O-live coordinator assisted by the whole consortium members.

**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).**

Data will be stored and preserved permanently in the repositories such as Zenodo.

## **6. Data security**

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

We will perform data backup weekly through two methods: (i) using an external hard drive device and (ii) creating backup copies of data in virtual spaces like Google Drive®. UJA, our coordinator institution, already has free institutional accounts on Google Drive® with several spaces for data archiving (200 Gb each). Each Taskleader will also create additional data backups locally for added security.

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

Yes.

## **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 the ethics chapter in the Description of the Action (DoA).**

N/A

**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)?**

N/A





## ANNEX I. DATA COLLECTION TEMPLATE (one sheet for type of data)

Partner:
TYPE OF DATA:
DATA DESCRIPTION:
Relationship to WP and Tasks:
Source (observational, experimental, simulation, pre-existing data, etc.)
Form and format of the data (text, numeric, audio-visual, computer-software codes, instrument-specific, etc.):
Software/Program to read the data:
Size of the data (and estimation of the extension in bytes along the project):
Metadata standard (see Annex II):
Keywords:
Dissemination Level / Data Sharing (public, private, embargo):
Licence for the dataset (CC0, PDDL, CC-BY-4.0, ODbL, Other):
Other research outputs (software, code, protocols, new materials, samples):
Funding for data management:
Data security (including data recovery as well as secure storage/archiving and transfer of sensitive data):

Ethics aspects:
Other aspects (use of other national/funder/sectorial/departmental procedures for data management):