



## D6.4 Data Management Plan

Deliverable title	Data Management Plan
Version	5
Deliverable Lead	KU Leuven
Related Work Package	WP6
Author(s)	Marc Rosiers

Project acronym	SoilValues
Project title	Enhancing Soil health through Values-based business models
Project number	101091308
Call identifier	HORIZON-MISS-2021-SOIL-02
Project start date	1 January 2023
Project duration	48
Project end date	31 December 2026
Deliverable due date	30/06/2023
Actual delivery date	30/06/2023
Lead	KU Leuven
Author(s)	Marc Rosiers
Contributors	Kato Van Ruymbeke (KU Leuven), Bert Smit (WR), Els Lemeire (EV ILVO)
Reviewers	Erik Mathijs (KU Leuven), Helmut Saatkamp (WU)
Dissemination level	Public (PU)
DOI	10.5281/zenodo.14204759

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them. Grant Agreement: 101091308

## Table of content

1	Introduction.....	6
2	Data Summary .....	7
2.1	Purpose .....	7
2.2	Research data .....	7
2.3	Management data .....	9
3	Fair data.....	10
3.1	Making data findable, including provisions for metadata.....	10
3.2	Making data accessible.....	10
3.3	Making data interoperable .....	11
3.4	Making data reusable .....	11
4	Allocation of resources .....	12
5	Data security.....	13
6	Ethics .....	14
7	Acknowledgements .....	15

# List of Tables

Table 1. Overview of the type of data collected within each WP ..... 8

## Summary

The Horizon Europe Model Grant Agreement requires that a data management plan (DMP) is established and regularly updated. We used the template which is recommended for Horizon Europe beneficiaries. In completing the sections of the template we address the requirements for research data management of Horizon Europe as described and analysed in the Annotated Grant Agreement, article 17. The Data Management Plan will contribute to the research and management data collected or generated in the course of SoilValues work and how these data are going to be stored, published, cited and made FAIR (findable, accessible, interoperable and reusable) beyond the project life. Our goal is to meet the requirements of excellent scientific practice and to allow for accessibility, interoperability and reproducibility of SoilValues research results. All operational costs of data production/collection and intra-organisation management are part of the budget allocated to each consortium beneficiary.

## Acronyms and abbreviations

CO	Confidential
CoP	Community of practice
DOI	Digital Object Identifiers
DMP	Data Management Plan
EC	European Commission
ESS	Ecosystem Services
EU	European Union
FADN	Farm Accountancy Data Network
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
GGN	Global Gap number
GLN	Global Location Number
GP	Good Practice
HE	Horizon Europe
KPI	Key performance indicator
MRV	Monitoring, reporting and verification
OA	Open Access
PID	Persistent identifiers
PU	Public
SHBM	Soil health business models
SESS	Soil-based ecosystem services
TG	Testing Ground
WP	Work Package

## Project consortium

No.	Participant organisation name	Country
1	Katholieke Universiteit Leuven (KU Leuven)	BE
2	Eigen Vermogen van het Instituut voor Landbouw- en Visserijonderzoek (EV ILVO)	BE
3	Stichting Wageningen Research (WR)	NL
4	Wageningen University (WU)	NL
5	European Landowners' Organization (ELO)	BE
6	Consultoria Agroindustrial (CONSULAI)	PT
7	Aarhus Universitet (AU)	DK
8	KOIS Invest (KOIS)	BE
9	MR F&A Consult (MFRA)	BE
10	Instytut Rozwoju Wsi i Rolnictwa Polska Akademia Nauk (IrWiR PAN)	PL
11	Thuenen Institut (THUENEN)	DE
12	Udruzenje Eko-Inovacija na Balkanu (ABE)	RS
13	Institute Navarro de Tecnologias e Infraestructuras Agroalimentarias (INTIA)	ES
14	Lietuvos Misko ir Zemes Savininku Asociacija (FOAL)	LT

## Overview of the SoilValues project

SoilValues: Enhancing Soil health through Values-based business models (HORIZON-MISS-2021-SOIL02-05)

Project duration: 1 January 2023 – 31 December 2026 (48 months)

Total project budget: € 4 999 922.50

EU Grant: € 4 999 922.50

Land managers combine man-made resources with natural resources to produce marketable products like food, feed, fibre and wood, but at the same time produce ecosystem services that are generally not marketed or compensated. However, land managers generally have little incentive to invest in healthy soils, as they cannot sufficiently capture the value generated by these ecosystem services. SoilValues aims to contribute to the conditions for developing successful soil health business models. These are models in which land managers make production decisions that result in higher levels of soil-based ecosystem services (SESS) and in which they are paid for the non-marketed services they generate. In order for such business models to function, three important conditions need to be fulfilled: (1) the outcomes of SESS need to be measured, thus requiring knowledge, indicators and models, (2) the data and information generated by these indicators and models need to be exchanged to facilitate monitoring, reporting and verification (MRV), and (3) all these activities should be governed by an appropriate institutional framework consisting of the necessary legislation, standards and incentive schemes. To enhance the conditions for developing successful soil health business models, SoilValues will: (1) provide a comprehensive assessment framework addressing all factors influencing the development of business models for investing in soil health, (2) establish 6 testing grounds across Europe to test and improve emerging and designing new soil health business models, (3) establish 12 communities of practice of land managers, value chain actors, investors and public authorities for soil health business models, (4) design a comprehensive toolbox of incentives and policy recommendations to facilitate soil health business models and (5) raise awareness and exchange knowledge for soil health business models.



## 1 Introduction

A well devised data management plan is essential for designing and producing a high quality research project. The data management plan covers every stage of the data lifecycle and provides guidelines and procedures to use datasets generated in the project. The plan specifies which type of repositories the partner institutions will use, and which procedures will be put in place for long-term and safe preservation of the data.

For SoilValues, task 6.2 (WP6) is responsible for the development of a data management plan (DMP), a key element of good data management. The DMP will ensure that partners comply with EU Regulation 2016/679 and includes guidelines for task managers to support the archiving and storing of data in accordance with project protocols and the DMP.

The data manager (KU Leuven, supported by WR) is responsible for setting up the DMP and for ensuring that all partners are aware of, educated in and adhere to the FAIR principles of data management. The data administrator (KU Leuven), collaboratively with all consortium partners, will ensure the relevant data, including sound metadata, are created, systematically stored, accessible to all partners during the project and archived conveniently to be reused by other researchers or citizens after the project.

This DMP establishes the underlying principles, methods and standards to direct the data collection/generation in relation to the objectives of the multiple components of the project. The DMP is intended to be a living document that will be updated as the implementation of the project progresses and when significant changes occur.

## 2 Data Summary

### 2.1 Purpose

The purpose of data collection is strongly linked with the objectives and the planned outputs of the SoilValues project.

Land managers combine man-made resources with natural resources to produce marketable products like food, feed, fibre and wood, but at the same time produce ecosystem services (ESS) that are generally not marketed or compensated.

However, land managers generally have little incentive to invest in healthy soils, as they cannot sufficiently capture the value generated by these ES. Through the testing and validation of a science-based and easy-to-use assessment framework SoilValues aims to contribute to the conditions for developing successful soil health business models (SHBMs). These are models in which land managers make production decisions that result in higher levels of soil-based ecosystem services (ESS) and in which they are paid for the non-marketed services they generate.

SoilValues involves carrying out data collection, including personal data and metadata in a set of case studies across Europe to assess the relevance and effectiveness of the proposed framework in real life conditions.

### 2.2 Research data

SoilValues will collect data from real-world environments and aggregate it into information and knowledge as well as other data in several types such as numerical data, texts, images, tables and other formats such as recordings. SoilValues will establish six testing grounds (TGs) across Europe to test and improve emerging and designing new SHBMs and establish twelve communities of practice (CoP) of land managers, value chain actors, investors and public authorities for SHBMs. During the project it is estimated that data will be produced specifically from the case studies in the TGs and during the CoP workshops with the stakeholders. The produced data will be enhanced and combined with the data acquired from existing repositories, GIS/EO data, unique identifiers (GGN, GLN) in order to analyse and aggregate the data as applicable. This will specifically require special attention with respect to the pseudonymisation of data and limitation of public access to critical data. However, for wider publication data will be abstracted in a way that core messages can be communicated, not revealing trade secrets and personal identifiers.

SoilValues will also use public FADN and Eurostat data for e.g. benchmarking, in compliance with the data policy applied by the European Commission and the different Member States.

More specifically

- Primary data on soil health business models will be collected by the consortium members including:
  - o raw data, that will be collected through various means, including a survey of farmers and other stakeholders; qualitative interviews of farmers and other stakeholders (e.g. processors, policy makers, investors, ...); and the structured consultation of stakeholders that will take place in stakeholder workshops. All these collected data are expected to be personal data, except when opinions only are collected without being linked to respondents' characteristics. Much of this data will be collected across the case studies;

- generated data, that will be computed on the basis of the collected raw data, namely various indicators or proxies. Such data will be generated at the individual (farmers' or other stakeholders') level and will therefore fall within the personal data category if they are linked to respondents' characteristics. They may also be generated at an aggregated level, e.g. regional level.
- Secondary data will be collected from multiple sources, including public databases, surveys, reports, scientific studies and other European projects, and across multiple levels: European, national, regional and local.

Table 1. Overview of the type of data collected within each WP

Work package	Description
WP1	WP1 will create an assessment framework for SHBMs. Data on how TGs provide which kind of ES (commodities, services, ...) along with the farm practices used as the basis for their SHBMs. As these ESS can be regarded as activities of the SHBM, we will collect economic data on the farm business cycle and structure. Moreover a soil indicator database will be included in order to assess if the current soil health level is sufficient to provide the current ES.
WP2	WP2 will establish six testing grounds in 6 EU member states to test and improve emerging and designing new SHBMs. Data to contact TG stakeholders: name of contact person, telephone, email, website, ... Descriptive information: any information related to the soil health case i.e. general description (sector, region, farming type,...), description of the soil health practices and purpose, involved actors, initiator(s), other stakeholders, resources and rewarding mechanisms. Soil health data and accounting data.
WP3	WP3 will establish twelve communities of practice for SHBMs. Collection of data on the selected SHBMs can be gathered through interviews, surveys within the CoPs.
WP4	WP4 will design a comprehensive toolbox of incentives and policy recommendations to facilitate SHBMs. Collection of data related to incentives can be gathered through interviews, surveys and/or workshops.
WP5	WP5 will increase awareness of the importance of SHBMs towards targeted audiences including the general public. Developing communication and dissemination materials.
WP6	Meeting Minutes

To avoid losing valuable information during interviews or other research methods, stakeholder interactions will often be (audio or video) recorded. In particular, SoilValues aims to produce animated videos of the scaled-up CoPs in WP3. For data security concerns we refer to section 5.

When research (primary or secondary) data will encompass personal data, only pseudonymous data will be centralised and made available for use by SoilValues partners. Next, each SoilValues partner will commit to complying with the EU and national legal rules governing the use, transfer, storage and erasure of secondary data. Such data will be obtained from authorities or organisations that are knowledgeable as regards the rules on data protection and that will transfer data to SoilValues partners in compliance with these rules.

## 2.3 Management data

Management data in SoilValues, including contact details of persons, will be collected and used during the project for management (in WP6 Management) and communication (in WP5 Communication) purposes.

More specifically, management data include:

1. the project's management data collected and processed by KU Leuven and MR F&A Consult (for WP6) and any other partners, such as contact details and other personal information of SoilValues consortium partners, including personal data collected for organising consortium meetings or WP workshops;
2. the project's communication data collected and processed by ELO (the partner responsible for WP5) for the project's communication and dissemination activities, such as contact details of stakeholders targeted by these activities;
3. specific communication data collected and processed by any SoilValues partner for the CoP workshops and local communication and dissemination activities of WP3, such as contact details and other personal information of stakeholders included in these activities;
4. human resources data from each partner.

Specific rules will be enforced as regards data collection, management, processing, storage and erasure, whether they are primary data, secondary data or management data.

When management data processed during the project contain personal data, the application of the EU General Data Protection Regulation (GDPR) 2016/679 will be strictly followed. The five principles relating to processing of personal data (i.e. Purpose limitation, Accuracy and data minimisation, Storage limitation, Security, Integrity and confidentiality, Rights of the data subject) will be accordingly addressed. This is explained in the ethics section (section 6) of the DMP.

## 3 Fair data

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

The findability of the data will be succeeded as 1) the data will have Persistent identifiers (PIDs) (Digital Object Identifiers (DOIs)) which are important because they unambiguously identify the data and facilitate data citation. Furthermore, the data will be deposited in trusted repositories like Zenodo for the general public and Sharepoint and Adagio for the project partners, which assigns DOIs and 2) the data will have rich metadata which will support findability, citation and reuse. Rich metadata will provide important context for the interpretation of the data and make it easier for machines to conduct automated analysis. Standard metadata schemes will be followed. Indicatively the following schemes are presented: Dublin Core, CERIF, DDI (general data) and EML (specific for ecology disciplines), ISO 19115 and FGDC-CSDGM (for describing geospatial information), AgMES (Agricultural information), CDISC (clinical and non-clinical data).

### 3.2 Making data accessible

The shared data will be deposited in an Open Data repository which will be identified through the platform of re3data (<https://www.re3data.org/>). Zenodo, DRAYAD, Harvard Dataverse or the project website (<https://soilvalues.eu/>) for general project data, and Sharepoint and Adagio for specific project data. Sharepoint is a cloud-based service, hosted by Microsoft, for businesses of all sizes. KU Leuven deploys and manages a SharePoint Server. Adagio is a cutting-edge metadata-driven data management solution operated by Wageningen Research (WR). Adagio leverages metadata to generate the necessary infrastructure and data workflow automatically. Adagio goes beyond traditional data archiving and management solutions by offering a comprehensive suite of features that encompass data discovery, quality assurance, knowledge capture, data wrangling, extensive data descriptions, ontology support, metadata-driven workflows, data visualization, flexible data linking, and integration of diverse data sources.

However, due to the sensitive data that will be managed during the project, it is scheduled to request a total or partial opting-out of some specific data as they will be incompatible with the need for confidentiality in connection with security issues and with existing rules concerning the protection of personal data (GDPR, NIS Directive) and trade secrets.

To secure the anonymity of the stakeholders that provide input to the SoilValues project, restricted access will be assigned to sensitive data collected and generated within the project.

Any information related to a natural person that can be used to directly or indirectly identify the person will be pseudonymized and encrypted (via 7-zip) before storage on the sharepoint site of KU Leuven / the Adagio platform of WR.

1. The obtained personal data (through interviews, workshops, surveys, ...) will be first pseudonymized as soon as possible by the SoilValues partner on protected and backed-up network disks of the partner organisations themselves in folders to which only the researchers involved in the SoilValues project have access. With pseudonymisation, raw data is processed until one can no longer link the pseudonymised data (i.e. data used in the analysis) to the data subject without using the data subject's additional data. The link between the identity data of the subject (names, specific occupational information, physical, electronic addresses, identification numbers, ...) and the pseudonym (i.e., identifying data) is stored in a separate file: the key file. The key file is highly protected and only accessible to a limited number of people.

2. When storing/processing data, identification and analysis data should be stored separately with a common code. In this way, access to the identification data can be strictly limited and monitored. Only the person, who has the key to the code, can retrieve the person behind the code number. The key is kept on protected and backed-up network drives of institutions themselves in folders to which only the relevant researchers of the SoilValues project have access. The key file should preferably be encrypted.
3. Files will be named based on the following elements: source country, date, data type, ID number and data language (e.g. BE20170125Interview001dutch).
4. A metadata template describing the nature and contents of the data will be provided. It is up to the partner responsible for the data to implement the template.
5. The analyses are conducted only on the basis of pseudonymised data. The pseudonymised data are stored on protected and backed-up network drives of the institutions themselves in folders to which only the researchers involved in the SoilValues project have access and in encrypted folders (e.g. via 7-zip) on SoilValues sharepoint site of KU Leuven / Adagio platform of WR (where it will be checked whether everyone needs access to pseudonymised data).
6. The password of encryption is communicated through another channel (e.g. phone).
7. The key file is encrypted with another password than the pseudonymized data in the encrypted folders on SoilValues sharepoint site of KU Leuven / Adagio platform of WR.
8. Personal data obtained in the context of research will be kept for up to 5 years after the study's closing date (or longer (until the legal retention obligation has expired) if this would be required by law) and destroyed afterwards. If applicable, when permitted by law or when explicit consent is given (via informed consent) or when permissions or protocols granted allow it, personal data will be kept for possible follow-up studies.

### 3.3 Making data interoperable

The produced data will use common formats and standards and community agreed schemes, controlled vocabularies, keywords, thesauri or ontologies where possible in order to be interoperable and be integrated with other data, applications and workflows. Indicatively, the data files generated will be processed and submitted to the repositories mentioned above with the following formats according to the data type: Containers: TAR, GZIP, ZIP, Databases: XML, CSV, Geospatial: SHP, DBF, GeoTIFF, NetCDF, Moving images: MOV, MPEG, AVI, MXF, Sounds: WAVE, AIFF, MP3, MXF, Statistics: ASCII, DTA, POR, SAS, SAV, Still images: TIFF, JPEG 2000, PDF, PNG, GIF, BMP, Tabular data: CSV, Text: XML, PDF/A, HTML, ASCII, UTF-8, Web archive: WARC.

### 3.4 Making data reusable

The generated data will be well-documented and they will have clear license and provenance information. Concerning the documentation, a README file to ensure that the data can be correctly interpreted and reanalysed by others will be created. This file will include i) a short description of what data it includes along with tables, figures, or sections; ii) for tabular data: definitions of column headings and row labels, data codes (including missing data) and measurement units; iii) any data processing steps that may affect interpretation of results; iv) a description of what associated datasets are stored elsewhere, if applicable; and v) contact information. Referring to the license issues, data will have a clear license to govern the terms of its reuse (e.g., Public Domain, Attribution, Non-commercial, No Derivatives, or other). Indicatively, for the open access data SoilValues will use the Creative Commons licenses and more specifically, the Attribution (CC BY) license and Creative Commons Zero (CC0) which is for dedicating works to the public domain.

## 4 Allocation of resources

Each SoilValues partner is responsible for collecting and storing their own data, whether primary research data, secondary research data or management data.

Each partner who will collect primary and secondary research data will transfer it to partner WR who is responsible for managing primary and secondary data for the whole project as part of WP1. Each partner with management data which can be shared, will transfer this to partner KU Leuven who is responsible for managing management data for the whole project as part of WP6.

Once the activities of the project are terminated, long-term preservation of data is preferred, where possible, to data destruction for historical and scientific interests of future generations.

- Nevertheless, secondary data provided by public or private sources may be governed by specific preservation or destruction conditions, and such conditions, agreed on with the owner of the data, will be strictly followed.
- Data that can be made open access will be deposited on a data repository as explained in section 3.2, the cost of which is born by the repository provider and not by SoilValues partners.
- Five years after the project has ended, primary data (including accompanying documents such as signed consent forms) will be destroyed or archived to comply with national laws. The place for archiving will be explored during the project and discussed among SoilValues partners.
- Management and communication data will not be made publicly available, unless based on explicit consent, and will be destroyed 3 years after the end of SoilValues or archived if requested by EU or national laws.

The costs incurred for data management during the project will be supported by each partner with the SoilValues budget (e.g. purchase of storage devices, training of data managers) or their own institutional budget (e.g. salary of dedicated people for data management).

## 5 Data security

Data storage will be performed in a secured form (e.g., data encrypted with a strong cryptographic protocol) in servers indicated by the case studies or the technology providers, and agreed within the consortium.

Raw data will not be shared because it is subject to privacy laws. Pseudonymised data will be shared with all members of SoilValues, but will not be made public. However, in accordance with Horizon policy (pseudonymised) data will be made available upon request.

Audio files are highly privacy-sensitive information and, in the context of data minimisation, should be made 'safe' as soon as possible by pseudonymising. After all, one should not process more data than is strictly necessary for one's intended purpose. Audio recordings are therefore pseudonymized within three months (e.g. transcription of interview where directly identifiable elements have been replaced by a code). Data minimisation ensures that the privacy risk of the data subject (in this case, interviewees/participants of focus group discussions) is reduced: after all, what you do not own, you cannot lose or end up in the wrong hands (unauthorised persons, hackers, ...).

Regarding video recordings, before finalizing and publishing, a consent document will be signed by whom is concerned (cf. ethics section of the DMP).

If necessary, data transfer to and from end-users (including transfer of sensitive data if allowed) is performed encrypted, either sent by encrypted ZIP or RAR files, or by download directly as web-based services from servers. In any case strong passwords are required for accessing transferred datasets and passwords must be sent separately from the dataset (preferably using also different channels of communication e.g. SMS, Viber, WhatsApp).



## 6 Ethics

All the activities carried out in the SoilValues project shall comply with ethical principles and relevant national, EU and international legislation, including the Charter of Fundamental Rights of the European Union, the European Convention on Human Rights and its Supplementary Protocols.

Particular attention shall be paid to the principle of proportionality, the right to privacy, the right to the protection of personal data, the right to the physical and mental integrity of a person, the right to non-discrimination and the need to ensure high levels of human health protection (Horizon Europe Framework Programme Regulation 2021/695, Article 19).

Both online and live interviews will be conducted in accordance with the EU Law (no.97/2008, 104/2009, 68/2012 and 107/2012). The prior information will be provided to the interviewees in accordance with the Article 15 and their consent (i.e. authorisation to collect, process, use data, preserve on a long term and share) will be asked for: in writing in case of oral interviews, and by clicking an “I Agree” button at the bottom of the page in case of online questionnaires, which will contain all the information included in the informed consent form and the information sheet (described in more detail in D6.5 Ethics Requirements).

## 7 Acknowledgements

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them. Grant Agreement: 101091308