Organisation Wageningen University
Main Authors Abhishek, Abraham; Rezaei, Mahrooz; Ritsema, Coen; Steinbuch, Luc

SOILPROM

Data Management Plan

Deliverable 7.4

Date: 27/02/2025
Doc. Version: V1.1

Document Control Information

Settings	Value
Deliverable Title	Data Management Plan
Work Package Title	Project management and ethics
Deliverable number	7.4
Description	The data management plan compiles all rules with regards to management of data and measures to comply with the FAIR principles
Lead Beneficiary	Wageningen University
Lead Authors	Abhishek, Abraham; Rezaei, Mahrooz (WU)
Contributors	Ritsema, Coen; Steinbuch, Luc (WU)
Submitted by	Wageningen University
Doc. Version (Revision number)	V1.1
Sensitivity (Security):	Public
Date:	28/02/2025

Document Approver(s) and Reviewer(s):

NOTE: All Approvers are required. Records of each approver must be maintained. All Reviewers in the list are considered required unless explicitly listed as Optional.

Name	Role	Action	Date
Liu, Jian (NIBIO)	Reviewer	*Approve*	19/02/2025
Bajard, Manon (NIBIO)	Reviewer	Approve	19/02/2025

Document history:

The Document Author is authorized to make the following types of changes to the document without requiring that the document be re-approved:

- Editorial, formatting, and spelling
- Clarification

To request a change to this document, contact the Document Author or Owner.
order (latest version first).

Revision	Date	Created by	Short Description of Changes

Configuration Management: Document Location

The latest version of this controlled document is stored in the project SharePoint Folder 'Submitted Deliverables' and in the Document Library of the EU Funding and Tenders Portal.

Nature of the deliverable		
DMP	Data Management Plan	

Dissemination level		
PU	Public	

ACKNOWLEDGEMENT

This report is part of the deliverables from the project "SOILPROM" which has received funding from the European Union's Horizon Europe Research and Innovation program under grant agreement No 101156589.

More information on the project can be found at: http://www.soilprom.eu.

SOILPROM

TABLE OF CONTENTS

I. INTRODUCTION

The overall objective of SOILPROM is to deliver upgraded and integrated models for simulating transport and fate of soil pollutants. The models help assess the impact of soil pollution on soil functions and related ecosystem services, by considering a selection of pollutants that pose a high risk to the environment and to humans.

The updated and integrated models will be able to better support a wide range of stakeholders in implementing sustainable land management strategies and formulating policies for healthy soils all over Europe. In order to facilitate them to do so, an open access Modelling Platform (MP) and a Decision Support Tool (DST) will be developed in course of the project.

This Data Management Plan summarizes the data that will be collected during the project, and outlines the envisioned data sources and collection purposes. It describes the project's approach to compliance with the FAIR principles of data management, allocation of resources for that purpose, data security measures, and ethical considerations. The plan will be updated twice - in months 24 and 48. The updates will document the state of affairs with respect to data management mid-way and at the end of the project, highlighting and explaining any deviations from the original plan in course of its implementation.

2. DATA SUMMARY

2.2.1. DATA OVERVIEW AND ORIGIN

The aforementioned models will be updated by applying them in seven use cases of soil pollution; located in The Netherlands, Germany, Spain, Norway, Poland, and Belgium. They will cover five distinct pollutants (microplastics, pesticides, metals/metalloids, PFAS, and nutrients such as phosphorus and nitrogen) and 13 processes through which they propagate, namely:

1. Colloidal transport of microplastics through the soil
2. Metals biogeochemistry in the soil
3. Sorption and desorption of phosphorus
4. Wind erosion, atmospheric transport and deposition of dust and particulate phase of pesticides
5. Wind erosion and atmospheric transport and deposition of dust and microplastics
6. Atmospheric PFAS inputs to soil
7. Adsorption and transport of PFAS

8. Wind erosion and atmospheric transport and deposition of dust polluted by metals
9. Water erosion and run-off transport of dissolved and particulate phase pesticides

10. Hydrological transportation of metals by water erosion, runoff, and infiltration into soils
11. Plant uptake of PFAS
12. Flow of water and transport of pesticides in soils and groundwater
13. Transport of nutrients in soil, groundwater, and surface water with marine discharge

The data to be collected and fed into the models can be classified under the following three categories, corresponding to three work packages:

1. Existing data from the use cases to identify knowledge gaps and characterise the background information for the seven use-cases, the models under consideration; and pollutants, transport mechanisms, and management gaps in general (Work Package 1): This will include secondary data collected between months 1 and 8 of the project, in order to:
    a. investigate the current state of knowledge and identify knowledge gaps in understanding soil pollution processes
    b. define use cases; and define needs of end users of the modelling platform and the decision support tool
    c. define relevant existing databases (corresponding to Task 1.3)
    d. prepare for modelling

2. Experimental and technical data related to pollutant behaviour in the use cases (Work Package 3): This pertains to the primary and secondary data to be collected for the purposes of demonstration, validation, and optimization of the selected models.
   Examples include soil samples analysis, meteorological data, water-level and pollutant concentration data.

3. Data from questionnaires and surveys targeted at local stakeholders pertaining to the identified use cases (Work Package 5): All personal data gathered through these surveys will be anonymised and treated to be compliant with European GDPR standards (Regulation (EU) 2016/679)

The seven use cases will be defined and specified in accordance with the scope and focus of the project, as well as the needs and priorities of stakeholders, and identified knowledge gaps. This process is being guided by the 'Methodology and Data Collection Framework' which was prepared and validated by all consortium partners in month 3 (Milestone 1). Information related to environmental contexts and local/regional/national policies will be collected for the use cases by month 8. The work will inform what kind of data are already available and what else will be collected and fed into the models, and what kind of data will be generated by the models as output. At present (month 6), what is known is the range of data that the models can possibly be fed, and the outputs they are capable of generating. These are tabulated in Table 1.

Table 1: Models under consideration by SOILPROM, their data needs and outputs

	Model Name	Pollutants covered	Processes/ sub-processes covered	Data input	Output data
1.	HYDRUS	- Microplastics - Nutrients - PFAS - Metals	- Simulating water, heat, and solute movement in 1, 2, and 3-dimensional variably saturated media	- Transport geometry - Domain properties - Initial conditions - Boundary conditions - Flow and transport parameters	- Pressure heads - Water contents - Water fluxes - Solute concentrations - Solute fluxes - Temperatures - CO2 concentrations - Neutron fluxes - Water and solute mass balances
2.	ICECREAM	- Nutrients	- Field level phosphorus loss	- Soil physical and chemical properties for different layers - Weather inputs: temperature, precipitating, humidity, wind speed, and cloudiness or solar radiation. - Land and crop management inputs: type of crop, yield, fertilization, soil tillage operations, and dates for the different actions	- Soil: values of different P pools. - Crop: nutrient update and yield. - Water: flow and concentrations of different P forms in surface runoff and subsurface drainage. - Also possible to get N and C results but the model has not been tested for N.

3.	IFDM	- PFAS	- Atmospheric transport: calculation of pollutant dispersion in the atmosphere based on meteorological data	- Meteorology data - Emission and source characteristics - Receptor grid parameters - Background concentrations	- Concentration and deposition at receptor locations
4.	MicroHH	- Microplastics - Pesticides - Metals	- Wind erosion and atmospheric transport	- Meteorology data - Land-use data - Soil properties	- Emissions, airborne concentrations, and deposition of pollutants - Meteorological parameters (temperature, wind, et cetera).
5.	MODFLOW_ MT3D_SWAT	- Nutrients (Nitrogen, Phosphorus)	- Pollutant transport in groundwater	- Digital elevation - Land use data - Soil data - Weather data - Aquifer data - Atmosphere forcing - River discharge	- Streamflow - Organic nitrogen - Organic phosphorus - Phosphate - Nitrate - Pesticide - Groundwater head - $N\text{-}NO_3$ concentration - Nutrient spread - Hydrodynamic data - Biochemical data
6.	OPENLISEM	- Pesticides - Metals	- Water erosion and runoff transport of dissolved	- Precipitation events	- Timeseries at outlet - Hydrograph

			and particulate phase pesticide and metals	- Maps describing catchment topography and land use - Soil characteristics maps - Line elements: channels, roads - Pesticide concentration (upper soil)	- Sediment load - Pesticide transport - Maps with spatial results of catchment processes - Erosion - Infiltration - Pesticide redistribution
7.	PEARL/ GEOPEARL	- Pesticides - PFAS - Microplastics - Organic pollutants	- Transport of organic pollutants in soil to groundwater and to surface water, uptake by plants, volatilisation to atmosphere	- Groundwater depth - Soil hydraulic properties - Irrigation + drainage systems data - Weather data - Crop properties - Soil properties - Substance application method and timing - Substance properties	- Water content, pressure heads - Water fluxes - Pollutant concentrations/ masses - Mass fluxes
8.	SWAP	- Pesticides - PFAS	- Field scale water flow and solute transport in variably saturated, cultivated soils - Adsorption and desorption PFAS - Plant uptake of PFAS	- Daily potential evapotranspiration - Rainfall/ irrigation data - Soil hydraulic properties - Drainage conditions - Groundwater levels - Water content measurements - Soil water pressure heads	- Actual crop development - Water and solute balances of cultivated soil - Water and solute fluxes of cultivated soil - Soil profile data on moisture and solute concentrations

				- Groundwater levels - Drainage fluxes - Light use efficiency - Maximum assimilation rate - Sensitivity to water and salt stress - Partitioning assimilates to plant organs - Senescence - Development stages: emergence, start flowering, full grown, harvest - Leaf area index or soil cover during the growing period - Rooting depth during the growing period - Crop yield - Initial pesticide concentrations in the soil - Amounts of pesticide applied - Pesticide adsorption isotherm - Pesticide decomposition rate - Pesticide uptake by plants	- Crop growth reduction due to unfavourable soil conditions

				- Pesticide concentrations in the root zone - Pesticide concentrations in drainage water	

2.2.2. DATA UTILITY

SOILPROM's strategy to ensure a high degree of visibility of the data generated within the project, and for the data to become (and remain) useable, hinges on (a) effectively engaging end users of project results throughout the project, (b) the exploitation and dissemination measures, and (c) adherence to Open Science practices.

- Effective engagement of end users: Use-case leaders will perform a stakeholder mapping at two levels:
    - (1) for MP and DST users, through a Modelling Board consisting of scientists, agencies, European organisations (e.g., JRC, ISMC, EUMETSAT, ESA, ECMWF) and consultancy firms (created at M8), and
    - (2) for use-case stakeholders, organised in 7 local stakeholders networks consisting of local policy makers, land-users (i.e., farmers, foresters, landowners, advisors), urban planners, and local companies.

  Their needs and requirements will be gathered at the beginning of the project (until month 12) through a survey to gather MP and DST users (task 5.1), and through a workshop (under task 5.2) to gather the needs of local stakeholders. Both the Modelling Board and the seven local stakeholders' networks will be setup at the beginning of the project to ensure proper and targeted engagement activities.

  SOILPROM considers EU policymakers to be a key stakeholder group. For their benefit, recommendations for use-cases will be derived from selected scenarios which present the best trade-offs for reduced levels of pollution. EU policy recommendations will be based on rigorous policy research carried out in SOILPROM. To this end, various legal instruments that exist or can be implemented to prevent soil pollution and related biodiversity loss will be compiled, systematised, and evaluated.

  Continuous engagement of stakeholders- starting right at the beginning of the project- will ensure that the data and results generated are in line with their needs, and that they are visible to potential end users among the stakeholders.

- Exploitation and dissemination measures: A range of activities will be carried out to facilitate the exploitation of project results and dissemination of its activities and outputs. These will include workshops at use-case level to present and explain recommendations to the associated stakeholders, and a webinar targeted at EU policymakers. Additionally, dissemination activities will be carried out regularly throughout the project, as detailed in the Plan for Dissemination, Exploitation and Communication (Deliverable 6.1).

- Open Science Practices: SOILPROM is committed to implementing open science practices that align with the general principles of the Horizon Europe Work Programme.

The project will share knowledge and tools as early and widely as possible in an open and cooperative way, and engage citizens, civil society, and end users where appropriate. To ensure responsible management of research outputs, the partners will adopt the 'gold' open access model, making scientific publications immediately accessible online free of charge or by having each affiliated partner cover the relevant cost. In cases where the 'gold' access model is not applicable, the 'green' model will be used by supplementary publishing of relevant articles to an online repository in consultation with the publisher, in case an embargo period is needed. Partners will deposit technical and other professional information in the European Open Science Cloud (EOSC) and/or the European Soil Observatory (EUSO) to enable storage, sharing, processing, analysis, and re-use of research outputs (data and software) across borders and disciplines. The usage of EOSC and other trusted repositories will enhance the reproducibility of SOILPROM research outputs, limit waste of resources, increase the reusability and reliability of research, and thus strengthen European R&I. SOILPROM's Open Science strategy will promote innovation and transfer through reusable and reproducible results, enable increased quality and efficiency of research, and create more trust in science. The tools developed during the project i.e., the Modelling Platform and the Decision Support Tool will be hosted by the JRC's platform to ensure not only the openness but also the easy access to these tools.

3. FAIR DATA

This data management plan (and its updated versions) will be used as a handbook regarding the numerous aspects related to data management, and will be structured around the findability, accessibility, interoperability and reusability (FAIR) principles.

- Findability of data/research outputs: The data collected and outcomes generated in the project targeting a wider public will be made available through the SOILPROM website. The discoverability of data will be enhanced through two channels: project website and social media.

- Accessibility of data/research outputs: The access to sensitive raw data collected from stakeholders, interviews and similar primary data generated within the project will be limited to the partners in charge. Processed, finalised data will be made accessible through EU platforms such as DestinE and EUSO. Regarding the models used in the project, most of them are in open access with copyright-free source code. These models will be accessible through the Modelling Platform.

- Interoperability of data/research outputs: To allow data exchange between researchers, institutions, organisations, stakeholders, countries, etc., the project will assure that interoperable formats, like those supported by Microsoft Excel and Microsoft Word, will be used. Both for internal smooth collaboration and for comprehensibility by third parties when accessing the data and files collected and created by SOILPROM, meta data will be provided. Moreover, the data produced during the project will be tightly controlled by the

JRC to ensure the compliance with EUSO. Additionally, the upgraded models will be developed with the view to being reused for the Destination Earth initiative.

- Reusability of data/research outputs: Data generated by SOILPROM will remain re-usable by interested third parties after the end of the project and as soon as possible. In all cases, the consortium will have authorship of re-usable data and documents. Raw data collected in WP3 and WP5 will be gathered in databases that will remain stored by WU, VITO, FZJ, GUT, UPCT, NIBIO and SAV and on a shared online server (e.g., Zenodo). Third parties that want to use these data will be able to send a request to get access. If the work conducted is to be published, SOILPROM as the origin of the data will be acknowledged in the publication, along with specific project partners/ team members as applicable. Dissemination material will be open to re-use as soon as they are made available on the project website, for an unlimited period. During the project implementation, they will be downloadable free of charge from the website. Scientific publications based on SOILPROM work will be published with open access according following lines set in the Open Research Europe Guidelines on Open Access to Scientific Publications and Research Data in Horizon Europe programmes. Copyright-free models and modules that are developed in SOILPROM will be made open-access on the Modelling Platform.

4. Data Storage and Security

4.1. Data storage

Following are the broad plans with respect to storage of the three distinct categories of data identified in 2.2.1.

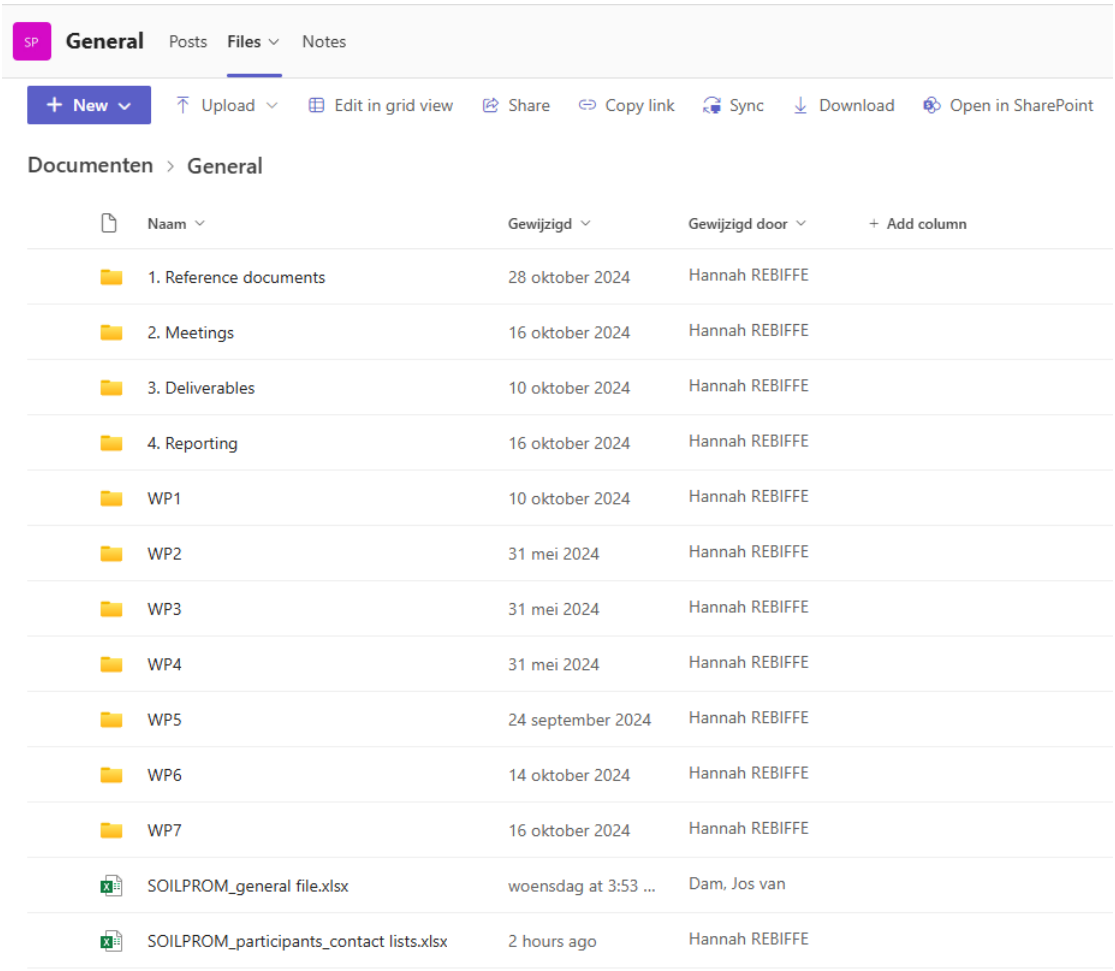
Table 2: Data storage outlook in SOILPROM

Data type	Format(s)	Storage location
1. Environmental data to characterize the eight use cases	Various	- SOILPROM Microsoft Teams space
2. Experimental and technical data related to pollutant behaviour in the use cases (modelling inputs)	Various (see table 1)	- Local computers, servers, networks of various partners (with backups created as per standards and practices of local system/network administrators)
3. Data from questionnaires and surveys targeted at local stakeholders pertaining to the identified use cases	.xls files	- In online survey systems until the survey is active - Survey results to be transferred to project Microsoft Teams space after survey is closed

In addition, datasets considered to be project results/outputs will be stored on the project website and various EU portals (see section 3- FAIR data).

For storage and sharing of day-to-day data within the team (documents such as agreements, meeting minutes, deliverables, reference documents, reports, etc.) the project will utilise a cloud-based data management system (Microsoft Teams) for the safe and accessible storage of data. This storage system is compatible with the data storage policies of all the SOILPROM partners. It is being managed by partner organisation Euroquality. Documents in the project Teams space are organised in folders corresponding to Work Packages; outside of which there are separate folders for basic reference documents (such as grant agreement and consortium agreements), project meetings, final deliverables, and reporting templates. Also included in the root folder are a project-wide contact list; and a spreadsheet called 'SOILPROM_general file' with various project management tools such as a deliverables and milestone tracker, a deviation tracker, the risk matrix, and the GANTT chart.

As data collection, processing, and modelling tasks commence in course of the project, separate folders will be created for (1) raw data (2) processed data, 3) outputs, such as graphs and tables, and 4) the eventual code used to process all the data.



A screenshot of the SOILPROM Microsoft Teams space

Data collected as input to the models, and raw data generated by the models, will be stored on local computers, servers and networks as per the standard procedures of partner organisations that will eventually be in charge of running the models. All batches and sets of data will be backed up. In each instance of data storage, it will be ensured that multiple SOILPROM team members have access to the raw/ processed data in question, as preparation for the possibility of one of them exiting the team unexpectedly. (Final arrangements in this regard will be documented in and shared through the updated Data Management Plan in month 24).

Individuals providing data to SOILPROM partners for the purposes of the project will have the right to self-determination, i.e., having the right to control their own data, knowing what information is stored and how that information will be used and communicated. No participants will be discriminated on the basis of ability, race or ethnicity, religion, upper age, gender, financial or marital status, or any other potential discriminatory factor.

The resources for making the DMP, updating it throughout the project and monitoring the data management are foreseen in the project budget in the form of PMs allocated to WU, which is responsible for the data management. The costs arising from long-term preservation of the project outcomes on the project website will be covered by EQY and by partner institutions for the data backed-up on their servers. Moreover, as explained in previous sections, arrangements will be made to make sure that the Modelling Platform and the Decision Support Tool will eventually be hosted by the JRC, in order to make sure that they are available long after the project is over (the JRC is already apprised of the project's plans to this effect. Discussions are underway to come to specific arrangements).

The Grant Agreement with the consortium partners lays down that the partners shall cooperate in order to enable one another to fulfil legal obligations arising under applicable data protection laws (the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and relevant national data protection law applicable to said Party) within the scope of the performance and administration of the Project and of this Consortium Agreement. In particular, the Parties shall, where necessary, conclude a separate data processing, data sharing and/or joint controller agreement before any data processing or data sharing takes place.

4.2. File naming convention

Naming convention for project reports will follow the pattern: SOILPROM_ Deliverable#_FileName_version So, for example, the first DMP is titled *'SOILPROM_D7.4_DMP_V1.1'*.

File naming conventions maybe updated as the project progresses to keep up with the demands of the project.

4.3. Guidelines for README files

Data folders will contain README files containing information about the dataset that will help ensure that the data can be correctly interpreted by all when accessing the folder, and when sharing or publishing data.

Best practices to be followed while creating README files will include:

- Creating one README file for each dataset
- Ensuring that the files is named 'README' (not 'readme', 'read_me', ABOUT, etc)
- The README document will be a plain text file (.txt)

Recommended minimum content for README files

1. Introductory information
    a. Title of the dataset
    b. For each file or group of similar files, a short description of what data it contains
2. Methodological information
    a. Method description for collecting or generating the data, as well as the methods for processing data, if data other than raw data are being contributed
3. Data specific information
    a. Full names and definitions (spell out abbreviated words) of column headings for tabular data
    b. Units of measurement
    c. Definitions for codes or symbols used to record missing data
    d. Specialized formats or abbreviations used
4. Sharing and access information
    a. Licenses or restrictions placed on the data

5. ETHICAL ASPECTS

The data security measures being taken by the project (detailed in the previous section) have been designed to minimize the possibility of legal issues arising in the course of data collection and sharing.

Ethical considerations have been codified into the project in WP7. This is guided by the European Code of Conduct for Research Integrity (2023). An Ethics Issues table (see below; as included in the Grant Agreement) guiding the overall ethics compliance of the project, will also guide ethical aspects of data collection and dissemination. Seeking and documenting the consent of respondents will be an integral part of the data collection process.

Compliance with the ethical principles and the applicable EU, international, and national laws for the ethics issues identified in the Ethics Summary Report and any additional ethics issues that may emerge in the course of the grant will be ensured. For any applicable ethics issue, the guidance provided in the European Commission Ethics Self-Assessment Guidelines will be rigorously followed.

Table 3: SOILPROM Ethics Issues Table

1. Human Embryonic Stem Cells and Human Embryos			Page
Does this activity involve Human Embryonic Stem Cells (hESCs)?	○ Yes	◉ No	
Does this activity involve the use of human embryos?	○ Yes	◉ No	
2. Humans			**Page**
Does this activity involve human participants?	◉ Yes	○ No	3
Are they volunteers for non medical studies (e.g. social or human sciences research)?	○ Yes	◉ No	
Are they healthy volunteers for medical studies?	○ Yes	◉ No	
Are they patients for medical studies?	○ Yes	◉ No	
Are they potentially vulnerable individuals or groups?	○ Yes	◉ No	
Are they children/minors?	○ Yes	◉ No	
Are they other persons unable to give informed consent?	○ Yes	◉ No	
Does this activity involve interventions (physical also including imaging technology, behavioural treatments, etc.) on the study participants?	○ Yes	◉ No	
Does this activity involve conducting a clinical study as defined by the Clinical Trial Regulation (EU 536/2014)? (using pharmaceuticals, biologicals, radiopharmaceuticals, or advanced therapy medicinal products)	○ Yes	◉ No	
3. Human Cells / Tissues (not covered by section 1)			**Page**
Does this activity involve the use of human cells or tissues?	○ Yes	◉ No	
4. Personal Data			**Page**
Does this activity involve processing of personal data?	○ Yes	◉ No	
Does this activity involve further processing of previously collected personal data (including use of preexisting data sets or sources, merging existing data sets)?	○ Yes	◉ No	
Is it plannec to export personal data from the EU to non-EU countries?	○ Yes	◉ No	
Is it plannec to import personal data from non-EU countries into the EU or from a non-EU country to another non-EU country?	○ Yes	◉ No	
Does this activity involve the processing of personal data related to criminal convictions or offences?	○ Yes	◉ No	
5. Animals			**Page**
Does this activity involve animals?	○ Yes	◉ No	
6. Non-EU Countries			**Page**
Will some of the activities be carried out in non-EU countries?	◉ Yes	○ No	34
NIBIO as Norway partner will perform activities as non-EU countries. NIBIO is leader of WP1.			

Human participants will be involved in the project through the participation to project events, surveys, workshops and conferences (all WPs). The project partner collecting the data will ensure that there is no sensitive information contained in it, which could be used to identify individuals. Free Informed consent will be requested and obtained from every participant participating in the events and workshops planned within the project. In order to minimize the potential for perspectives to be dismissed or disrespected within the context of cross-stakeholder groups, a terms of reference process will be undertaken at the initiation of each group. This will involve group members setting out the conditions under which they engage, and respect will be emphasized as a central theme within these agreements.

6. Contact Persons

The main contact person for SOILPROM for all matters with regards to data management will be Abraham Abhishek (WU), who is also the Project Manager. The setup, monitoring, periodic evaluation and execution of the data management plan will be carried out by a larger team which currently includes Luc Steinbuch, Mahrooz Rezaei, and Coen Ritsema from WU. One person from each partner organisation will soon be assigned the role of Data Officer for the purposes of the project. The SOILPROM data officers will be identified and listed in the deliverable *Data Management Plan (First Update)*, due month 24 (August 2026).