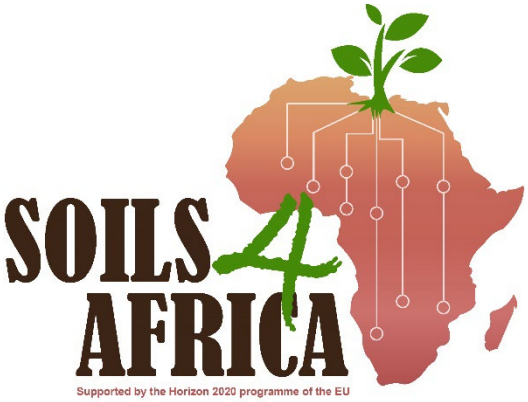

Soils4Africa Data Management Plan (Third Update)

Project No. 869200

Version 4.0

15 June 2023

(Deliverable: Soils4Africa_D1-1_DMP_v04)

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 869200

Contact details

Person in charge: Rik van den Bosch (Director)
Project coordinator: Mary Steverink-Mosugu
Address: Droevendaalsesteeg 3, 6708 PB Wageningen (Building 101), the Netherlands
Post: PO Box 353, 6700 AJ Wageningen, the Netherlands
Phone: +31 317 48 7634
E-mail: mary.steverink-mosugu@isric.org

Document details

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

Project Acronym	Soils4Africa
Project Name	Soil Information System for Africa
Grant Agreement Number	862900
Work Package	WP1
Deliverable	D1.1 Data Management Plan (Second Update)
Version	4
Filename	*Soils4Africa_D1-1_DMP_v04.pdf*
Dissemination Level	Public
Partner in charge	ISRIC
Contributing partners	WU, FARA, SZIU, ARC, IITA, I-BEC, SU, ICRAF, RCMRD , IFA-YANGAMBI, BUNASOLS, IRA , KALRO, SGS Hungária, JRC, and MetaMeta (see Table 1 for full names)
Author	Niels H. Batjes
Contributors	Mary Steverink-Mosugu, Bas Kempen, Oluwole Fatunbi, Erika Micheli, Jeroen Huising, Garry Paterson, Andrew Silas, and Rik van den Bosch
Due date	n/a (not a deliverable itself, for information project partners)
Revision date	15 June 2023

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

Version History

History of Changes		
Version	Publication date	Change(s)
1.0	15/09/2020	• Initial version
2.0	06/08/2021	• Added version history table • Added URL to project website • Added URL to 'Soils4Africa maps of agricultural land in Africa' • Reference to deliverable D1.2 for PEDR and Communication plan added • Reference to 'Use cases plus supporting soil quality indicators' added • Soils4Africa Review procedure for deliverables revised/added • Ethical aspects, added deliverables • Updates for WP3, design SIS, were added • INSPIRE compliance metadata added • Updated file formats • Updated Abbreviations • Added name of Data Protection Officers • Added a table for partners to fill in the name of their respective DPO's (Appendix 2). • Other issues, updated • Appendix 1 – List of project deliverables (up to 1 June 2021) • Processed feedback received from project partners (up to 19/07/2021) • List of DPO's resp. contact persons (App. 2) has been actualised based on responses up to October 2021
3.0	03/06/2022	• Processed feedback from EU Project Officer (04/2022): A precise definition of how data outputs are to be quality assured and approved for external release (see p. 12); Arrangements for data management

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

		after the current project ends (see p. 16-17) and how this management is scheduled to be funded (see p. 18). • List of deliverables up to June 2022 (App. 1) • Information on the field campaigns was added. • The need for documented/archived permissions from the various 'data owners', with a reference to the corresponding field protocol (see p. 16-17), to be stored in the SIS. • Updated the names and contact information for some DPO's (see App. 2). • Added information about the GeoSpatial Content Management System (GCMS) of the Soils4Africa SIS. An open-source web-based platform GeoNode will be used for the project (see p. 17) • All feedback received before 31 May 2022 has been processed in the current version of the DMP.
4.0	15/06/2023	• List of deliverables, and reports, updated up to June 2023 (App. 1) • Progress with field campaigns was updated. • Initial batch of samples was shipped to central laboratory (ARC, Pretoria). • Soil sample processing and analysis at ARC has started. • Initial tests in importing soil data into the evolving SIS have been carried out. • New open-source web-based platform (TerriaJS) proposed instead of GeoNode. • The contact person at MetaMeta is now Mr. Long Hoang. • A strategy for sustainable serving and maintenance of the SIS at a renowned institute in Africa, as well as for raising awareness of the SIS amongst relevant stakeholders, was discussed during the Third Annual Project meeting in Ghana (Accra, 22-24 May 2023). • All feedback received from partners before 15 June 2023 has been processed in the current version of the DMP.

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

Table of Contents

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

List of tables

List of figures

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

Abbreviations

CC: Creative Commons license
DMP: Data Management Plan
DOI: Digital object identifier
DPO: Data Protection Officer
EC: European Commission
EO: Earth Observation platforms
EUDPR: EU Data Protection Supervisory Authorities
FAIR: Findable, accessible, interoperable and re-usable (i.e., FAIR principles)
FARA: Forum for Agricultural Research in Africa
FNSSA: Food, Nutrition, Security and Sustainable Agriculture
GDPR: EU General Data Protection Regulation
GeoNode: Web-based application and platform for developing geospatial information systems (GIS) and for deploying spatial data infrastructures (SDI)
GIS: Geographical information system
GLOSIS: Global Soil Information System
GSMS: GeoSpatial Content Management System
GSP: Global Soil Partnership
H2020: European Funding Programme for research and innovation
ICO: Information Commissioner's Office
INSPIRE: Infrastructure for spatial information in Europe
ISO: International Organization for Standardization
IT: Information technology
LUCAS: Land Use and Coverage Area Frame Survey (EC)
OpenAIRE: European Open Science Infrastructure, for open scholarly and scientific communication
PEDR: Plan for the Exploitation and Dissemination of Results
PET: Project Expert Team
SDI: Spatial data infrastructure
SIS: Soil Information System
SOP: Standard Operating Procedure
UUID: Universally unique identifier
WP: Work Package
WPL: Work Package Leader

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

Abstract

The aim of Soils4Africa is to provide an open-access soil information system with a set of key indicators and underpinning data, accompanied by a methodology for soil monitoring across the African continent. The soil information system will become part of the knowledge and information system of FNSSA (Food, Nutrition, Security and Sustainable Agriculture) and will be hosted by an African institute that has undergone the capacity building necessary for this task.

To address these goals, a wide range of data will be collected and produced across seven work packages. The data management plan aims to describe the data to be collected and analysed, and how the data will be stored, shared, and protected.

Soils4Africa will produce a soil information system (SIS) to store and process data collected and created in the project. This SIS will also act as a repository for data collection and will store the data beyond the life of the project with access appropriate for the type of data as set by the owner of the raw data. Soils4Africa will create a publicly accessible website for communication and dissemination to the public.

The Soils4Africa Data Management Plan (DMP) is based on the H2020 FAIR Data Management Plan template designed to be applicable to any H2020 project that produces, collects, or processes data for research. The purpose of the DMP is to define the data to be processed, collected, and produced during the lifetime of the project and designate plans for data storage, data sharing, and data visibility.

This DMP is a living document which will be updated over the lifetime of the Soils4Africa project. Deliverables from the various packages will update the DMP at months 13, 25 and 37, or earlier should any major changes occur, with the final version submitted to the EU in month 48 when the project ends.

All partners were sent a copy of the draft DMP for their comments as well as general information. Version 4.0 of the DMP incorporates main changes since publication of version 3.0. EU-deliverables completed so far are listed in Appendix 1 and referred to in the body of the report when appropriate.

Importantly, contact details for the partner's data protection officers (DPO) are listed in Appendix 2.

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

Introduction

Background and motivation

Over the lifetime of the Soils4Africa project soil data will be collected according to standard methods for field sampling and analysis. This DMP will direct the way in which the data collected is to be controlled within and beyond the project. This plan will also define management practices including use of metadata, file naming conventions, and fostering interoperability.

This is a living document to be updated throughout the project when significant changes occur (Figure 1) with the final version submitted before the final review of the project (month 48). Review procedures for project deliverables, and a detailed schedule for the internal and external submission thereof, were formalised in March 2021 ('Soils4Africa_Procedure_Deliverables_review.pdf') as a follow up to a recommendation of the Project Executive Team (PET).

Soils4Africa follows the principles of the Open Research Data Pilot (OpenAIRE), an EC-supported initiative to foster Open Science in Europe to accelerate research and boost innovation, as well as the FAIR guiding principles for scientific data management and stewardship[1].

Organisation of the Plan

The Soils4Africa DMP is constructed using the H2020 FAIR Data Management Plan template[2] designed to be accessible to any H2020 project that produces, collects, or processes research data. This plan is referred to by the OpenAIRE[3] guidance materials.

Main sections of the DMP include a) Data summary, b) Making data FAIR, c) Allocation of resources, d) Data security, e) Ethical aspects, and f) any other issues arising. These aspects are addressed in the following sections. The DMP applies to all partners (Table 1) of Soils4Africa and these are required to inform the consortium as any important changes arise.

[1] Wilkinson, M.D. et al (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi:10.1038/sdata.2016.18

[2] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[3] www.OpenAIRE.eu

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

Table 1. Partner Details

No	Short name	Name	Country
1	ISRIC	Stichting International Soil Reference and Information Centre	Netherlands
2	WU	Wageningen University	Netherlands
3	FARA	Forum for Agricultural Research in Africa, Ghana	Ghana
4	SZIU (MATE)	Szent Istvan University	Hungary
5	ARC	Agricultural Research Council	South Africa
6	IITA	International Institute of Tropical Agriculture	Nigeria
7	I-BEC	Diabalkaniko Kentro Periballontos	Greece
8	SU	Stellenbosch University	South Africa
9	ICRAF	International Centre for Research in Agroforestry	Kenya
10	RCMRD	Regional Centre for Mapping of Resources for Development	Kenya
11	IFA-YANGAMBI	Institut Facultaire des Sciences agronomiques (IFA) de Yangambi	Congo (Democratic Republic of)
12	BUNASOLS	Bureau National des Sols	Burkina Faso
13	IRA	Institut des Regions Arides	Tunisia
14	KALRO	Kenya Agricultural and Livestock Research Organisation	Kenya
15	SGS HUNGÁRIA	SGS Hungaria Minosegellenorzo Kereskedelmi Es Szolgaltato Korlatolt Felelossegu Tarsasag	Hungary
16	JRC	JRC - Joint Research Centre, European Commission	Belgium
17	METAMETA	Metameta Research BV	Netherlands

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

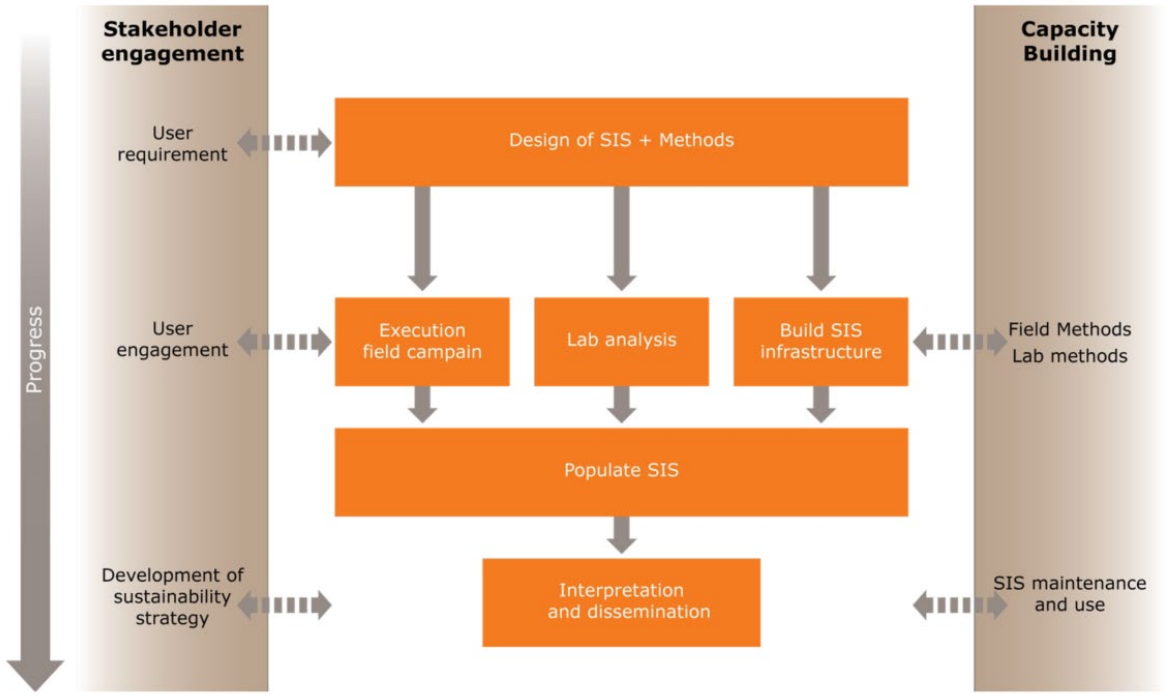

Figure 1. Schematic representation of the proposed methodology (Source: Soils4Africa Grant Agreement).

Data Summary

The overarching goal of the Soils4Africa project is to provide an open-access soil information system with a set of key indicators and underpinning data, accompanied by a methodology for soil monitoring across the African continent. The soil information system will become part of the knowledge and information system of FNSSA and will be hosted by an African institute. Ultimately, the SIS will allow broad-scale applications aimed at informing policy and planning at high level (national or pan-African), with respect to land potential in terms of both the yield and environmental sustainability (including adaptation to climate change and environmental resilience).

Main activities are:

(i) define use cases and indicators in consultation with stakeholders

(ii) make a functional design of the soil information system

(iii) develop detailed procedures and tools for the field activities based on the LUCAS methodology and collect 20 000 soil samples

(iv) develop detailed procedures for laboratory work and analyse the collected soil samples at one reference laboratory located in Africa, and

(v) develop the technical infrastructure for the soil information system and serve the results as open data linked with open EO data.

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

Soils4Africa addresses the work programme of Horizon 2020's Societal Challenge[4] 2 in the following ways. First, it contributes to priority 2 (Fostering functional ecosystems) because the SIS is a tool to target interventions that improve soil quality and provides insight in the impact of these interventions. Secondly, it contributes to priority 1 (Addressing climate change and resilience on land and sea), as the SIS will contribute to the assessment of carbon losses from soil and the identification of areas with high potential for soil carbon sequestration. Finally, the SIS provides a platform for the development of sustainable business models by service companies aiming at the development of sustainable food systems, contributing to priority 3 (Boosting major innovations on land and sea). Soils4Africa is linking with relevant H2020 projects and Copernicus on EO data use. It actively connects organizations across Africa and Europe for synergies and promotes an open science approach.

Data Overview

This research will produce a range of data and technical guidelines for collecting/analysing these data (Table 2). It is expected that the total data and reports generated by the project, and stored within the SIS, uses preferred data formats[5]. This is expected to include files including, but not limited to, the following formats:

- Portable document format (.pdf; for final reports)
- Comma-separated Values (.csv; for soil sample and analytical data)
- Plain text (.txt)
- Microsoft Word (.docx; for draft reports)
- Tab-separated values (.tsv; for soil sample and analytical data)
- Geostationary earth orbit tagged image file format (.GeoTIFF; for gridded maps)
- Joint photographic experts group (.jpeg)
- Portable network graphics (.png)
- JavaScript Object Notation (.json; for web-based data exchange and transfer)
- Extensible Markup Language (.xml; for web-based data exchange and transfer)
- Shapefiles (.shp; for polygon GIS layers).
- OPUS spectral files (.0-9; for storing Bruker spectral data files)

Technical details about the soil data collected and submitted (e.g. nature, size, format, licenses) to the SIS has been updated now that the field and laboratory work have progressed. In conjunction with this, procedures defining how data outputs are to be quality assured and approved for external release have been developed.

[4] https://ec.europa.eu/programmes/horizon2020/en/h2020-section/societal-challenges
[5] https://dans.knaw.nl/en/about/services/easy/information-about-depositing-data/before-depositing/file-formats.

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

Table 2. Nature, size, format and licenses of Soils4Africa deliverables

Nature	Size	Format	License	Location
Map of Agricultural Land in Africa (100m x 100m) - version 1	195 MB (zipped); 283 MB (unzipped)	GeoTIFF	CC BY	URL
Map of Agricultural Land in Africa (100 x 100 m) – version 2	152 MB (zipped); 167 MB (unzipped)	GeoTIFF	CC BY	URL
Map of Agricultural Land in Africa (100 x 100 m) – version 2: agricultural land classes	396 MB (zipped); 405 MB (unzipped)	GeoTIFF	CC BY	URL
Technical reports	< 1 MB (each)	PDF	CC BY	URL
Videos of protocols and guidelines (YouTube)	-	-	CC BY	URL

Data Origin

Initially, the project will make an inventory of historic (existing) datasets available for Africa. Ownership of the historic data may vary from open access (Creative Commons[6], CC0) to restricted (e.g. CC BY NC SA) depending on the purposes of the underpinning surveys. Generally, this data is owned by national soil survey organisations and sometimes private companies. Personal data will be handled in accordance with the EU General Data Protection Regulation (GDPR)[7], as indicated on the Soils4Africa website.

Most soil data underpinning the SIS for Soils4Africa (WP6), however, will come from novel data collection work (WP5). As indicated in the Grant Agreement, it is the intention that these new geo-referenced data will be made available under a Creative Commons, CC BY license 4.0, license thereby providing the basis for their future distribution following FAIR principles (see 'making data openly available', p. 16-17).

As the project progresses, the various work packages have produced protocols/reports, generated novel data, and prepared communication materials as further detailed in the overarching Grant Agreement:

- Work package 1 will generate a range of reports (e.g. DPM, PEDR) and communication materials, accessible through the Soils4Africa website. The project website[8] was launched in September 2020 and includes a special section (i.e., Documents) with public deliverables resulting from the project (see Appendix 1).
- Work package 2 will develop use cases, user requirements and identify associated soil quality indicators; this will inform the development of the SIS and methods for field and laboratory development in WP3. These aspects have been documented in deliverable D2.1, 'Soils4Africa_D2.1_Use case_v01.pdf'. Personal data for stakeholders identified as potential users of the SIS will be managed in accord with GDPR regulations, informing the communication strategy.
- Work package 3 will produce a series of reports describing methods for deriving soil quality indicators, an overview of existing datasets for Africa, the sampling protocol, guidance for the fieldwork and laboratory analyses, as well as user requirement for the IT infrastructure. Deliverables to date include 'Methods for deriving selected soil quality indicators' (D3.1) and 'Inventory of soil

[6] https://creativecommons.org/
[7] https://gdpr-info.eu/
[8] https://www.soils4africa-h2020.eu/

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

data for Africa represented in the ISRIC-WDC holdings & Soils4Africa Sampling Design' (D3.2); 'Detailed guidance for field work' (D3.3); 'Guidance for laboratory analysis' (D3.4) and 'User requirements for the IT-infrastructure' (D3.5).

- Work package 4 will prepare reports documenting protocols, Standard Operating Procedures (SOPs) and a toolset to guide and support the field survey, as well as the collection, handling and preparation of soil samples (see Appendix 1). It will also deliver reports documenting the training of all field teams and field campaign itself. All field observations and data related to the collection, pre-processing and handling will be uploaded in the SIS.

In view of the COVID situation, preparations for the field campaign were delayed. From April 2022, the Soils4Africa team organised a series of workshops for country supervisors from West Africa and later East Africa. The workshops served to develop a blue-print for the methodology that will be applied by all the regional hubs. Ultimately, the country supervisors will coordinate the field campaigns—which will include obtaining permissions, a range of assessments, and collection of soil samples in their respective countries. Training workshops for field surveyors were held in Tanzania and Zambia (March and April, 2023). The first batch of samples, from Nigeria, has been shipped to ARC in March 2023. All samples will be analysed for a wide range of parameters at ARC, as central laboratory (see Work Package 5).

- Work package 5 focuses on Soil Analysis and Capacity Building. The reference laboratory (ARC) will generate several reports on laboratory upgrade (e.g., ISO 45001) and quality control procedures to be followed during the project, with supplementary reports on periodic quality controls carried out (see Appendix 1). So far, 716 samples have been shipped to ARC. Of these, 35 samples have been analysed with wet chemistry and 155 samples with spectroscopy (status: 31 May 2023). During the project, the number of samples analysed will eventually increase to 30 000 (approximately 20 000 topsoil and 10 000 subsoil samples). Following thorough quality control results will be regularly added to the central data management system, with appropriate metadata documenting data lineage. Metadata for the spatial and soil data collected will be made INSPIRE compliant; this aspect will be addressed during the capacity building component of the project.
- Work package 6 will generate an operational SIS containing all data collected, providing access through a user-friendly web portal. So far, field data from over 2000 plots from various countries have been harvested from the API and ingested in the database. A range of reports/manuals describing the technical design of the SIS, results of the analysis of heavy metals and pesticides residues at selected sites[9], and SIS use and soil information product development from the SIS will be issued (see Documents). Further, a report documenting training of technical staff at the SIS-hosting institute; it will provide all training materials, as well as a description of the training and records of the trainees in accord with GDPR will be issued.
- Work package 7 sets out the 'ethics requirements' that the project must comply with. Four deliverables were submitted for this work package as indicated in the section on 'Ethical Aspects'.

[9] In view of the often-sensitive nature of pollution and pesticide-related data such data may need to be anonymised, i.e. presented without georeferencing.

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

Data Utility

Soils4Africa data management aims to ensure the visibility of the data generated within the project, and for the data to become (and remain) useable. The 'Plan for the Exploitation and Dissemination of Results (PEDR) and Communication Activities' (deliverable 'Soils4Africa_D1-2_PEDR_v01.pdf') describes the strategy for sharing the project results with potential users (dissemination) and the use of project results. The PEDR is subject to annual review by the PET and updated as necessary.

All public documents, guidelines, manuals arising from the project have been published on the website established for the project and will be updated as necessary after EU-review. The website will also provide news items and announce events to inform Soils4Africa partners and other communities.

Soils4Africa plans to publish results of the investigations in open access peer-reviewed journals, ensuring visibility of the data generated by the project.

The project will develop a strategy for sustainable serving and maintenance of the SIS at a renowned institute in Africa, as well as for raising awareness of the SIS amongst relevant stakeholders. These important issues have been discussed during the Third Annual Project meeting in Ghana (Accra, 22-24 May 2023) as documented in the meeting report by the Soils4Africa project coordinator.

FAIR Data

Making data Findable, including provision for Metadata

All data collated and generated by the project must be findable, the first criterion for data being FAIR. Metadata describing the various resources will be created and stored in the SIS's metadata catalogue using a standard format, such as ISO 19115:1-2014, and be INSPIRE compliant. Typically, the metadata will describe the datasets concisely, including title of the materials, name of the author(s), year the materials were generated or published, information on stratum (e.g., soil science), region (e.g., Africa, Kenya) and other domain specific keywords such as pH, organic carbon and CEC (cation exchange capacity), language of the resource, unique identifier of the resource (UUID or DOI), license (e.g., CC BY 4.0, CC BY NC or Embargo), contact for the resource (e.g., metadata catalogue contact), status of the resource (e.g., ongoing and completed), the provenance/lineage of the materials. For spatial data, provision of the coordinate reference system and precise coordinates is essential.

All the data sets available within the project must provide a detailed level of metadata on their description, including contact information. The following guidance, based on guidelines prepared by the NERC[10], will be followed:

[10] Natural Environment Research Council http://www.nerc.ac.uk

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

Table 3 Guidance on metadata topics to address

Metadata topic to address	Details
Contact information	Metadata should be provided that documents the provenance of the data so that these are fully traceable
Experimental Design / Sampling Regime	Metadata should be provided which details the experimental design and/or sampling regime, where applicable
Collection / Generation / Transformation Methods	Metadata should be provided covering methods used for collection of samples/observations. Alternatively, where data values are derived/generated/ transformed, then details of how this is achieved should be provided.
Fieldwork and / or Laboratory Instrumentation	Information should be supplied on instruments/machines used for collection/analysis of samples/observations where relevant.
Calibration Steps and Values	Details of the steps taken to calibrate any instruments/machines used, including use of any blanks, and the values used for calibration should be provided.
Nature and Units of Recorded Values	Metadata should be provided describing the nature of the recorded values contained and the units used sufficient to unambiguously define what has been measured and recorded in the dataset.
Analytical Methods	Full descriptions of any analytical methods used to generate the data values contained in the dataset should be included
Quality Control	Any quality control measures undertaken to ensure the quality of the data values included in the dataset should be detailed.
Format of Stored Data	The format which was used to store the dataset during the lifetime of the project, and the format in which the dataset is made publicly available, if different, should be named in the contextual metadata.
Miscellaneous	Any additional information necessary to expand on that given in the discovery metadata record.

Naming convention for project reports will follow the pattern:

Soils4Africa_ Deliverable#_FileName_version

So, for example, the first DMP is titled 'Soils4Africa_D1-1_DMP_v01'. File naming conventions are expected to update as the project progresses to keep up with the demands of the project.

The metadata model is intended to associate the various file types which form the basis of the data. Metadata will detail the associations between files and their relevance within the project. This is intended to safeguard the openness and interoperability of the data. The detail of the metadata accompaniment will be updated as the project progresses to incorporate the expansion of the data.

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

Making data openly accessible

During the lifetime of the project the deliverables stated by the grant agreement as having public dissemination will be openly available, these will be distributed through the website. The website is fully accessible by all members of the public using modern web browsers and mobile devices. In addition, it is available in two languages (English and French) to maximise the accessibility. The website provides a page for downloadable materials including the recently produced 'Soils4Africa maps of agricultural land in Africa'[11]. Where relevant, deliverables will also be shared through the SIS or directly disseminated to stakeholders, this in accord with the licenses (permissions) stipulated by the various data providers/ landowners.

In practice, it is unrealistic to expect specific licenses for each of the 20 000 sampling points. Some of these may be located on privately owned land, while others will be located on communal lands. IITA, in particular, is looking at these complex issues as in many situations it will be unclear to whom the land to be sampled actually belongs. The recommended procedure for 'seeking permissions' has been documented in the 'Protocol for Field Survey: Guidelines for Field Surveyors on Soil Sample Collection and Field Assessment of Agricultural Lands in Africa'. The country supervisors themselves will make the arrangements needed for his/her particular country and provide the instructions to the surveyor on what to do in alignment with the corresponding protocols. Importantly, an online data management tool (ODK) has been developed to store the information in conjunction with the fieldwork itself, so that all licenses/permissions can be uploaded to the evolving SIS.

Ultimately, key research data generated by the project will be made available in an open repository (e.g., https://www.re3data.org/search?query=isric). This repository will be the Geospatial Content Management System (GCMS) of the Soils4Africa SIS (under development; URL will follow at a later stage). The open-source web-based platform TerriaJS will be the most likely candidate for this purpose. Freely accessible data will be needed to validate the results to be presented in scientific publications; the associated metadata (i.e., data describing the deposited research data themselves) will be managed in the Soils4Africa GCMS with back-up copies likely to be retained in the ISRIC WDC-Soils GeoNetwork metadata catalogue.

Besides soil sample data collected by the Soils4Africa project itself, the SIS will include third party datasets that are judged relevant in the context of sustainable intensification of agriculture in Africa. Such datasets may include, for example, maps of land cover, farming systems and agro-ecological zones. The third party datasets themselves will be made available following licenses/permissions of the data owners.

Formal approval and operational processes for linkages between Soils4Africa SIS and existing large data instruments or repositories hosted by different organizations, such as FARA's 'Datainforms', are being worked out by the consortium. Contacts with EO data providers for Africa have been formalized (e.g.,

[11] https://www.soils4africa-h2020.eu/s4a-maps-agricultural-land-in-africa

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

AfriGEO and GMV1 AfriCultureS); the corresponding 'letters of agreement' will be stored in the SIS by the data processor.

Making data interoperable

Soils4Africa will aim to adhere to the most commonly used file formats to be supported by the scientific and agricultural communities, using defined vocabularies (or ontologies) and uniform ISI units of measurement. The project will review existing ontologies that are applicable and acceptable to the Soils4Africa project and SIS.

All data should be available as *open* data and according to FAIR[12] (Findable, Accessible, Interoperable and Reusable) principles through user-friendly web services allowing data to be reviewed and downloaded. To ensure inter-operability, an adequate multi-lingual ontology will be identified, possibly drawing on the recommendations for the Online Collaborative Platform (OCP3) being developed for the EU CIRCASA[13] project. The SIS will be compatible with (and possibly become a node in) the *Global Soil Information System* (GLOSIS)[14] being developed in Pillar 4 of the Global Soil Partnership (GSP).

Increase data re-use (through clarifying licenses)

Licensing is defined in the consortium agreement and will dictate the level of data re-use including the period beyond the end of the project. An 'open' Creative Commons license (i.e., CC 0 or CC BY) will be pursued for the novel data resulting from the project, to encourage data re-use. This information will always be visible so that existing and prospective partners may pursue business opportunities and/or options for new research collaboration that build on the results of the Soils4Africa project, this in line with the overall terms of the Consortium Agreement.

It is foreseen that the system will have its in-built analytics that can provide summarized data for the end users.

Allocation of Resources

Costs and partner responsibilities are covered in the budget defined in the consortium agreement. This aspect is the responsibility of the lead partner, ISRIC, and is within the remit of Work Package 1, Project Coordination and Management.

The costs of long-term data management and maintenance of the SIS after the project ends are not covered by the project budget. At this stage, it is anticipated these will be borne by the African Institute that will host the SIS after the project ends. Developing a sustainability strategy is part of the project

[12] https://www.openaire.eu/how-to-make-your-data-fair

[13] https://www.ocp.circasa-project.eu/en/1/home.html

[14] http://www.fao.org/global-soil-partnership/areas-of-work/soil-information-and-data/en/

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

(task 2.5 under work package 2); this strategy will not only entail the selection of a hosting institute in Africa, but also the safeguarding of funding to cover the costs of hosting the SIS for a reasonable amount of years. In view of long-term safeguarding of the project deliverables (> 10 years), a copy of the data will be kept by ISRIC - World Soil Information.

Data Security

Partners hold responsibility for the recoverability of the data they generate. Importantly, data and documentation resulting from the research will be stored on institutional network drives with defined access restrictions and regular back-ups. A centralized data management system will be developed as part of the SIS, enabling cross-institutional data exchange in a protected cloud environment. A mechanism should be in place to allow partners backup data being generated on a weekly basis to a centralized repository.

To minimize the risk of data loss or corruption, external devices such as hard disks or USB drives will not be used. A protocol will be set up to ensure that files are consistently named and versioned following identical standards by all participating members (see above).

Final drafts of documents will be stored by the partner responsible for the task on a project-specific Microsoft Teams folder, managed by the coordinator of WP1. This folder is accessible only by project partners. Final documents will be verified by at least one member of the project Executive Team (PET) prior to their submission to the EU as part of the regular and final reporting. As indicated, final public deliverables/documents will be uploaded to the Soils4Africa website to ensure wide accessibility.

Importantly, as indicated earlier, in view of long-term safeguarding of the project deliverables (> 10 years) and sustained quality-assurance, a back-up of the data will be kept by ISRIC - World Soil Information. ISRIC is a CoreTrustSeal certified repository (CTS)[15] and a regular member of the ICS (International Council for Science) World Data System where it is accredited as ISRIC WDC-Soils[16].

Ethical Aspects

Ethical considerations are defined in Work Package 7, Ethics Requirements, as described in the Grant Agreement. In particular, templates for informed consent procedures covering voluntary participation and data protection issues (in language and terms intelligible to the participants) will be collated and archived by the respective partners (see Appendix 2). Scanned copies of the filled-in templates (as pdf files, English version) will be submitted as a deliverable to the WP1 lead for central registration in the SIS by the data processor. This will ensure compliance with the 'European Code of Conduct for Research Integrity, Revised Edition 2017'[17].

[15] https://www.coretrustseal.org/

[16] https://www.isric.org/about/world-data-centre-soils-wdc-soils

[17] https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics_code-of-conduct_en.pdf

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

The following documents have been produced and shared with the partners to make them aware of the 'ethical requirements' of the project:

- Soils4Africa_D7-1_H_v01.pdf – Humans (H), Requirement No. 1
- Soils4Africa_D7-2_H_v01.pdf – Processing of personal data (European countries)
- Soils4Africa_D7.2-POPD_v01 – Protection of personal data (POPD)
- Soils4Africa_D7.3_NEC-v01.pdf – Processing of personal data, Non-European Countries (NEC)
- Soils4Africa_D7.4_EPQ-v01.pdf – Environmental Protection and Safety (EPQ), Requirement No. 4.

Soils4Africa has appointed Niels Batjes (ISRIC) as the Data Protection Officer (DPO) to monitor that the project is processing all personal data in compliance with applicable data protection rules. The DPO carries out the functions for the Project with respect to data protection obligations, as described in project deliverable D7.2-POPD.

Each data handling partner (or beneficiary) should appoint a DPO to ensure awareness of and adherence to the principles of the GDPR; written confirmation for each beneficiary (i.e., indicating whether they require a DPO or not) must be submitted to ISRIC as a deliverable to be stored in the SIS. Requests for this have been mailed to all partners in early May 2021, in the framework of D7.2. The final reminders were sent after the First Annual Project Meeting (9 September 2021). Responses are listed in Appendix 2; see also Annex 4 of the Grant Agreement. The main tasks of the 'regional' DPO's are outlined in deliverable D7.2 'Processing of Personal Data', which has been sent to the various partners for further circulation.

If any individual or partner within Soils4Africa is aware of a (personal) data-related incident or breach they should report it to their DPO, as well as the DPO at ISRIC, as soon as possible and within 72 hours of discovering the breach. The ISRIC DPO is obliged to report any breaches which result in a risk to individual's rights and freedoms to the EU Data Protection Supervisory Authorities (EUDPR).

Other Issues

No other issues have been raised at present by the consortium. This is a living document to be updated with outstanding issues as the project progresses. The present version (no. 4) has been updated to reflect the changes to the project up to June 2023 (Month 37). Deliverables from the various work packages will update the DMP at month 48 (version 5, June 2025) of the project. The final DMP, itself a project deliverable, will be submitted to the EU in month 48 when the project ends.

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

Appendix 1 – List of project deliverables and field protocols (up to 15 June 2023)

Deliverable	Title	Submission date	Status[18]
Soils4Africa_Procedure_Deliverables_review	Procedure_Deliverables_review	2/3/2021	Internal
Soils4Africa_D1-1_DMP_v01	Data Management Plan	15/9/2021	R
Soils4Africa_D1-1_DMP_v02	Data Management Plan (First Update)	20/10/2021	R
Soils4Africa_D1-1_DMP_v03	Data Management Plan (Second Update)	3/06/2022	R
Soils4Africa_D1-2_PEDR_v01	Plan for the Exploitation and Dissemination of Results (PEDR) and Communication Activities	15/9/2020	R
Soils4Africa_D1-3_WEBSITE_v01	Website	15/9/2020	R
Soils4Africa_D2.1_Use cases_v01	Set of use cases plus supporting soil quality indicators	15/1/2021	R
Soils4Africa_D3.1_SoilQualityIndicators_v01	Methods for deriving selected soil quality indicators	31/1/2021	R
Soils4Africa_D3-2A_v01	Inventory of soil data for Africa represented in the ISRIC-WDC holdings	29/5/2021	R
Soils4Africa_D3.2B_Sampling_design_v01	Soils4Africa sampling design	29/5/2021	R
Soils4Africa_D3.3_Guidance_Fieldwork_v01	Detailed guidance for field work	31/5/2021	R
Soils4Africa_D3.4_Guidance_LaboratoryAnalysis_v01	Guidance for the laboratory analysis	29/5/201	R
Soils4Africa_D3.5_UserRequirements_v1.0	User requirements for the IT-infrastructure	28/5/2021	R
Soils4Africa_D4.1_v01	Map of agricultural land of continental Africa	7/6/2021	R
Soils4Africa_D4.1_Annex	Annex to D4.1: Map of agricultural land of continental Africa	1/06/2023	R
Soils4Africa_D4.2_Fieldsurvey-Protocols-SOPs and tools_v03	Protocols, SOPs and tools to guide field survey, collection, handling and preparation of samples	30/11/2021	R
Soils4Africa_D5.1_LabUpgrade-HealthSafetyPlan_v01	Report on the laboratory upgrade including health and safety plan (ISO 45001)	30/11/2021	R
Soil4Africa_D6.1_Technical design of the SIS_v01	Technical design of the SIS	30/11/2021	R
Soils4Africa_D7.1_H_v01	Humans(H), Requirement No.1	30/11/2021	R
Soils4Africa_D7.2_POPD_v01	Processing of Personal Data (POPD), Requirement No.2	7/6/2021	R
Soils4Africa_D7-3_NEC_v01	Non-European Countries (NEC), Requirement No. 3	30/11/2020	R
Soils4Africa_D7-4_EPQ_v01	Environmental Protection and Safety (EPQ), Requirement No.4	30/11/2021	R

[18] R: *This deliverable is under review by the EU and subject to changes if deemed necessary. The European Commission and European Research Executive Agency are not responsible of the content of the documents. Any subsequent updates will be shared and flagged as such (final versions will be made available at:* https://www.soils4africa-h2020.eu/documents*.)*

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

Protocols for field campaign:			
Report	Standard operating procedure (SOP) Sample Preparation and Shipment	02/2023	URL
Report	Detailed guidance for fieldwork	05/2021	URL
Report	Soil sample collection and field observations	07/2022	URL
Report	Protocol for Field Survey Management (Instructions for the country supervisor)	05/2022	URL
Report	Protocol for Field Survey (Guidelines for Field Surveyors on Soil Sample Collection and Field Assessment of Agricultural Lands in Africa)	05/2022	URL
Report	Survey data management tool (SDMT) user manual	05/2021	URL
Report	Guidance for the Laboratory Analysis	05/2021	URL
Report	Central Concept for Minimalistic Reference Site	05/2021	URL
Report	Soil4Africa sampling design	05/2021	URL
Report	SOP for soil sample collection and observation	07/2022	URL

Note: The most recent list of project documents is available at https://www.soils4africa-h2020.eu/documents.

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

Appendix 2 – Data Protection Officers for the Soils4Africa project

Partner no.	Partner, short name [a]	EU partner	Has DPO [b]	Name of DPO [d]	E-mail	Confirmation	Online privacy and contact details
1	ISRIC	Y	Y	Niels H. Batjes	niels.batjes@isric.org	Y	https://www.isric.org/privacy-and-personal-data
2	WU	Y	Y	Coen Ritsema [b]	privacy@wur.nl	N	https://www.wur.nl/nl/Privacy-Cookie-verklaring.htm
3	FARA	N	Y	Benjamin Abugri	Babugri@faraafrica.org	Y	-
4	SZIU (MATE)[c]	Y	Y	Bence Györe	dpo@uni-mate.hu	Y	http://sziu.hu/szius-privacy-statement
5	ARC	N	Y	Khangwelo Rathogwa	RathogwaK@arc.agric.za,	Y	http://www.arc.agric.za/Pages/Terms-and-Conditions.aspx
6	IITA	N	Y	Olatunbosun Obileye	O.Obileye@cgiar.org	Y	https://www.iita.org/privacy-policy/
7	I-BEC	Y	N	Nikos Tsakiridis	tsakirin@auath.gr	Y	-
8	SU	N	?	Andrei Rozanov [b]	dar@sun.ac.za	N	-
9	ICRAF	N	Y	Thomas Zschocke	cifor-icraf-dpo@cgiar.org	Y	https://www.worldagroforestry.org/node/36303
10	RCMRD	N	?	Emmanuel Nkurunziza [b]	enkurunziza@rcmrd.org	N	-
11	IFA-YANGAMBI	N	?	Alongo Longomba [b]	sylvainalongo@yahoo.fr	N	https://www.auf.org/mentions-legales/
12	BUNASOLS	N	N	Traore Mamoudou	tramadalbela@yahoo.fr	Y	-
13	IRA	N	?	Mohamed Ouessar [b]	Ouessar.Mohamed@ira.rnrt.tn	N	-
14	KALRO	N	Y	Patricia Ngutu [b]	Patricia.Ngutu@kalro.org	N	https://www.kalro.org/code-of-conduct-and-ethics
15	SGS HUNGÁRIA	Y	Y	Ferenc Medgyessy	privacy.hu@sgs.com	Y	https://www.sgs.hu/en/privacy-at-sgs
16	JRC	Y	Y	Premys Spicar (JRC.A.4)	JRC-DATA-PROTECTION-COORDINATOR@ec.europa.eu	Y	https://ec.europa.eu/dpo-register/detail/DPR-EC-00933

Partner no.	Partner, short name [a]	EU partner	Has DPO [b]	Name of DPO [d]	E-mail	Confirmation	Online privacy and contact details
17	METAMETA	Y	Y	Simon Chevalking	schevalking@metameta.nl	Y	-

[a] For long names of consortium partners see Table 2.
[b] For beneficiaries that have no DPO, the name and contact email for receiving GDPR-related information (in case of no reaction, the name of the primary contact person has been added).
[c] Name and contact information for SZIU (now MATE) changed subject to updates from the EU Project Adviser.
[d] List of DPO's resp. contact persons based on feedback received from project partners up to 30 May 2022.

This project has received funding from the Europea
Innovation Programme under Grant

European Commission | Horizon 2020 European Union funding for Research & Innovation