



**Sustainable intensification of food production through
resilient farming systems in West & North Africa**

Deliverable D7.4
Data management plan

Due date of deliverable: M6
Actual submission date: 28th February 2021

GENERAL DATA

Grant Agreement: 861924

Project acronym: SustInAfrica

Project title: Sustainable intensification of food production through resilient farming systems in West & North Africa

Project website: www.sustinafrica.com

Start date of the project: September 1st, 2020

Project duration: 60 months

Organisation name of lead contractor: Luke

- Funding source: SFS-35-2019-2020 - Sustainable Intensification in Africa
Type of action: Research and Innovation Action

DELIVERABLE NUMBER:	D7.4
DELIVERABLE TITLE:	Data management plan
DELIVERABLE TYPE:	ORDP
WORK PACKAGE N:	WP7
WORK PACKAGE TITLE:	Coordination and administration
DELIVERABLE LEADER:	Luke
AUTHOR:	Giles Young and Nils Borchard
CONTRIBUTORS:	
DISSEMINATION LEVEL:	PU

ABSTRACT

The data management plan (DMP) describes principles and protocols for data collection (see also Ethics described D8.1 and D8.2), data transfer, data processing, data storage, and data retention and sharing. The principles and protocols will be regularly reviewed and updated throughout the project, with the guidance of an expert on data management (Luke), and will ensure that data management and protection is compliant with the EU's Guidelines on Findable, Accessible, Interoperable, and Reusable (FAIR) Data Management in Horizon 2020, and with relevant national data protection laws and institutional data management policies. Produced data will be treated according to management guidelines and implemented through WP7.



Table of Contents

List of abbreviations and acronyms.....	3
1 Introduction.....	4
1.1 Reasoning for data collection/generation.....	4
1.2 Type, format and origin of data	4
1.3 Metadata	5
1.4 Accessibility and interoperability	5

List of abbreviations and acronyms

DMP	Data management plan
IPR	Intellectual property rights
OA	Open access



1 Introduction

This document presents SustInAfrica's data management plan (DMP) and describes the project's principles on how to manage collected data and metadata collected of various processing stages. The project's data management and protection is compliant with the EU's Guidelines on Findable, Accessible, Interoperable, and Reusable ((FAIR) Data Management in Horizon 2020), the H2020 open access to research data and scientific publications (see for more details D6.3). Major objectives of this DMP are:

- Reasoning of data collection/generation;
- Describing type, format and origin of data;
- Introducing methodology and standards for metadata;
- Ensuring accessibility and interoperability;
- Explaining principles addressing data storage, security and protection.

2 Managing data

2.1 Reasoning for data collection/generation

Based on the project's objectives, complex nature of agro-ecological systems and expected impacts described in the grant agreement the project's parties will collect, gather, process and disseminate data and information based on standardized and harmonized procedures (summarized by WP; D1.9 in M48). Thus, data collection and processing of biogeophysical data and socio-economic information are essential towards ensuring impact generation. Collecting data and disseminating of thereon generated knowledge via the project's multi-actor approach requires planning, design, development and implementation of data management standard and procedures.

2.2 Type, format and origin of data

The lead contact at each participant organisation will be responsible for the collection and transfer of data from its own research activities, within the framework of the DMP. The SustInAfrica project will generate data varying in type, format and their origin:

1. Primary data: Based on a jointly designed methodology (see also D1.1 and D1.9); Biogeophysical data (e.g. yield, soil organic carbon), data on ecosystem services (e.g. carbon sequestration potential), socio-economic information (e.g. income, education), and innovativeness
2. Audio-visual data: Video, photos, text messages, voice mails, etc. (see D6.4).
3. Text data: Literature reviews, reports, news etc. (see D6.4).
4. Processed data (e.g. numerical data, databases, meta-data)

Expected data formats are: Microsoft Office file types, imagery (e.g. tiff, jpg), audio/video (mp3, mp4, etc.), etc.

SustInAfrica will use existing data, which includes text and numeric data from scientific and grey literature, accessible results from other projects and initiatives, expert assessments from multiple sources; most relevant are listed in table 1.



Table 1: Selected existing data used in and novel data generated by SustInAfrica.

Source	Data, type, topic & value
COPERNICUS	Global data from satellites; open access; European Aerospace Agency
HWSD	Harmonised World Soil Database; open access; UN FAO
FAOStat	Agricultural data; open access, UN FAO
World Bank	Demographic data; open access, World Bank
GPDD	Global Pest and Disease Database; open access after registration; NSF Centre for integrated pest management
EPPO	Global pest and disease database; open access after registration; European and Mediterranean Plant Protection Organisation
Smart-AKIS	Database and network on smart farming; open access, EU
AG4impact	References on sustainable intensification of farming systems in Africa
TWIGA	Open innovation platform for IoT and big-data in Sub-Saharan Africa
Soil4Africa	H2020 project: Soil data and analytical results
EJPSOIL	H2020 project: Soil data and knowledge on sustainable soil management
Leap4FNSSA	H2020 project: Knowledge and networking

Data for internal use will be stored and preserved at Luke and made available after registration to all participants through Tiimeri, which automatically generates back-ups. Data and information for external use will be stored and preserved at Farmerline and made available after registration to all participants through Mergdata platform, which automatically generates back-ups. The expected size of the data will vary and will be evaluated during the course of the programme implementation. However, largest data and data sets are expected to be generated by WP2 as collected and used imagery are typically large; the consortium will use a big data SharePoint with an initial size of 10TB operated by www.csc.fi.

2.3 Metadata

For transparency and traceability each data set either of its type and degree of processing will get described in more detail to ensure that most relevant meta-data are available:

- Contact information of the collector and processor
- Time of collection and processing
- Methods used during collection and processing
- Additional information concerning repeated sampling, geographical information etc.
- See also the INSIRE directive¹

2.4 Accessibility and interoperability

According to the publication and data policies of the consortium data produced in the project will be made fully available to the scientific community and wider society. The right of the data producer (members of the research team) to the use of research data is reserved when providing Open Access (OA). “Right to use” here refers to the right of the data producer to execute the original project plan before releasing data for further use. Metadata concerning collected data will be documented through Luke’s OA Research Data Descriptions Recovery Service (RADAR, <http://radar.luke.fi>), which allows flexible searching for metadata descriptions (according to Open Research Data Pilot of Horizon 2020

¹ <https://inspire.ec.europa.eu/document-tags/metadata>



[ORDP], ISO19115 and INSPIRE standards), which include core information about datasets, facilitating their future use. Reasons for data not being shared may include commercial exploitation, protection of IPR, securing of future data production, project agreements, privacy policies, or other legislation. In order to secure the continuation of research and commercial developments, EU regulations on variety use and preservation of IP rights will be followed. Scientific publications will—whenever possible—be published OA in peer-reviewed journals and/or through self-archiving in Luke’s repository JUKURI (<http://jukuri.luke.fi>). After the end of the project data and results will be archived at Luke (RADAR, JURKI), made available through FMP, ResearchGate, and other databases (e.g. SmartAKIS) to ensure accessibility and maintenance of SustInAfrica’s data and information after granted period. Geo-referenced data will be stored at the data owner’s resources and made accessible through SustInAfrica’s GIS portal operated and maintained by Farmerline.

The SustInAfrica consortium will make all the necessary efforts to make data collected and processed interoperable and re-used by the relevant actors, including broader researcher community, policy makers, farmers, industry and civil society and other key stakeholders with a specific interest in agricultural soil and agricultural soil management (WPs 4 and 6). Licensing of data and knowledge is not foreseen.

3 Knowledge management and ownership

Knowledge management is a central component bridging the management and exploitation strategies of the project. It seeks to define and assess exploitation opportunities of each project result, facilitating a creative process to define new products and services based on project prototypes and knowledge obtained, as well as ideas for new business development, further research activities and scientific publications. Some activities—such as market insights analysis and intellectual property intelligence—have already started during the proposal preparation phase and will be continued through workshops, meetings and product life cycle analyses during project implementation. Innovation will be managed throughout the project via WPs 4, 5 and 6 and overseen by the Steering Committee. All information produced during the project will be assessed for the need of IPR protection by the consortium through the General Assembly, which will also discuss strategic issues, ethics, and the exploitation of results, in order to ensure joint understanding, and facilitate creative collaboration.

Work package leaders will play a key role in gathering information and ensuring that protocols, methods and results are exploited commercially, made accessible to the public, or both, whenever the opportunity arises. The objective is to ensure fair and transparent means by which to exploit and protect background information and results generated through the project. Ownership of results, access rights and IPR protection will be defined in the Grant Agreement (GA) and Consortium Agreement (CA), following Horizon 2020 principles. The principles commonly agreed by all partners are summarised as follows:

- Background information and knowledge contributed to the project by each participant will be listed in the CA. When planned in the Description of Action (DoA), access to background will be provided royalty-free to other participants for the implementation of the project’s tasks.
- Results shall be owned by the participant who generated them. Each participant will be responsible for ensuring fulfilment of their obligations under the GA regarding results, by making arrangements with any third parties that could claim rights to them.
- Whenever results have been produced jointly between two or more partners, the ownership of the results will be shared among the participants who carried out the work. The terms of joint ownership, protection, and share of ownership, and costs for possible protection will be agreed upon in writing via a joint ownership agreement.



- Each participant will be responsible for examining the possibility to protect any results, which can be expected to be commercially or industrially exploited. When deciding on protection, the participant must consider its own legitimate interests and the interests of the other participants. Participants will ensure that adequate steps towards protection are taken prior to DEC activities, preventing unapproved public disclosure of results, tools, products and services.
- Access rights to results will be granted on a royalty-free basis for further research, and on fair and reasonable conditions if needed for commercial exploitation.
- Authors, affiliated institutions, publishers will receive appropriate credit via a persistent data identifier (for authorships guidelines see also: [DFG](#)² and the [CRediT tool](#)³). Relevant metadata would include (but may not be limited to): Dataset Persistent ID (i.e. DOI), Publication Date, Title, Author Name, Contact of Data Author, Description, Subject, Associated Keyword, Notes, Depositor Name, Deposit Date.

4 Resources, data security and ethics

The financial and technical resources of the SustInAfrica consortium are sufficient to ensure proper implementation and application of the data management procedures described in this document. Data storage is described in section 2.2 and supported by Luke`s management support team and ICT teams. Accessing data requires a registration and the use of a user name and password.

Protection of personal data in the project will be regulated by the General Data Protection Regulation (GDPR, Regulation (EU) 2016/679) that is binding for all EJP SOIL beneficiaries from the European Union and non-EU countries (see also D8.2).

²https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gw_p_en.pdf

³ <https://casrai.org/credit/>

