



NESTLER

**One hHealth SusTainability partnership between
EU-AFRICA for food security**

Deliverable D7.2

Data Management Plan

Authors	P. Nyabinwa, J. Munyemana, A. Ndayambaje
Nature	Report
Dissemination	PU-Public
Version	1.0
Status	Draft
Delivery Date (DoA)	M04
Actual Delivery Date	31/01/2023

Keywords	Data Management Plan, Repository, open science, DOI, Metadata, FAIR data
Abstract	This Deliverable provides the NESTLER Data Management Plan (DMP) version 1. This Task outlines how the research data collected or generated by the NESTLER Consortium will be handled during and after the end of the project. It describes which standards and methodology for data collection and generation will be followed, and whether and how the data will be shared/made open access. It will also describe how best practices in terms of metadata and archiving will be used to ensure that the data will be findable, accessible, interoperable, and reusable for other potential users. Moreover, the DMP provides information about what datasets the consortium is aiming to preserve and in which format. Legal and ethical issues related to the NESTLER's collecting and/or processing of personal data are identified and practically considered, taking into consideration the different methods by which data are collected such as interviews, online surveys, workshops, questionnaires, etc.).



DISCLAIMER

This document is a deliverable of the NESTLER project funded by the European Union under Grant Agreement no.101060762. 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, while neither the European Union nor the granting authority can be held responsible for any use of this content.

This document may contain material, which is the copyright of certain NESTLER consortium parties, and may not be reproduced or copied without permission. All NESTLER consortium parties have agreed to the full publication of this document. The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the NESTLER consortium as a whole, nor a certain party of the NESTLER consortium warrant that the information contained in this document is capable of use, nor that use of the information is free from risk and does not accept any liability for loss or damage suffered using this information.

	Participant organisation name	Short	Country
01	SYNELIXIS SOLUTIONS S.A.	SYN	EL
02	CloudEO AG	CEO	DE
03	RINIGARD DOO ZA USLUGE	RINI	HR
04	EBOS TECHNOLOGIES LIMITED	eBOS	CY
05	STICHTING IDH SUSTAINABLE TRADE INITIATIVE	IDH	NL
06	ZANASI ALESSANDRO SRL	Z&P	IT
07	AGRIX TECH SARL	AGRI	CM
08	CONSERVATION THROUGH PUBLIC HEALTH	CTPH	UG
09	THE INTERNATIONAL CENTRE OF INSECT PHYSIOLOGY AND ECOLOGY LBG	ICIPE	KE
10	ETHIOPIAN INSTITUTE OF AGRICULTURAL RESEARCH	EIAR	ET
11	RWANDA AGRICULTURE AND ANIMAL RESOURCES DEVELOPMENT BOARD	RAB	RW
12	INTERNATIONAL INSTITUTE OF TROPICAL AGRICULTURE	IITA	NG
13	MANA BIOSYSTEMS LIMITED	MANA	UK
14	UNIVERSITY COLLEGE LONDON	UCL	UK

ACKNOWLEDGEMENT

This document is a deliverable of NESTLER project. This project has received funding from the European Union's Horizon Research and innovation programme under grant agreement N° 101060762. 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, while neither the European Union nor the granting authority can be held responsible for any use that may be made of the information it contains.

Document History

Version	Date	Contributor(s)	Description
V0.1	16/12/2022	RAB	Initial Version
V0.2	29/12/2022	RAB	Integration of other inputs
V0.3	26/01/2023	RAB	Feedback integration
V0.4	31/01/2023	RAB	Feedback integration

Document Reviewers

Date	Reviewer's name	Affiliation
11/01/2023	Stavroula Bourou	SYN
24/01/2023	Rika Lostarakou	Project Ethics Advisor
24/01/2023	Eniola	IDH
26/01/2023	Paulina Polak	RiniGARD
26/01/2023	Georgia Pantelide	eBOS
30/01/2023	Alkyoni	Cloudeo

Table of Contents

Definitions, Acronyms and Abbreviations	5
Executive Summary.....	6
1. Introduction	7
2. Data Summary.....	8
2.1 Data Capture.....	8
2.2 Organize data.....	9
2.3 Assure data quality	9
2.4 Access data	9
2.5 Data sharing and preservation	10
2.6 Data security.....	10
3. Data that will be used in the project	11
3.1 Missions	11
3.1.1 Multispectral Instruments Inputs.....	11
3.1.2 SAR inputs.....	11
3.1.3 Weather	11
3.2 Climate data.....	11
3.3 Elevation	13
3.4 Connectivity	14
3.5 Historical cases studies	14
3.6 Risks on food security roadmap	15
4. FAIR (Findable, Accessible, interoperable and reusable) data	16
4.1 Making data findable, including provisions for metadata.....	16
4.2 Making data accessible	16
4.2.1 Repository	16
4.2.2 Data	17
4.2.3 Metadata	18
4.3 Making data interoperable	18
4.4 Increase data re-use	19
5. Other research outputs	20
6. Allocation of resources	21
7. Data security	22
8. Ethics	23
9. Conclusion.....	24
10. References	25

Definitions, Acronyms and Abbreviations

AgMES	Agricultural Metadata Element Set
CERIF	Common European Research Information Format
CIM	Common Information Model
DMP	Data Management Plan
DOI	Digital Object Identifier
DPO	Data Protection Officer
FAIR	Findable, Accessible, Interoperable and Reuse
GDPR	General Data Protection Regulation
GIS	Geographic Information System
GRASS	Geographic Resources Analysis Support System
LGPL	Lesser General Public License
MPL	Mozilla Public License
OSS	Open-Source Software
QGIS	Quantum Geographic Information System
RDBMS	Relational Database Management System
SQL	Structured Query Language
CC0	Creative Commons Public Domain Dedication

Executive Summary

The purpose of the NESTLER Data Management Plan (DMP) is to define key elements that will facilitate the potential reuse of the data collected and processed during and after NESTLER. Additionally, the purpose of DMP is also to describe the data, its intended use, how will be managed, stored. Therefore, the DMP will ensure that the data will be findable, preferably via Digital Object Identifier (DOI); accessible; assessable; and intelligible; re/usable beyond the original purpose for which it was collected and interoperable to specific quality standards, in accordance with the Horizon Europe Open Research Data pilot. Mostly, Consortium will be the first candidates for open data re-use. NESTLER will offer open access to the data gathered through the process of evaluation of project results based on the pilots. Special care will be taken to preserve anonymity as the interest is in providing scientists with valuable data while not disclosing personal information. Additionally, the purpose of DMP is also to describe the data, its intended use, how will be managed, and stored.

In summary, this deliverable provides the NESTLER Data Management Plan (DMP) version 1. This deliverable outlines how the research data collected or generated by the NESTLER Consortium will be handled during and after the end of the project. It describes which standards and methodology for data collection and generation will be followed, and whether and how the data will be shared/made open access. It will also describe how best practices in terms of metadata and archiving will be used to ensure that the data will be findable, accessible, interoperable, and reusable for other potential users. Moreover, the DMP provides information about what datasets the consortium is aiming to preserve and in which format. Legal and ethical issues related to the NESTLER's collecting and/or processing of personal data are identified and practically considered, taking into consideration the different methods by which data are collected such as interviews, online surveys, workshops, questionnaires, etc.).

The NESTLER project will develop an open access platform where training material, online lessons, guidelines and tools will be available for partners. Besides, the NESTLER methodology will be integrated in the platform with the objective to design a specific capacity building program for every partner accessing the NESTLER toolkit, which will result in ad-hoc training adapted to the partner's needs and opportunities. The NESTLER methodology and the toolkit will be both tested and validated in 6 pilot partners. The DMP will allow these data to be aligned with the Horizon Europe open Science, for which NESTLER opted in.

1. Introduction

NESTLER is a joint project between the EU and African member states designed to promote One-Health sustainable partnership. The project aims to bring together interdisciplinary technological advances to effectively monitor the well-being of animals, plants, and humans in a holistic approach. NESTLER project brings together the vision of eradicating hunger and promote One-Health programme initiatives. To achieve this vision, the project consortium partners recognize the importance and significance of handling the research data during and after the end of the project, data to be collected and processed, the methodology and standards to be applied, whether data will be shared or made open access and how data will be curated and preserved including after the end of the project.

It should be noted that this DMP is a “living document” where information will be continuously added and revised as the implementation progresses. But, it is the responsibility of each partner to notify the NESTLER’s Coordinator or DPO of changes in the data they are collecting during the project.

2. Data Summary

This section provides a brief description of the main datasets identified so far by the NESTLER consortium that will be collected and processed within the project to evaluate the project outcomes. **Error! Reference source not found.** summarizes the type, format and estimated size of the data. Note that this is a initial list of datasets that may be expanded during the project as the field trial needs are further clarified.

Table 1. Type, Format and Estimated Size of data collected within NESTLER

Type	Details - Format	Estimated Size
System requirements. Design notes	reports (.docx, .xls, .csv), photos (.jpeg, .png) training videos (.mp4, .mov, .avi)	Depending on the file format (> 50 GB)
Heterogeneous measurements of IoT devices	Air, soil, leaf, chemical compositions of crops, location, temperature, heart rates of animals JSON, XML, text files (.json, .csv, .txt)	> 600MB
Visual/ (multi-)spectral images	Drones, Copernicus, Satellite imagery. videos (.mp4, .mov, .avi), photos (.jpeg, .png)	Depending on experiment duration (10-50 GB)
Machine Learning training models – AI methodology and results	ML algorithms and trained models as reports (.docx, .xls, .csv), or in JSON, XML	>250 MB
Demonstration material	videos (.mp4, .mov, .avi), photos (.jpeg, .png)	> 50 GB
Use case validation	Policies, reports (.docx, .xls, .csv)	> 800 MB

This section also describes the management of NESTLER data through the types of data that will be generated or gathered during the project, the standards that will be used, the ways that the data will be exploited, shared for verification or reuse, and how data will be preserved (Figure 1).

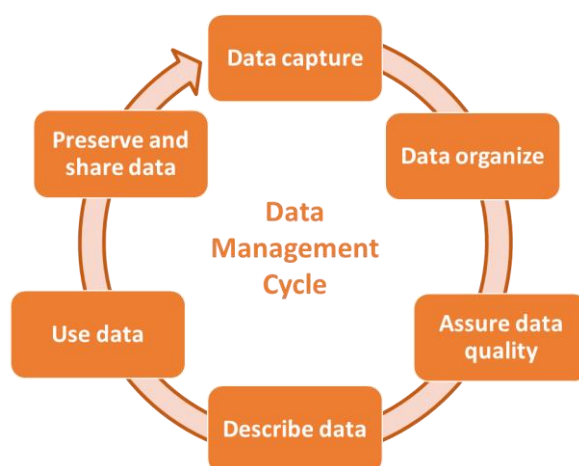


Figure 1. Data Management Cycle

2.1 Data Capture

Most of the NESTLER’s routine data collection will be done via the web-based application software that is set up in servers hosted at NESTLER server clusters. This advanced technology provides

excellent environmental and data security conditions for continuous data entry and use. The software also ensures secure access to data based on role-based user profiles and secure individual passwords. Backups are scheduled nightly to an off-site server in the NESTLER small data centre. Based on the objectives of NESTLER project, the types of data will be collected are both structured and unstructured data. structured data refers to data that has a defined length and format (Examples include numbers, dates, and groups of words and numbers called strings). Unstructured data is data that does not follow a specified format (satellite images, photographs and videos and sonar data). Due to the volume of data, we will need the right amount of computational power and speed as well as the storage capacity

2.2 Organize data

A growing amount of data comes from a variety of sources that aren't quite as organized or straightforward. And because of the large amount of data generated for every second, the use of relational database management systems (RDBMS) for linked data tables like Oracle or MySQL, or a Geographic Information System (GIS) for geospatial data layers like ArcGIS, GRASS, or QGIS will be required.

2.3 Assure data quality

This refers to the messiness or trustworthiness of the data. With many forms of big data quality and accuracy are less controllable but big data technology now allows us to work with this type of data. The NESTLER project partners will be required to make routine review in order to validate and verify data that will be reported. The purpose of the routine data quality assessment is to ensure that data collected are reliable, valid, complete, comparable, and timely.

The validated and verified data will improve reporting and will provide project partners with assurance that data are credible and consistently collected and reported in accordance with standard procedures and guidelines

2.4 Access data

Various categories of people can interact with NESTLER database. Let us briefly describe the most important.

- **The database administrator** is the person responsible for the design, control and administration of the database. The database administrator has the task of mediating among the various requirements, often conflicting, expressed by the users, ensuring centralized control over the data. In particular, he or she is responsible for guaranteeing services, ensuring the reliability of the system, and managing the authorizations for access to the data.
- **The application designers and programmers** define and create programs that access the database. They use the data manipulation language or various support tools for the generation of interfaces for the database such as allowing the user to edit, export, and enter the records
- **The users** employ the database for their own activities. Each user in the project platform, including the data sets accessed, is registered in order to track and detect harmful behaviour of users with access to the platform. A public API will be provided to registered users allowing them the access to the platform. The database compliance aims to ensure the correct implementation of the security policy on the databases verifying vulnerability and incorrect data. The target is to identify excessive rights granted to users, too simple passwords (or even the lack of password) and finally to perform an analysis of the entire database

2.5 Data sharing and preservation

The digital data created by the project will be diversely curated depending on the sharing policies attached to it. For both open and non-open data, the aim is to preserve the data and make it readily available to the interested parties for the whole duration of the project and beyond.

2.6 Data security

Data will be kept on the responsible partner's storage system for the duration of the project. Each partner is in charge of making sure the data are stored legally, safely, and securely in accordance with all EU data protection legislation. All data files will be transmitted over secure connections while being password- and encryption-protected.

3. Data that will be used in the project

3.1 Missions

3.1.1 Multispectral Instruments Inputs

Name	Instrument	Resolution	Bands	Level	Coverage	Cadence	Format
Sentinel-2	MSI	10m	B1-12	L2	World	5 days	SAFE

3.1.2 SAR inputs

Name	Instrument	Resolution	Bands	Level	Coverage	Cadence	Format
Sentinel-2	IW	Varies	C	L1	World	6 days	SAFE

3.1.3 Weather

Name	Instrument	Level	Coverage	Cadence	Format
GOES-FP	SEISS	L1	World	Daily	netCDF
MeteoSAT	SEVIRI, MVIRI	L2	EU, Africa	Daily	netCDF
Sentinel-3	SLSTR	L2	World	4 days	netCDF

3.2 Climate data

Source	Variable	Unit	Resolution	Temp Res	Format
Worldclim.org	Minimum Temp	C	2.5 min	Month	GEOTIFF
	Maximum Temp	C		Month	
	Average Temp	C		Month	
	Precipitation	mm		Month	
	Solar Radiation	kJ m ⁻² day ⁻¹		Month	
	Wind Speed	m s ⁻¹		Month	
	Water vapor pressure	kPa		Month	

Source	Variable	Unit	Resolution	Temp Res	Format
AgERA5 (C3S)	Wind Speed	m s ⁻¹	10m	Daily	netCDF
	Dewpoint Temperature	K	0.1°	Daily	netCDF
	Temperature	K	0.1°	Daily	netCDF
	Cloud Cover	-	0.1°	Daily	netCDF
	Precipitation Flux	mm day ⁻¹	0.1°	Daily	netCDF
	Snow Thickness	Cubic cm	0.1°	Daily	netCDF
	Solar Radiation Flux	J m ⁻² day ⁻¹	0.1°	Daily	netCDF

Source	Variable	Unit	Resolution	Temp Res	Format
Answr	Cloud Cover	%	Km	Month	JSON
	Consecutive Dry Days	# Days	Km	Month	JSON
	Consecutive Frost Days	# Days	Km	Month	JSON
	Consecutive Summer Days	# Days	Km	Month	JSON
	Consecutive Wet Days	# Days	Km	Month	JSON
	Dew Point Temperature	C	Km	Month	JSON
	Frost Days	# Days	Km	Month	JSON
	Heating Days	# Days	Km	Month	JSON
	Ice Days	# Days	Km	Month	JSON
	Maximum Temperature	C	Km	Month	JSON
	Mean Temperature	C	Km	Month	JSON
	Minimum Temperature	C	Km	Month	JSON
	Precipitation Days (10mm)	Mm	Km	Month	JSON
	Precipitation Days (50mm)	Mm	Km	Month	JSON
	Precipitation Flux	mm	Km	Month	JSON
	Snow Thickness	cm	Km	Month	JSON
	Solar Radiation	Watt * sqkm-1	Km	Month	JSON
Volumetric Soil Moisture	%	Km	Month	JSON	

Source	Variable	Severity	Probability	Resolution	Temp Res	Format
Answr	Cold wave risk	L,M,H	%	Km	Month	JSON
	Drought Probability	%		Km	Month	JSON
	Flood Severity	Meters	N/A	Km	Month	JSON
	Heat Wave Probability	% LoO		Km	Month	JSON
	Wind Storm Probability	% LoO		Km	Month	JSON

3.3 Elevation

Name	Projection	H. Datum	V. Datum	Unit	Resolution	Coverage	Format
SRTM	Geo	WGS84	EGM96	Meter	1 arc sec	World	GEOTIFF

3.4 Connectivity

APIs and cloud repository connections will be used to facilitate access to the satellite imagery. Datasets that are not available via APIs will be download and stored in the cloud using AWS S3 Bucket, which includes access to all AWS tools. Datasets that are available via the API through the bucket are following:

Dataset	Access	Location
Sentinel-1	API	https://scihub.copernicus.eu/
Sentinel-2	API	https://scihub.copernicus.eu/
Sentinel-3	API	https://scihub.copernicus.eu/
GOES-FP	API	https://power.larc.nasa.gov/docs/services/api/
MeteoSAT	API	http://api.eumetsat.int/data/download/
WorldClim.org	Cloud	https://worldclim.org ; https://docs.aws.amazon.com/AmazonS3/latest/API/Type_API_Reference.html
AgERA5	API	https://cds.climate.copernicus.eu/
Answer	API	https://docs.answer.space/
SRTM	API, Cloud	https://opentopography.org/developers
Dataset	Access	Location

3.5 Historical cases studies

History and case study review are two diachronic research strategies that will commonly be used by partners to create a repository of historical case-studies and challenges that resulted in the disruption of food supply from Africa to EU (and vice-versa). The following is NESTLER- Working format for bulking information on cases-studies that have affected food supply chain between EU-Africa member states.

Description	Use Case 1	Use Case 2	Use Case 3
Case No			
Country			
Case type/ name*			
Date/year or season of occurrence of the case			
Region/ area affected			
Affected category (crop, livestock, poultry, fishery, etc)			
Severity for the affected category (estimated loss of the product)			
Effect on the ecosystem			
Severity for humans			
Effects on economy			
Description of the case			
Possible mitigation actions/ countermeasures			
Possible prevention measures			

3.6 Risks on food security roadmap

This will help in gathering insights into the challenges often encountered by the farmers during cultivation (to be include in NESTLER Deliverable D1.1). The following template will be used:

Stakeholder information			
Country			
Type of stakeholder			
Role of stakeholder			
Challenges			
Description of the challenge			
Type of crop			
Name of the particular crop			
Region			
Season			
Severity			
Effects to the Ecosystem			
Severity for Human			
Effects to the Economy			
Possible Mitigation actions			
Possible ways to prevent it			

4. FAIR (Findable, Accessible, interoperable and reusable) data

4.1 Making data findable, including provisions for metadata

All data will have an associated metadata document which describes key aspects of the data. Event listings are stored in a central spreadsheet and individual events are assigned a unique identifier of the formal format in order to avoid any confusion with contributions, deliverables and internal documents. Therefore, official deliverables and internal documents/reports will have the following formal format:

NESTLER_Dw.d_ACR_Vx.y_YYYYMMDD.ext and
NESTLER_Ww_TTT_ACR_Vx.y_YYYYMMDD.ext respectively.

Where

w: is the work package number,
d: is the deliverable number,
ACR: is the partner acronym (e.g. SYN, CEO, RAB, eBOS,),
x: is the version major number,
y: is the version minor number,
YYYY: is the year,
MM: is the month,
DD: is the day,
TTT: topic title,
ext: is the extension (.docx, .pdf, .pptx, .xlsx, .zip).

In more details, the partners acronyms to be used are: SYN, CEO, RINI, EBOS, IDH, Z&P, AGRI, CTPH, ICIPe, EIAR, RAB, IITA, MANA and UCL. Photographs and audio/visual recordings are:

NESTLER [event] [date of event] [description of event/ content e.g. workshop/WP meeting].

NESTLER provides search keywords in the metadata to optimize the possibility for discovery and then potential re-use.

4.2 Making data accessible

4.2.1 Repository

NESTLER will provide early access to its results, through informal publications to reputable repositories, including the Computing Research Repository (CoRR) (a partnership of ACM, arXiv.org and other), and the Electronic Colloquium on Computational Complexity (ECCC). Many repositories (e.g., IEEE, MDPI, ACM, DBLP, Google Scholar) index publications of these repositories increasing their outreach. Additionally, most of the developments of the academic partners will be made available at open source software, published in Github and OFair Data Marketplace. The open-source strategy foresees offering OSS results in a business-friendly way, thus NESTLER commits to selecting a business-friendly license (MPL/LGPL), whereas consortium partners already have a track record of OSS contributions and experience.

Also, the project's datasets will be anonymized and made available as open data in various FAIR repositories (e.g., EOSC, re3data.org, DataHub) and OFair Data Marketplace. The project will also share the constructed AI models and data analysis flows in open repositories (e.g., OpenML). General awareness and wider access to the NESTLER research data will be ensured by including the repository in registries of scientific repositories. DataCite [1] offers access to data via DOI and metadata search

and has recently merged with Re3data [2] and Databib [3], the most popular registries for digital repositories. These sites are now collaborating to provide open research data services. The repositories and platforms that provide public access to all information products are backed up locally and off-site (cloud). Images and photos, audio, video, presentations from training/workshops or other events that involved NESTLER Project are published on social media platforms (Twitter, LinkedIn, facebook, etc) and where possible on project's website.

4.2.2 Data

The data will be made openly available. In contrast, the only data which will not be made openly accessible will be data which contains personally identifiable information (e.g. individual evaluation forms) and data underlines deliverables that are covered by confidentiality. The personal data processed in the project are not made publicly accessible but kept closed and inaccessible to third parties. Furthermore, Data will be published using standard file formats (Acrobat PDF/A: .pdf; Comma-separated values: .csv; Open Office formats: .odt, .ods, .odp; plain text: .txt; and XML: .xml). All data will be accessed using standard tools. Software relevant to access the data would be made available, but it is not seen as being a requirement. Should it be needed we will provide the required open source to access and analyse the data.

For this, NESTLER consortium is committed towards contribution to open science. The project outcomes generated in collaboration with partners will rely on the knowledge reuse and transparency for promoting the outcomes to EU, research communities, and citizens based on the following Open Science practices:

- (i). **Open Access Publications:** Open Access refers to a practice of giving access to all scholarly disciplines information that is free of charge to the end-user. In this way data becomes re-usable, and the benefit of public investment in the research will be improved. Open access to all peer-reviewed scientific publications of the project will be provided. Selected publications will be made available with the highest standard (Gold Open Access). The rest publications will be made open in the project's website and in OpenAIRE's Zenodo.org.
- (ii). **Open Peer Review:** NESTLER will allow self-selected reviewers to provide comments on the project's scientific outputs (i.e., publications, blueprints), beyond reviewers selected by the Open Access journals (or other forums), where the results will be made available. Likewise, NESTLER partners will participate as open peer reviewers to open peer reviews of results from related projects such as projects funded under the same call.
- (iii). **Open Research Europe Publishing Platform (OREPP):** NESTLER will publish a minimum of three (≥ 3) articles in the OREPP, including one project overview article at the beginning of the project and a concluding article providing a summary of the project's main research offerings and achievements at the end of the project.
- (iv). **Open Innovation:** NESTLER will promote open innovation in national/international networks. It will utilize collaboration with partners' academic and business networks, as well as with projects from other Horizon EU clusters, and collaborate with stakeholders across diverse domains (e.g., health, environment, ICT), aiming to adapt and/or extend the developed systems and techniques whilst sharing ideas, knowledge, for their adoption.
- (v). **Open Citizen Science:** NESTLER will reinforce citizens knowledge of research upon climate change and its human health impact by exposing incentive programs and material, ensuring citizens' involvement, whereas via the trinity visualization citizens will adjust their prior awareness and strengthen their research knowledge. Also, liaisons with data marketplaces from

partners academic/business networks will be supported, engaging environmental/health authorities, SMEs, and citizens for using the project's results. Also, NESTLER's co-creation methodology will engage third-party individuals and stakeholders in the development, validation, and evaluation of project's outcomes. For the duration of the project, personal data will be stored on local secured server of the partner responsible of taking care of this.

4.2.3 Metadata

Metadata of deposited publications must be open under a Creative Commons Public Domain Dedication (CC0) or equivalent legal tool, in accordance with the FAIR principles, particularly the requirement for machine-actionability. Following the Horizon Europe guidelines [4] The metadata should include at least the following information:

- publication (author(s))
- title
- date of publication
- publication venue
- grant project name
- acronym and number
- licensing terms
- persistent identifiers for the publication
- the authors involved in the action and their organisations
- the grant.

Where applicable, the metadata must include persistent identifiers for any research output or any other tools and instruments needed to validate the conclusions of the publication.

4.3 Making data interoperable

The data driven NESTLER platform will enable ingestion of heterogeneous data points from relevant sensors and offer advanced AI enabled capabilities to improve the quality of food production. The platform will incorporate the use of Big-data analytics framework components interfaced with streaming endpoints to ingest realtime sensory information collected from several data sources. The ingested information resources will be further subjected to enhancements using advanced algorithms capable of extracting intelligence and assist in the economic sustainability of the food producers.

The overall novelty of NESTLER platform relies in the integration of heterogenous data formats generated and captured such as (satellite images, drone images, images from fixed cameras, location information, device for monitoring crops, and data collected from humans using smartphone) from different and disparate data sources as highlighted in Figure 2. The ingestion of these different data sources will be processed by the NESTLER backend services, powered by AI and ML tools. The backend service implementation will include the design and development of GIS systems, which allows for accurate mapping and data models obtained from sensors to be plotted against geographical maps. As the project progresses and data is identified and collected, further information on making data interoperable will be outlined in subsequent versions of the DMP. But, Consortium will focus on making the data interoperable on the metadata that will be provided. In the project we use standard file formats as noted above in 3.2.2 to make NESTLER data interoperable to allow data exchange and re-use within and across disciplines.

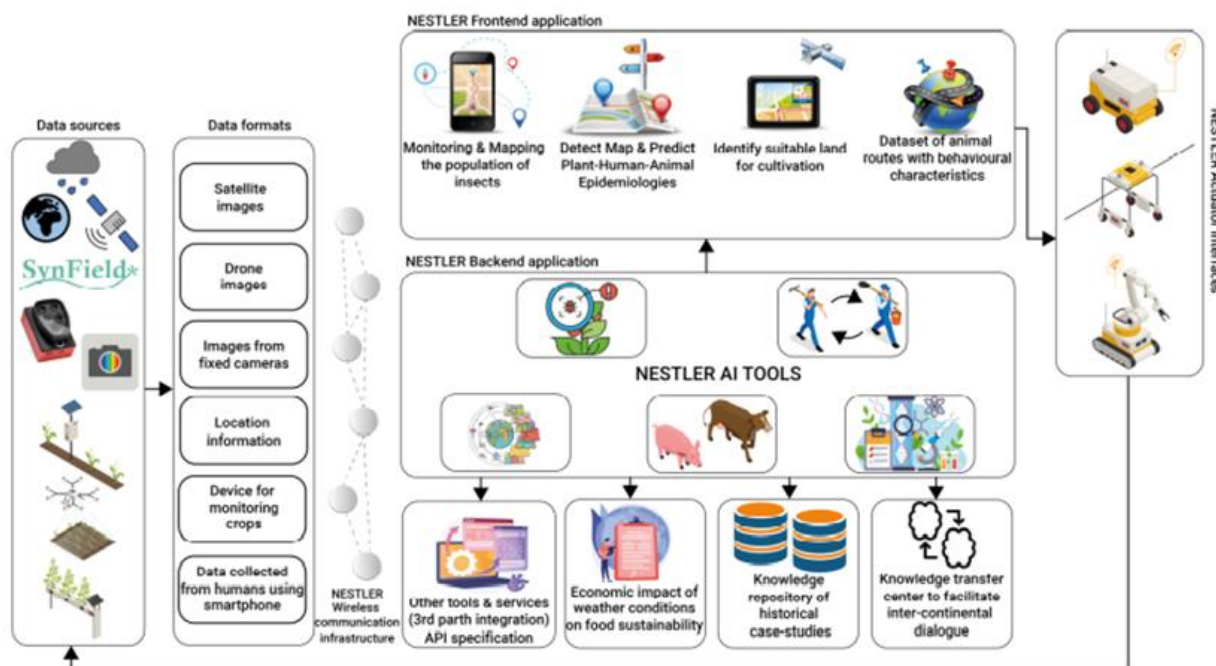


Figure 3. NESTLER conceptual design and data workflow

4.4 Increase data re-use

All Personal Identifiable Information will be restricted to internal usage and not going to be shared with third parties. For shared information, standard format, open source software, and proper documentation will guarantee re-usability by third parties. Creative Commons Licenses such as CC BY-NC will be used for all data to be preserved and reusers to distribute, remix, adapt, and build upon the material in any medium or format for noncommercial purposes only. Furthermore, a quality assurance process will be pursued within the duration of the project.

All data generated and collected during NESTLER can be categorized into types that will be used by different groups but the Consortium will be the first candidates for open data re-use:

- Data from pilots regarding the effectiveness of NESTLER technology intervention for securing food safety will be exploitable by the farmers, food supply chain service providers, logistics operators, governmental agencies, healthcare professionals, and other stakeholders of One Health programme.
- Data from insect production will be exploitable by the suppliers to feed industry, insect protein manufacturers, healthcare professionals, regional and national healthcare authorities, and farmers.
- Scientific results from all areas tackled by NESTLER will be exploitable by the scientists from Big-data analytics, deep-learning communities, IoT infrastructure operators, environmental sensor manufacturers, healthcare professionals, intelligent transport system developers, mobility service providers and routing service operators.

5. Other research outputs

In addition to the management of data, beneficiaries will also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects/pilot demonstration sites.

Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.). Also, Most of the developments of the academic partners will be made available at open- source software, published in Gitlab [5] and OFair Data Marketplace.

The open-source strategy foresees offering Open-Source Software (OSS) results in a business-friendly way, thus NESTLER commits to selecting a business-friendly license (MPL 2.0 [6]/LGPL [7]), whereas consortium partners already have a track record of OSS contributions and experience.

6. Allocation of resources

Cost for making data FAIR are estimated to be zero, which means archiving data in the NESTLER repository will be free of charge. However, repositories of specific partners will not be free. Consortium partners may use where possible their own budgets to archive personal data in their own repositories during their retention time period.

Maybe it would be better that each partner of the consortium (that generates data) will be responsible for preparing the datasets to be FAIR. Furthermore, consortium partners have the responsibility to make sure their activities are in line with all applicable local, government and international laws, regulations and guidelines. Length of time for which the de-identified data will remain re-usable is at least 10 years after the end of the project.

RAB will be responsible for data management in NESTLER project, while SYN will provide the Project Coordination, Technical Management and IPR Management.

The NESTLER Data Protection Officer (DPO) is **Dr. Pascal Nyabinwa** from RAB. Dr. Nyabinwa is a RAB staff with 12 years of experience in animal resources research and extension. Though he is very experienced, he will contact the EC and national supervisory authorities whenever needed.

7. Data security

With regards to personal data, the Consortium shall ensure that data on individuals are transmitted and used in a secure environment; that the use of the data complies with ethical and legal requirements (Ref NESTLER D8.1) and that the use of both existing and new data is agreed with the data provider/owner. Data records containing personal data will be managed in accordance with the General Data Protection Regulation (GDPR, EU: 2016/679).

NESTLER will take measures to preserve anonymity and appropriately curate the collected data. The aim is to gather data valuable for technological and scientific evaluation of the project achievements respecting privacy-related issues and legislation. To enable further validation, mining and re-use, the collected data will be annotated, and the metadata will also be published along with the pilot data. The aim is to render the research data discoverable, accessible, assessable, and intelligible, usable beyond the original purpose for which it was collected and interoperable to specific quality standards. The NESTLER project activities will evaluate and implement issues related to data protection & privacy and evaluate informed consent (to guarantee the voluntary participation in research as it is one of the most important procedures to address privacy issues in research).

For the duration of the project, datasets will be stored on the responsible partner's storage system. Every partner is responsible to ensure that the data are stored safely and securely and in full compliance with European Union data protection laws. Adequate institutional level network security will be applied, including security systems, firewalls, and safe storage places. All data files will be transmitted over secure connections while being password- and encryption-protected.

After the end of the project, the project's datasets will be anonymized and stored in various FAIR repositories (e.g., EOSC, re3data.org, DataHub) and Fair Data Marketplace.

8. Ethics

Data protection and good research ethics are major topics for the consortium of this project. NESTLER partners will comply with the ethical principles as set out in Article 34 of the Grant Agreement which states that all activities must be carried out in compliance with: • Ethical principles (including the highest standards of research integrity as set out, for instance, in the European Code of Conduct for Research Integrity and including, in particular, avoiding fabrication, falsification, plagiarism or other research misconduct).

With regards to personal data, the Consortium shall ensure that the use of the data complies with ethical and legal requirements (Ref. D8.2 [8]) and that the use of both existing and new data is agreed with the data provider/owner. Data records containing personal data will be managed in accordance with the General Data Protection Regulation (GDPR, EU: 2016/679).

The NESTLER project activities will evaluate and implement issues related to data protection & privacy and address informed consent procedure for communication with stakeholders in order to guarantee the voluntary participation in research as it is one of the most important procedures to address privacy issues in research (Ref D8.1 [9]).

The NESTLER DPO will also lead the Privacy, Ethical, Legal & Regulatory Compliance monitoring tasks. NESTLER DPO will be compliant to the GDPR (EU 2016/679, EU 2016/680). More specifically, he will ensure that the following rules as described at the NESTLER project Grant Agreement are followed:

- ➔ Personal Data are properly anonymized/pseudo-anonymized and processed legally and fairly (Refer to D8.2 [8])
- ➔ It must be collected for explicit and legitimate purposes and used accordingly
- ➔ It must be adequate, relevant and not excessive in relation to the purposes for which it is collected and/or further processed
- ➔ It must be accurate, and updated where necessary
- ➔ Each pilot will assign an Ethical & Ecosystem Chair (Pilot Data Controller), who must ensure that data subjects can rectify, remove or block incorrect data about themselves
- ➔ Data that identifies individuals (personal data) must not be kept any longer than strictly necessary and always in an encrypted format
- ➔ Data controllers must protect personal data against accidental or unlawful destruction, loss, alteration and disclosure, particularly when processing involves data transmission over networks. They shall implement the appropriate security measures.

9. Conclusion

The NESTLER Data Management Plan (DMP) is this deliverable (D7.2), which describes the way in which the NESTLER consortium will manage, store, secure and retrieve, disseminate, archive and dispose of the datasets that will emerge from the project for the purposes for evaluating the project outcomes, and how best practices in terms of metadata and archiving will be used to ensure that the data will be findable, accessible, interoperable, and reusable (FAIR) for other potential users. Furthermore, the DMP provides information about what datasets the consortium is aiming to preserve and in which format. The DMP will allow these data to be aligned with the Horizon Europe Open Science, for which NESTLER opted in.

10. References

- [1] “DataCite,” [Online]. Available: <http://www.datacite.org>.
- [2] “Registry of Research Data Repositories (Re3data),” [Online]. Available: <http://www.re3data.org/>.
- [3] “Databib,” [Online]. Available: <http://databib.org/>.
- [4] European Commission, “DMP template for Horizon Europe,” European Commission, [Online]. Available: <https://rdm.mpdl.mpg.de/2022/04/04/data-management-plans-for-horizon-europe-and-erc-grants>.
- [5] “GitLab: the DevSecOps platform,” [Online]. Available: <https://about.gitlab.com/>.
- [6] “MPL 2.0,” [Online]. Available: <https://www.mozilla.org/en-US/MPL/2.0/FAQ/>.
- [7] “GNU Lesser General Public License,” [Online]. Available: <https://www.gnu.org/licenses/lgpl-3.0.html>.
- [8] NESTLER Consortium, POPD – Requirement No. 2, 30 Dec. 2022.
- [9] NESTLER Consortium, H – Requirement No. 1, 30 November 2022.