This project has received funding from the European Union's Horizon 2022 Research & Innovation Actions – Project. 101086355 – HORIZON-CL6-2022-GOVERNANCE-01

OPEN SCIENCE AND DATA MANAGEMENT PLAN, V1

Updated version, 27 January 2025

Author		
Name	**Organization**	**Draft release date**
Koen Van Rossum	VITO (coordinator)	15/09/2023
Panos Ilias	ILVO	15/09/2023
Nuno Grosso	Deimos	15/09/2023
Valantis Tsiakos	ICCS	15/09/2023

Approval on behalf of the Executive Board		
Name	**Organization**	**Date of approval**
Marios Vlachos	ICCS	21/09/2023
Maria P. Gonzalez-Dugo	IFAPA	21/09/2023
Isabelle Piccard	VITO	26/09/2023

Revision records			
Version	**Date**	**Changes**	**Authors**
1.0	27/09/2023	Original document	VITO
1.1	27/01/2025	Updated version of the document, outlining how the feedback from the EC and external reviewers on version 1.0 of the document has been addressed.	VITO

Acronyms and Abbreviations

Acronyms and Abbreviations	
AI	Artificial Intelligence
AgroSTAC	global repository on key agronomy observations
BIOPAR	Biophysical Parameters
CA	Consortium Agreement
CEADS	Common European Agricultural Data Space
CSV	Comma Separated Values
CSW	Catalogue Service for the Web
DES	Deimos Spain
DGA	Data Governance Act
DME	DEIMOS ENGENHARIA SA
DSS	Decision Support System
DSSC	Data Space Support Centre
EC	Electrical Conductivity
EnMAP	Environmental Mapping and Analysis Program
EO	Earth Observation
EOSC	European Open Science Cloud
ERP	Enterprise Resource Planning
ESA	European Space Agency
ET	EvapoTranspiration
ETa	Actual EvapoTranspiration
EU	European Union
EV ILVO	Eigen Vermogen van het Instituut voor Landbouw en Visserij Onderzoek
FAIR	Findable, Accessible, Interoperable, and Reusable
Fcover	Fraction of vegetation cover
GA	Grant Agreement
GB	Giga Byte
GEOGLAM	Group on Earth Obserations Global Agricultural Monitoring Initiative
GIS	Geographic Information Systems
fAPAR	fraction of Absorbed Photosynthetically Active Radiation
GDPR	General Data Protection Regulation
GeoJSON	Geospatial Java Script Object Notation
GeoTIFF	Geographic Tagged Image File Format
ICCS	Institute of Communications and Computer Systems
iCM	Integrated Crop Management
IoT	Internet of Things

IPR	Intellectual Property Rights
LAI	Leaf Area Index
LUE	Light Use Efficiency
ML	Machine Learning
MSI	MultiSpectral Instrument
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NIR	Near Infra Red
NP	NeuroPublic
OAuth	Open Authorization protocol
OGC	Open Geospatial Consortium
OpenAIRE	Open Access Infrastructure for Research in Europe
OpenDEI	Open Digitizing European Industry
PPP	Plant Protection Product
PSNC	Instytut Chemii Bioorganicznej Polskiej Akademii Nauk
R&D	Research and Development
R&I	Research and Innovation
RIL	Research and Innovation Lab
RS	Remote Sensing
SciHub	Scientific Hub
SITRA	Finnish innovation fund
SOC	Soil Organic Carbon
STAC	Spatio Temporal Assets Catalog
VITO	Vlaamse Instelling voor Technologische Onderzoek
VNIR	Visible and Near Infra Red
VPO	Vlaams Planbureau voor Omgeving
WP	Work Package

Table of Contents

List of figures

List of tables

1. Introduction

1.1. Project overview

ScaleAgData is a response to the call HORIZON-CL6-2022-GOVERNANCE-01-11 Upscaling (real-time) sensor data for EU-wide monitoring of production and agri-environmental conditions. The ScaleAgData project will run from January 2023 till December 2026 and consists of a consortium of twenty-six partners from fourteen countries. The vision of ScaleAgData is two-fold. On one hand, it wants to obtain insights in how the complex data streams should be governed and organized (governance call). On the other hand, it aims to develop the data technology needed to scale data collected at the farm level to regional datasets, agri-environmental monitoring, and the management of agricultural production.

To do so, ScaleAgData has five objectives:

- Developing innovative approaches for collecting in-situ data and applying data technologies.
- Enabling and promoting data sharing along the entire data value chain.
- Demonstrating how the sensor data can be scaled to agri-environmental data products at the national, regional or European level.
- Demonstrating the benefit of the improved monitoring capacities in a precision farming context.
- Demonstrating the benefit of upscaled regional datasets for the agricultural sector in general.

During its lifecycle, the project will explore seven innovation areas: innovative sensor technology, edge processing, data sharing architecture and data governance, satellite data augmentation, from data assimilation to service development, privacy-preserving technology, and data integration methodologies.

Six Research and Innovation Labs (RIL) have been identified within the project, across various bio-geographical regions of Europe, where different data upscaling and integration models or approaches will be evaluated and demonstrated. The six RILs are: water productivity, crop management, yield monitoring, soil health, grasslands and sustain dairy.
Recommendations will be formulated on how such integrated datasets can be capitalized to help national and regional policy making to strengthen both the competitiveness and sustainability of European agriculture.

1.2. Scope of the document

This Data Management Plan describes the procedures, resources, and security rules applicable to the data assets handled by the ScaleAgData project and addresses the relevant aspects of making the data products FAIR (Findable, Accessible, Interoperable, and Reusable).

1.3. Document structure

This document is structured as follows:

- Chapter 1 provides a project overview and then goes on to describe the scope, responsibilities, and structure of this deliverable.
- Chapter 2 presents a summary of the data involved in the services
- Chapter 3 discusses the approach to the FAIR principles, whenever possible in the project, as well as the allocation of resources

- Chapter 4 introduces data security issues
- Chapter 5 describes the ethical aspects

1.4. Evolution of the document

This Open Science and Data Management Plan is intended to be a living document in which information will be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur. Therefore, each data management plan will have a clear version number and include a timetable for updates.

Version 1.0 of this document, submitted on 27 September 2023, shows the initial approach to data management within ScaleAgData.

In December 2024, at the end of the first reporting period of the ScaleAgData project, the EC and external reviewers formulated a number of recommendations to improve the Open Science and Data Management Plan. This included, among others, adding information about the initial planning and monitoring of the research, scientific and technical progress of the project and activities related to the management of R&D aspects, including the monitoring of research objectives and KPIs, technology selection, standards adoption, scientific and research issues concerning methodologies and approaches via milestones and active risk management. The requested information will be provided in version 2.0 of this document, deliverable D1.3, which will be submitted at the end of February 2025.

A clarification about how to get access to the data managed by Neuropublic has already been added to Table 1 of the present document.

Additional updates will take place if necessary due to changed circumstances that require alterations to the approaches presented herein.

2. Data summary

ScaleAgData will deal with different types of data, including: EO imagery (raw, intermediate and processed data), data collected via in-situ sensors covering: soil data (i.e. soil moisture, temperature, salinity, EC, pH, clay and Soil Organic Carbon), weather data (i.e., air temperature, air humidity, wind direction and speed, solar radiation, precipitation, atmospheric pressure), machinery data (i.e., crop type, yield, dates of management practices, yield quality), key biophysical parameters (i.e. LAI, fAPAR), reflectances from cameras and spectrometers and various added value data products such as yield maps, crop phenology, crop development/field activities, growth rate, crop type, surface soil moisture, root zone soil moisture, soil maps, Eta, and methane concentration.

This chapter describes the input and output data ingested/needed by ScaleAgData as source data in the project (section 2.1 and 2.2) and discusses the broader context of data access and data sharing and related European initiatives and regulations (section 2.3).

2.1. Data inputs

Table 1 below provides an overview of the input data that will be collected and used by ScaleAgData to develop improved products for agri-environmental monitoring. It is the first version of a living table that will be completed / modified during the project's lifetime.

Data inputs are defined in terms of:

- **Data input:** identification of the data.
- **Spatial scale:** at what geographical scale will the data be available.
- **Purpose and relation to the project:** what is the purpose of the data collection/generation and its relation to the objectives of the project?
- **Types and formats:** what types and formats of data will the project generate/collect?
- **Licensing:** licenses required and how?
- **Origin:** what is the origin of the data?
- **Size (GB):** what is the expected size of the data?
- **Data utility:** for what might it be useful?
- **Service:** service or services using the data.
- **Partner:** partner or partners involved.

Table 1: ScaleAgData data inputs

Data input	Spatial scale	Purpose and relation to the project	Types and formats	Licencing	Origin	Size (DB)	Data utility	Service	Partner
Sentinel-2 NDVI	EU	input for crop yield estimation in RIL3	geoTIFF, geojson	yes, available via Terrascope for all Labs	ESA SCIHUB, processed by VITO	1GB/100k m²	broad range of users from agricultural sector (farmers, advisors, processors, suppliers, downstream service providers…)	Terrascope NDVI service	VITO
Sentinel-2 LAI	EU	input for crop yield estimation in RIL3	geoTIFF, geojson	yes, available via Terrascope for all Labs	ESA SCIHUB, processed by VITO	1GB/100k m²	broad range of users from agricultural sector (farmers, advisors, processors, suppliers, downstream service providers…)	Terrascope BIOPAR service	VITO
Sentinel-2 fAPAR	EU	input for crop yield estimation in RIL3	geoTIFF, geojson	yes, available via Terrascope for all Labs	ESA SCIHUB, processed by VITO	1GB/100k m²	broad range of users from agricultural sector (farmers, advisors, processors, suppliers, downstream service providers…)	Terrascope BIOPAR service	VITO
Sentinel-2 Fcover	EU	input for crop yield estimation in RIL3	geoTIFF, geojson	yes, available via Terrascope for all Labs	ESA SCIHUB, processed by VITO	1GB/100k m²	broad range of users from agricultural sector (farmers, advisors, processors, suppliers, downstream service providers…)	Terrascope BIOPAR service	VITO
Sentinel-1 & 2 CropSAR fAPAR	EU	input for crop yield estimation in RIL3	geojson	yes, available via Terrascope for all Labs	ESA SCIHUB, processed by VITO	1GB/100k m²	broad range of users from agricultural sector (farmers, advisors, processors, suppliers, downstream service providers…)	Terrascope CropSAR service	VITO
evapotranspiration	regional	input for crop yield estimation in RIL3							DHI
soil moisture	regional	input for crop yield estimation in RIL3							DHI

Net primary production	EU		geojson	yes, available via Terrascope for all Labs	ESA SCIHUB, processed by VITO	1GB/100k m²	broad range of users from agricultural sector (farmers, advisors, processors, suppliers, downstream service providers…)		VITO
Dry matter production	EU		geojson	yes, available via Terrascope for all Labs	ESA SCIHUB, processed by VITO	1GB/100k m²	broad range of users from agricultural sector (farmers, advisors, processors, suppliers, downstream service providers…)		VITO
Standing biomass grassland									
Start, peak, end of season	EU	input for crop yield estimation in RIL3	geojson	yes, available via Terrascope for all Labs	ESA SCIHUB, processed by VITO	1GB/100k m²	broad range of users from agricultural sector (farmers, advisors, processors, suppliers, downstream service providers…)	Terrascope Phenology – TIMESAT service	VITO
Emergence	EU	input for crop yield estimation in RIL3	geojson	yes, available via Terrascope for all Labs	ESA SCIHUB, processed by VITO	1GB/100k m²	broad range of users from agricultural sector (farmers, advisors, processors, suppliers, downstream service providers…)	Terrascope Phenology service	VITO
Harvest	EU		geojson	yes, available via Terrascope for all Labs	ESA SCIHUB, processed by VITO	1GB/100k m²	broad range of users from agricultural sector (farmers, advisors, processors, suppliers, downstream service providers…)	Terrascope Phenology service	VITO
Crop type	EU		geoTIFF, geojson	yes, available via Terrascope for all Labs	ESA SCIHUB, processed by VITO	1GB/100k m²	broad range of users from agricultural sector (farmers, advisors, processors, suppliers, downstream service providers…)	Terrascrope Crop type classification service	VITO
Top soil	local								
Root zone	local								
Soil texture	regional	input for crop yield estimation in RIL3							
Actual Evapotranspiration	local								
Usage of pesiticides and fertilizers	local								
Methane concentration	regional								

Dairy maps	local								
Weather data	EU	input for crop yield estimation in RIL3							
Weather data	Local	input for DSS models, sub-lab2			weather stations in selected farms in North Italy		Temperature, Rain, Relative Humidity, Leaf wetness data from in situ weather station		Horta
potato yield sensor data	local	input for crop yield estimation in RIL3			AVR harvesters			AVR API	AVR
wheat yield sensor data	local	input for crop yield estimation in RIL3			CNH harvesters			CNH API	CNH
wheat NIR sensor data	local	input for crop yield estimation in RIL3			CNH harvesters			CNH API	CNH
Weather sensor data	local	input for crop yield estimation in RIL3			CNH/Ugent			CNH API	CNH, Ugent
Soil scan data	local	input for crop yield estimation in RIL3			CNH/Ugent			CNH API	CNH, Ugent
Crop scan data	local	input for crop yield estimation in RIL3			CNH/Ugent			CNH API	CNH, Ugent
Soil analysis data	local	input for crop yield estimation in RIL3			CNH/Ugent			CNH API	CNH, Ugent
Soil data	Local	input for DSS models, sub-lab2			compiled in the DSS		Texture from in situ analysis		Horta
Farm log	Local	input for DSS models, sub-lab2			compiled in the DSS		Register of operations carried out in field		Horta

Sentinel 2	Local	input for DSS models, sub-lab2			external provider				Horta
Field characteristics	Local	input for DSS models, sub-lab2			compiled in the DSS		Location (coordinates); agronomic characteristics		Horta
Management of PPP	local	input for sustainability calculation, sub-lab2			compiled in the DSS		Calculate sustainability indicators, with particular reference to the reduction of risk exposure to Plant Protection Product		Horta
Sentinel-2 NDVI	Global	Crop Management, sub-lab1	geojson	yes, available via (NDVI) sense-web.neuropublic.gr on request	ESA proccessed by NP's earth insight Engine		Policy Makers, Farmers, Agronomists, Scientists	NP's earth Insight service	NP
Sentinel-2 LAI	Global	Crop Management, sub-lab1	geojson	yes, available via (LAI) sense-web.neuropublic.gr on request	ESA proccessed by NP's earth insight Engine		Policy Makers, Farmers, Agronomists, Scientists	NP's earth Insight service	NP
Sentinel-2 NDWI	Global	Crop Management, sub-lab1	geojson	yes, available via (NDWI) sense-web.neuropublic.gr on request	ESA proccessed by NP's earth insight Engine		Policy Makers, Farmers, Agronomists, Scientists	NP's earth Insight service	NP
gaiaTron Temperature	local	Crop Management, sub-lab1	geojson	yes, available via gaiasense.neuropublic.gr/ gaiaexmon on request	NP's Telemetric Station (gaiTron)		Local Weather and Soil Data from farmers' parcel	NP's gaiatron Station	NP
gaiaTron rainfall	local	Crop Management, sub-lab1	geojson	yes, available via gaiasense.neuropublic.gr/ gaiaexmon on request	NP's Telemetric Station (gaiTron)		Local Weather and Soil Data from farmers' parcel	NP's gaiatron Station	NP

gaiaTron relative Humidity	local	Crop Management, sub-lab1	geojson	yes, available via gaiasense.neuropublic.gr/ gaiaexmon on request	NP's Telemetric Station (gaiTron)		Local Weather and Soil Data from farmers' parcel	NP's gaiatron Station	NP
gaiaTron soil moisture 10cm-60cm	local	Crop Management, sub-lab1	geojson	yes, available via gaiasense.neuropublic.gr/ gaiaexmon on request	NP's Telemetric Station (gaiTron)		Local Weather and Soil Data from farmers' parcel	NP's gaiatron Station	NP
gaiaTron soil salinity 10cm-60cm	local	Crop Management, sub-lab1	geojson	yes, available via gaiasense.neuropublic.gr/ gaiaexmon on request	NP's Telemetric Station (gaiTron)		Local Weather and Soil Data from farmers' parcel	NP's gaiatron Station	NP
gaiaTron windDirection	local	Crop Management, sub-lab1	geojson	yes, available via gaiasense.neuropublic.gr/ gaiaexmon on request	NP's Telemetric Station (gaiTron)		Local Weather and Soil Data from farmers' parcel	NP's gaiatron Station	NP
gaiaTron windSpeed	local	Crop Management, sub-lab1	geojson	yes, available via gaiasense.neuropublic.gr/ gaiaexmon on request	NP's Telemetric Station (gaiTron)		Local Weather and Soil Data from farmers' parcel	NP's gaiatron Station	NP
iCM-FarmBook_Irrigation	local	Crop Management, sub-lab1	geojson	yes, available via (Events) sense-web.neuropublic.gr on request	NP's iCM app		FarmBook for each parcel	NP's iCM app (FarmBook)	NP
iCM-FarmBook_Fertilization	local	Crop Management, sub-lab1	geojson	yes, available via (Events) sense-web.neuropublic.gr on request	NP's iCM app		FarmBook for each parcel	NP's iCM app (FarmBook)	NP

iCM-FarmBook_sprays	local	Crop Management, sub-lab1	geojson	yes, available via (Events) sense-web.neuropublic.gr on request	NP's iCM app		FarmBook for each parcel	NP's iCM app (FarmBook)	NP
iCM-FarmBook_phenological-stages	local	Crop Management, sub-lab1	geojson	yes, available via (Events) sense-web.neuropublic.gr on request	NP's iCM app		FarmBook for each parcel	NP's iCM app (FarmBook)	NP
iCM-FarmBook_pests	local	Crop Management, sub-lab1	geojson	yes, available via (Events) sense-web.neuropublic.gr on request	NP's iCM app		FarmBook for each parcel	NP's iCM app (FarmBook)	NP
iCM-FarmBook_harvest	local	Crop Management, sub-lab1	geojson	yes, available via (Events) sense-web.neuropublic.gr on request	NP's iCM app		FarmBook for each parcel	NP's iCM app (FarmBook)	NP
iCM-FarmBook_LandManagement	local	Crop Management, sub-lab1	geojson	yes, available via (Events) sense-web.neuropublic.gr on request	NP's iCM app		FarmBook for each parcel	NP's iCM app (FarmBook)	NP
Soil moisture data	local	input for water productivity assessment, RIL1	geojson	yes, available to all partners on request	Soil probes		Local IoT sensors in test fields		IES, MIGAL
Weather sensor data	local	input for water productivity assessment, RIL1	geojson	yes, available to all partners on request	Weather stations		Local IoT sensors in test fields		IES, MIGAL

Yield assessment	local	input for water productivity assessment, RIL1	geojson	yes, available to all partners on request	In situ measurements		Measurement of pepermint and quinoa yields		IES, MIGAL
Thermal data	local	input for water productivity assessment, RIL1	geoTIFF	yes, available to all partners on request	Airborne sensors		Several airborne data acquisition campaigns using uTABI		IES, MIGAL
VNIR hyperspectral data	local	input for water productivity assessment, RIL1	geoTIFF	yes, available to all partners on request	Airborne sensors		Several airborne data acquisition campaigns using uCASI		IES
Sentinel 2	regional	input for analysis of dairy productivity and environment	geoTIFF	yes, available to all partners on request	ESA SciHub		analysis of productivity and precision farming at regional level	under construction by OHB-DS	OHB
EnMAP	regional	input for analysis of dairy productivity and environment	geoTIFF	yes, at moment for non-commercial use	EnMAP Portal		analysis of productivity and precision farming at regional level	under construction by OHB-DS	OHB
Milk quantity & quality data	regional	input for analysis of dairy productivity and environment	csv	tbd.	DMK ERP system	12,9MB/100km²	Laboratory analysis data and milk delivery details	DMK data export	DMK
Grass yield aggregation	regional	input for analysis of dairy productivity and environment	tbd.	tbd.	Claas farm equipment		Harvester data		Claas

Grassland biomass, fPAR and LAI measurementslocal	local	input for grasslands lab	geojson	yes, available to all partners on request	In situ measurements		development of RS-based models		IFAPA
Grasslands biomass estimations	regional	input for grasslands lab	geoTIFF	yes, available to all partners on request	LUE model output		broad range of users from agricultural sector (farmers, advisors, processors, suppliers, downstream service providers…)		IFAPA
Soil moisture data	local	ET models validation	geojson	yes, available to all partners on request	Soil probes		remote sensing ET model analysis		IFAPA
Weather sensor data	local	input for grasslands lab	geojson	yes, available to all partners on request	Weather stations		remote sensing ET and biomass model analysis		IFAPA
Energy, water an carbon fluxes	local	input for grasslands lab	geojson	yes, available to all partners on request	Flux tower measurements		remote sensing ET and biomass model analysis		IFAPA
Sentinel-2 MSI L2A reflectances	regional	input for grasslands lab	.SAFE	available according to the consortium agreement	ESA scihub, processed by EURAC		estimation of grassland biophysical parameters		EURAC
Sentinel-1 backscattering	regional	input for grasslands lab	openEO collection on EURAC backend	available according to the consortium agreement	ESA scihub, processed by EURAC		development of a datasusion algorithm between Sentinel-1 and Sentinel-2		EURAC
Grassland LAI and yield ground measurements		input for grasslands lab	.csv	available according to the consortium agreement	ground measurements		study the relationship between LAI and yield		EURAC
Soil Quality Map - SOC Indicators	regional	Output of RILab for Soil related to Soil Quality Indicators - Flemish Region	Geopackage	yes, available to all partners on request	EV ILVO product	2 GB	Indicator created using one of the four selected approaches	EV ILVO Soil Quality Indicators Web Services	EV ILVO

Soil Texture	regional	Input of RILab for Soil - Flemish Region	Geopackage	public available	VPO		https://www.vlaanderen.be/datavindplaats/catalogus/wrb-soil-units-40k-bodemkaart-van-het-vlaamse-gewest-volgens-het-internationale-bodemclassificatiesysteem-world-reference-base-op-schaal-140000	DOV service	EV ILVO
Soil organic carbon stock map	national	Input of RILab for Soil - Flemish Region	geoTIFF	public available	VPO		https://www.dov.vlaanderen.be/geonetwork/srv/api/records/37aaa10d-50bd-431e-b924-209c2e88b9d7	DOV service	EV ILVO
Soil Associations	regional	Input of RILab for Soil - Flemish Region	Geopackage	public available	Flemish Landscape		https://metadata.vlaanderen.be/srv/dut/catalog.search#/metadata/C4F51B28-51BF-4189-8E98-72B94AE8DA13	Geopunt	EV ILVO
Sentinel 2 and other Copernicus data	regional	Input of RILab for Soil - Flemish Region	STAC services	Open data plus the Licence for the use of the STAC services	https://stacindex.org		Satellite data		EV ILVO
Soil scans and sensor data	local	Input of RILab for Soil - Flemish Region	various (GIS files and reports)	no, to be used by EV ILVO for the RILab activities	ILVO		remote sensing data at local level, to be combined with EO data		EV ILVO
Soil measurements	local	Input of RILab for Soil - Flemish Region	xls	no, to be used by EV ILVO for the RILab activities	ILVO		laboratory analysis data on soil properties		EV ILVO

2.2. Data outputs

In this section the data outputs are listed (Table 2).

Data outputs are defined in terms of:

- **Data output:** identification of the data.
- **Types and formats:** what types and formats of data will the project generate?
- **Size (GB):** what is the expected size of the data?
- **Service:** service or services using the data.
- **Partner:** partner or partners involved.

Table 2: ScaleAgData data outputs

Data output	Types and formats	Size (DB)	Service	Partner
EO based yield / variability maps	geoTIFF, geojson			VITO
yield estimates at field level	geojson			VITO
yield estimates at regional level	geojson			VITO
improved sensor based yield maps for potatoes	geoTIFF, geojson			AVR
improved sensor based yield maps for wheat	geoTIFF, geojson			CNH
Expected deviation of milk quality & quantity	txt			DMK
Regional productivity of dairy farms	txt			DMK
yield estimates at regional level	tbd.			Claas
gap-filled grasslands LAI maps at parcel level	geoTIFF, geojson			EURAC, IFAPA
Estimated grassland yield at parcel level	geojson			EURAC, IFAPA
improved grassland yield maps based on flux tower sensors	geojson			IFAPA, DEIMOS
field water status for target crops	geojson, geoTIFF			IES, MIGAL
satellite data-based field water status for target crops	geoTIFF			IES, MIGAL
predicted yield for target crops	geojson, geoTIFF			IES, MIGAL
satellite data-based predicted yield for target crops	geoTIFF			IES, MIGAL
EO based regional Soil Organic Carbon maps	Cloud-optimized geoTIFF	ca. 0.5 – 1.5 GB / region	tbd	EV ILVO / AUTH
soil health indicator estimates at field level	Cloud-optimized geoTiff / geojson or equivalent vector format	ca. 300 – 700 MB / region	tbd	EV ILVO / AUTH
soil moisture aggregation at LAU / Commune Level	geoTIFF		tbd	NP
calculated indicators (aggregates) based only on the ground truth evidence	json		tbd	NP
calculated indicators (LUKE) based on data assimilation mechanisms along with the respective annotations	txt		tbd	NP
aggregated pesticide use for policy makers	txt		tbd	NP
Calculated sustainability indicators	txt			Horta
DSS model outputs (fetilisation rate, disease risks, crop yield and quality, soil moisture)	txt			Horta
Vegetation indexes, crop nutrient status	txt			Horta
statistical data on the accuracy of observations of the occurrence of agrophages	json			WODR/PSNC

improved predicted agrophage occurence data based on geolocation	json			WODR/PSNC
predicted overall level of agrophage occurence risk for the selected region	json			WODR/PSNC

2.3. Broader context

As presented in Table 1 but also Table 2, ScaleAgData partners are dealing with different types of data, including IoT data, and aiming to develop data products and services by performing transactions within loose or more organized (in terms of governance) data ecosystems. For this reason, ScaleAgData will address the issue of fair access to and use of data, accessing different elements of the technical and non-technical building blocks[1] of data spaces, in a way to facilitate data sharing and enable the creation of value from data within the vertical domains defined by the RI Labs.

ScaleAgData will follow the Building Blocks Taxonomy suggested by the Data Space Support Center[2] (DSSC) (Figure 1), targeting the:

- Governance building block (Organizational Governance within WP2 and Data Governance within WP3),
- Data Interoperability building block (Data Models and Formats within WP3),
- Data Sovereignty & Trust building blocks (Access and Usage Policy within WP3 and WP4),
- Data Value Creation building blocks (Data and Services within WP2, WP3, WP4 and WP5)
- Business building blocks (with focus on the Data products within WP6).

ScaleAgData considers the recent EC legislation of the Data Governance Act (DGA)[3], the Data Act[4] that have as primary objectives:

- Ensure fairness in the allocation of value from data among actors.
- Foster access to and use of data.

and sub-objectives:

- To ease the switching of providers of data processing services,
- To put in place safeguards against unlawful data transfers by cloud service providers.
- The development of interoperability standards for data to be reused between sectors.

To align with the current developments and the convergencies related to the development of the data spaces but also with the new EC legislation, ScaleAgData will adopt, when needed, new taxonomies like the one coming from DSSC (a new effort inspired by the OpenDEI but also consider and reference the SITRA Rule book), or new terms that describe the role of each actor in the development or use of new products with data coming from sensors (Data Act).

[1] Legal, Business and Governance

[2] Data Spaces Support Centre (dssc.eu)

[3] The European regulation that aims to create a framework to facilitate European data spaces and increase trust between actors in the data market. The DGA entered into force in June 2022 and applies from Sept 2023. The DGA defines the European Data Innovation Board.

[4] Data Act: EU institutions finalise agreement on industrial data law – EURACTIV.com

Additionally, to align with the support and coordination activities of the specific sector, ScaleAgData will also monitor and follow the Common European Agricultural Data Space[5] CSA project, AgriDataSpace, that aims to set up to share and access agricultural data.

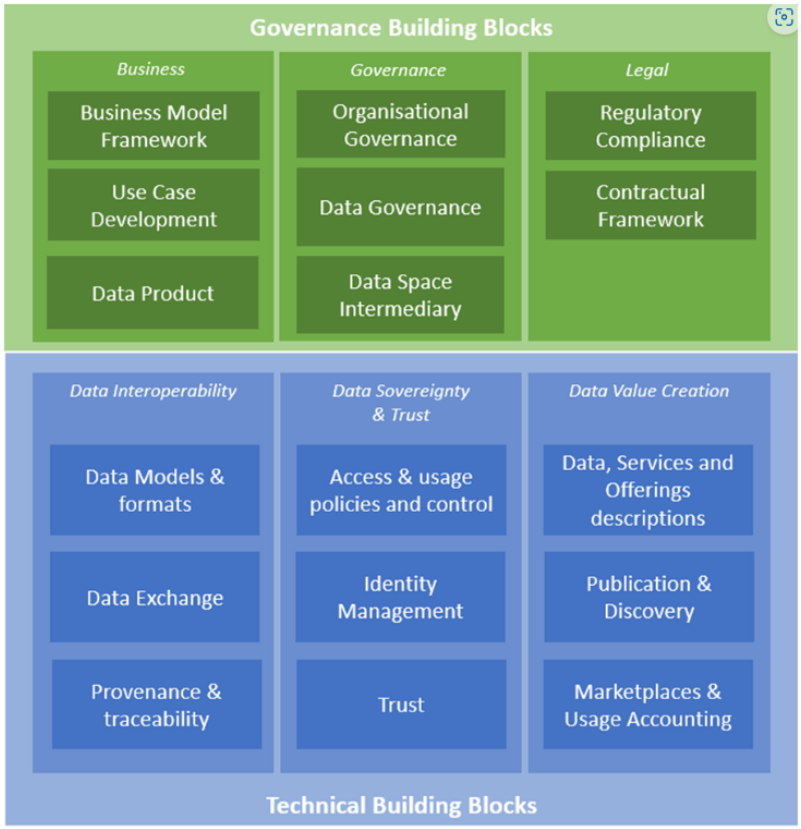

Figure 1 Taxonomy of Building Blocks v0.5 of the Data Spaces Blueprint.

For the development of a Governance framework[6] (DSSC prefers the use of framework instead of model) which is a set of principles, standards, policies (rules/regulations) and practices that apply to the governance, management, and operations within a particular scope (e.g., a data space, a data space initiative, or data spaces blueprint) as well as to the enforcement thereof, and the resolution of any conflicts, WP2 will identify, analyse, and map the current status and level of governance frameworks in the data ecosystem of the vertical domains of the RI labs.

More specific and to achieve this, T2.4, which started on July 23 (M7), is planning to use:

- The results of T2.1 & T.2.2 (D2.1) that reveals existing legal, operational and functional agreements as well as technical standards [7] widely adopted by RI labs actors, stakeholders and potential users of the ScaleAgData results,
- The evaluation outcomes of WP5, T5.3 (D5.3).
- The results of WP3, related to the data governance in WP3 T3.4 (D3.1, D3.4).
- The results of WP6, T6.4, because Governance typically includes business aspects. (D6.4)

[5] The common European data spaces is a subclass of data spaces that adheres to European rules and values. The common European data spaces were introduced in the EU data strategy and referenced in the Data Governance Act and Data Act. The Agricultural data space, it's the data space that refers to the Agricultural sector.

[6] DSSC Glossary | Version 1.0 | March 2023 - Glossary - Data Spaces Support Centre

[7] SITRA Rulebook for a fair data economy.

After the analysis of the collected material, at the second iteration WP2, T2.4 will proceed to the initiation and tailoring of a governance framework, considering the DSSC Starter Kit[8], DSSC Glossary and DSSC Blueprint[9] and the results of the AgriDataSpace[10] project, aiming to support the RI labs partners to realise the status of different aspects of Governance within their data ecosystems[11], business, legal, and organisational, supporting the adoption of rules, practices and processes needed for effective data sharing and innovation within and cross the vertical domains of the RI labs.

[8] Starter Kit for Data Space Designers | Version 1.0 | March 2023 - Starter Kit - Data Spaces Support Centre (dssc.eu)

[9] Public Consultation - Data Spaces Blueprint v0.5 - Blueprint (external-share.com)

[10] AgriDataSpace (agridataspace-csa.eu)

[11] A loosely coupled set of autonomous parties engaging in data sharing

3. FAIR data

Through the RI environment, the ScaleAgData project partners will have access to all data and processing infrastructure on the Deimos EO thematic exploitation solution (services4EO) during the development and testing stage of the project.

The RI environment is a computing infrastructure where all the relevant data, methods and data products are centralised. In this development and testing environment, the RI Labs, together with the technology providers, can evaluate which of the provided solutions are fit for the RI Labs. The Labs will combine datasets and technologies already available in the field with the developments performed by the project aiming to have their own architectural blueprint and adequate services to tackle their needs.

Further it needs to be mentioned that:

- Partners already have the prerequisite infrastructure, facilities and equipment to conduct the activities envisioned by ScaleAgData.
- Partners VITO, DME/DES, ICCS, EV ILVO, PSNC, VTT, DHI and NP have access to all the cloud computing infrastructure that is needed to deploy the core datasets, services and common components of the project.
- In addition, all the partners responsible for building project demonstrators, have all the infrastructure necessary to build and deploy the ScaleAgData-powered solutions.

3.1. Making data findable, including provision of metadata

The findability of the data will be succeeded through the use of Persistent Identifiers, which are important because they unambiguously identify the data and facilitate data citation, as well as through rich metadata (i.e., Dublin Core, ISO19115, etc) that provide an important context for the interpretation of the data and make it easier for machines to conduct automated analysis. The catalogue solution offered by Deimos supports several standard interfaces to discover the data, namely CSW and openSearch so all users can query the catalogue directly. Alternatively those users can use a Python library made available in the RI environment that encapsulates the discovery methods provided by those two standard interfaces to query the catalogue.

3.2. Making data openly accessible

ScaleAgData embraces both Responsible Research and Responsible Innovation to promote systemic change in the practices of the RI systems towards inclusiveness, “opening up” and sharing, aiming to promote co-development of research-based responsible innovations and by including a wider spectrum of responsible innovation processes, and not only those that are research-driven. Responsible Innovation meets economic, social, ethical and environmental goals and its constructed complexes with material (technologies, resources), organisational (management, networking, policy) and discursive dimensions (ideas, narratives). The EC set off the ambition of the European Open Science Cloud: a trusted environment for hosting and processing research data to support EU science in its global leading role. Resonating with the principle *'as open as possible*', ScaleAgData wholeheartedly supports the goals of the Open Science Policy under Horizon Europe and thus, appropriate open science actions will be implemented as an integral part of its proposed methodology as described below.

ScaleAgData aims to use the Open Research Europe scholarly fully open access publishing service for Horizon Europe to enable rapid publication times and publication outputs that support research integrity, reproducibility, transparency and enable open science actions. Open access to every ScaleAgData publication (towards sustaining also either self-archiving / 'Green' open access or open access publishing / 'gold' open access) will be ensured for all interested persons, mainly through the project's website and through the use of the EU open access publishing services (e.g. European Open Science Cloud (EOSC), OpenAIRE), but also via services for research communities like ResearchGate or Academia. Moreover, ScaleAgData will facilitate the uptake and accessibility of resources collected and generated in the context of the project, through the use of reference data models and the provision of interfaces data access and exchange according to well stablished standards (i.e., OGC).

Open access clauses are added to both the Consortium Agreement (CA) and the Grant Agreement (GA). For every publication, no later than 6 months after the publication is released, a machine-readable copy of the manuscript will be available on the website. Finally, and to make project knowledge publicly available, the Dissemination leader will exploit EC open access portals and tools.

Open access to research data will be provided (whenever there is no conflict with the protection of the results) and shared data will be deposited in open data repositories (i.e., Open Research Europe, Zenodo, GEOSS, AgroSTAC and GEOGLAM in situ data repository). In addition, open access will be provided to the software needed to validate the scientific conclusions of the project's publications (if not protected).

The RI environment provides the capabilities to manage access permissions to each of the different datasets used/produced, according to what has been defined by the project in general and by each user, during the different stages of the project.

3.3. Making data interoperable

The initial plan is for all partners to be aligned in order to make all project data interoperable thus facilitating the exchange, sharing and reuse of data across various research institutions and organizations. In cases where this will be proved very difficult, such as when dealing with project-specific technical datasets, the consortium will examine the possibility to provide mappings to commonly used vocabularies and standards. For this purpose, open data standards will be used when possible, including the ones that will be used for captured data. The produced data will use community agreed schemas and well-defined data models compliant with OGC standards, standardized encodings, controlled vocabularies, keywords, thesauri or ontologies where possible in order to be interoperable and be integrated with other data, applications and workflows. All necessary API interfaces will be prepared to handle the designed data schemas, including schema validation functionalities.

As mentioned in previous sections of this document, all datasets, used as input or produced as output, in the RI environment will be catalogued using standard metadata models (i.e., Dublin Core, ISO19115, etc.) and will be discoverable via CSW and openSearch, assuring full interoperability with systems that use similar standard interfaces.

3.4. Increase data re-use

The generated data will be well-documented and they will have clear licence and provenance information. By default, open access data will be made available to the public for reuse following appropriate licensing schemes (i.e., Open Data Commons licences). The CA will also optimize the

creation of networks, interaction, exchange, and access to internal information/technologies and maximize the technology transfer.

The scripts generated by the different WP3 and WP4 partners in the RI environment will be available in dedicated GIT repositories, managed by each of them.

3.5. Allocation of resources

The cost of making data FAIR is covered by multiple work packages.

- WP1 takes the lead in drafting the open science and data management plan and received resources to this end to have a first version at month 6 and an updated version later in the project at month 26.
- WP3 focuses more on data governance, architecture and integration as well as the edge processing enabling technologies, real time data processing and privacy. Resources have been allocated to this work package for the development of edge processing enablers, NRT processing, data rate transmission automatic adaptation, data fusion from Copernicus and other missions, preprocessing and the creation of datasets to be used by the RI labs, as well as the development of components needed for the data technologies, and the creation of a blueprint architecture to be used in the RI labs.
- WP4 received resources to focus on the Research and Innovation environment. To this end, a centralized data store will be implemented, where all input datasets will be available to the techniques/product development teams, and accessible via dedicated Jupyter notebooks. This virtual lab will be based in the Deimos EO thematic exploitation solution, services4EO, a collaborative EO Exploitation Platform relying on a set of core IT and geospatial-based application development modules with standardized interfaces, which can be chained together to produce a number of different services easily tailored by the platform user.
- WP 5 has resources to focus on integration and knowledge sharing. It will focus on integrating ScaleAgData services to existing applications as well as designing an approach for sharing knowledge with a larger external target audience. This will be applicable for sharing soft facts concerning experience gained and lessons learnt as well as for sharing tangible knowledge/project results, considering the FAIR principles.
- WP6 will focus both on the exploitation and capacity building of products and services, as well as on the IPR management and business models.

The consortium will handle all project collected and generated data and required resources to ensure FAIR management and effective and secure storage. The consortium has the knowledge and tools to make data FAIR and appropriate resources have been foreseen to cover costs for data harmonization, integration, sharing and publication. During the implementation of the project, the consortium members will collect data in various forms, i.e., photos, videos, electronic documents. For the purpose of documentation, data management and governance rules and regulations will be applied with respect to data storage and security.

4. Data security

Collected and generated data will be curated and preserved in secure data repositories that will follow up to date quality, security, privacy standards and relevant legislation of EU (i.e., GDPR). Also, a system for data security is going to be used as a reference model for curation and preservation. ScaleAgData reference architecture will support access control policies for private data and/or file sharing restrictions based on partners' data licensing decisions. These are versatile and fine-grained means to assign permissions in such a way that access to resources can be sufficiently regulated for any eventuality. Moreover, all data that can encompass any personal data protection or privacy and IPR will be stored in the project repository, given the stakeholder's consent, and will not be publicly disclosed, unless otherwise agreed by the stakeholder.
The RI Environment includes an identity management solution that provides Federated Identity Management functions that include data access authorization. Identity Management is accessible online via state-of-the–art standards promoted via OGC innovation programmes (i.e., OAuth, Open ID Connect and User Managed Access) and a browser that provides a user-friendly interface for managing the access to user resources.

The secure management of project data throughout their lifecycle will be achieved by using a strong data protection strategy. In this context, the consortium partners will determine specific security controls to apply in each phase, evaluating the level of compliance during project evolution. Apart from privacy and sensitive data management, another important aspect for data security is data recovery. Thus, additional data recovery strategies will be developed for protecting data from loss that can take place due to several events (natural disasters, software bugs and infrastructure failures).

5. Ethical aspects

Within the ScaleAgData project, Artificial Intelligence (AI) will be applied for integrating sensor data in the data products on agro-environmental conditions, as well as for innovative solutions to enhance the characteristics of the input data (EO and others). The algorithms will involve AI techniques, such as machine learning (ML), for fusion of raw data, fusion between EO and in situ data as well as the development of services for the RI Labs. In the past few years, ethical questions associated with the usage of ML have been the subject of academic and public scrutiny. The EC has recently published guidelines on "Requirements of Trustworthy AI". Based on these guidelines and the active experience of the consortium researchers and experts in data technologies, remote sensing and decision-making R&D,that the ScaleAgData consortiumconfirms that there are no ethics concerns in the project developments.

The project will not collect human data such as self-identified ethnicity, or age from the participants. The fundamental principles outlined in the Charter of European Fundamental Rights complemented by the GDPR, such as human dignity, the integrity of the person, the protection of personal data to ensure privacy, will be fully respected and promoted in the project.

Management of the personal data by the consortium members and participating stakeholders will follow GDPR guidelines. Personal data will be collected and processed only if, and to the extent, necessary. Prior to any interaction, respondents and participants of interviews will receive information on what will happen with their personal data, and what their rights are in this respect. If participants think that certain information should not be used, or if they consider some information incorrect, they can contact the designated data protection officers with their requests at the visibly marked contact e-mail address. If the participants consider that their personal information has not been handled in a correct way, they will also have the right to make a complaint.

The project will not include any research on vulnerable populations. The machine learning and artificial intelligence parts of the project will apply to remote sensing prototypes and will not apply to population monitoring. Our AI systems will not subordinate, coerce, deceive or manipulate people, and will not create attachment or stimulate addiction. All datasets and processes associated with AI & machine learning decisions will be well communicated and appropriately documented. Best possible efforts will be made to avoid unfair bias. No possible risk or harm is anticipated.