Research and Innovation Action
HORIZON-CL6-2022-GOVERNANCE-01
TOPIC: HORIZON-CL6-2022-GOVERNANCE-01-11

AgriDataValue

Smart Farm and Agri-environmental Big Data Value

Deliverable D3.4
Smart Farming Pilots &
Data Management Plan (DMP) V2

Authors	Tim Van De Gucht (ILVO), Nico Peiren (ILVO), Theodore Zahariadis, Alexandra Lakka, Kostas Railis, Apostolos Skias (SYN), Panagiotis Athanasoulis, Stamatia Rizou (SLG), Jan Kiers(Delphy), Eva Ampe(Inagro), Sarah Bossuyt(Inagro), Vicky Inglezou (NILEAS), Elena Ilie (BioRO), Ioannis Katris (TBA), Inga Berzina (ZSA), Antonio Del Saz (TEC)
Nature	Report
Dissemination	PU-Public
Version	V1
Status	Final
Delivery Date (DoA)	24
Actual Delivery Date	31/01/2025

Keywords	Data Management Plan, Repository, Open Science, Open Data, Metadata, FAIR data
Abstract	This Deliverable provides the AgriDataValue Data Management Plan (DMP) version 2. The document initially describes the pilot sites and the government structure and outlines the process for collecting or generating research data by the AgriDataValue Pilot owners and the Consortium and how these data will be handled during and after the project's completion. It also 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. The document provides 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 the datasets the consortium aims to preserve and in which format. Legal and ethical issues related to the AgriDataValue's collecting and/or processing of personal data are identified and practically considered.

Funded by the European Union under Grant Agreement no. 101086461

ACKNOWLEDGEMENT

The AgriDataValue project is funded by the European Union under Grant Agreement No. 101086461. 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. No part of this document may be used, reproduced and/or disclosed in any form or by any means without the prior written permission of the AgriDataValue consortium.

	Participant organisation name	Short	Country
01	SYNELIXIS SOLUTIONS S.A.	SYN	EL
02	ATOS IT SOLUTIONS AND SERVICES IBERIA SL (Terminated)	ATOS	ES
03	SIXENSE ENGINEERING	SIXEN	FR
04	NETCOMPANY-INTRASOFT SA	INTRA	LU
05	SIEMENS SRL	SIEM	RO
06	SINERGISE LABORATORIJ ZA GEOGRAFSKEINFORMACIJSKE SISTEME DOO	SINER	SI
07	ALMAVIVA - THE ITALIAN INNOVATION COMPANY SPA	ALMA	IT
08	INTERNATIONAL DATA SPACES EV	IDSA	DE
09	SOFTWARE IMAGINATION & VISION SRL	SIMAVI	RO
10	SINGULARLOGIC S.A.	SLG	EL
11	EIGEN VERMOGEN VAN HET INSTITUUT VOOR LANDBOUW- EN VISSERIJONDERZOEK	EV ILVO	BE
12	ETHNIKO KAI KAPODISTRIAKO PANEPISTIMIO ATHINON	NKUA	EL
13	INAGRO, PROVINCIAAL EXTERNVERZELFSTANDIGD AGENTSCHAP IN PRIVAATRECHTELIJKE VORM VZW	Inagro	BE
14	UNIWERSYTET LODZKI	UL	PL
15	FUNDACION PARA LAS TECNOLOGIAS AUXILIARES DE LA AGRICULTURA	TEC	ES
16	DELPHY BV	Delphy	NL
17	INSTITUTO TECNOLOGICO DE ARAGON	ITAIN	ES
18	ZEMNIEKU SAEIMA	ZSA	LV
19	SOCIEDAD ARAGONESA DE GESTION AGROAMBIENTAL SL	SARGA	ES
20	AGROTIKOS KTINOTROFIKOS SYNETAIRISMOS KATOUNAS TO VIOLOGIKO AGROKTIMA	TBA	EL
21	SOCIETA ITALIANA DI VITICOLTURA ED ENOLOGIA	SIVE	IT
22	NILEAS-SYNETAIRISMOS PISTOPOIIMENON AGROTIKON PROIONTON DIMOU NESTOROS MESSINIAS	NILEAS	EL
23	CONSEIL DES VINS DE SAINT-EMILION	CVSE	FR
24	ASOCIATIA OPERATORILOR DIN AGRICULTURA ECOLOGICA BIO ROMANIA	BIORO	RO
25	RI.NOVA SOCIETA COOPERATIVA	RI.NO	IT
26	AGRO DIGITAL SOLUTIONS	AgroDS	LT
27	NATIONAL PAYING AGENCY	NPA	LT
28	AGENZIA PROVINCIALE PER I PAGAMENTIDELLA PROVINCIA AUTONOMA DI TRENTO	APPAG	IT
29	AGENTIA DE PLATI SI INTERVENTIE PENTRU AGRICULTURA	APIA	RO
30	QUEEN MARY UNIVERSITY OF LONDON	QMUL	UK

DISCLAIMER

Document History

Version	Date	Contributor(s)	Description
V0.1	10/12/2O24	ILVO	Template
V0.2	27/01/2025	ILVO	First draft DMP V2
V1.0	31/01/2025	ILVO	Final version

Document Reviewers

Date	Reviewer's name	Affiliation
28/01/2025	Aleksejs Zacepins	ZSA
31/01/2025	Javier Sancho	SARGA

Table of Contents

Definitions, Acronyms and Abbreviations

ADS	Agri-Environment Data Space
ADV	AgriDataValue
CERIF	Common European Research Information Format
DMP	Data Management Plan
DPIA	Data Protection Impact Assessment
DPO	Data Protection Officer
EO	Earth Observation
FAIR	Findable, Accessible, Interoperable and Reuse
FTIR	Fournier Transform Infra-Red
GDPR	General Data Protection Regulation
GHG	Greenhouse Gas
LGPL	Lesser General Public License
LMAA	Lean Multi-Actor Approach
ML	Machine Learning
MPL	Mozilla Public License
NDIR	Non-Dispersive Infra-Red
OSS	Open-Source Software
PGC	Pilot Governance Committee
ppm	Part per million
TL	Task Leader
TM	Technical Manager
TDLS	Tuneable Diode Laser Spectrometers
UC	Use Case
VWC	Volumetric Water Content
WP	Work Package
WPL	Work Package Lead

Executive Summary

The pilots play an essential role in the AgriDataValue project as suppliers of data which will later be used by the technology partners to train the various Machine learning (ML) models to realize the project Use Cases (UCs) and finally create a platform of platforms.

Because data generation, capturing, description, and handling are of utmost importance, this Deliverable entitled "Smart Farming pilots & Data Management Plan (DMP) V2" has twofold objectives: on the one hand, it describes the pilots, which data they will deliver and how they are linked to the project Use Cases (UCs) at the current stage in the project timeline. It also describes which additional data are necessary to complete the UCs and model them. On the other hand it contains a section with the second version of the Data Management Plan (DMP) where the FAIR principles are incorporated. 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. The document provides 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 the datasets the consortium aims to preserve and in which format. Legal and ethical issues related to the AgriDataValue's collecting and/or processing of personal data are identified and practically considered.

1 Introduction

Modern farms create a huge amount of data, more and more based on IoT sensors and drones, while vast earth observation (EO) data become available via Copernicus Hubs. AgriDataValue aims to establish itself as the "Game Changer" in Smart Farming and agri-environmental monitoring, and strengthen the smart-farming capacities, competitiveness and fair income by introducing an innovative, intelligent, and multi-technology, fully distributed platform of platforms. To achieve technological maturity and massive acceptance, AgriDataValue adopts and adapts a multidimensional approach that combines state-of-the-art big data and dataspaces technologies (BDVA/IDSA/GAIA-X) with agricultural knowledge, new business models, and agri-environment policies, leverages on existing platforms and edge computing, and introduces novel concepts, methods, tools, pilots and engagement campaigns to go beyond today's state of the art, perform breakthrough research and create sustainable innovation in upscaling (real-time) sensor data, already evident within the project lifetime. AgriDataValue is validated via 24 Use cases in 23 pilots in 9 countries, representing more than 181,000 ha with 25 types of crops that span from southwest to northeast Europe, outdoor and greenhouse crops, organic and non-organic production, and more than 2,000 animals of 5 types. More than 4,200 farmers will provide insights and more than 89,000 will be directly informed. More than 1,600 sensors will be utilized and more than 4,500 additional sensors will be installed to measure (real-time) data, including more than 2,500 RFID tags.

The ADV pilots play an essential role in AgriDataValue, as they supply a significant amount of on-site/in-situ data which will be used by the technology partners to train the various ML models to realize the project Use Cases (UCs). The pilots are a very heterogeneous group providing a general overview of different Agricultural sectors in Europe. AgriDataValue represents 30 partners from 13 countries. Eleven of these partners, representing 9 countries, are pilot owners. The pilots are active in the main areas of agriculture going from arable crops to vegetables to orchards and vineyards to livestock. Some pilots cover several different sectors. The pilots include experimental or commercial farms or test facilities that already have or will install more sensors for capturing data relevant to the specific crop, product, or animal type. These pilots are hence the primary source of relevant raw data to be used for model development to create added value from the data. The pilots will hence also be the first to validate the constructed models using newly captured data.

The purpose of this deliverable is to provide an updated description of the initial AgriDataValue pilots description, the data that they are generating or will generate in the near future, and the second version of the AgriDataValue Data Management Plan (DMP), which defines key elements that facilitates the potential reuse of the data collected and processed during and after AgriDataValue. Additionally, the purpose of this DMP is also to describe the data, its use or intended use, and how it is and will be managed and 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, the project consortium will be the first candidates for open data re-use. AgriDataValue 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.

In summary, this deliverable provides the AgriDataValue Data Management Plan (DMP) version 2. It describes which standards and methodology for data collection and generation is and will be followed, and whether and how the data will be shared and made open access. It also describes 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 the datasets the consortium aims to preserve and in which format. Legal and ethical issues related to AgriDataValue'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.

This document evolves during the lifespan of the project and is now updated (V2) 18 months after the first version to provide additional information on datasets created throughout the project. Due to the complex nature of the project and the long project duration, future unplanned updates of this document may be performed when deemed necessary by the project partners, to keep the document up to date.

1.1 Data Management Plan

The Data Management Plan (DMP) describes the data management life cycle for the data to be collected, processed and generated by the AgriDataValue project. This DMP follows the Horizon FAIR protocol making data findable, accessible, interoperable, and re-usable (FAIR). The DMP includes information on:

- the handling of research data during & after the end of AgriDataValue
- the collection, processing and, generation of data
- the methodology & standards that will be applied
- the data sharing and open access stipulations
- the way data will be curated & preserved (including after the end of the project).

This document represents the second version of the DMP, as a better view has been obtained on how the data and data handling evolved within the project during the first years of the project's lifetime.

Within this DMP, the term 'data' refers mainly to the following, non-exhaustive categories of information that will be used or produced during the project:

1. Datasets that are collected or generated as a result of project activities including ADV Pilots
2. The deliverables within the WPs that will be either openly accessible or sensitive
3. Communication and dissemination materials, including mailing lists, media items, and reports

AgriDataValue project files are made accessible to the project partners through the ownCloud platform (https://owncloud.synelixis.com/) which is governed by the project coordinator Synelixis. File structure and file names are therefore intended to be clear, logic, descriptive, and consistent to make it obvious where to find specific data or documents. Documents are ordered and new documents will be ordered using the project work package structure, supplemented with folders for administration, deliverables, dissemination, meetings, etc.

1.2 Structure of the document

The deliverable is structured as follows:

- Chapter 1 provides the introduction of the deliverable,
- Chapter 2 presents the AgriDataValue pilots and associated datasets,

- Chapter 3 presents Earth Observation datasets and datasets that are available or will be generated within AgriDataValue,
- Chapter 4 presents the AgriDataValue pilots, the data management strategy, thereby exposing classification, archiving, performance, safety, and security, FAIR data, and ethics requirements and procedures for data management,
- Chapter 5 considers Data Security,
- Chapter 6 considers Ethics and GDPR compliance,
- Chapter 7 presents recommendations and further steps.

2 AgriDataValue Pilots

The ADV Pilots are quite important for the success of the AgriDataValue project. It should be underlined that the Horizon Call focus and the main objective of the AgriDataValue project is to establish itself as the "Game Changer" in Smart Farming digital transformation and agri-environmental monitoring, as an innovative, open source, intelligent and multi-technology, fully distributed Agri-Environment Data Space (ADS). In detail, AgriDataValue aims to design and develop a flexible, open, federated, and fully distributed ADS as a "Platform of Platforms" to collect, process, enhance, and combine fair, diverse, heterogeneous, and multi-sources (sensor) data, and develop approaches and tools to upscale EU-wide data sets, while maintaining their sovereignty, trust and quality (Project Objective O1) and design, develop, deploy and establish confidence to innovative edge cloud and AI tools that collect and process at the source, combine, map and analyse in real-time in-situ, regional and global terrestrial, aerial and satellite agri-environment data to enhance production and food chain traceability, while supporting climate policies (Project Objective O2). To achieve the above Objectives O1 and O2 it is quite important to collect and process (on-site or in an aggregated way) crop, livestock, and environmental data, which are mandatory for ML training so that it may also demonstrate and prove the value of upscaling and combining agri-environmental data in agricultural (both arable and livestock) and climate change informed decision making in real demonstrators (Project Objective O4).

2.1 Pilots Governance Committee (PGC)

Due to its importance, since the project proposal time, we have specified that each pilot site will be coordinated by a ***Pilots Governance Committee (PGC)***. The PGC will consist of the following persons:

- **Pilot Coordinator**. This person will have the overall responsibility of the pilot and will be the contact person for all pilot-related activities. The Pilot Coordinator will be the owner of the pilot, the president of a cooperative, or the person who has the decision-making authority related to pilot-related strategic decisions. The pilot coordinator will be assisted by three Chairs, namely:
- **Technical Chair.** This person will be responsible for coordinating the technological adaptation and upscale of the data. He or she should understand and collaborate with the AgriDataValue technological partners, both the ones related to the AgriDataValue platform and the project ML team, to identify where the data are available, in which format they are stored, what their characteristics are (e.g. frequency, accuracy, error percentage), which adaptations need to be applied to them before they may be used and under which conditions they may become available, either on-site or aggregated at the edge or the cloud, and who the target audience for these data are (i.e. only the project consortium, the scientific community, or the general interested audience and stakeholders).
- **Ethical & Ecosystem Chair.** This person will be responsible to coordinate the multi-stakeholder and multi-actor involvement and ensure ethical and legal compliance in collaboration with the project Data Protection Officer (DPO). The Ethical & Ecosystem chair will ensure that all relevant laws, regulations, and EC directives concerning livestock and humans are met.
- **Business Chair.** This person will focus on the market impact of the experimentation. In detail, he or she will be responsible for analysing the benefits of the experimentation, performing cost-benefit analysis of

the new sensors to be installed, and cost-benefit analysis of the actual implementation and deployment of the AgriDataValue solution to the pilot.

Though the abovementioned structure is highly recommended by the AgriDataValue consortium, for simplicity reasons and to lower any additional overhead, a person may have more than one role in a pilot.

Table 1: Overview of the pilot numbers with their coordinators, technical, ecosystem, and business chairs

#	Pilot partner	Pilot coordinator	Technical Chair	Ecosystem Chair	Business Chair
1	UL	Patrycja Grzyś	Paulina Tobiasz-Lis	Paulina Tobiasz-Lis	Patrycja Grzyś
2	Delphy	Jan Kiers	Jan Kiers	Jan Kiers	Jan Kiers
3	ZSA	Inga Berzina	Aleksejs Zacepins	Inga Berzina	Maira Dzelzkaleja-Burmistre
4	TBA	Ioannis Katris	Spiros Karamplianis	Spiros Karamplianis	Ioannis Katris
5	Inagro	Jan Vanwijsberghe	Jan Vanwijsberghe	Sarah Bossuyt	Eva Ampe
6	TEC	Guadalupe Lopez	Begoña Camacho	Carolina Martinez	Mari Carmen Galera
7	Inagro	Emma Vandenbogaerde	Emma Vandenbogaerde	Sarah Bossuyt	Eva Ampe
8	Inagro	Tim De Cuypere	Tim De Cuypere	Sarah Bossuyt	Eva Ampe
9	Inagro	Sarah Bossuyt	Eva Ampe	Sarah Bossuyt	Eva Ampe
10	Inagro	Sarah Bossuyt	Tim De Cuypere	Sarah Bossuyt	Eva Ampe
11	Delphy	Jan Kiers	Jan Kiers	Jan Kiers	Jan Kiers
12	SARGA	Javier Sancho	Javier Sancho	Javier Sancho	Javier Sancho
13	TBA	Ioannis Katris	Spiros Karamplianis	Spiros Karamplianis	Ioannis Katris
14	CVSE	Nawel Aouadi	Franck Binard	Franck Binard	Franck Binard
15	RINO	Giovanni Nigro	Domenico Bossio	Giovanni Nigro	Domenico Bossio
16	NILEAS	Vicky Inglezou	Nikos Tsotsolas	Vicky Inglezou	George Kokkinos
17	RINO	Giovanni Nigro	Domenico Bossio	Giovanni Nigro	Domenico Bossio
18	BioRO	Avraham Marian Cioceanu	Alexandra Tudorascu	Avraham Marian Cioceanu	Elena Ilie
19	EV ILVO	Nico Peiren	Tim Van De Gucht	Nico Peiren	Leen Vandaele
20	ZSA	Aleksejs Zacepins	Aleksejs Zacepins	Inga Berzina	Maira Dzelzkaleja-Burmistre
21	TBA	Ioannis Katris	Spiros Karamplianis	Spiros Karamplianis	Ioannis Katris
22	EVILVO	Nico Peiren	Tim Van De Gucht	Nico Peiren	Sarah De Smet
23	TBA	Ioannis Katris	Spiros Karamplianis	Spiros Karamplianis	Ioannis Katris

2.2 Pilots Description

Based on lessons learned, success methodologies, and open innovation ecosystems supported by the EC (both DG Agriculture and DG Connect) and Local Action Groups, AgriDataValue will follow the Lean Multi-Actor Approach (LMAA). LMAA was introduced by the H2020 project IoF2020 (Internet of Food & Farm) [1], most of AgriDataValue consortium members have been IoF2020 partners, and further developed by EIP-AGRI Network [2] to foster the

development of research, uptake innovations into operational applications and create real impact in the agri-environment domain.

AgriDataValue LMAA (Figure 1) will fertilize new ideas and create a wide range of field tools, thanks to interactions between complementary actors and sharing knowledge, expertise, and capabilities. We plan to adopt LMAA as a fully demand-driven approach, involving during each iterative cycle of AgriDataValue's lifetime, various actors (i.e. farmers, farmers' groups and cooperatives, foresters and forestry groups, advisors, stakeholders, researchers, CAP paying authorities, decision making bodies, etc.) to demonstrate on one hand how the project fulfils the proposed objectives, needs, problems and opportunities of the full chain, from farmers to service advisors, suppliers, and stakeholders, and on the other hand, how it complements existing research, innovation, and best practices. The AgriDataValue LMAA is implemented through requirements and specifications extracted from comprehensive use cases originating from farmers and cooperatives (WP1), a complete set of tools and AI models co-created with technological partners and end-users (WP2-WP3) and fully tested and validated through pilots (WP4). Feedback from pilots via human interaction with all stakeholders is further utilized to extend and upgrade sensors and pilots (WP5) and create a real impact, not only to agriculture, but also to the greater public (WP6).

Figure 1: LMAA adapted to AgriDataValue concept

The AgriDataValue tools, use cases and LMAA approach will be fully validated in **23 pilots in 9 EU countries (Figure 2)**. The pilots will also collect sensor data and feedback from the involved end-users, and monitor and adapt the pilots over the project lifetime. Following the overall project-phased approach, the 1st phase (M1-M12) will prepare the pilots and capture new sensor data, extending the already available datasets. The activity will continue throughout the project's lifetime (and at least 3 years after); yet, captured datasets will be used in the 2nd phase (M13-M36) for the ML training and evaluation of the evolving impact in each use case. In this phase, mature AgriDataValue tools and solutions will be delivered evaluating the impact and collecting feedback. The 3rd phase (M37-M72), based on sensor market analysis, pilot evaluation, and feedback from the LMAA, will drive the overall pilots' expansion with new sensors, experimentation, feedback, and impact assessment.

Figure 2: AgriDataValue Pilots Geographical Distribution

The following section describes the pilots available to the AgriDataValue project, which will be used to collect data and validate various use cases.

2.2.1 Pilot 1: Orchard - Lodzkie, Poland

Pilot Owner:	UL
Number of ha	17 ha
Farmers involved Directly:	1
Farmers involved indirectly:	> 100
Available Data:	Soil and weather data

Pilot #1 refers to the "Wiatrowy Sad" (*Wind Orchard*). It is a 17-hectare family-owned apple orchard located in Kałęczew, within the Lodz Heights Landscape Park (Lodz province) in Poland. As part of the AgriDataValue project, Wiatrowy Sad integrates the SynField smart agriculture system, which includes a meteorological station and soil sensor to monitor real-time environmental data. This technology helps optimize orchard management by providing precise information on ambient temperature, humidity, wind speed, and soil conditions. Gathered data supports precision agriculture by allowing for more informed and sustainable decisions, optimizing fruit production and orchard management. The integration of this technology represents a step towards a digital orchard model, where data-driven solutions ensure efficiency and sustainability in response to market demands.

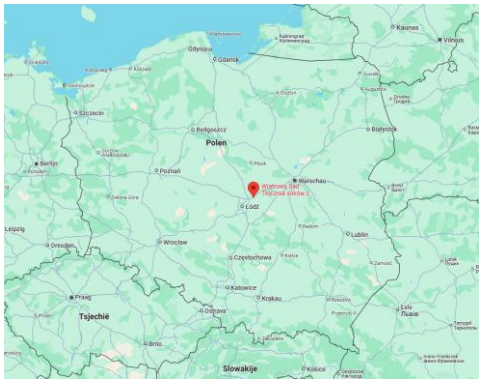


Figure 3: AgriDataValue Pilot 1 Location

Figure 4: Figures from Pilot 1.

Dataset #1	
Overview	
Dataset Name	UL Orchard
Dataset Category	Primary data collected by partner in ADV
Data Owner	UL
Description	The datasets contain air temperature, humidity, wind speed and direction, rainfall, and soil moisture levels that are collected in real time

Already before the ADV project?	Yes. Additional data concerning the use of water, plant protection products and fertilizers will be captured.
Data purpose in ADV	Train and challenge ML models to support precision irrigation (1), ensure that fertilizers are applied efficiently, reducing waste and improving overall crop health (2), predict when pest outbreaks are likely to occur and advise on the best timing for pesticide application (3)
Use beyond ADV	Could be helpful in efficient data linking

Measuring technologies	**Measuring**	**Units**	**Frequency**
Thermometer	Air Temperature	°C	15-20 minutes
Air Sensor	Air Humidity	%	15-20 minutes
Anemometer	Wind Direction	degrees	15-20 minutes
Anemometer	Wind Speed	m/sec	15-20 minutes
Rain Gauge	Rainfall	mm	15-20 minutes
Soil Sensor	Soil Moisture (VWC)	%	15-20 minutes
Soil Sensor	Electrical Conductivity	mS/m	15-20 minutes
Soil Sensor	Soil Temperature	°C	15-20 minutes
Other Data	**Measuring**	**Units**	**Frequency**
Field observation data	Irrigation water volume	m3	Daily
Field observation data	Spraying volume	m3	Daily

Storage and access details	
Data format type	CSV, XLS
Storage Location	Cloud based
Storage Duration	Permanent
Is the data public or confidential?	Confidential
Data harmonisation activities required?	Probably
Data merging activities required?	Probably
Can data be used after the ADV project?	Yes

Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes. AI data are needed, or it is allowed to be used. For each specific demand/sharing, permission should be approved.
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes.
Does dataset contain previously collected personal data?	No

Dataset #1: Data use per pilot use case				
Data variable	**UC 1.1**	**UC 1.2**	**UC 1.3**	**UC 2.5.**
Air Temperature	Yes	Yes	Yes	Yes
Air Humidity	Yes	No	Yes	Yes
Wind Direction	Yes	No	No	Yes
Wind Speed	Yes	No	Yes	Yes
Rainfall	Yes	Yes	Yes	Yes
Soil Moisture (VWC)	Yes	Yes	Yes	Yes
Electrical Conductivity	Yes	Yes	No	Yes

Soil Temperature	Yes	Yes	Yes	Yes
Irrigation water volume	Yes	No	No	No
Spraying volume	No	No	Yes	No

2.2.2 Pilot 2: Onions - Flevoland, The Netherlands

Pilot Owner:	Delphy
Number of ha	A parcel of 6 ha.
Farmers involved Directly:	1
Farmers involved indirectly:	> 100
Available Data:	Soil and weather data

Pilot 2 is situated at an arable farm in Swifterbant, Flevoland, The Netherlands. The involved partner is Delphy. The pilot is being conducted in onion cultivation. On a 3-year trial plot of Delphy at this grower's farm, experiments are being conducted with the use of 13 soil sensors and multiple weather and rain sensors. The data from these sensors is used in QMS water, which is a platform developed by Delphy to control drip irrigation and fertigation. Within QMS-water, the sensor data is supplemented with other data, such as crop type, crop growth, weather forecasts, among others, and this is used to control drip irrigation. The aim of the setup is to irrigate as precisely as possible and thus avoid improper use of precious water.

Figure 5: AgriDataValue Pilot 2 Location

Figure 6: AgriDataValue Pilot 2 Photos

Dataset #2			
Overview			
Dataset Name	DELPHY-Onions		
Dataset Category	Primary data collected by partner in ADV		
Data Owner	DELPHY		
Description	The datasets contain the output data of IoT sensors (soil and weather). It's optional that we use QMS-Water data (calculated irrigation advice) for ADV as well		
Already before the ADV project?	Delphy's 3-year pilot started at the same time as ADV. Both will be operational for 3 years		
Data purpose in ADV	Train and challenge ML models, possible improvement for QMS-water		
Use beyond ADV	Could be helpful in efficient data linking		
Measuring technologies	**Measuring**	**Units**	**Frequency**

Soil sensor (Farm21)	Soil moisture at several depths	%	Continuously
	Soil temperature	°C	Continuously
	Air temperature	°C	Continuously
	Air humidity	%	Continuously
Rain sensor (RMA)	Rain	Mm	Each visit
Other Data	**Measuring**	**Units**	**Frequency**
QMS-Water data	n.a. (calculated advice)	n.a	Daily
Crop yield data	Tonnes/ha	Kg	Once a year
Crop size grading	Numbers/category	Numbers	Once a year
Plant emergence	Number of plants/meter	Numbers	Once a year

Storage and access details	
Data format type	CSV, XML, XLSX
Storage Location	Cloud based
Storage Duration	Permanent
Is the data public or confidential?	Sensor data is confidential for owner, this data may be used publicly. QMS-water is developed and owned by Delphy. The calculated data may be used publicly, however calculation rules are the property of Delphy. Annual crop data can be used.
Data harmonisation activities required?	Probably
Data merging activities required?	Probably
Can data be used after the ADV project?	Yes

Ethics and Data Protection	
Includes Personal data?	Yes, however may be used publicly
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes. All data are needed, or it is allowed to be used. For each specific demand/sharing, permission should be approved.
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes. All data to be used will be anonymized or pseudonymised or synthetic data
Does dataset contain previously collected personal data?	No

Dataset #2: Data use per pilot use case	
Data variable	**UC 1.1**
Soil moisture at several depths	Yes
Soil temperature	Yes
Air temperature	Yes
Air humidity	Yes
Rain	Yes
QMS-Water data	Yes
Crop yield data	Yes
Crop size grading	Yes
Plant emergence	Yes

2.2.3 Pilot 3: Wheat and hard wheat - Zemgale, Latvia

Pilot Owner:	ZSA
Number of ha	A parcel of 60 ha
Farmers involved Directly:	1
Farmers involved indirectly:	> 800
Available Data:	Soil and weather data

Pilot 3 is situated at an arable farm in Zemgale, Latvia. ADV Involved partner ZSA is the biggest agricultural organisation of producers in Latvia. The ZSA plans to inform more than 800 producers on AgriDataValue results. The pilot is currently equipped with an Agro-weather station, and an additional weather station with soil moisture sensor is installed. The current monitoring system can detect air temperature, air humidity, rainfalls and wind. The additional weather station is used to monitor the same main weather parameters like temperature, rain, wind and is installed in different field locations to analyse the differences in the measurements within one field if any. The soil sensor for soil temperature, water volume and conductivity is provided by Synelixis and used to monitor soil parameters.

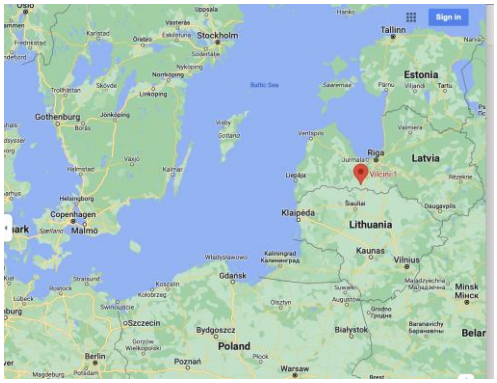


Figure 7.1: AgriDataValue Pilot 3 Location

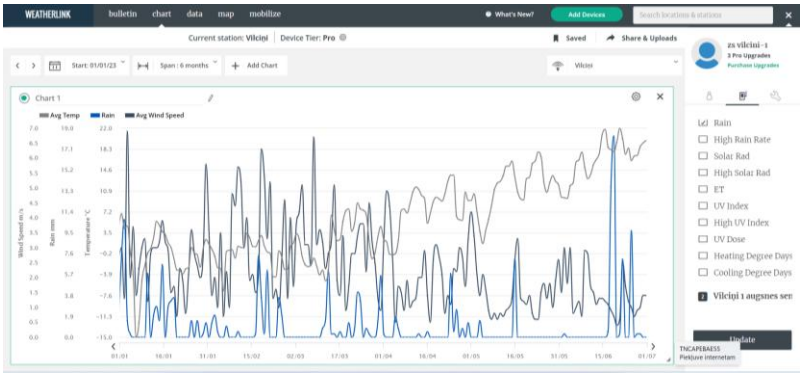


Figure 8: AgriDataValue Pilot 3 Photos

Dataset #3	
Overview	
Dataset Name	Weather data and plant monitoring/observation data
Dataset Category	Primary data collected by partner in ADV

Data Owner	Vilcini 1 Farm
Description	The datasets contain air humidity and temperature, weather data and will be expanded with soil sensor to measure additional soil parameters. A decision support system that helps to assess the risk of the development of various diseases of winter wheat leaves and ears under specific condition.
Already before the ADV project?	Yes. Additional data will be captured.
Data purpose in ADV	Train/challenge ML models to detect prediction of pest spraying
Use beyond ADV	Could be helpful in efficient data linking

Measuring technologies	**Measuring**	**Units**	**Frequency**
Thermometer	Air Temperature	°C	15 minutes
Air Sensor	Air Humidity	%	15 minutes
Anemometer	Wind Direction		15 minutes
Anemometer	Wind Speed	m/sec	15 minutes
Rain Gauge	Rainfall	mm	15 minutes
Other Data	**Measuring**	**Units**	**Frequency**
Field Observation Data	Observation date	Date	Weekly
Field Observation Data	Wheat BBCH	text	Weekly
Field Observation Data	Plant diseases rate	%	Weekly

Storage and access details	
Data format type	CSV, XLS
Storage Location	Cloud Server
Storage Duration	Permanent
Is the data public or confidential?	Confidential
Data harmonisation activities required?	Yes
Data merging activities required?	Yes
Can data be used after the ADV project?	Yes, but permission needs to be granted

Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #3: Data use per pilot use case	
Data variable	**UC 1.1**
Air Temperature	Yes
Air Humidity	Yes
Wind Direction	Yes
Wind Speed	Yes
Rainfall	Yes
Observation date	Yes
Wheat BBCH	Yes
Plant diseases rate	Yes

2.2.4 Pilot 4: Forages (clovers, corn) - Agrinio, Greece

Pilot Owner:	TBA
Number of ha	60 ha
Farmers involved Directly:	15
Farmers involved indirectly:	> 5,000
Available Data:	Weather, irrigation and soil data from forage crops for livestock feeding

Pilot 4 is located in Agrinio, Etoloakarnania (Greece) and covers forage (clovers and corn crops) for livestock feeding. The pilot is owned by TBA and directly involves 15 farmers. However, TBA via the mother company (Agrinio Union Cooperative) plans to inform more than 5,000 producers on AgriDataValue results, mainly on irrigation and fertilization reduction. The pilot is currently equipped with only 1 Agro-weather station, but additional stations will be installed. The current monitoring system can detect air temperature, air humidity, rainfalls and wind. Additional weather stations are planned to be included.

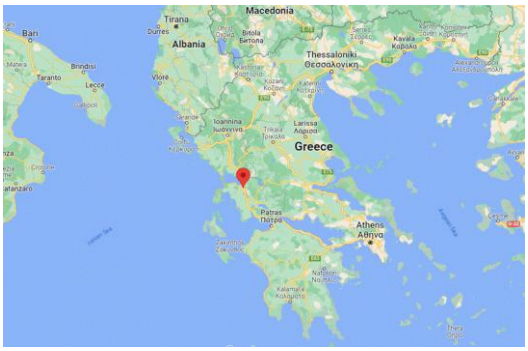


Figure 9: AgriDataValue Pilot 4 Location


Figure 10: Figures from AgriDataValue Pilot 4 forages

Dataset #4			
Overview			
Dataset Name	TBA Forage production		
Dataset Category	Primary data collected by partner in ADV		
Data Owner	TBA		
Description	The datasets contain wind, rain and temperature weather and soil data		
Already before the ADV project?	Limited. Additional data will be captured.		
Data purpose in ADV	Train/challenge ML models to irrigation and fertilization optimization.		
Use beyond ADV	Could be helpful in efficient data linking		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Thermometer	Air Temperature	°C	15 minutes
Air Sensor	Air Humidity	%	15 minutes
Anemometer	Wind Direction	degrees	15 minutes
Anemometer	Wind Speed	m/sec	15 minutes
Rain Gauge	Rainfall	mm	15 minutes
Soil Sensor	Soil Moisture (VWC)	%	15 minutes
Soil Sensor	Electrical Conductivity	mS/m	15 minutes

Soil Sensor	Soil Temperature	°C	15 minutes
Other Data	**Measuring**	**Units**	**Frequency**
Calculated Data	Evapotranspiration	mm	Daily
Other Data	Irrigation Water volume	m^3	Daily
Other Data	Organic Fertilization volume	m^3	Daily

Storage and access details	
Data format type	CSV, XML, JSON
Storage Location	Cloud Server
Storage Duration	Permanent
Is the data public or confidential?	Confidential
Data harmonisation activities required?	Yes
Data merging activities required?	Yes
Can data be used after the ADV project?	Yes, but permission needs to be granted

Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #3: Data use per pilot use case		
Data variable	**UC 1.1**	**UC 1.2**
Air Temperature	Yes	Yes
Air Humidity	Yes	Yes
Wind Direction	Yes	Yes
Wind Speed	Yes	Yes
Rainfall	Yes	Yes
Soil Temperature	Yes	Yes
Soil Moisture	Yes	Yes
Soil Electrical Conductivity	Yes	Yes
Evapotranspiration	Yes	Yes
Irrigation Water volume	Yes	Yes
Organic Fertilization volume	Yes	Yes

2.2.5 Pilot 5: Vegetables and arable crops - Flanders, Belgium

Pilot Owner:	Inagro
Number of ha	3.5 ha
Farmers involved Directly:	2
Farmers involved indirectly:	> 1000
Available Data:	Weather, soil and drone data

Pilot 5 is situated in Belgium (Flanders). It covers both vegetables and arable crops. The involved partner is Inagro. The pilot consists of a field with six crops that are rotated yearly called the Optifarm. New technologies and management practices are applied, and data is collected. The field simulates a real farmers' field while still allowing more risky and experimental treatments. The data acquired includes weather station data, laboratory analyses of soil samples, soil scans and multispectral drone data. Additional IoT sensors may be included during the AgriDataValue lifetime.

Figure 11: AgriDataValue Pilot 5 Location

Figure 12: Figures from AgriDataValue Pilot 5 Optifarm organic matter map

Dataset #5			
Overview			
Dataset Name	Optifarm		
Dataset Category	Primary data collected by partner in ADV		
Data Owner	Inagro		
Description	Inagro has a parcel in which the newest smart-farming techniques are demonstrated on five crops in a crop rotation. One of these techniques is spot spraying in celeriac to eliminate weeds. Both temporal and spatial data is collected from this field and could provide new insights in IPM and fertilization.		
Already before the ADV project?	Yes. Additional data will be captured.		
Data purpose in ADV	The data can be used to improve algorithms such as weed detection by remote sensing and machine learning.		
Use beyond ADV	When the algorithms are successful, others can use this to reduce inputs. Researchers can use our methods to expand the machine learning to other applications.		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Thermometer	Air Temperature	°C	15 minutes
Air Sensor	Air Humidity	%	15 minutes
Rain Gauge	Rainfall	mm	15 minutes
Soil Sensor	Soil Temperature	°C	15 minutes
Drone images	Spectral indices, reflectance	-	Multiple flights
Satellite images	Spectral indices, reflectance	-	Daily
Soil scan (Veris)	Organic carbon	%	1 scan
Soil scan (Veris)	pH		1 time
Soil scan (Veris)	Electrical conductivity (EC)	mS/m	1 time

Soil scan (Interra)	Nutrients	ppm	1 time
Soil scan (Interra)	Texture		1 time
Soil scan (Interra)	Organic carbon	%	1 time
Soil scan (Interra)	Plant available water	%	1 time
Yield mapping	Winter wheat yield	Ton/ha	1 per season
Yield mapping	Field bean yield	Ton/ha	1 per season
Other Data	**Measuring**	**Units**	**Frequency**
Soil samples	Chemical or NIR analysis	Mg/100mg	Daily
Field Observation Data	Pest, Phenology	-	Few times during season

Storage and access details	
Data format type	CSV, shapefile, geotif
Storage Location	Sharepoint, Terrascope, climate sensor app
Storage Duration	At least 6 years
Is the data public or confidential?	Confidential
Data harmonisation activities required?	Yes
Data merging activities required?	Yes
Can data be used after the ADV project?	Data are anonymised and can be used to challenge ML models

Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #5: Data use per pilot use case			
Data variable	**UC 1.2**	**UC 1.3**	**UC 2.1**
Air Temperature	yes	No	yes
Air Humidity	Yes	No	Yes
Rainfall	Yes	No	Yes
Soil Temperature	Yes	No	Yes
Drone Spectral indices, reflectance	Yes	Yes	Yes
Satellite Spectral indices, reflectance	Yes	No	Yes
Organic carbon	Yes	No	Yes
pH	Yes	No	Yes
Electrical conductivity (EC)	Yes	No	Yes
Nutrients	Yes	No	Yes
Texture	Yes	No	Yes
Organic carbon	Yes	No	Yes
Plant available water	Yes	No	Yes
Winter wheat yield	Yes	No	Yes
Field bean yield	Yes	No	Yes
Soil samples	Yes	No	Yes
Field Observation Data	Yes	No	Yes

2.2.6 Pilot 6: Greenhouse vegetables - Andalusia, Spain

Pilot Owner:	TEC
Number of ha	12 ha
Farmers involved Directly:	1
Farmers involved indirectly:	> 500
Available Data:	Climatic and soil data Weather/Climate, irrigation, fertilizers and pesticides

Pilot 6 is situated in the desertic area of Almeria, Andalusia Region (Spain) and covers 12ha of greenhouses. The pilot is currently equipped with 5 agro-weather stations and some IoT sensors (measuring wind, rain, air, soil parameters). Additional 4-6 sensors are planned to be included during the project lifetime.

The Experimental Center for technology and agronomic trials is located just 10 minutes from TECNOVA main headquarters. This agronomic research center has 12 hectares of facilities among which: 14 different types of greenhouse infrastructures, an office building and training and transfer rooms, 3 reservoirs for irrigation and 6 irrigation booths. It is one of the most advanced of its kind and was opened in 2014.The Center offers to the business sector an area for research, development and validation of technology of the Agriculture Auxiliary Industry and Services.

Figure 13: AgriDataValue Pilot 6 Location

Figure 14: Pictures from AgriDataValue Pilot 6

Dataset #6	
Overview	
Dataset Name	Greenhouse vegetables
Dataset Category	Primary data collected by partner in ADV
Data Owner	TEC
Description	This pilot is being conducted since 2014. It is a research site used by many crop producers, seeds manufactures and other agri-research private bodies. All the data is collected by digital platforms operated under license.
Already before the ADV project?	Yes

Data purpose in ADV	Train and challenge ML models to improve greenhouse practices		
Use beyond ADV	To enhance use of this kind of data		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Soil sensors	Soil moisture	%	5 minutes
Weather station	Rainfall	mm	5 minutes
Weather station	Wind speed	m/sec	5 minutes
Weather station	Wind direction	degrees	5 minutes
Thermometer	Air Temperature	°C	5 minutes
Other Data	**Measuring**	**Units**	**Frequency**
Calculated Data	Evapotranspiration	mm	Daily
Calculated Data	Expected Rainfall	%	Daily
Field Observation Data	crop development		Daily
Other Data	Fertilizers	Kg	Administered amount
Other Data	Pesticides	m^3	Administered amount
Storage and access details			
Data format type	To be determined		
Storage Location	To be determined		
Storage Duration	At least 1 year		
Is the data public or confidential?	Public		
Data harmonisation activities required?	Probably		
Data merging activities required?	Probably		
Can data be used after the ADV project?	Yes		
Ethics and Data Protection			
Includes Personal data?	No		
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes. AI data are needed, or it is allowed to be used. For each specific demand/sharing, permission should be approved.		
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes		
Does dataset contain previously collected personal data?	No		

Dataset #6: Data use per pilot use case	
Data variable	**UC 2.1**
Soil moisture	Yes
Rainfall	Yes
Wind speed	Yes
Wind direction	Yes
Air Temperature	Yes
Evapotranspiration	Yes
Expected Rainfall	Yes
crop development	Yes
Fertilizers	Yes
Pesticides	Yes

2.2.7 Pilot 7: Endives - Flanders, Belgium

Pilot Owner:	Inagro
Number of ha	50 ha
Farmers involved Directly:	5
Farmers involved indirectly:	> 500
Available Data:	Weather and satellite data

Pilot 7 is situated in Belgium (Flanders) covers vegetables. The involved partner is Inagro. In Belgian Endives, the emergence is a very important step in the production process. Low or heterogenous emergence can affect not only the yield of Belgian Endive roots but also the quality. With low emergence counts, the roots grow too thick or heterogeneous. The timing of cold storage and forcing becomes more difficult for the farmer when the roots are heterogeneous, and thick roots yield lower quality crops. Sufficient soil moisture is important for emergence. In this pilot we will help farmers to make use of soil moisture content satellite imagery to make improved irrigation decisions. The soil moisture satellite images can also be used in other crops that are sensitive to droughts such as spinach, beans and potatoes.

Figure 15: AgriDataValue Pilot 7 Location

Figure 16: Figures from AgriDataValue Pilot 7 Irriwatch satellite platform

Dataset #7			
Overview			
Dataset Name	Belgian Endives		
Dataset Category	Primary data collected by partner in ADV		
Data Owner	Inagro		
Description	Belgian Endives emergence is highly dependent on soil water content. Historical data are available but additional will be gathered. This new emergence data can be merged with soil water content satellite images.		
Already before the ADV project?	No		
Data purpose in ADV	To find the minimum amount of soil moisture content needed.		
Use beyond ADV	The minimum soil moisture content can be useful information for farmers.		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Thermometer	Air Temperature	°C	15 minutes

Air Sensor	Air Humidity	%	15 minutes
Rain Gauge	Rainfall	mm	15 minutes
Soil Sensor	Soil Temperature	°C	15 minutes
Soil moisture sensor	Soil moisture	%	15 minutes
Satellite images	Soil moisture		Daily
Satellite images	Spectral indices, reflectance		Daily
Other Data	**Measuring**	**Units**	**Frequency**
Field Observation Data	Number of plants	Plants/m²	Few times during emergence
Soil data	Bulk density	g dry soil/cm³	1 time per field
Storage and access details			
Data format type	CSV, shapefile, geotif		
Storage Location	Sharepoint, Terrascope, Irriwatch platfrom, Fieldclimate platform		
Storage Duration	At least 6 years		
Is the data public or confidential?	Confidential		
Data harmonisation activities required?	Yes		
Data merging activities required?	Yes		
Can data be used after the ADV project?	Data are anonymised and can be used to challenge ML models		
Ethics and Data Protection			
Includes Personal data?			No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?			Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?			Yes
Does dataset contain previously collected personal data?			No

Dataset #7: Data use per pilot use case		
Data variable	**UC 1.1**	**UC 5.4**
Air Temperature	Yes	Yes
Air Humidity	Yes	Yes
Rainfall	Yes	Yes
Soil Temperature	No	Yes
Soil moisture	Yes	Yes
Soil moisture (from satellite)	Yes	Yes
Spectral indices, reflectance	Yes	Yes
Number of plants	Yes	No
Bulk density	Yes	No

2.2.8 Pilot 8: Leek - Flanders, Belgium

Pilot Owner:	Inagro
Number of ha	40 ha
Farmers involved Directly:	10
Farmers involved indirectly:	> 1000
Available Data:	Weather, soil and drone data

Pilot 8 is situated in Belgium (Flanders) and covers vegetables. The involved partners are Inagro and ILVO. Special focus will be put on Leek Fertilization. Nitrogen leaching is a major problem in the cultivation of leeks. In the past Inagro has done on farm trials for place specific fertilization in leeks. This data will be further analysed by the project consortium to improve the current growth model and provide new insights in leek quality and fertilizer reduction. Multispectral data, soil scans, laboratory analyses of soil samples, yield and quality data will be provided. Improved advice on leek fertilization will be applied on the Inagro Optifarm field.

Figure 17: AgriDataValue Pilot 8 Location

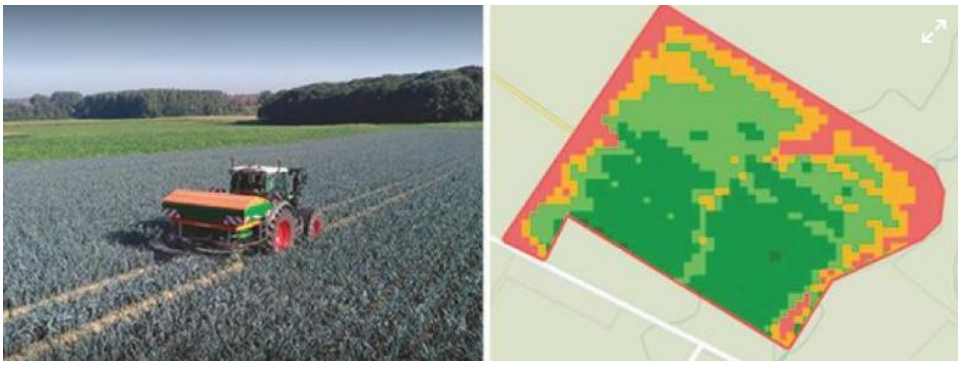


Figure 18: Figures from AgriDataValue Pilot 8 task map for leek fertilizer spreading

Dataset #8	
Overview	
Dataset Name	Leek Fertilization
Dataset Category	Primary data collected by partner in ADV and in previous project
Data Owner	Inagro and project WikiLeeks consortium
Description	Within the Flemish project Wikileeks, data was collected from leek fertilizer trials. This dataset includes soil scans, field measurements, analyses of tissue- and soil samples and remote sensing data.
Already before the ADV project?	Yes during the Wikileeks project. Additional data will be captured.
Data purpose in ADV	The historical leek fertilization data will be compiled, and the crop growth will be taken into consideration by using the remote sensing imagery. This allows us to improve the current model.
Use beyond ADV	Improvement of the current model can help researchers to simulate the optimal fertilizer amount for leeks and use this in precision farming applications.

Measuring technologies	Measuring	Units	Frequency
Drone images	Spectral indices, reflectance		Multiple flights
Satellite images	Spectral indices, reflectance		Daily
Soil scan (Veris)	Organic carbon	%	1 time
Soil scan (Veris)	pH		1 time
Soil scan (Veris)	Electrical conductivity (EC)	mS/m	1 time
Weather station	Temperature	°C	During season (for some fields)

Weather station	Soil moisture	%	During season (for some fields)
Weather station	Precipitation	mm	During season (for some fields)
Other Data	**Measuring**	**Units**	**Frequency**
Soil samples	Chemical or NIR analysis	Mg/100mg	During the season
Field Observation Data	Pest, Phenology	-	Few times during season
Interpolated weather data	Temperature	°C	Daily
Interpolated weather data	Precipitation	mm	Daily
Crop data	Biomass	ton/ha	During season
Crop data	Yield	ton/ha	At harvest
Crop data	Dry matter plants harvested	g DM/ plant	At harvest
Crop data	Nitrogen uptake	kg N/ ha	During season
Crop data	Crop quality		At harvest
Crop data	Crop size		At harvest

Storage and access details	
Data format type	CSV, shapefile, geotif
Storage Location	Sharepoint, Terrascope
Storage Duration	At least 6 years
Is the data public or confidential?	Confidential
Data harmonisation activities required?	Yes
Data merging activities required?	Yes
Can data be used after the ADV project?	Data are anonymised and can be used to challenge ML models

Ethics and Data Protection	
Includes Personal data?	Yes
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	Yes, it includes data of farmers under agreement

Dataset #8: Data use per pilot use case			
Data variable	**UC 1.2**	**UC 2.1**	**UC 2.2**
Spectral indices, reflectance	Yes	Yes	Yes
Organic carbon	Yes	Yes	Yes
pH	Yes	Yes	Yes
Electrical conductivity (EC)	Yes	Yes	Yes
Temperature	Yes	Yes	Yes
Soil moisture	Yes	Yes	Yes
Precipitation	Yes	Yes	Yes
Soil samples	Yes	Yes	Yes
Pest, Phenology	Yes	Yes	Yes
Temperature	Yes	Yes	Yes
Precipitation	Yes	Yes	Yes

Biomass	Yes	Yes	Yes
Yield	Yes	Yes	Yes
Dry matter plants harvested	Yes	Yes	Yes
Nitrogen uptake	Yes	Yes	Yes
Crop quality	Yes	Yes	Yes
Crop size	Yes	Yes	Yes

2.2.9 Pilot 9: Potato - Flanders, Belgium

Pilot Owner:	Inagro
Number of ha	2 ha
Farmers involved Directly:	2
Farmers involved indirectly:	2
Available Data:	Hyperspectral line scan data

Pilot 9 is situated in Belgium (Flanders) covers arable crops. The involved partners are Inagro and ILVO. This pilot will be focussed on post-harvest quality of potatoes. Defects such as black spot are not always visible without destructive measurements and so these products often end up with the consumer. The screening of these defects could be done with sensors and AI. Will we use hyperspectral line scanning combined with models.

Figure 19: AgriDataValue Pilot 9 Location

Figure 20: Figures from AgriDataValue Pilot 9 hyperspectral imaging of potatoes

Dataset #9			
Overview			
Dataset Name	Potato quality		
Dataset Category	Primary data collected by partner in ADV		
Data Owner	Inagro & ILVO		
Description	This dataset will contain hyperspectral measurements of post-harvest potatoes and quality scores of trained evaluators.		
Already before the ADV project?	Yes, during the Spectrofood project. Additional data will be captured.		
Data purpose in ADV	This data will be used to improve the current PLS model with machine learning models that are able to detect defects and diseases.		
Use beyond ADV	Fast and non-destructive disease detection could be useful in researching the causes of these defects/diseases		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Hyperspectral scanner FX10	400-1000 nm reflectance	nm	Post harvest: before and after impact

Hyperspectral scanner FX17	900-1700 nm reflectance	nm	Post harvest: before and after impact
RGB camera	Color picture of potato		Post harvest: before and after impact
Other Data	**Measuring**	**Units**	**Frequency**
Post harvest assessment	Scab score	score 1-9	once
Post harvest assessment	Blackspot score	Score 1-9	once
Laboratory assessment	Dry matter percentage	%	once

Storage and access details	
Data format type	CSV, images
Storage Location	Sharepoint
Storage Duration	At least 6 years
Is the data public or confidential?	Confidential
Data harmonisation activities required?	Yes
Data merging activities required?	Yes
Can data be used after the ADV project?	Data are anonymised and can be used to challenge ML models

Ethics and Data Protection	
Includes Personal data?	Yes
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #9: Data use per pilot use case	
Data variable	**UC 1.4**
400-1000 nm reflectance	Yes
900-1700 nm reflectance	Yes
Colour picture of potato	Yes
Scab score	Yes
Blackspot score	Yes
Dry matter percentage	Yes

2.2.10 Pilot 10: Vegetables - Flanders, Belgium

Pilot Owner:	Inagro
Number of ha	30 ha
Farmers involved Directly:	8
Farmers involved indirectly:	> 500
Available Data:	Satellite, drone, soil and weather data

Pilot 10 is situated in Belgium (Flanders) covers Western Flanders. The involved partner is Inagro. Pilot 10 will build a learning network of farmer applying precision agriculture. The farmers are very enthusiastic about this initiative. The goal is to apply precision techniques, reduce fertilizer and pesticide use. But above all to increase farmers digital independence. The learning network exist out of 5 farmers. Each year some farmers will experiment on their own field. Our researchers will guide them and follow up the trials where needed. We will also teach them how to use the technology without our help. The farmers will use satellite, drone, soils scan and weather station technology.

Figure 21: AgriDataValue Pilot 10 Location


Figure 22: Figures from AgriDataValue Pilot 10 task map with sampling points for farmers of learning network

Dataset #10			
Overview			
Dataset Name	Learning network smartfarming		
Dataset Category	Primary data collected by partner in ADV		
Data Owner	Inagro and farmers		
Description	The learning network for precision farming will bring farmers together that want to try out new technologies on their own farms. The data collected could include soil scans, satellite imagery, soil analysis and yield.		
Already before the ADV project?	No.		
Data purpose in ADV	The data can be used to improve algorithms such as weed detection by remote sensing and machine learning.		
Use beyond ADV	When the algorithms are successful, others can use this to reduce inputs. Researchers can use our methods to expand the machine learning to other applications.		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Thermometer	Air Temperature	°C	15 minutes
Air Sensor	Air Humidity	%	15 minutes
Rain Gauge	Rainfall	mm	15 minutes
Soil Sensor	Soil Temperature	°C	15 minutes
Drone images	Spectral indices, reflectance		Multiple flights
Satellite images	Spectral indices, reflectance		Daily
Soil scan	Organic carbon	%	1 scan

Other Data	Measuring	Units	Frequency
Soil samples	Chemical or nir analysis	Mg/100mg	Daily
Yield	Yield sampling	Ton/hectare	At harvest

Storage and access details	
Data format type	CSV, shapefile, geotif
Storage Location	Sharepoint, Terrascope
Storage Duration	At least 6 years
Is the data public or confidential?	Confidential
Data harmonisation activities required?	Yes
Data merging activities required?	Yes
Can data be used after the ADV project?	Data are anonymised and can be used to challenge ML models

Ethics and Data Protection	
Includes Personal data?	Yes
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	Yes, data owned by the farmers under sharing agreement. Data can by use anonymized

Dataset #10: Data use per pilot use case				
Data variable	**UC 1.2**	**UC 1.3**	**UC 2.1**	**UC 5.4**
Air Temperature	No	No	No	Yes
Air Humidity	No	No	No	Yes
Rainfall	No	No	No	Yes
Soil Temperature	No	No	No	Yes
Spectral indices, reflectance	Yes	Yes	Yes	Yes
Organic carbon	Yes	No	Yes	Yes
Soil samples	Yes	No	Yes	Yes
Yield sampling	Yes	No	Yes	Yes

2.2.11 Pilot 11: Fruit orchards - Gelderland, The Netherlands

Pilot Owner:	Delphy and/or a commercial grower
Number of ha	Several orchards (100 ha)
Farmers involved Directly:	1
Farmers involved indirectly:	>20
Available Data:	Vigour maps by drone, Soil, and weather data

Pilot 11 is situated at Proeftuin Randwijk in Gelderland, the Netherlands. The involved partner is Delphy, one the major agricultural advisory companies in The Netherlands. Delphy has dozens of Dutch Fruit farmers in its network of customers

The pilot is being conducted in apple and pear. Randwijk is the central location of Fruit research in The Netherlands and co-owned by Delphy. Not only at Randwijk, but also for a dozens of commercial Fruit growers, Delphy uses its QMS-Fruit platform to enhance the yield and to minimize the use of chemical pesticides. The core of QMS-Fruit is data per tree and the major source are drone images to calculate the number of flowers, vigour and autumn colouring.

Figure 23: AgriDataValue Pilot 11 Location

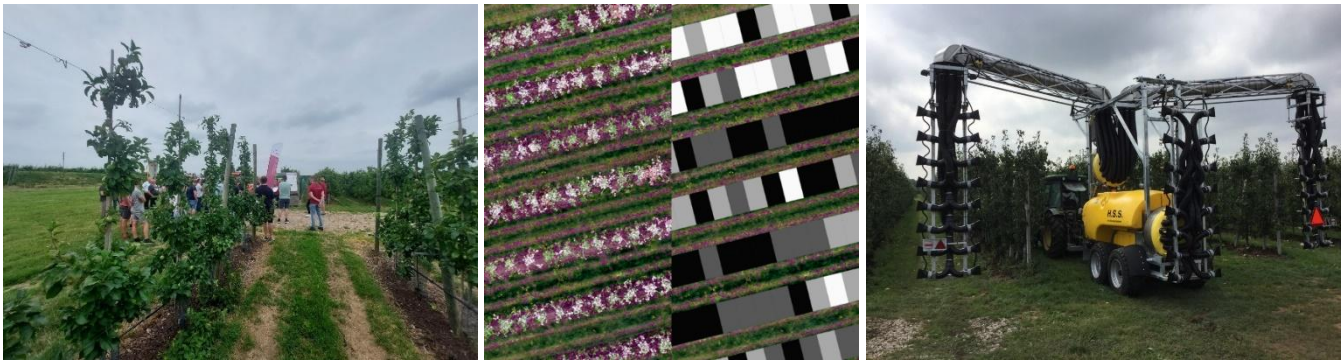


Figure 24: Photo's Pilot 11 fruit orchards

Dataset #11			
Overview			
Dataset Name	DELPHY-Fruit		
Dataset Category	Primary data collected by partner in ADV		
Data Owner	DELPHY		
Description	The datasets contain the data collected by a drone. It's optional that we use QMS-Fruit data (models for root cutting, fruit pruning and pests & diseases) for ADV as well		
Already before the ADV project?	Yes, Delph already has a lot of historical data and future data can be used		
Data purpose in ADV	Train and challenge ML models, possible improvement for QMS-Fruit		
Use beyond ADV	Could be helpful in efficient data linking		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Drone images	Multi spectral images		On demand/ historical
Other Data	**Measuring**	**Units**	**Frequency**
QMS-Fruit data	n.a. (calculated advice)	n.a	On request
Storage and access details			
Data format type	CSV, XML		
Storage Location	Cloud based		
Storage Duration	Permanent		
Is the data public or confidential?	Drone images may be used anonymously.		

	QMS-Fruit is developed and owned by Delphy. The calculated data may be used publicly (anonymously), however calculation rules are the property of Delphy.	
Data harmonisation activities required?	Probably	
Data merging activities required?	Probably	
Can data be used after the ADV project?	Yes	
Ethics and Data Protection		
Includes Personal data?		Yes, however may be used publicly
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?		Yes. All data are needed, or it is allowed to be used. For each specific demand/sharing, permission should be approved.
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?		Yes. All data to be used will be anonymized or pseudonymised or synthetic data
Does dataset contain previously collected personal data?		No

Dataset #11: Data use per pilot use case	
Data variable	**UC 1.3**
Drone images	Yes
QMS-Fruit data	Yes

2.2.12 Pilot 12: Non-Citrus Fruit Trees - Aragon Region, Spain

Pilot Owner:	SARGA/ITAINNOVA
Number of ha	160.000 ha
Farmers involved Directly:	2,000
Farmers involved indirectly:	> 10,000
Available Data:	Disease prediction models, Weather/Climate data, Satellite images

Pilot 12 is situated in Aragon Region (Spain) and covers 160.000 ha of non-citrus fruit trees. The pilot involves directly 2,000 producers, however SAGRA plans to inform more than 10,000 producers on AgriDataValue results. The pilot is currently equipped with more than 49 agro-weather stations and 500 IoT sensors (measuring wind, rain, air, soil parameters), while 9 suction probes are also available. Additional 8-12 agro-weather stations are planned to be included during the project lifetime.

Figure 25: AgriDataValue Pilot 12 Location


Figure 26: Figures from AgriDataValue Pilot 12 Non-Citrus Fruit Trees

Dataset #12.1			
Overview			
Dataset Name	Red FARA dataset		
Dataset Category	Secondary data (not publicly available)		
Data Owner	ITAINNOVA		
Provider (if different from partner)	Centro de Salud y Control Vegetal (CSCV) del Gobierno de Aragón / Plant Health and Control Center (CSCV) of the Government of Aragon		
Description	Field observation samples of the phenology evolution and pest presence in woody crops.		
Already before the ADV project?	Yes. Additional data will be captured.		
Data purpose in ADV	Train/challenge ML models to detect and improve pest prediction.		
Use beyond ADV	Could be helpful in efficient data linking		
Other Data	**Measuring**	**Units**	**Frequency**
Field Observation Data	Pest (adult mots)	Number	Weekly
Field Observation Data	Pest (%/number of affected leaves)	%	Weekly
Storage and access details			
Data format type	Postgres database with REST API		
Storage Location	Cloud Server		
Storage Duration	Permanent		
Is the data public or confidential?	Public but restricted access		
Data harmonisation activities required?	Yes		
Data merging activities required?	Yes		
Can data be used after the ADV project?	Yes, but permission needs to be granted		
Ethics and Data Protection			
Includes Personal data?			No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?			Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?			Yes
Does dataset contain previously collected personal data?			No

Dataset #12.2	
Overview	
Dataset Name	Spanish Cadastral Registry
Dataset Category	Secondary data (not publicly available)

Data Owner	ITAINNOVA		
Provider (if different from partner)	Spanish Treasury		
Description	Geographical description of the rustic farms in Aragón		
Already before the ADV project?	Yes. Additional data will be captured.		
Data purpose in ADV	Train/challenge ML models		
Use beyond ADV	Could be helpful in efficient data linking		
Other Data	**Measuring**	**Units**	**Frequency**
Earth Observation Data	Locate control parcels and obtain satellite images	Parcel geographical description	Weekly
Storage and access details			
Data format type	Postgres database with REST API		
Storage Location	Cloud Server		
Storage Duration	Permanent		
Is the data public or confidential?	Public but restricted access		
Data harmonisation activities required?	Yes		
Data merging activities required?	Yes		
Can data be used after the ADV project?	Yes, but permission needs to be granted		
Ethics and Data Protection			
Includes Personal data?			No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?			Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?			Yes
Does dataset contain previously collected personal data?			No

Dataset #12.3			
Overview			
Dataset Name	Copernicus Sentinel 2 Images Derived Data		
Dataset Category	Secondary data (not publicly available)		
Data Owner	ITAINNOVA		
Provider (if different from partner)	ESA		
Description	Geographical description of the rustic farms in Aragón		
Already before the ADV project?	Yes. Additional data will be captured.		
Data purpose in ADV	Train/challenge ML models		
Use beyond ADV	Could be helpful in efficient data linking		
Other Data	**Measuring**	**Units**	**Frequency**
Earth Observation Data	NDVI and derived data for control parcels	0 – 0.9	Weekly
Storage and access details			
Data format type	Postgres database with REST API		
Storage Location	Cloud Server		
Storage Duration	Permanent		
Is the data public or confidential?	Public but restricted access		
Data harmonisation activities required?	Yes		
Data merging activities required?	Yes		
Can data be used after the ADV project?	Yes, but permission needs to be granted		
Ethics and Data Protection			

Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #12.4			
Overview			
Dataset Name	Climatic data real observation		
Dataset Category	Secondary data (not publicly available)		
Data Owner	SARGA		
Provider (if different from partner)	AEMET/SIAR		
Description	Real climatic data observation (T, Humidity, Precipitation, etc.)		
Already before the ADV project?	Yes		
Data purpose in ADV	Train/challenge ML models to better un crop environmental conditions		
Use beyond ADV	Could be helpful in efficient data linking		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Weather station	Air Temperature	°C	15 minutes
Weather station	Air Humidity	%	15 minutes
Weather station	Wind Direction	degrees	15 minutes
Weather station	Wind Speed	m/sec	15 minutes
Weather station	Solar Radiation	W/m^2	15 minutes
Weather station	Rainfall	mm	15 minutes
Other Data	**Measuring**	**Units**	**Frequency**
Earth Observation Data	NDVI and derived data for control parcels	0 – 0.9	Weekly
Storage and access details			
Data format type	Postgres database with REST API		
Storage Location	Cloud Server		
Storage Duration	Not defined		
Is the data public or confidential?	Public but restricted access		
Data harmonisation activities required?	Yes		
Data merging activities required?	Yes		
Can data be used after the ADV project?	Yes, but permission needs to be granted		
Ethics and Data Protection			
Includes Personal data?			No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?			Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?			Yes
Does dataset contain previously collected personal data?			No

Dataset #12.5	
Overview	
Dataset Name	Climatic data estimations of ECMWF Reanalysis v5 (ERA5)
Dataset Category	Publicly available dataset
Data Owner	SARGA /ITAINNOVA

Provider (if different from partner)	ECMWF Reanalysis v5 (ERA5)		
Description	Real climatic data observation (T, Humidity, Precipitation, etc.)		
Already before the ADV project?	Yes		
Data purpose in ADV	Train/challenge ML models to better un crop environmental conditions		
Use beyond ADV	Could be helpful in efficient data linking		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Weather station	Air Temperature	°C	15 minutes
Weather station	Air Humidity	%	15 minutes
Weather station	Wind Direction	degrees	15 minutes
Weather station	Wind Speed	m/sec	15 minutes
Weather station	Solar Radiation	W/m^2	15 minutes
Weather station	Rainfall	mm	15 minutes
Other Data	**Measuring**	**Units**	**Frequency**
Earth Observation Data	NDVI and derived data for control parcels	0 – 0.9	Weekly
Storage and access details			
Data format type	Postgres database with REST API		
Storage Location	Cloud Server		
Storage Duration	Not defined		
Is the data public or confidential?	Public but restricted access		
Data harmonisation activities required?	Yes		
Data merging activities required?	Yes		
Can data be used after the ADV project?	Yes, but permission needs to be granted		
Ethics and Data Protection			
Includes Personal data?			No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?			Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?			Yes
Does dataset contain previously collected personal data?			No

Dataset #12.6			
Overview			
Dataset Name	Open Meteo Forecast		
Dataset Category	Publicly available dataset		
Data Owner	ITAINNOVA		
Provider (if different from partner)	Open Meteo		
Description	Real climatic data observation (T, Humidity, Precipitation, etc.)		
Already before the ADV project?	Yes		
Data purpose in ADV	Train/challenge ML models to better un crop environmental conditions		
Use beyond ADV	Could be helpful in efficient data linking		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Weather station	Air Temperature	°C	15 minutes
Weather station	Air Humidity	%	15 minutes
Weather station	Wind Direction	degrees	15 minutes
Weather station	Wind Speed	m/sec	15 minutes
Weather station	Solar Radiation	W/m^2	15 minutes

Weather station	Rainfall	mm	15 minutes

Other Data	**Measuring**	**Units**	**Frequency**
Earth Observation Data	NDVI and derived data for control parcels	0 – 0.9	Weekly

Storage and access details	
Data format type	Postgres database with REST API
Storage Location	Cloud Server
Storage Duration	Not defined
Is the data public or confidential?	Public but restricted access
Data harmonisation activities required?	Yes
Data merging activities required?	Yes
Can data be used after the ADV project?	Yes, but permission needs to be granted

Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #12.1-12.6: Data use per pilot use case		
Data variable	**UC 1.3**	**UC 3.1**
Pest (adult mots)	Yes	Yes
Pest (%/number of affected leaves)	Yes	Yes
Locate control parcels and obtain satellite images	Yes	Yes
NDVI and derived data for control parcels	Yes	Yes
SARGA Air Temperature	Yes	Yes
SARGA Air Humidity	Yes	Yes
SARGA Wind Direction	Yes	Yes
SARGA Wind Speed	Yes	Yes
SARGA Solar Radiation	Yes	Yes
SARGA Rainfall	Yes	Yes
SARGA NDVI and derived data for control parcels	Yes	Yes
ECMWF Reanalysis Air Temperature	Yes	Yes
ECMWF Reanalysis Air Humidity	Yes	Yes
ECMWF Reanalysis Wind Direction	Yes	Yes
ECMWF Reanalysis Wind Speed	Yes	Yes
ECMWF Reanalysis Solar Radiation	Yes	Yes
ECMWF Reanalysis Rainfall	Yes	Yes
ECMWF Reanalysis NDVI and derived data for control parcels	Yes	Yes
Open Meteo Forecast Air Temperature	Yes	Yes
Open Meteo Forecast Air Humidity	Yes	Yes
Open Meteo Forecast Wind Direction	Yes	Yes
Open Meteo Forecast Wind Speed	Yes	Yes
Open Meteo Forecast Solar Radiation	Yes	Yes
Open Meteo Forecast Rainfall	Yes	Yes
Open Meteo Forecast NDVI and derived data for control parcels	Yes	Yes

2.2.13 Pilot 13: Vineyards – Amfilochia, Greece

Pilot Owner:	TBA
Number of ha	10 ha
Farmers involved Directly:	2
Farmers involved indirectly:	> 12,000
Available Data:	Weather and soil data from vineyards/Wine production

Pilot 13 is located in Amfilochia, Etoloakarnania (Greece) and covers vineyards/wine production pilot. The pilot is owned by 2 produces, however TBA via the mother company (Agrinio Union Cooperative) plans to inform more than 12,000 producers on AgriDataValue results. The pilot is currently equipped with only 1 Agro-weather station, but additional stations will be installed. The current monitoring system can detect air temperature, air humidity, rainfalls and wind. Additional weather stations are planned to be included along with at least 100 RFID tags to measure time and duration of transfer from field to the winery.

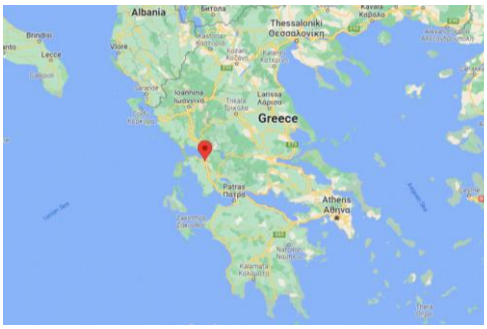


Figure 27: AgriDataValue Pilot 13 Location

Figure 28: Photos from AgriDataValue Pilot 13 Vineyard

Dataset #13			
Overview			
Dataset Name	TBA Vineyards		
Dataset Category	Primary data collected by partner in ADV		
Data Owner	TBA		
Description	The datasets contain air humidity and temperature, weather data and will be expanded with RFID tags to measure additional parameters		
Already before the ADV project?	Yes. Additional data will be captured.		
Data purpose in ADV	Train/challenge ML models for pest detection		
Use beyond ADV	Could be helpful in efficient data linking		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Thermometer	Air Temperature	°C	15 minutes
Air Sensor	Air Humidity	%	15 minutes
Anemometer	Wind Direction	degrees	15 minutes
Anemometer	Wind Speed	m/sec	15 minutes
Rain Gauge	Rainfall	mm	15 minutes
Soil Sensor	Soil Moisture (VWC)	%	15 minutes
Soil Sensor	Electrical Conductivity	mS/m	15 minutes
Soil Sensor	Soil Temperature	°C	15 minutes

Other Data	Measuring	Units	Frequency
Calculated Data	RFID location & time	Location & time	Daily
Field Observation Data	Pest	-	Daily
Storage and access details			
Data format type	CSV, XML, JSON		
Storage Location	Cloud Server		
Storage Duration	Permanent		
Is the data public or confidential?	Confidential		
Data harmonisation activities required?	Yes		
Data merging activities required?	Yes		
Can data be used after the ADV project?	Yes, but permission needs to be granted		
Ethics and Data Protection			
Includes Personal data?			No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?			Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?			Yes
Does dataset contain previously collected personal data?			No

Dataset #13: Data use per pilot use case		
Data variable	**UC 1.1**	**UC 5.2**
Air Temperature	Yes	Yes
Air Humidity	Yes	Yes
Wind Direction	Yes	Yes
Wind Speed	Yes	Yes
Rainfall	Yes	Yes
Soil Temperature	Yes	Yes
Soil Moisture	Yes	Yes
Soil Electrical Conductivity	Yes	Yes
RFID location & time	No	Yes
Pest (Mediterranean fruit fly) control	No	Yes

2.2.14 Pilot 14: Vineyards – Saint-Emilion, France

Pilot Owner:	CVSE
Number of ha	7500 ha
Farmers involved Directly:	900
Farmers involved indirectly:	> 5000
Available Data:	Weather data and multispectral data from vineyards/Wine production

Pilot 14 is located in Saint-Emilion (France) and covers vineyards/wine production pilot. The pilot is currently equipped with a large number of weather stations of which CSVE owns 37. The current weather monitoring system is able to detect a number of weather conditions such as: air temperature, air humidity, wind speed and direction. There are already historical data, but additional data will be collected.

Figure 29: AgriDataValue Pilot 14 Location

Figure 30: Figures from AgriDataValue Pilot 14 Vineyard

Dataset #14.1			
Overview			
Dataset Name	Climate data		
Dataset Category	Primary data collected by partner in ADV		
Data Owner	Winegrowers of CSVE		
Description	The datasets contain data of a large number of weather stations used by winegrowers and 37 weather stations of CVSE (selerys system)		
Already before the ADV project?	Yes. data acquisition in progress		
Data purpose in ADV	Train/challenge ML models to better understand the frost and climate change/ predictive models/warning system for frost management		
Use beyond ADV	More knowledge about the frost phenomenon/improve the prediction systems/ improve active and passive control techniques		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Weather station	Air Temperature	°C	15 minutes
Weather station	Air Humidity	%	15 minutes
Weather station	Wind Direction	degrees	15 minutes
Weather station	Wind Speed	m/sec	15 minutes
Weather station	Solar Radiation	W/m^2	15 minutes
Weather station	Rainfall	mm	15 minutes
Storage and access details			
Data format type	CSV, XLS,		
Storage Location	To be defined		
Storage Duration	To be defined		
Is the data public or confidential?	Public but restricted access		
Data harmonisation activities required?	Yes		

Data merging activities required?	Yes
Can data be used after the ADV project?	To be defined
Ethics and Data Protection	
Includes Personal data?	yes
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #14.2			
Overview			
Dataset Name	Climate data		
Dataset Category	Primary data collected by partner in ADV		
Data Owner	CSVE		
Description	37 weather stations of CVSE (selerys system)		
Already before the ADV project?	Yes. data acquisition in progress		
Data purpose in ADV	Train/challenge ML models to better understand the frost and climate change/ predictive models/warning system for frost management		
Use beyond ADV	More knowledge about the frost phenomenon/improve the prediction systems/ improve active and passive control techniques /		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Weather station	Air Temperature	°C	15 minutes
Weather station	Air Humidity	%	15 minutes
Weather station	Wind Direction	degrees	15 minutes
Weather station	Wind Speed	m/sec	15 minutes
Weather station	Solar Radiation	W/m^2	15 minutes
Weather station	Rainfall	mm	15 minutes
Other Data	**Measuring**	**Units**	**Frequency**
Hedges presence (yes/no)		To be determined	To be determined
Field observations: Hedges description (direction, permeability, species…)		To be determined	To be determined
Field observations on vine phenology (vine sensitivity)		To be determined	To be determined
Storage and access details			
Data format type	CSV, XLS,		
Storage Location	To be defined		
Storage Duration	To be defined		
Is the data public or confidential?	Public but restricted access		
Data harmonisation activities required?	Yes		
Data merging activities required?	Yes		
Can data be used after the ADV project?	To be defined		
Ethics and Data Protection			
Includes Personal data?	No		

Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #14.3			
Overview			
Dataset Name	Thermal maps		
Dataset Category	Secondary data		
Data Owner	CSVE		
Description	Thermal map of Saint-Emilion area in 2022 (via aircraft with thermal camera). Thermal maps of the whole area will be performed in the future during an episode of frost without protection systems functioning		
Already before the ADV project?	Yes. data acquisition in progress		
Data purpose in ADV	Identify cold area to install temperature sensors and humidity sensors		
Use beyond ADV	Understand cold air flows / study factors that lead to frost/improve active and passive control techniques (ie. Wind machine)		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Thermic camera	Temperature	°C	map
Storage and access details			
Data format type	jpeg		
Storage Location	local		
Storage Duration	To be defined		
Is the data public or confidential?	Public with restricted access		
Data harmonisation activities required?	Yes		
Data merging activities required?	Yes		
Can data be used after the ADV project?	To be defined		
Ethics and Data Protection			
Includes Personal data?			Yes
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?			Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?			Yes
Does dataset contain previously collected personal data?			No

Dataset #14.4	
Overview	
Dataset Name	Climate data
Dataset Category	Secondary data
Data Owner	CSVE
Description	Synfield devices have been installed to complete the existing network: 2 weather stations, 4 leaf wetness sensor , 4 soil moisture sensors and 1 pyranometer
Already before the ADV project?	No. in ADV project
Data purpose in ADV	Train/challenge ML models to better understand the frost and climate change/ predictive models/warning system for frost management

Use beyond ADV	More knowledge about the frost phenomenon/improve the prediction systems/ improve active and passive control techniques / study the impact of cultural practices on the frost risk		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Weather station	Air Temperature	°C	15 minutes
Weather station	Air Humidity	%	15 minutes
Weather station	Wind Direction	degrees	15 minutes
Weather station	Wind Speed	m/sec	15 minutes
Weather station	Solar Radiation	W/m^2	15 minutes
Weather station	Rainfall	mm	15 minutes
Leaf wetness sensors	Leaf wetness	mm	15 minutes
Soil moisture sensors	Soil temperature	°C	15 minutes
Soil moisture sensors	Water content	mm	15 minutes

Storage and access details	
Data format type	CSV, XLS,
Storage Location	To be defined
Storage Duration	To be defined
Is the data public or confidential?	Public but restricted access
Data harmonisation activities required?	Yes
Data merging activities required?	Yes
Can data be used after the ADV project?	To be defined

Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #14.5			
Overview			
Dataset Name	Topography		
Dataset Category	Secondary data		
Data Owner	CSVE		
Description	map that represents the slope expressed in degree in the Saint-Emilion region. Topography is a factor that impacts frost Technical partner worked on the data: SIXENS		
Already before the ADV project?	No. in ADV project		
Data purpose in ADV	Train/challenge ML models to better understand the frost and climate change/ predictive models/warning system for frost management		
Use beyond ADV	More knowledge about the frost phenomenon/improve the prediction systems/ improve active and passive control techniques /		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Data of the National Institute of Geographic and Forest Information processed using the slope algorithm of Google Earth Engine	slope	degrees	Determined once

Data of the National Institute of Geographic and Forest Information processed using the slope algorithm of Google Earth Engine	Elevation/altitude	meter	Determined once
Data of the National Institute of Geographic and Forest Information processed using the slope algorithm of Google Earth Engine	Pedology	morphology	Determined once

Storage and access details	
Data format type	.shp, .tif
Storage Location	To be defined
Storage Duration	To be defined
Is the data public or confidential?	Public but restricted access
Data harmonisation activities required?	Yes
Data merging activities required?	Yes
Can data be used after the ADV project?	To be defined

Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #14.6			
Overview			
Dataset Name	Land use		
Dataset Category	Secondary data		
Data Owner	CSVE		
Description	Map showing the land use in Saint-Emilion region: vineyards (low/high density), forest, vegetation, etc… Technical partner worked on the data: SIXENS		
Already before the ADV project?	No. in ADV project		
Data purpose in ADV	Train/challenge ML models to better understand the frost and climate change/ predictive models/warning system for frost management		
Use beyond ADV	More knowledge about the frost phenomenon/improve the prediction systems/ improve active and passive control techniques /		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Satellite data (Sentinel)-10m	Land use		

Storage and access details	
Data format type	.shp, .tif
Storage Location	To be defined
Storage Duration	To be defined
Is the data public or confidential?	Public but restricted access

Data harmonisation activities required?	Yes
Data merging activities required?	Yes
Can data be used after the ADV project?	To be defined
Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #14: Data use per pilot use case	
Data variable	**UC 5.2**
Winegrowers Air Temperature	Yes
Winegrowers Air Humidity	Yes
Winegrowers Wind Direction	Yes
Winegrowers Wind Speed	Yes
Winegrowers Solar Radiation	Yes
Winegrowers Rainfall	Yes
CVSE Air Temperature	Yes
CVSE Air Humidity	Yes
CVSE Wind Direction	Yes
CVSE Wind Speed	Yes
CVSE Solar Radiation	Yes
CVSE Rainfall	Yes
Thermal camera temperature	Yes
Synfield Air Temperature	Yes
Synfield Air Humidity	Yes
Synfield Wind Direction	Yes
Synfield Wind Speed	Yes
Synfield Solar Radiation	Yes
Synfield Rainfall	Yes
Synfield Leaf wetness	Yes
Synfield Soil temperature	Yes
Synfield Water content	Yes
Google Earth slope	Yes
Google Earth Elevation/altitude	Yes
Google Earth Pedology	Yes
Sentinel land use	Yes
Hedges presence (yes/no)	Yes
Field observations: Hedges description (direction, permeability, species…)	Yes
Field observations on vine phenology (vine sensitivity)	Yes

2.2.15 Pilot 15: Vineyards - Emilia-Romagna, Italy

Pilot Owner:	RI.NOVA
Number of ha	7 ha
Farmers involved Directly:	20
Farmers involved indirectly:	> 500
Available Data:	Weather and soil data from vineyards/Wine production

Pilot 15 is situated in Tebano, Emilia-Romagna (Italy) and covers vineyards/wine production pilot. The pilot is currently equipped with 1 Agro-weather station and 1 soil station. The current monitoring system is able to detect: air temperature, air humidity, evapotranspiration potential, rainfalls, solar radiation, wind, leaf wetness, soil temperature, soil humidity, Volumeter water potential (20 cm and 40 cm), electrical conductivity and irrigation. Additional weather stations are planned to be included.

Figure 31: AgriDataValue Pilot 15 Location

Figure 32: Figures from AgriDataValue Pilot 15 Vineyard

Dataset #15			
Overview			
Dataset Name	RI.NOVA Vineyards		
Dataset Category	Primary data collected by partner in ADV		
Data Owner	RI.NOVA		
Description	The datasets contain soil moisture and temperature, leaf wetness, air moisture and temperature, weather data and prediction		
Already before the ADV project?	Yes. Additional data will be captured.		
Data purpose in ADV	Train/challenge ML models to improve disease prediction.		
Use beyond ADV	Could be helpful in efficient data linking		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Thermometer	Air Temperature	°C	15 minutes
Air Sensor	Air Humidity	%	15 minutes
Anemometer	Wind Direction	degrees	15 minutes
Anemometer	Wind Speed	m/sec	15 minutes
Leaf Sensor	Leaf Wetness	%	15 minutes
Pyranometer	Solar Radiation	W/m^2	15 minutes
Rain Gauge	Rainfall	mm	15 minutes

Soil Sensor	Soil Moisture (VWC)	%	15 minutes
Soil Sensor	Electrical Conductivity	mS/m	15 minutes
Soil Sensor	Soil Temperature	°C	15 minutes
Soil Sensor	Water potential at 20cm and 40 cm	%	15 minutes
Other Data	**Measuring**	**Units**	**Frequency**
Calculated Data	Evapotranspiration	mm	Daily
Predicted Data	Air Humidity	%	Daily
Field Observation Data	Pest, Phenology	-	Daily

Storage and access details	
Data format type	CSV, XML, JSON
Storage Location	Local and Cloud Server
Storage Duration	Permanent
Is the data public or confidential?	Confidential
Data harmonisation activities required?	Yes
Data merging activities required?	Yes
Can data be used after the ADV project?	Yes, but permission needs to be granted

Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #15: Data use per pilot use case	
Data variable	**UC 3.2**
Air Temperature	Yes
Air Humidity	Yes
Wind Direction	Yes
Wind Speed	Yes
Leaf Wetness	Yes
Solar Radiation	Yes
Rainfall	Yes
Soil Moisture (VWC)	Yes
Electrical Conductivity	Yes
Soil Temperature	Yes
Water potential at 20cm and 40 cm	Yes
Evapotranspiration	Yes
Air Humidity	Yes
Field observation dataPest, Phenology	Yes

2.2.16 Pilot 16: Olive Grove – Messinia (Greece)

Pilot Owner:	NILEAS
Number of ha	300 ha
Farmers involved Directly:	20

Farmers involved indirectly:	> 800
Available Data:	Weather and soil data from olive production

Pilot 16 is located in Chora, Messinia, in the Peloponnese Region in Greece and covers an Olive grove. The pilot already features 35 soil and air sensors that monitor soil moisture and temperature, as well as air moisture and temperature. Additionally, it includes a weather station to monitor rainfall gauge, temperature, humidity, wind speed and direction, soil (moisture, and temperature) to aid in weather predictions.

Recently, 6 new air and soil sensors have been installed, utilizing SYN's SynField devices, which consist of 2 SynField X5 Head Nodes and 4 SynOdos peripheral devices. Furthermore, 2 additional weather stations have been added to enhance data collection, including irrigation data.

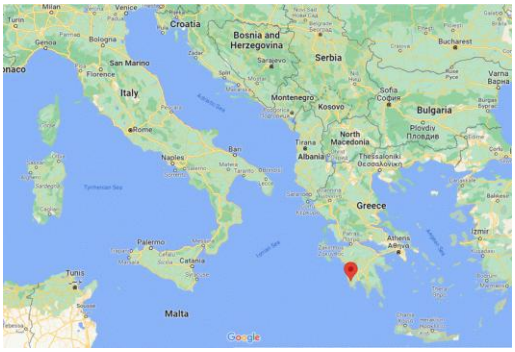


Figure 33: AgriDataValue Pilot 16 Location

Figure 34: Photos from AgriDataValue Pilot 16 Olive Groves

Dataset #16			
Overview			
Dataset Name	NILEAS Olive Grove		
Dataset Category	Primary data collected by partner in ADV		
Data Owner	NILEAS		
Description	The datasets contain soil moisture and temperature, air moisture and temperature, weather data and prediction		
Already before the ADV project?	Yes. Additional data will be captured.		
Data purpose in ADV	Train/challenge ML models to detect pest and Olive fly prediction.		
Use beyond ADV	Could be helpful in efficient data linking		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Thermometer	Air Temperature	°C	15 minutes
Air Sensor	Air Humidity	%	15 minutes
Anemometer	Wind Direction	degrees	15 minutes
Anemometer	Wind Speed	m/sec	15 minutes
Rain Gauge	Rainfall	mm	15 minutes
Soil Sensor	Soil Moisture (VWC)	%	15 minutes
Soil Sensor	Electrical Conductivity	mS/m	15 minutes

Soil Sensor	Soil Temperature	°C	15 minutes
Other Data	**Measuring**	**Units**	**Frequency**
Calculated Data	Evapotranspiration	mm	Daily
Predicted Data	Air Humidity	%	Daily
Field Observation Data	Pest, Phenology	-	Daily
Field Observation Data	Olive Fly Traps	number of olive flies	Manual input Weekly from June until the end of harvest
Field Observation Data	Systematic monitoring for Olive anthracnose (Colletotrichum gloeosporioides): samples of olive fruit and/or olive leaves and branches for analysis	**Scale of presence**: a) low: monitor, b) high: intervene with plant protection methods (IPM)	Manual input during the critical stages: I) 2 times during April (beginning of flowering) and II) Every 10 days from the first autumn rainfalls until the end of harvest

Storage and access details	
Data format type	CSV, XML, JSON
Storage Location	Remote Server (Azzure Platform and Digital Ocean cloud)
Storage Duration	At least 5 years
Is the data public or confidential?	Confidential
Data harmonisation activities required?	Yes
Data merging activities required?	Yes
Can data be used after the ADV project?	Data are anonymised and can be used to challenge ML models

Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #16: Data use per pilot use case			
Data variable	**UC 3.1**	**UC 3.2**	**UC 3.4**
Air Temperature	Yes	Yes	Yes
Air Humidity	Yes	Yes	Yes
Wind Direction	No	Yes	No
Wind Speed	No	Yes	No
Rainfall	Yes	No	Yes
Soil Moisture (VWC)	Yes	Yes	Yes
Electrical Conductivity	Yes	Yes	Yes
Soil Temperature	Yes	Yes	Yes
Evapotranspiration	Yes	No	Yes
Pest, Phenology	Yes	Yes	Yes

Olive fly traps	No	No	Yes
Olive anthracnose (Colletotrichum gloeosporioides): when necessary, send samples of olive fruit and/or olive leaves and branches for analysis	Yes	No	No

2.2.17 Pilot 17: Olive Grove – Emilia-Romagna (Italy)

Pilot Owner:	RI.NOVA
Number of ha	3.64 ha
Farmers involved Directly:	100
Farmers involved indirectly:	> 1500
Available Data:	Olives fly detection

Pilot 17 is an olive grove comprising approximately 2,000 olive trees, located in the municipality of Roncofreddo, in the province of Forlì-Cesena, Emilia-Romagna region (Northeastern Italy). The pilot includes a weather station and a soil sensor installed as part of the ADV project, provided by Synelixis. There is no historical meteorological data available, but historical data related to olive fly monitoring and is present. Additionally, soil analysis data is available, providing detailed information about the grove's pedological characteristics.

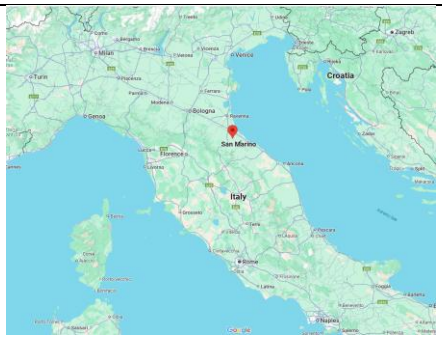


Figure 35: AgriDataValue Pilot 17 Location

Figure 36: Figures from AgriDataValue Pilot 17 Olive Grove

Dataset #17	
Overview	
Dataset Name	RI.NOVA Olive Grove
Dataset Category	Primary data collected by partner in ADV
Data Owner	RI.NOVA
Description	The datasets contain soil moisture and temperature, air moisture and temperature, weather data and prediction. Olive flies monitoring

Already before the ADV project?	Yes, olive fly monitoring data existed before ADV. Additional real-time and predictive data are being collected during ADV.
Data purpose in ADV	Train/challenge ML models to improve disease prediction and Pest control on Olive Fly and improve data-driven decision-making for pest and disease management
Use beyond ADV	Could be helpful in efficient data linking

Measuring technologies	**Measuring**	**Units**	**Frequency**
Thermometer	Air Temperature	°C	15 minutes
Air Sensor	Air Humidity	%	15 minutes
Anemometer	Wind Direction	degrees	15 minutes
Anemometer	Wind Speed	m/sec	15 minutes
Pyranometer	Solar Radiation	W/m^2	15 minutes
Rain Gauge	Rainfall	mm	15 minutes
Soil Sensor	Soil Moisture (VWC)	%	15 minutes
Soil Sensor	Electrical Conductivity	mS/m	15 minutes
Soil Sensor	Soil Temperature	°C	15 minutes
Other Data	**Measuring**	**Units**	**Frequency**
Calculated Data	Evapotranspiration	mm	Daily
Predicted Data	Air Humidity	%	Daily
Field Observation Data	Pest, Phenology	-	Daily

Storage and access details	
Data format type	CSV, XML, JSON
Storage Location	Local and Cloud Server
Storage Duration	Permanent
Is the data public or confidential?	Confidential
Data harmonisation activities required?	Yes
Data merging activities required?	Yes
Can data be used after the ADV project?	Yes, but permission needs to be granted

Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #17: Data use per pilot use case		
Data variable	**UC 3.1**	**UC 3.4**
Air Temperature	Yes	Yes
Air Humidity	Yes	Yes
Wind Direction	Yes	Yes
Wind Speed	Yes	Yes
Solar Radiation	Yes	Yes
Rainfall	Yes	Yes
Soil Moisture (VWC)	Yes	Yes
Electrical Conductivity	Yes	Yes

Soil Temperature	Yes	Yes
Evapotranspiration	Yes	Yes
Air Humidity	Yes	Yes
Field Observation Data Pest, Phenology	Yes	Yes

2.2.18 Pilot 18: Bio cereals - Romania

Pilot Owner:	BIORO
Number of Animals	30 ha
Farmers involved Directly:	1
Farmers involved indirectly:	> 10000
Available Data:	Soil and weather data

Pilot 18 is situated in Teremia Mare, Jud Timis (Romania) and covers cereals. Two parcels of approximately 15 ha each. One of the parcels will be used in conventional agriculture in a system of classical works, and the other will include the two experimental methods depending on the crop rotation, also in the system of conventional agriculture.

Both plots are surrounded on four sides by irrigation/drainage channels. Development of agricultural procedures and practices for obtaining sustainable and healthy production generated by conventional agriculture in the Minimum tillage system and strip-till system. An analysis will be made between the classic tillage system and the two new systems, respectively minimum tillage, and strip-till in conventional agriculture system. 6 sensors for each of the parcels will gather data: soil humidity sensors, soil temperature sensors, plant air humidity sensors, air temperature sensors, precipitation sensors, nitrogen sensors.

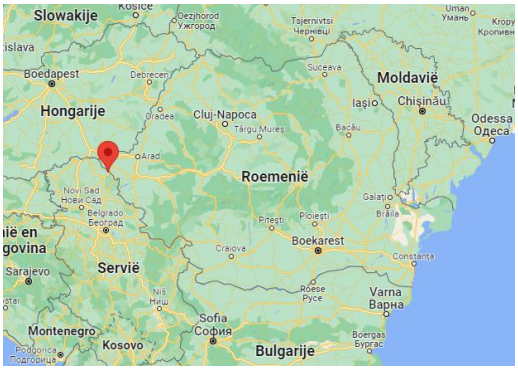


Figure 37: AgriDataValue Pilot 18 Location

Figure 38: Photos from AgriDataValue Pilot 18

Dataset #18	
Overview	
Dataset Name	BIORO/CEREALS
Dataset Category	Primary data collected by partner in ADV
Data Owner	BIORO

Description	The datasets contain the output data of IoT sensors (6 sensors for each of the parcels: Soil humidity sensor, soil temperature sensor, plant air humidity sensor, air temperature sensor, precipitation sensor, nitrogen sensor).
Already before the ADV project?	No
Data purpose in ADV	Train/challenge ML models to detect pest and cereals prediction.
Use beyond ADV	To enhance use of this kind of data

Measuring technologies	**Measuring**	**Units**	**Frequency**
Air Sensor	Air Temperature	°C	Continuously
Air Sensor	Air Humidity	**%**	Continuously
Nitrogen sensor	Nitrogen	kg	Continuously
Precipitation sensor	Rain	mm	Continuously
Soil Sensor	Soil humidity	%	Continuously
Soil sensor	Soil temperature	°C	Continuously

Storage and access details	
Data format type	CSV
Storage Location	Database
Storage Duration	At least 6 years
Is the data public or confidential?	Public but restricted access
Data harmonisation activities required?	Yes
Data merging activities required?	Yes
Can data be used after the ADV project?	Yes, but permission needs to be granted

Ethics and Data Protection	
Includes Personal data?	could
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #18: Data use per pilot use case	
Data variable	**UC 1.1**
Air Temperature	Yes
Air Humidity	Yes
Nitrogen	Yes
Rain	Yes
Soil humidity	Yes
Soil temperature	Yes

2.2.19 Pilot 19: Dairy Barn – Flanders, Belgium

Pilot Owner:	EV ILVO
Number of Animals	164 Holstein Friesian cows (144 lactating and 20 dry cows)
Farmers involved Directly:	1
Farmers involved indirectly:	> 200
Available Data:	Feed, production, health and emission data

Pilot 19 is situated in Melle, Flanders (Belgium) near Ghent and covers a Dairy barn at the ILVO Animal Science Unit research farm. The involved partner is ILVO. The created data are complete barn data typical for a state-of-the-art farm in Northwestern Europe equipped with individual cow recognition. The decision to utilise this farm is based on the fact that the specific farm, next to normal dairy production data, also generates additional data related to monitoring of the greenhouse gas (GHG) and ammonia emissions. Moreover, in this barn there are extra sensors that can be used for behaviour and health monitoring. The production data are supplied feed data, milk data (production, fat, protein, lactose, urea), weight, general cow data etc.

Figure 39: AgriDataValue Pilot 19 Location

Figure 40: Figures from AgriDataValue Pilot 19 Dairy Barn

Dataset #19			
Overview			
Dataset Name	ILVO Dairy barn		
Dataset Category	Primary data collected by partner in ADV		
Data Owner	ILVO		
Description	The datasets contain the output data of the daily functioning of a state-of-the-art dairy barn, with extra feed intake and methane parameters		
Already before the ADV project?	Yes. Additional data will be captured during the project.		
Data purpose in ADV	Train and challenge ML models for more efficient milk production with less emissions and deposition, and better animal health		
Use beyond ADV	Could be helpful in efficient data-linking		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Milk robot/parlour	Milk Production	kg/d or kg/year	2-3 times a day
Feed Load cells/volumetric sensors	Feed uptake (roughage/concentrate)	kg/d	Each visit
Greenfeed and NDIR/TDLS Gas Sensors	CH4 (Methane) emission	g/d or kg/y	1-6 times a day
NDIR/TDLS Gas Sensors	NH3 (Ammonia) in barn	ppm	1 second
GreenFeed NDIR/TDLS Gas Sensors	CO2 emission	g/d	1-6 times a day
Animal body condition score	Health monitoring	score (1-4)	daily

Animal Collar/RFID ear tag	Animal identification	presence	continuous
Load cells	Animal weight	kg	Twice a day
Temperature driven ventilation	Climate control	rpm	Continuous
Rumination collar	Rumination	Min/day	Daily if available
Synair sensor	Air humidity	%	Once/20 min
Synair sensor	Air temperature	°C	Once/20 min
Synair sensor	Ammonia concentration	ppm	Once/20 min
Synair sensor	CO2 concentration	ppm	Once/20 min
Synair sensor	Fine particulate matter PM1	µg/m³	Once/20 min
Synair sensor	Fine particulate matter PM2.5	µg/m³	Once/20 min
Synair sensor	Fine particulate matter PM4	µg/m³	Once/20 min
Synair sensor	Fine particulate matter PM10	µg/m³	Once/20 min
Other Data	**Measuring**	**Units**	**Frequency**
Data from milk analysis	Milk fat	mg/100ml	Daily
Data from milk analysis	Milk Lactose	mg/100ml	Daily
Data from milk analysis	Milk Protein	mg/100ml	Daily
Data from milk analysis	Milk Urea	mg/100ml	Daily
Data from feed analysis	Nutritional values	mg/kg	Periodically
Birth date	Age	Days/years	Manual input
Lactation number	Calving	Number	Manual input
Lactation stage	Days in milk	Days	Manual input
Gestation stage	Days in gestation	Days	Manual input
Health observations	Health status	Event	Manual input
Lab feed analysis	Dry matter content	mg/kg	periodically
Lab feed analysis	Crude protein content	mg/kg	periodically
Lab feed analysis	Fat content	mg/kg	periodically
Lab feed analysis	Crude ash content	mg/kg	periodically
Lab feed analysis	NDF content	mg/kg	periodically
Lab feed analysis	Organic digestibility	mg/kg	periodically
Storage and access details			
Data format type	CSV, XLS, JSON		
Storage Location	Local Server		
Storage Duration	Permanent		
Is the data public or confidential?	Confidential		
Data harmonisation activities required?	Yes		
Data merging activities required?	Yes		
Can data be used after the ADV project?	Data can still be used if they are anonymised, to challenge ML models; The data are kept on internal servers		
Ethics and Data Protection			
Includes Personal data?	Only Indirectly.		
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes. All data are needed, or it is allowed to be used. For each specific demand/sharing, permission should be approved.		
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes. All data to be used will be anonymized or pseudonymised or synthetic data		

Does dataset contain previously collected personal data?	Pilot DPO will check data when there is doubt.

Dataset 19: Data use per pilot use case				
Data variable	**UC 4.1**	**UC 4.2**	**UC 4.3**	**UC 4.4**
Milk Production	Yes	Yes	Yes	No
Feed uptake (roughage/concentrate)	Yes	Yes	Yes	Yes
Dry matter content	Yes	Yes	Yes	Yes
Crude protein content	Yes	Yes	Yes	Yes
Fat content	Yes	Yes	Yes	Yes
Crude ash content	Yes	Yes	Yes	Yes
NDF content	Yes	Yes	Yes	Yes
Organic digestibility	Yes	Yes	Yes	Yes
CH4 (Methane) emission	Yes	No	No	No
NH3 (Ammonia) in barn	No	Yes	No	No
CO2 emission	Yes	No	No	No
Health monitoring	No	No	Yes	Yes
Animal identification	Yes	Yes	Yes	Yes
Animal weight	Yes	Yes	Yes	Yes
Climate control	Yes	Yes	No	No
Milk fat	Yes	No	Yes	No
Milk Lactose	Yes	No	Yes	No
Milk Protein	Yes	Yes	Yes	No
Milk Urea	No	Yes	Yes	No
Rumination data	No	No	Yes	Yes
Animal age	Yes	Yes	Yes	Yes
Lactation number	Yes	Yes	Yes	Yes
Lactation stage	Yes	Yes	Yes	Yes
Gestation stage	No	No	Yes	Yes
Health observations	Yes	Yes	Yes	Yes

2.2.20 Pilot 20: Dairy barn - Vecauce, Latvia

Pilot Owner:	ZSA
Number of Animals	1000 cows
Farmers involved Directly:	1
Farmers involved indirectly:	600- 800
Available Data:	Farm statistical data

Pilot 20 is situated in Vecauce (Latvia), and covers a Dairy barn at the LBTU teaching and research farm with ZSA as a project partner. Vecauce is a multidisciplinary farm that combines student training, research, crop farming, dairy farming, biogas production, fruit growing, forestry and has created one of the largest and most productive herds of dairy cattle in Latvia. The barn houses more than 1000 cows. In this dairy barn, various research trials are performed on topics such as feed components and feed efficiency, enteric emissions, colostrum quality, youngstock management etc. The involved partner is ZSA. The Farm "Bruzilas", Kurzeme region. ADV Involved partner ZSA is the biggest agricultural organisation of producers in Latvia. The ZSA plans to inform more than 600-800 producers on ADV results. Pilot 20 is situated in Vecauce (Latvia), and covers a Dairy barn at the LBTU teaching and research farm with ZSA as a project partner. Vecauce is a multidisciplinary farm that combines student training, research, crop farming, dairy farming, biogas production, fruit growing, forestry and has created one of the largest and most productive herds of dairy cattle in Latvia. The barn houses more than 1000 cows. In this dairy barn, various research trials are performed on topics such as feed components and feed efficiency, enteric emissions, colostrum quality, youngstock management etc.

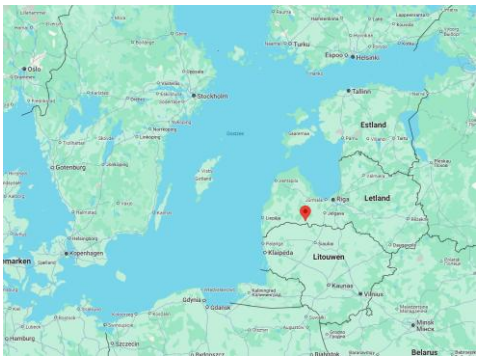


Figure 41: AgriDataValue Pilot 20 Location

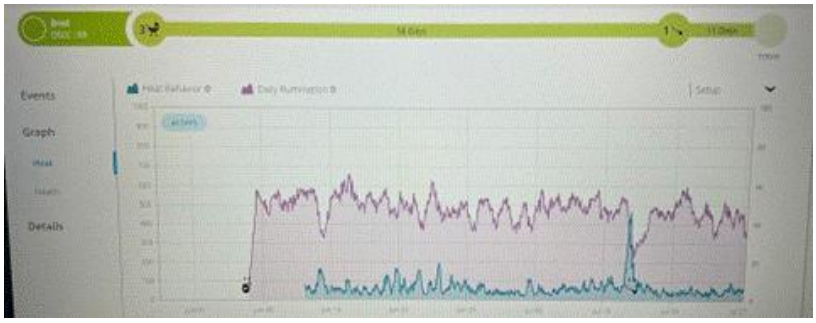


Figure 42: Figures of AgriDataValue Pilot 20

Dataset #20	
Overview	
Dataset Name	Primary data collected by partner ZSA in ADV
Dataset Category	Primary data collected by the Farm
Data Owner	Farm
Description	The datasets contain the output data of the daily functioning of a dairy barn. The data will help to identify the situation on the farm and the calving monitor follows the farmer could take decisions of the help necessity.
Already before the ADV project?	Yes. Additional data will be captured.

Data purpose in ADV	Train/challenge ML models to detect problem situations for calving		
Use beyond ADV	The data can be used to improve existing technologies and models on livestock farms		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Activity/tracking sensor AXI sensors from Medria company	Calving activity	Number of moves	15 minutes
Commercial farm equipment	Livestock data (ID number, pregnancy report)	varying units	Continuous
Load cells	Animal weight	kg	After each milking
Other Data	**Measuring**	**Units**	**Frequency**
Statistical data	Calving data	-	-
Data from Milk analysis	Milk fat	%	Daily
Data from Milk analysis	Milk Lactose	mg/100ml	Daily
Data from Milk analysis	Milk Protein	%	Daily
Data from Milk analysis	Milk Urea	mg/100ml	Daily
Data from Milk analysis	Milk yield	ml	After each milking
Health observations	Health status	-	Manual input
Lactation number	Lactations	-	Manual input
Lactation stage	Days in milk	Days	Manual input
Storage and access details			
Data format type	CSV, XLS, API		
Storage Location	Local Server		
Storage Duration	Permanent		
Is the data public or confidential?	Confidential		
Data harmonisation activities required?	Yes		
Data merging activities required?	Yes		
Can data be used after the ADV project?	Yes, but permission needs to be granted		
Ethics and Data Protection			
Includes Personal data?	Yes, we will use data from actual livestock farm that are owned by farmers. Indirectly, can be anonymized		
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Data minimization is not conducted yet.		
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes		
Does dataset contain previously collected personal data?	No		

Dataset 20: Data use per pilot use case		
Data variable	**UC 4.3**	**UC 4.4**
Milk Production	Yes	No
Health observation	Yes	Yes
Animal identification	Yes	Yes
Animal weight	Yes	Yes
Milk fat	Yes	No

Milk Lactose	Yes	No
Milk Protein	Yes	No
Milk Urea	Yes	No
Animal age	Yes	Yes
Lactation number	Yes	Yes
Lactation stage	Yes	Yes

2.2.21 Pilot 21: Organic Cattle Farm – Katouna, Etoloakarnania, Greece

Pilot Owner:	TBA
Number of Animals	320 Limousine cows
Farmers involved Directly:	10
Farmers involved indirectly:	> 12,000
Available Data:	Air quality data for cattle monitoring

Pilot 21 is located in Katouna, Etoloakarnania (Greece) near Agrinio and covers an Organic Cattle Farm (TBA means the Biologiko Agroktima or Organic Farm in Greek). The involved partner is TBA. Within this pilot, data are farm data of a cattle fattening system. All involved cattle belong to the Limousine race, a French breed of beef cattle from the Limousin and Marche regions of France. It was formerly used mainly as a draught animal, but in modern times is reared for beef. The TBA has breeds more than 900 cattle animals, but around 320 are involved. The production data are the number of cattle, the air quality and the collars in selected animals.

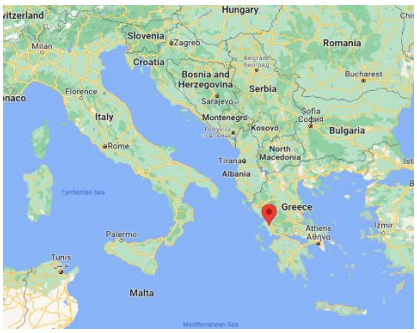


Figure 43: Pilot 21 Location

Figure 44: Figures from AgriDataValue Pilot 21 Beef cattle farm

Dataset #21	
Overview	
Dataset Name	TBA Cattle Air quality Data
Dataset Category	Primary data collected by partner in ADV
Data Owner	TBA
Description	The datasets contain the output data of the daily functioning of a state-of-the-art cattle barn, with air scrubber
Already before the ADV project?	No. Data will be captured starting at Q3 2023.
Data purpose in ADV	Train and challenge ML models

Use beyond ADV	Could be helpful in efficient data linking		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Air Quality Sensor SynAir	NH3 (Ammonia) emission	ppm	15 minutes
Air Quality Sensor SynAir	OC2	ppm	15 minutes
Air Quality Sensor SynAir	Total volatile organic compounds (VOC)	µg/m³	15 minutes
Air Quality Sensor SynAir	Particulate Matter ($PM_{1.0}$, $PM_{2.5}$, PM_4, PM_{10})	µg/m³	15 minutes
Thermostats	Temperature	°C	15 minutes
Other Data	**Measuring**	**Units**	**Frequency**
Load Cells/volumetric	Feed uptake	kg	Daily
Lab Manure analysis	Volume	kg	Daily
Lab Manure analysis	Organic Matter (Estimated)	kg/1000kg	Weekly
Lab Manure analysis	Total nitrogen (N) (Estimated)	kg/1000kg	Weekly
Lab Manure analysis	Mineral NH4 & NO3 (Estimated)	kg/1000kg	Weekly
Lab Manure analysis	Phosphorus (P2O5) (Estimated)	kg/1000kg	Weekly

Storage and access details	
Data format type	CSV, XML, JSON
Storage Location	Cloud Server
Storage Duration	Permanent
Is the data public or confidential?	Confidential
Data harmonisation activities required?	Yes
Data merging activities required?	Yes
Can data be used after the ADV project?	Data can still be used if they are anonymised, to challenge ML models; The data are kept on internal servers

Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #21: Data use per pilot use case		
Data variable	**UC 4.3**	**UC 5.1**
NH3 (Ammonia) in barn	Yes	No
CO2 emission	Yes	No
Total volatile organic compounds (VOC)	Yes	No
Particulate Matter ($PM_{1.0}$, $PM_{2.5}$, PM_4, PM_{10})	Yes	No
Temperature	Yes	No
Feed uptake	Yes	Yes
Manure Volume	Yes	Yes
Manure Organic Matter (Estimated)	Yes	Yes
Manure Total nitrogen (N) (Estimated)	Yes	Yes
Manure Mineral NH4 & NO3 (Estimated)	Yes	Yes
Manure Phosphorus (P2O5) (Estimated)	Yes	Yes

2.2.22 Pilot 22: Pigs Campus – Flanders, Belgium

Pilot Owner:	EV ILVO
Number of Animals	1450 pigs (106 sows, 576 piglets and 768 fattening pigs)
Farmers involved Directly:	1
Farmers involved indirectly:	> 1.000
Available Data:	Pigs feed, production, health and emission data

Pilot 22 is situated in Melle, Flanders (Belgium) near Ghent and covers a Pigs Campus. The involved partner is ILVO. The data are complete farm data of a closed pig rearing and fattening system including sows, piglets and fattening pigs. Like most newly built pig barns in this region the ILVO barn is equipped with air scrubbers and low emission floor types that provide data on emissions. The production data are the commonly used pig farm indicators such as the number of piglets per litter per sow, the survival rate, temperature in the different barn compartments and ventilation data. Abovementioned data are mostly gathered using commercially available farm equipment including feeders, ventilation systems and farm software systems as well as air scrubbers. The barn has sensors for monitoring animals such as eating pattern, drinking pattern, water consumption, weight, lameness detection, behaviour, location in the compartments. Sensors monitor different types of nutrition with feeding stations, weather conditions (heat stress), management, stable environment, welfare and, behaviour, physiology and morphology of pigs.

Figure 45: AgriDataValue Pilot 22 Location

Figure 46: Figures from AgriDataValue Pilot 22 Pigs Campus

Dataset #22			
Overview			
Dataset Name	ILVO Pig Campus		
Dataset Category	Primary data collected by partner in ADV		
Data Owner	ILVO		
Description	The datasets contain the output data of the daily functioning of a state of the art pig barn, from insemination till finishing phase, with air scrubber		
Already before the ADV project?	No		
Data purpose in ADV	Train and challenge ML models for more efficient pig production with less emissions and deposition, and a better health		
Use beyond ADV	Could be helpful in efficient data linking		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Air scrubber	NH3 concentration	ppm	15 minutes
Ventilation sensors	Air flow control	m^3/h	1-5 minutes

Thermostat steering	Temperature control per compartment	°C	1 minute
Volumetric	Feed uptake	kg	Daily
FTIR/ basic sensors	NH3 (Ammonia) emission	g/d	1 minute
Temperature/humidity sensors	Temperature/humidity	°C/%	1 minute
Weighing scale	Animal body weight	kg	periodically
Synair sensor	Air humidity	%	Once/20 min
Synair sensor	Air temperature	°C	Once/20 min
Synair sensor	Ammonia concentration	ppm	Once/20 min
Synair sensor	CO2 concentration	ppm	Once/20 min
Synair sensor	Fine particulate matter PM1	µg/m³	Once/20 min
Synair sensor	Fine particulate matter PM2.5	µg/m³	Once/20 min
Synair sensor	Fine particulate matter PM4	µg/m³	Once/20 min
Synair sensor	Fine particulate matter PM10	µg/m³	Once/20 min
Other Data	**Measuring**	**Units**	**Frequency**
Lab Manure analysis	Dry matter	kg/1000kg	Monthly
Lab Manure analysis	Organic Matter	kg/1000kg	Monthly
Lab Manure analysis	Total nitrogen (N)	kg/1000kg	Monthly
Lab Manure analysis	Mineral Nitrogen (NH4 & NO3)	kg/1000kg	Monthly
Lab Manure analysis	Phosphorus (P2O5)	kg/1000kg	Monthly
Lab Manure analysis	Potassium (K2O)	kg/1000kg	Monthly
Lab Manure analysis	Magnesium (MgO)	kg/1000kg	Monthly
Lab Manure analysis	Calcium (CaO)	kg/1000kg	Monthly
Lab Manure analysis	Sodium (Na2O)	kg/1000kg	Monthly
Lab feed analysis	Dry matter content	mg/kg	periodically
Lab feed analysis	Crude protein content	mg/kg	periodically
Lab feed analysis	Fat content	mg/kg	periodically
Lab feed analysis	Crude ash content	mg/kg	periodically
Lab feed analysis	NDF content	mg/kg	periodically
Lab feed analysis	Organic digestibility	mg/kg	periodically
Faecal consistency score	Manual observations	scale	periodically
Sow and litter data	Manual observations	diverse	Daily

Storage and access details	
Data format type	CSV, XLS, JSON
Storage Location	Local Server
Storage Duration	Permanent
Is the data public or confidential?	Confidential
Data harmonisation activities required?	Yes
Data merging activities required?	Yes
Can data be used after the ADV project?	Data can still be used if they are anonymised, to challenge ML models; The data are kept on internal servers

Ethics and Data Protection	
Includes Personal data?	Only Indirectly. Can be anonymised
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes. All data are needed, or it is allowed to be used. For each specific demand/sharing, permission should be approved.

Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes. All data to be used will be anonymized or pseudonymised or synthetic data
Does dataset contain previously collected personal data?	Pilot DPO will check data when there is doubt

Dataset #22: Data use per pilot use case			
Data variable	**UC 4.1**	**UC 4.2**	**UC 4.3**
NH3 concentration	No	Yes	Yes
Air flow control	Yes	Yes	Yes
Temperature control per compartment	Yes	Yes	Yes
Feed uptake	Yes	Yes	Yes
NH3 (Ammonia) emission	No	Yes	Yes
Temperature/humidity	Yes	Yes	Yes
Animal body weight	Yes	Yes	Yes
Dry matter content	Yes	Yes	Yes
Crude protein content	Yes	Yes	Yes
Fat content	Yes	No	Yes
Crude ash content	Yes	No	Yes
NDF content	Yes	No	Yes
Organic digestibility	Yes	No	Yes
Manure dry matter content	Yes	Yes	Yes
Manure organic Matter content	Yes	Yes	Yes
Manure total nitrogen (N) content	No	Yes	Yes
Manure mineral nitrogen (NH4 & NO3) content	No	Yes	Yes
Manure phosphorus (P2O5) content	No	No	Yes
Manure potassium (K2O) content	No	No	Yes
Manure magnesium (MgO) content	No	No	Yes
Manure calcium (CaO) content	No	No	Yes
Manure sodium (Na2O) content	No	No	Yes
Faecal consistency score	No	No	Yes
Ammonia concentration	No	Yes	No
CO2 concentration	Yes	No	No
Sow and litter data	Yes	Yes	Yes

2.2.23 Pilot 23: Biogas – Agrinio, Greece

Pilot Owner:	TBA
Capacity	5 MW
Farmers involved Directly:	10
Farmers involved indirectly:	> 22,000
Available Data:	Electricity production and temperature data in biogas digestion towers

Pilot 23 is located in Agrinio, Etoloakarnania (Greece) and covers electricity generation from biogas. The pilot is closely located and directly associated with pilot 4 (Forage production for livestock feeding) and pilot 21 (Organic Cattle Farming). The pilot is owned by TBA and directly involves 1o farmers. However, TBA via the mother company (Agrinio Union Cooperative) plans to inform more than 20,000 producers on AgriDataValue results, mainly on irrigation and fertilization reduction. Via the pilot, manure from the organic cattle farm is directly transferred (via underground pipes of more than 1km) to the anaerobe's digesters, where biogas is produced. The biogas is used by a 5MW generator that directly provides electricity to the smart grid. Solid and liquid remaining are directly used as fertilizers and irrigation at pilot 4, which produces in return livestock feeding to be used in pilot 21.

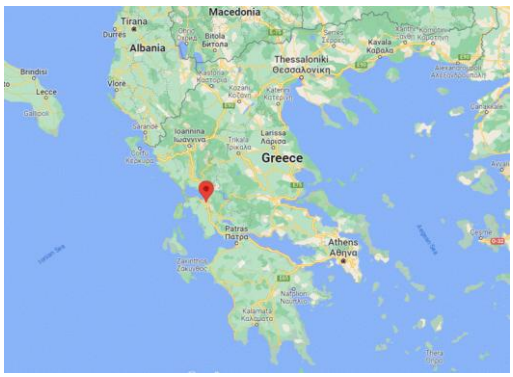
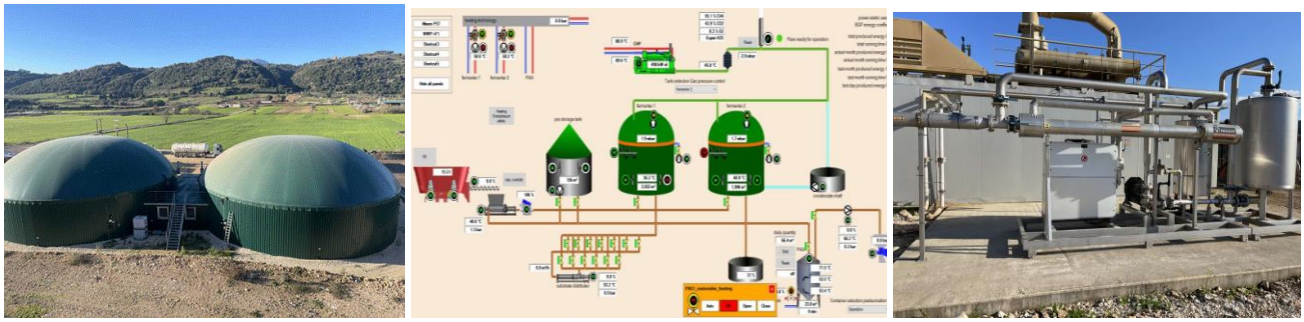


Figure 47: AgriDataValue Pilot 23 Location



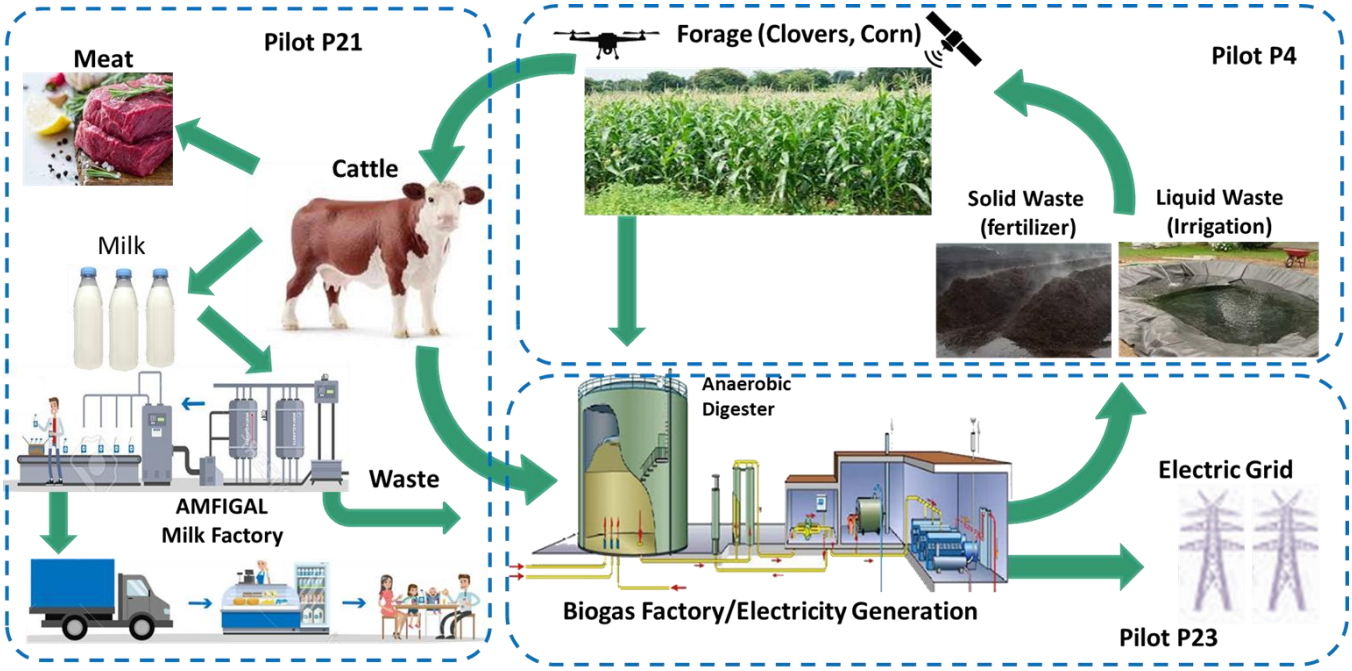


Figure 48: Figure from AgriDataValue Pilot 23 Biogas based electricity generation

Dataset #23			
Overview			
Dataset Name	TBA Biogas based electricity generation		
Dataset Category	Primary data collected by partner in ADV		
Data Owner	TBA		
Description	The datasets contain		
Already before the ADV project?	Limited. Additional data will be captured.		
Data purpose in ADV	Train/challenge ML models to circular economy.		
Use beyond ADV	Could be helpful in efficient data linking		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Electricity Generator	Power Generation	W	10 minutes
Electricity Generator	BGP coefficient	%	10 minutes
Electricity Generator	Actual month produced energy	MW	10 minutes
Sensors in Digesters	Temperature in the digesters	°C	10 minutes
Sensors in Digesters	CH4 (Methane)	%	10 minutes
Sensors in Digesters	CO2	%	10 minutes
Sensors in Tanks	Manure in tanks	m^3	10 minutes
Sensors in Tanks	Temperature in Tanks	°C	
Other Data	**Measuring**	**Units**	**Frequency**
Lab analysis Data	pH	-	Weekly Samples
Lab analysis Data	Total/Volatile Solids	%	Weekly Samples
Lab analysis Data	Volatile Organic Acids	mg/L	Weekly Samples
Lab analysis Data	Total Inorganic Carbonate	mg/L	Weekly Samples
Lab analysis Data	Conductivity	mS/cm	Weekly Samples
Storage and access details			
Data format type	CSV, XML, JSON		
Storage Location	Cloud Server		
Storage Duration	Permanent		
Is the data public or confidential?	Confidential		
Data harmonisation activities required?	Yes		
Data merging activities required?	Yes		
Can data be used after the ADV project?	Yes, but permission needs to be granted		
Ethics and Data Protection			
Includes Personal data?			No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?			Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?			Yes
Does dataset contain previously collected personal data?			No

Dataset #23: Data use per pilot use case	
Data variable	**UC 5.1**
Power Generation	Yes
BGP coefficient	Yes
Actual month produced energy	Yes
Temperature in the digesters	Yes

CH4 (Methane)	Yes
CO2	Yes
Manure in tanks	Yes
Temperature in Tanks	Yes
pH	Yes
Total/Volatile Solids	Yes
Volatile Organic Acids	Yes
Total Inorganic Carbonate	Yes
Conductivity	Yes

2.3 Captured data/information

This section maps the individual pilots to the Agricultural Use Cases and the data type to the agricultural Use Cases

The table 2 below gives an overview of the Agricultural Use Cases, indicating in which pilot which UC occurs.

Table 2: Overview of the Agricultural Use Cases per pilot

#	Pilot partner	UC 1.1	UC 1.2	UC 1.3	UC 1.4	UC 2.1	UC 2.2	UC 2.3	UC 2.4	UC 2.5	UC 3.1	UC 3.2	UC 3.3	UC 3.4	UC 4.1	UC 4.2	UC 4.3	UC 4.4	UC 5.1	UC 5.2	UC 5.3	UC 5.4
1	UL	X	X	X						X												
2	Delphy	X																				
3	ZSA			X																		
4	TBA	X	X																			
5	Inagro		X	X		X																
6	TEC	X	X	X				X	X													
7	Inagro	X																				X
8	Inagro		X			X	X															
9	Inagro				X																	
10	Inagro		X	X		X																X
11	Delphy			X																		
12	Sarga			X							X											
13	TBA												X								X	
14	CSVE											X										
15	RINOVA										X	X										
16	NILEAS										X	X		X								
17	RINO										X			X								
18	BioRo	X																				
19	ILVO														X	X	X	X				
20	ZSA																X	X				
21	TBA	X	X														X		X	X		
22	ILVO														X	X	X					
23	TBA																		X			

Table 3 shows an overview of all data/information collected and mapped to the agricultural Use Cases.

Table 3: Overview of collected data in relation to the use cases

Captured Information (part 1 of 2)	Use Case / Technology used	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	2.5	3.1	3.2	3.3	3.4	4.1	4.2	4.3	4.4	5.1	5.2	5.3	5.4
AI-algorithm 3rd party	Drone images			X																		
Air flow control	Ventilation sensors														X							
Air Humidity	Air sensor										X			X								
	Soil sensor	X																				
	Weather Forecast										X	X	X									
	Weather station	X	X	X		X			X	X	X	X	X	X							X	X
Air Temperature	Air sensor										X			X								
	Soil sensor	X																				
	Thermostates														X							
	Weather Forecast										X	X	X									
	Weather station	X	X	X		X			X	X	X	X	X	X							X	X
Ammonia content	Air scrubber															X						
Ammonia emission (g/day)	Ammonia sensor															X						
Ammonia emissons (g/day)	Ammonia sensor															X						
Animal behavior	Activity/tracking sensor																X					
	Camera system																X					
Animal identification	RFID														X	X	X					
Animal indentification	identity collar RFID															X						
As applied information	PWM-spot spray system			X																		
Blossom map	Drone images			X																		
Calving behavior	Activity/tracking sensor																X	X		X		
Calving date	Artificial insemination																	X				
CH4	Air sensor														X							
CO2	Air sensor														X							
Conformity	Body condition score																X					
Drinking uptake																	X					
EC	Soil scan		X			X	X														X	
	Soil sensor										X	X	X									
Energy	Energy Measurement sensor																		X			
Evapotransiration	Weather station										X	X	X									
FAPAR	Satellite images		X	X		X	X														X	
Feed uptake																	X					
Flow control	Ventilation sensors															X						
FTIR emisson (g/day)	Greenhouse FTIR sensor														X							
Health monitoring	Medication registration																X					
Heat detection	Activity/tracking sensor																X			X		
Hyperspectral signal	Harvest scanner				X																	
Irrigation	Soil sensor										X	X	X									
Leaf wetness	Weather station										X	X	X									

table continues...

Captured Information (part 2 of 2)	Use Case / Technology used	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	2.5	3.1	3.2	3.3	3.4	4.1	4.2	4.3	4.4	5.1	5.2	5.3	5.4
Meat Supply Chain data	RFID																			X		
Methane emission (g/day)	Methane/CO2 sensor														X							
Milk production	Milk robot/parlor														X							
N-content	Manure analysis															X						
NDRE	Drone images		X			X	X														X	
NDVI	Drone images		X			X	X														X	
	Satellite images	X	X	X		X	X				X										X	
NIR/Chem. Analysis	Feed Quality analysis														X	X						
OC%	Soil scan		X			X	X														X	
Pest	Field observation data										X											
pH	pH meter					X															X	
	Soil scan		X			X	X														X	
Phenology	Field observation data										X											
Plant nutrients	Selective ion sensor					X															X	
PLS model results	Harvest scanner				X																	
Production data	Biogas																		X			
Radiation	Weather station								X		X	X	X								X	
Rainfall	Soil sensor	X																			X	
	Weather Forecast										X	X	X									
	Weather station	X	X	X		X					X	X	X	X							X	X
Salinity	Conductivity meter					X															X	
Soil humidity	Soil sensor										X	X	X									
Soil moisture	Satellite images	X	X			X	X														X	
	Soil sensor	X	X	X		X				X	X		X	X							X	X
Soil pH	Soil sensor									X												
Soil temperature	Soil sensor	X	X	X							X	X	X	X							X	X
Solar radiation	Weather station											X										
Temperature	Thermal Measurement sensor																		X			
Thermal mag	Aircraft Thermal map											X										
TIF-files	Drone images			X																		
Total Soluble solids	Refractometer							X													X	
Ureum content	Ureum analysis															X						
Walking behavior	Pressure sensors																X					
Water potential 20cm	Soil sensor										X	X	X									
Water potential 40cm	Soil sensor										X	X	X									
weight of animals	Load cells														X							
Weight of feed	Load cells														X							
Wind direction	Weather station														X							
Wind speed	Weather station	X	X	X		X			X		X	X	X								X	X
Yield potential	Drone images		X			X	X														X	
	Satellite images		X	X		X	X														X	

end of table

3 Other Datasets

Beyond the pilots there are also other datasets that will be available in the AgriDataValue project. These can be summarised in the following categories:

- EO datasets coming from satellites
- Image datasets that contain other images then earth observation or drone data
- Synthetic datasets that are artificially generated from original data by a model that is trained to reproduce the characteristics and structure of the original data. This means that synthetic data and original data should deliver very similar results when undergoing the same statistical analysis.
- Machine Learned models created during the project to predict trends and even make decisions
- XAI explainers and explanation artefacts created during the explanation of ML models
- Hash of Trained AI models tracked into the blockchain to avoid any data tampering
- Compiled/aggregated data
- Project management data

Some examples are shown below. A good overview is given in deliverable D1.1 section 5.6.1 for EO data. The EO data are described more in detail in D4.1.

3.1 Earth Observation Datasets

Dataset #24.1	
Overview	
Dataset Name	Sentinel-1 data
Dataset Category	Publicly available dataset (e.g., training / benchmark data)
Data Partner	SINER
Provider (if different from partner)	ESA/Copernicus
Description	Copernicus Sentinel-1 imagery is provided by two polar-orbiting satellites, operating day and night performing C-band synthetic aperture radar imaging, enabling them to acquire imagery regardless of the weather. Main applications are for monitoring sea ice, oil spills, marine winds, waves and currents, land-use change, land deformation among others, and to respond to emergencies such as floods and earthquakes. Accessible through Sentinel-Hub services.
Already before the ADV project?	Yes
Data purpose in ADV	The data will be used for qualitative and quantitative monitoring of pilots.
Use beyond ADV	Already widely in use
Storage and access details	
Data format type	GeoTIFF, JP2, SAFE
Storage Location	Copernicus Data Access Service, DIAS, AWS Available through Sentinel-Hub: services.sentinel-hub.com/api/
Storage Duration	Permanent
Is the data public or confidential?	Public
Data harmonisation activities required?	No
Data merging activities required?	No

Can data be used after the ADV project?	The data is already publicly available through the Copernicus programme. Usage during and after the project will be facilitated through Sentinel Hub.
Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #24.2	
Overview	
Dataset Name	Sentinel-2 data
Dataset Category	Publicly available dataset (e.g., training / benchmark data)
Data Partner	SINER
Provider (if different from partner)	ESA/Copernicus
Description	Copernicus Sentinel-2 is a European wide-swath, high-resolution, multi-spectral imaging mission. Its high-resolution optical images have many applications, including land monitoring, emergency response and security services assistance. The satellite's multispectral imager provides a versatile set of 13 spectral bands spanning from the visible and near infrared to the shortwave infrared.
Already before the ADV project?	Yes
Data purpose in ADV	The data will be used for qualitative and quantitative monitoring of pilots.
Use beyond ADV	Already widely in use
Storage and access details	
Data format type	GeoTIFF, JP2, SAFE
Storage Location	Copernicus Data Access Service, DIAS, AWS Available through Sentinel-Hub: services.sentinel-hub.com/api/
Storage Duration	Permanent
Is the data public or confidential?	Public
Data harmonisation activities required?	No
Data merging activities required?	No
Can data be used after the ADV project?	The data is already publicly available through the Copernicus programme. Usage during and after the project will be facilitated through Sentinel Hub.
Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #24.3
Overview

Dataset Name	Sentinel-3 data
Dataset Category	Publicly available dataset (e.g., training / benchmark data)
Data Owner	SINER
Provider (if different from partner)	ESA/Copernicus
Description	Copernicus Sentinel-3 is a European wide-swath, medium-resolution, multi-spectral imaging mission designed to monitor ocean surface topography as well as land and sea surface temperature. The satellite hosts 4 instruments: the Sea and Land Surface Temperature Radiometer (SLSTR), the Ocean and Land Colour Instrument (OLCI), a Sar Radar Altimeter (SRAL) and a Microwave Radiometer (MWR). Sentinel-3A was launched on 16 February 2016 and its twin Sentinel-3B on 25 April 2018.
Already before the ADV project?	Yes
Data purpose in ADV	The data will be used for qualitative and quantitative monitoring of pilots.
Use beyond ADV	Already widely in use
Storage and access details	
Data format type	GeoTIFF, JP2, SAFE
Storage Location	Copernicus Data Access Service, DIAS, AWS Available through Sentinel-Hub: services.sentinel-hub.com/api/
Storage Duration	Permanent
Is the data public or confidential?	Public
Data harmonisation activities required?	No
Data merging activities required?	No
Can data be used after the ADV project?	The data is already publicly available through the Copernicus programme. Usage during and after the project will be facilitated through Sentinel Hub.
Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #24.4	
Overview	
Dataset Name	Sentinel-5P data
Dataset Category	Publicly available dataset (e.g., training / benchmark data)
Data Owner	SINER
Provider (if different from partner)	ESA/Copernicus
Description	The Sentinel-5P (P for precursor) mission aims at providing information and services on air quality and climate between 2017 and at least 2023. With the TROPOMI sensor on board it makes daily global observations of key atmospheric constituents, including ozone, nitrogen dioxide, sulphur dioxide, carbon monoxide, methane, formaldehyde as well as cloud and aerosol properties. The mission aims at ensuring data continuity between the retirement of the Envisat satellite and NASA's Aura mission and the launch of Sentinel-5.

Already before the ADV project?	Yes.
Data purpose in ADV	The data will be used for qualitative and quantitative monitoring of pilots.
Use beyond ADV	Already widely in use
Storage and access details	
Data format type	GeoTIFF, JP2, SAFE
Storage Location	Copernicus Data Access Service, DIAS, AWS Available through Sentinel-Hub: services.sentinel-hub.com/api/
Storage Duration	Permanent
Is the data public or confidential?	Public
Data harmonisation activities required?	No
Data merging activities required?	No
Can data be used after the ADV project?	The data is already publicly available through the Copernicus programme. Usage during and after the project will be facilitated through Sentinel Hub.
Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #24.5	
Overview	
Dataset Name	Global Land Cover data
Dataset Category	Publicly available dataset (e.g., training / benchmark data)
Data Partner	SINER
Provider (if different from partner)	European Commission Joint Research Centre (JRC)
Description	Global Land Cover products at 100 m resolution are delivered annually by the global component of the Copernicus Land Service. The most recent collection 3 (version 3.0.1) of 100 m Land Cover products for the years 2015 - 2019 were generated from the PROBA-V 100 m and 300 m satellite observations and several other ancillary datasets with global coverage. These Land Cover products provide a main discrete land cover classification map according to UN-FAO Land Cover Classification System LCCS. Additional continuous fractional layers for all basic land cover classes which give the percentage of a 100 m pixel that is filled with a specific land cover class, are also included in the Land Cover products to provide more detailed information on each land cover class. Global Land Cover products at 100 m resolution are delivered annually by the global component of the Copernicus Land Service. The most recent collection 3 (version 3.0.1) of 100 m Land Cover products for the years 2015 - 2019 were generated from the PROBA-V 100 m and 300 m satellite observations and several other ancillary datasets with global coverage. These Land Cover products provide a main discrete land cover classification map according to UN-FAO Land Cover Classification System LCCS. Additional continuous fractional layers for all basic land cover classes which give the percentage of a 100 m pixel that is filled with a

	specific land cover class, are also included in the Land Cover products to provide more detailed information on each land cover class.
Already before the ADV project?	Yes
Data purpose in ADV	The data will be used for qualitative and quantitative monitoring of pilots.
Use beyond ADV	Already widely in use
Storage and access details	
Data format type	GeoTIFF, JP2, SAFE
Storage Location	Copernicus Data Access Service, DIAS, AWS Available through Sentinel-Hub: services.sentinel-hub.com/api/
Storage Duration	Permanent
Is the data public or confidential?	Public
Data harmonisation activities required?	No
Data merging activities required?	No
Can data be used after the ADV project?	The data is already publicly available through the Copernicus programme. Usage during and after the project will be facilitated through Sentinel Hub.
Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #24.6	
Overview	
Dataset Name	Vegetation Phenology and Productivity Parameters
Dataset Category	Publicly available dataset (e.g., training / benchmark data)
Data Partner	SINER
Provider (if different from partner)	European Commission Joint Research Centre (JRC)
Description	Vegetation Phenology and Productivity Parameters (VPP) product is part of the Copernicus Land Monitoring Service (CLMS), pan-European High Resolution Vegetation Phenology and Productivity (HR-VPP) product suite. The VPP product is comprised of 13 parameters that describe specific stages of the seasonal vegetation growth cycle. These parameters are extracted from Seasonal Trajectories of the Plant Phenology Index (PPI) derived from Sentinel-2 satellite observations at 10m resolution. Since growing seasons can traverse years, VPP parameters are provided for a maximum of two growing seasons per year. The parameters include (1) start of season (date, PPI value and slope), (2) end of season (date, PPI value and slope), (3)length of season, (4) minimum of season, (4) peak of the season (date and PPI value), (5) amplitude, (6) small integrated value and (7) large integrated value. VPP parameters are generated over the EEA39 region on a yearly frequency from January 1 2017 onwards. Among other applications, the high-resolution phenology data provides a detailed assessment of the impacts of human or climate change on the ecosystem through monitoring

	of vegetation responses to disturbances, e.g. droughts, storms, insect infestations, and to human influence from global to local levels. More information about VPP product is outlined in the product user manual.
Already before the ADV project?	Yes
Data purpose in ADV	The data will be used for qualitative and quantitative monitoring of pilots.
Use beyond ADV	Already widely in use
Storage and access details	
Data format type	GeoTIFF, JP2, SAFE
Storage Location	Copernicus Data Access Service, DIAS, AWS Available through Sentinel-Hub: services.sentinel-hub.com/api/
Storage Duration	Permanent
Is the data public or confidential?	Public
Data harmonisation activities required?	No
Data merging activities required?	No
Can data be used after the ADV project?	The data is already publicly available through the Copernicus programme. Usage during and after the project will be facilitated through Sentinel Hub.
Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #24.7	
Overview	
Dataset Name	Global Surface Water
Dataset Category	Publicly available dataset (e.g., training / benchmark data)
Data Partner	SINER
Provider (if different from partner)	European Commission Joint Research Centre (JRC)
Description	The Global Surface Water dataset was developed by the European Commission's Joint Research Centre within the framework of the Copernicus Programme. It is derived from Landsat imagery and and shows different aspects of the spatio-temporal distribution of surface water between 1984 and 2021 at the global scale.
Already before the ADV project?	Yes
Data purpose in ADV	The data will be used for qualitative and quantitative monitoring of pilots.
Use beyond ADV	Already widely in use
Storage and access details	
Data format type	GeoTIFF, JP2, SAFE
Storage Location	Copernicus Data Access Service, DIAS, AWS Available through Sentinel-Hub: services.sentinel-hub.com/api/
Storage Duration	Permanent
Is the data public or confidential?	Public
Data harmonisation activities required?	No

Data merging activities required?	No
Can data be used after the ADV project?	The data is already publicly available through the Copernicus programme. Usage during and after the project will be facilitated through Sentinel Hub.
Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #24.8	
Overview	
Dataset Name	MODIS
Dataset Category	Publicly available dataset (e.g., training / benchmark data)
Data Partner	SINER
Provider (if different from partner)	United States Geological Survey
Description	MODIS (Moderate Resolution Imaging Spectroradiometer) is the main instrument operating on both NASA's Terra and Aqua satellites. It acquires images of the earth in 36 bands within the visible and the infrared regions of the spectrum at low to medium spatial resolutions. MODIS is designed to provide at least daily observations of land, oceans and lower atmosphere that contribute to local or global scale land or water applications. There are several products derived from MODIS which include land, atmosphere, cryosphere and ocean products. MCD43A4 Nadir BRDF-Adjusted Reflectance (NBAR) product Version 6 is a MODIS land product available in Sentinel Hub. It is computed daily for each of MODIS bands 1 - 7 and it contains the most representative pixels from 16 day period temporarily weighted to the ninth day. MCD43A4.006 is available since February 2000 on a daily basis but with a 8 days delay.
Already before the ADV project?	Yes
Data purpose in ADV	The data will be used for qualitative and quantitative monitoring of pilots.
Use beyond ADV	Already widely in use
Storage and access details	
Data format type	GeoTIFF, JP2, SAFE
Storage Location	Copernicus Data Access Service, DIAS, AWS Available through Sentinel-Hub: services.sentinel-hub.com/api/
Storage Duration	Permanent
Is the data public or confidential?	Public
Data harmonisation activities required?	No
Data merging activities required?	No
Can data be used after the ADV project?	The data is already publicly available through the Copernicus programme. Usage during and after the project will be facilitated through Sentinel Hub.
Ethics and Data Protection	

Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

Dataset #24.9	
Overview	
Dataset Name	Envisat
Dataset Category	Publicly available dataset (e.g., training / benchmark data)
Data Partner	SINER
Provider (if different from partner)	ESA
Description	The purpose of the Medium Resolution Imaging Spectrometer (MERIS) on the ESA ENVISAT mission was primarily, to aid in Ocean Colour Observations, and secondary, to aid the understanding of the atmospheric parameters associated with clouds, water vapour and aerosols. Additionally, MERIS broadened its scope to land surface parameters, in particular, vegetation processes. Therefore, MERIS monitors changes in ocean parameters (phytoplankton, yellow substance, suspended matter), atmosphere (water vapour, CO2, clouds, aerosols), and land (vegetation index, global coverage, moisture etc.). MERIS has a high spectral and radiometric resolution and a dual spatial resolution of 260m x 290m over land and coastal regions and reduced resolution of 1040m x 1160m over ocean. It acquires data with 3 days revisit time and in 15 spectral bands whose width and location can be programmed by ground command. The data is available from June 2002 to April 2012.
Already before the ADV project?	Yes, no new data
Data purpose in ADV	The data will be used for qualitative and quantitative monitoring of pilots.
Use beyond ADV	Already widely in use
Storage and access details	
Data format type	GeoTIFF, JP2, SAFE
Storage Location	Copernicus Data Access Service, DIAS, AWS Available through Sentinel-Hub: services.sentinel-hub.com/api/
Storage Duration	Permanent
Is the data public or confidential?	Public
Data harmonisation activities required?	No
Data merging activities required?	No
Can data be used after the ADV project?	The data is already publicly available through the Copernicus programme. Usage during and after the project will be facilitated through Sentinel Hub.
Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes

Does dataset contain previously collected personal data?	No

3.2 Images Datasets

Dataset #25			
Overview			
Dataset Name	Woode dataset		
Dataset Categories	1. Publicly available dataset 2. Synthetic dataset generated using augmentation techniques 3. Primary data sources collected within the ADV project from pilot visits 4. Secondary sources collected from farmers.		
Data Owner	QMUL		
Description	The dataset is collected from a review of existing public datasets that have been published in the literature. The annotated dataset samples will be complemented with the associated knowledge and the types of pest that would affect the leaves and thus the health of the plant/crop. The data aggregation will include a set of images being captured from the field by the respective personnel.		
Already before the ADV?	Yes. Additional data will be captured.		
Data purpose in ADV	The dataset will be used to achieve the objectives of T3.2: Smart Farming support toolbox & CAP compliance and T4.4: Climate impact projection on soil, crops, livestock and biodiversity		
Use beyond ADV	The publication of the annotated dataset and the performance baseline that has been established will be useful for the research community to undertake further research in the development of advanced machine learning models		
Measuring technologies	**Measuring**	**Units**	**Frequency**
Camera system	NH3 (Ammonia) emission	pixels	NA
Storage and access details			
Data format type	ZIP file consisting of the images and the associated annotations released in standard format		
Storage Location	QMUL central repository		
Storage Duration	5 years		
Is the data public or confidential?	Public with restricted access		
Data harmonisation activities required?	Yes		
Data merging activities required?	Yes		
Can data be used after the ADV project?	The collected dataset will be released in the public domain, but licence agreement will be required for the users of the dataset to sign and comply.		
Ethics and Data Protection			
Includes Personal data?	No		
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes, only the relevant sample of images consisting of leaves and pests will be recorded. For data audit purposes, only the minimal representation of the users will be recorded		
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	No, the quality of the annotation and the validity of the results are important to trust the data source. However,		

	pseudo anonymisation process can be adopted prior to the release of the dataset.
Does dataset contain previously collected personal data?	No

3.3 Synthetic Datasets

Dataset #26	
Overview	
Dataset Name	AgriDataValue Synthetic Datasets
Dataset Category	Synthetic data
Data Owner	INTA, ALMA, SIEM
Description	Synthetic labelled datasets are going to be generated with generative models, with data coming from use cases or vulnerability databases
Already before the ADV project?	No
Data purpose in ADV	These Synthetic labelled datasets are going to be in order to assist with the evaluation of FDML algorithms
Use beyond ADV	AgriDataValue Synthetic Datasets could be used to train and evaluate more Deep Learning and Machine Learning algorithms for the considered use cases
Storage and access details	
Data format type	- tabular or sequential data: .csv - images: .png or .jpeg
Storage Location	To be determined
Storage Duration	During the length of AgriDataValue project
Is the data public or confidential?	Confidential
Data harmonisation activities required?	No
Data merging activities required?	No
Can data be used after the ADV project?	Synthetic datasets could be made publicly available beyond the project end, for the purpose of increasing the research in the field.
Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	NA
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	NA
Does dataset contain previously collected personal data?	No

3.4 Trained Machine Learned models

Dataset #27	
Overview	
Dataset Name	AgriDataValue ML Dataset
Dataset Category	(semi-) trained ML models
Data Owner	INTRA, ALMA, SIEM
Description	(semi-) trained ML models are going to be generated with data coming from those use cases that require AI solutions.

Already before the ADV project?	No
Data purpose in ADV	These ML models will be used to make predictions for different use cases and provide AI solutions for AgriDataValue end users.
Use beyond ADV	It is expected that the stored models represent an advance in the state of the art and can serve as a reference for future research
Storage and access details	
Data format type	YAML, JSON, H5, PICKLE, JOBLIB, ONNX
Storage Location	OwnCloud
Storage Duration	At least the length of AgriDataValue project
Is the data public or confidential?	Confidential
Data harmonisation activities required?	No
Data merging activities required?	No
Can data be used after the ADV project?	Synthetic datasets could be made publicly available beyond the project end, for the purpose of increasing the research in the field.
Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	NA
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	NA
Does dataset contain previously collected personal data?	No

3.5 XAI explainers for AI models

Dataset #28	
Overview	
Dataset Name	AgriDataValue XAI Dataset
Dataset Category	XAI explainer models
Data Owner	ALMA
Description	Files containing pre-trained XAI explainers that analyse and interpret the behaviour of AI models, providing visualisations and metrics to make predictions and decision pathways trustable by and understandable to end-users.
Already before the ADV project?	No
Data purpose in ADV	XAI explainers designed to enhance interpretability and transparency for specific AI models, clarifying decision-making processes and providing insights into feature importance, local/global explanations, and interaction effects for targeted applications.
Use beyond ADV	Persistently stored XAI explainers enable consistent model interpretability, reproducibility of insights, and validation of AI decisions over time, trust in AI systems
Storage and access details	
Data format type	YAML, JSON, H5, PICKLE, JOBLIB
Storage Location	OwnCloud
Storage Duration	At least the length of AgriDataValue project
Is the data public or confidential?	Confidential
Data harmonisation activities required?	No

Data merging activities required?	No
Can data be used after the ADV project?	Synthetic datasets could be made publicly available beyond the project end, for the purpose of increasing the research in the field.
Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	NA
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	NA
Does dataset contain previously collected personal data?	No

3.6 Hash of trained AI models

Dataset #29	
Overview	
Dataset Name	AgriDataValue Blockchain Hash Dataset
Dataset Category	Primary data collected by partner in ADV
Data Owner	ALMA
Description	Trained AI Models are hashed, using SHA256 algorithm, and sent to a SmartContract, alongside the URL in which they are saved in SECURESTORE, to track them into the blockchain.
Already before the ADV project?	No
Data purpose in ADV	To avoid any tamper with data and to allow the verification of the data integrity.
Use beyond ADV	
Storage and access details	
Data format type	HASH, STRING
Storage Location	Blockchain, SECURESTORE
Storage Duration	At least the length of AgriDataValue project
Is the data public or confidential?	Confidential
Data harmonisation activities required?	No
Data merging activities required?	No
Can data be used after the ADV project?	Yes
Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	NA
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	No

3.7 Climate Change Adaptation Strategies & Recommendation Data

Dataset #30.1
Overview

Dataset Name	Climate Change Adaptation Strategies & Recommendation Data
Dataset Categories	Text Data
Data Owner	ALMA
Description	Based on models and indicators from T4.1 and T4.3, these datas project the climate impact indicators to the agriculture activities and estimate how they are linked with the relevant agriculture and biodiversity indicators in order to provide recommendation for adaptation strategies
Already before the ADV?	No
Data purpose in ADV	The data will provide recommendation of adaptation strategies for the involved pilots agricultural system.
Use beyond ADV	Adaptation strategies would be useful under features climate change scenarios beyond ADV project life.
Storage and access details	
Data format type	Text
Storage Location	OwnCloud
Storage Duration	At least the length of AgriDataValue project
Is the data public or confidential?	Confidential
Data harmonisation activities required?	No
Data merging activities required?	No
Can data be used after the ADV project?	These data could be made publicly available
Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	NA
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	NA
Does dataset contain previously collected personal data?	No

Dataset 30.2	
Overview	
Dataset Name	APPAG data
Dataset Category	Institutional anonymised dataset
Data Partner	APPAG
Provider (if different from partner)	/
Description	APPAG's dataset includes the following elements limited to the territory of the Autonomous Province of Trento: - land use / land cover - graphic cultivation plan - livestock database - graphic pasture land register
Already before the ADV project?	Yes

Data purpose in ADV	Within the framework of WP4, APPAG will provide valuable data to help finalize ADV climate models. This initiative focuses on the Autonomous Province of Trento, which will act as a small-scale pilot region. The goal is to predict the impacts of climate change on local agriculture and specific cultivation areas. By analysing this data, we aim to develop comprehensive guidelines that will be useful for helping local farmers to understand future agricultural trends and implement targeted mitigation measures, in order to adapt to changing climate conditions.
Use beyond ADV	Original data, in its non-anonymised form, is usually used to perform controls on beneficiaries of EU funds as provided by EU legislation. It is also used to support policy decisions in the field of agriculture in the Autonomous Province of Trento.
Storage and access details	
Data format type	• Shape, jml, geojson, gpkg; • alphanumeric (csv, Excel)
Storage Location	To be defined
Storage Duration	During the length of AgriDataValue project
Is the data public or confidential?	Confidential
Data harmonisation activities required?	Yes
Data merging activities required?	Yes
Can data be used after the ADV project?	No. Any use must be requested and authorised by APPAG as it is issued, in the agreed form, according to the specific request related to the ADV project
Ethics and Data Protection	
Includes Personal data?	No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?	Yes
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?	Yes
Does dataset contain previously collected personal data?	Yes (the dataset is derived from anonymised personal data)

3.8 Management datasets

Dataset #31	
Overview	
Dataset Name	Produced documentation
Dataset Category	Project management data
Data Owner	All partners

Description	The data that will be generated during the project concerning management and documentation	
Already before the ADV project?	No	
Data purpose in ADV	To document the project's progress and/or disseminate results	
Use beyond ADV	Information flow	
Storage and access details		
Data format type	pdf, doc, txt	
Storage Location	OwnCloud and submission to EU portal	
Storage Duration	At least the length of AgriDataValue project	
Is the data public or confidential?	Depending on the deliverable type	
Data harmonisation activities required?	No	
Data merging activities required?	No	
Can data be used after the ADV project?	Yes	
Ethics and Data Protection		
Includes Personal data?		No
Has Data Minimisation review conducted to ensure that data are adequate, relevant and limited to the ones necessary in relation to the purposes of ADV research?		NA
Can research objective be achieved by using anonymized or pseudonymised data or synthetic data?		NA
Does dataset contain previously collected personal data?		No

3.9 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, ML models, smart Farming best practices etc.) or physical (e.g. new materials, reagents, samples, etc.). Also, most of the developments of the academic partners will be made available at open- source software, published in Gitlab **[3]**. In addition, scientific outputs such as abstracts, papers, conference posters etc. will be made open access as much as possible to provide free accessibility to the scientific community.

The open-source strategy foresees offering Open-Source Software (OSS) results in a business-friendly way, thus AgriDataValue commits to selecting a business-friendly license (MPL 2.0 **[4]**/LGPL **[5]**), whereas consortium partners already have a track record of OSS contributions and experience.

4 Data Management Plan

As has already been highlighted, the AgriDataValue consortium already owns a significant number of datasets and will further capture data from the 23 project pilots, which are distributed throughout Europe. Additional data will be captured from drones and satellites. Finally, data will be generated either as synthetic data, as ML models or as supporting data (e.g. video material).

The purpose of the Data Management Plan (DMP) is to set out the main elements of the AgriDataValue consortium data management policy for the datasets owned, captured or generated by the project, to present the procedures for the management of datasets and to describe the key data management principles, notably in terms of data standards and metadata, sharing, archiving, and preservation. As it is presented in [6], DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon Europe project. As part of making research data findable, accessible, interoperable and re-usable, the DMP includes information about the handling of research data during and after the end of the project:

- The kind of data that are and will be collected, processed, and generated and to whom they are useful
- The methodology and standards are and will be applied
- The metadata that are required to enable data to be found and understood, ideally according to the particular standards of a scientific discipline
- Which data are or will be shared or made open access
- The way to preserve data (including after the end of the project)
- How the open datasets of the project are and will be archived and preserved

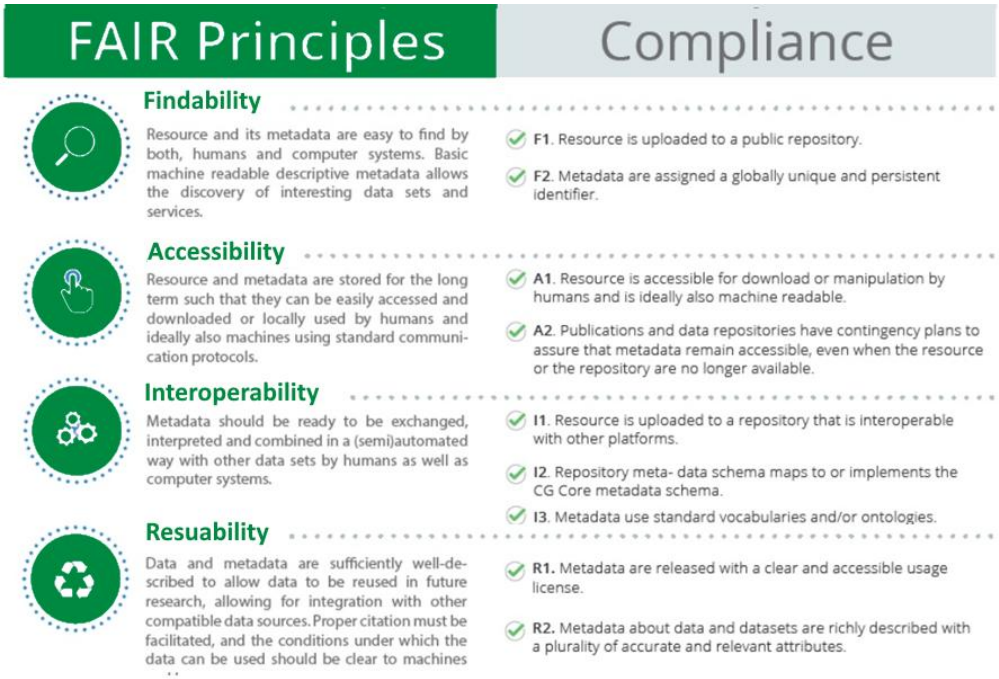


Figure 49: Research Data Management according to FAIR principles[1]

[1] Image source: CGIAR. Creative commons attribution-noncommercial 4.0 international license.

More specifically, for Horizon Europe projects like this one, a FAIR DMP template has been designed to apply to the project that produces, collects, and processes research data. Figure 49 and Figure 50 present the FAIR data principles for promptly disseminating the data outcomes of a research project.

Data should be **Findable**	F1. (meta)data are assigned a globally unique and persistent identifier (DOI) F2. data are described with rich metadata F3. metadata clearly and explicitly include the identifier of the data it describes F4. (meta)data are registered or indexed in a searchable resource
Data should be **Accessible**	A1. (meta)data are retrievable by their identifier using a standardized communications protocol A1.1 the protocol is open, free, and universally implementable A1.2 the protocol allows for an authentication and authorization procedure, where necessary A2. metadata are accessible, even when the data are no longer available
Data should be **Interoperable**	I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (meta)data use vocabularies that follow FAIR principles I3. (meta)data include qualified references to other (meta)data
Data should be **Reusable**	R1. meta(data) are richly described with a plurality of accurate and relevant attributes R1.1. (meta)data are released with a clear and accessible data usage license R1.2. (meta)data are associated with detailed provenance R1.3. (meta)data meet domain-relevant community standards

Figure 50 : Research Data Management according to the FAIR principles data source acquisition [7]

4.1 DMP Analysis

This section concerns the analysis of the captured and generated data.

4.1.1 Data generation and collection, and its relation to the objectives of the project

Basic data will initially be generated and collected by the pilot partners from their own sources or from farmers and service providers. The data includes historical data and newly generated data. These data are diverse as they are necessary to execute the different use cases. These data will cover different use cases. The basic data serve to feed the Machine Learning models, which will lead to the creation of a platform of platforms.

With the current knowledge, captured data, after relevant clearance and proper anonymization will be released if possible as open data, and if needed as closed data, but the application will be for Open Tools.

4.1.2 Origin of the data

The initial data are agricultural and farm-related data. These can be generated on the farm or on a research station. Some data are first processed by a service provider before they become available as interpretable data for the farmers or researchers. The different data flows are collected and uniformed by the partners, and after anonymising and minimisation given as input to the technology partners for model building.

4.1.3 Data Sources for acquisition and generation

As described in Chapters 2 and 3, data collected or generated in AgriDataValue can be either public, open data or internal data collected or generated from or by partners, including mainly AgriDataValue partners, ESA, and research organizations. Primary data will be generated or re-used by almost all partners, but especially by the pilot partners. For some partners, secondary datasets must be combined with the primary data to create a derived dataset, whether or not in combination with publicly available datasets. Some generated data can also be synthetic data for machine learning. A final category are the management data. Data collected in AgriDataValue involve the following data sources:

- Data types from farms other than sensor data,
- Data types captured from Sensors and drones from the various UCs,
- Data types received from satellites via various hubs,
- Sensitive data types, either personally or commercially sensitive,
- Data types that will be produced, for example ML (semi-)trained models, XAI explainer models,
- Data types from supply chain processes,
- Trained AI models to be hashed for blockchain tracking,
- Procedures to handle the data,
- Other data such as papers, abstracts, posters, project deliverables,

4.1.4 Types and formats of data that will be generated and collected

As already analyzed the data types can be classified in another way:

- Observational (e.g. survey results, sensor readings, sensory observations);
- Experimental (e.g. microscopy, spectroscopy, chromatograms, gene sequences);
- Compiled or aggregated data (e.g. text & data mining, derived variables, 3D modelling);
- Simulation data (e.g. synthetic data for Machine Learning models);
- Machine Learning Models;
- XAI explainers and explanation artifact data (e.g. XAI-trained models);
- Supply chain processes data (e.g. text data);
- AI models tracking data (e.g. hashes, strings);
- Software tools, etc.

The data formats are very diverse because of the different fields of expertise of the partners within the AgriDataValue consortium:

- tabular data (.csv, xls., .ods, .por,. spss, structured text or mark-up file XML, .tab,)
- textual data (.rtf, .xml, .txt, .doc)
- geospatial data (.dwg, .gml, .gpx)
- data ML models, XAI explanation artifact (.json, .yaml, .h5, .picle, .joblib, .onnx)
- visual/ multispectral/image data (.jpeg, .jpeg2000, .png, .gif, .tiff, .raw, .pdf, .hdf, .GeoTIFF, .jp2, .safe)
- audio data (mp3, .wav, .wma)
- video data (.mp4, .mov, .avi, avchd)
- documentation (.doc, .html, .pdf, .ppt, .xls, .odt)
- computational scripts (.js, .vbs, .scpt, .php)

- text data from supply chain operations (strings and numbers)
- hashed data from AI Models tracking
  (eg. 0x4355a46b19d348dc2f57c046f8ef63d4538ebb936000f3c9ee954a27460dd865)

In addition, a main point of the DMP is the definition of the open access type over the data. **Open Access (OA)** refers to the practice of providing online access to scientific information that is free of charge to the end-user and reusable. 'Scientific' refers to all academic disciplines. In the context of research and innovation, 'scientific information' can mean:

- peer-reviewed scientific research articles (published in scholarly journals),
- Scientific abstracts,
- Scientific posters,
- research data (data underlying publications, curated data, or raw data).

Open Access to scientific publications means free online access for any user. The two main routes to Open Access are:

- **Self-archiving / 'green' Open Access** – the author, or a representative, archives (deposits) the published article or the final peer-reviewed manuscript in an online repository before, at the same time as, or after publication. Some publishers request that open access be granted only after an embargo period has elapsed.
- **Open Access publishing / 'gold' Open Access** – an article is immediately published in open access mode. In this model, the payment of publication costs is shifted away from subscribing readers. The most common business model is based on one-off payments by authors.

Research data refers to information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings, and images. The focus is on research data that is available in digital form. Users can normally access, mine, exploit, reproduce, and disseminate openly accessible research data free of charge.

The open access mandate comprises two steps:

1) Depositing publications in repositories
2) Providing open access to them

4.1.5 Expected size of the data

The exact size of the captured and generated data is still unknown in this phase, as the partners do not know at this moment how much data will be created that feed AgriDataValue technology partners during the remaining project years. As the project progresses, it will become clearer what the data size will be, but we made an estimation of lower limits of the expected sizes of the different data types. Nevertheless, an initial estimation is provided in the following table:

Type	Details-Format	Estimated size
System requirements. Design notes and best practices.	reports (.docx, .xls, .csv) photos (.jpeg, .png) training videos (.mp4, .mov, .avi)	Depending on the file format (> 50 GB)
Remote sensing (>80 types of heterogeneous sensors and IoT devices)	Air, soil, leaf, location, electricity, solar radiation, animal-related data in JSON, XML, text files (. json, .csv, .txt)	> 600GB

Visual/ (multi-)spectral images/video	Drones, Sentinel and Landsat EO, GEOSS, videos (.mp4, .mov, .avi) photos (jpeg, JPEG2000, Fast-L7A, GeoTIFF, HDF)	Depending on experiment (8-15 TB) Access to >44 PB of EO
ML trained models – AI methodology and results	ML algorithms and trained models as reports (.docx, .xls, .csv), or in JSON, XML, YAML, JSON, H5, PICKLE, JOBLIB, ONNX	>2 GB
XAI explainers and explanation artifact data – XAI methodology and results	XAI explainer algorithms and XAI trained models as reports (.docx, .xls, .csv), or in JSON, XML, YAML, JSON, H5, PICKLE, JOBLIB, ONNX	> 2 GB
Supply chain operation data	Data identifying a specific operation in a supply chain	>1 MB
Hashed data for tracking	Data to be tracked into the blockchain saved as a hash	>96 bytes
Climate Change Adaptation Strategies & Recommendation Data	Text dataset	> 5 MB
Demonstration material	videos (.mp4, .mov, .avi) photos (.jpeg, .png)	> 50 GB
Strategies/Use case validation	Policies, guidelines, reports (.docx, .xls, .csv)	> 800 MB

4.1.6 Re-use of existing data and, if so, how?

Data of existing datasets and databases will be also used. Data within the project will be shared (in some cases with data sharing agreement).

4.1.7 Data utility

In first instance the data will flow to the technology developers for model learning and tool development. Data from the pilot partners are separate and have another category than the open data generated from the technology platforms on, which are open access.

4.2 FAIR Data

The main purpose of this section is to describe the methods that will make data **FAIR** (Findable, Accessible, Interoperable, and Reusable) [8]

4.2.1 Findable

A standard project-specific identification mechanism or Digital Object Identifier (DOI) will be assigned to datasets for effective and persistent citation when it is uploaded to a repository. This DOI can be used in any relevant document or publication to access the data.

Since the datasets are being deposited in mostly external repositories, a dataset registry record will be created. The registry record will include relevant metadata explaining which data exist (cf. metadata files) and a DOI linking to where the data are available in the external repository. SYN is responsible for the AgriDataValue dataset registry

implementation and management. A list of keywords will be implemented following the project partners for search purposes. This list of keywords will improve the findability and make the data better available for reuse within and outside of the project.

Specific internal document identification on OwnCloud of SYN during the project that is linked to the work packages, deliverables, and tasks will be used with clear versioning. For each WP, task, or deliverable there is a leader. The partners can adjust versions, but only the leader can create a new version. All changes/adjustments made by a partner or leader are identified with the acronym of that partner.

All data will have an associated metadata document that describes key aspects of the data. A metadata file will be created and linked within each dataset. It will include the following information:

- General information
- Title of the dataset
- Dataset identifier
- Responsible partner
- Involved partners
- Title of the project and Funding sources that supported the collection of the data, if applicable

Moreover, every dataset collection will include:

- Specific methods used for data collection: links or references to relevant documents, if available
- Specific methods for processing the data: describe how the submitted data were generated from the raw or collected data
- Instruments, sensors, and software used in data collection
- Standard format to read data and make these data usable, if applicable
- Specific parameters will be supplied and converted to a standard format to make them generally usable for the platforms.

Beyond datasets, ***event listings*** are stored in a central spreadsheet and individual events are assigned a unique identifier of the formal format to avoid any confusion with contributions, deliverables, and internal documents. Therefore, official deliverables and internal documents and reports will have the following formal format:

ADV_Dw.d_ACR_Vx.y_YYYYMMDD.ext and

ADV_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, ATOS, ILVO, TBA, ….),

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 detail, the partners' acronyms to be used are: SYN, ATOS, SIXEN, INTRA, SIEM, SINER, ALMA, IDSA, SIMAVI, SLG, ILVO, NKUA, Inagro, UL, TEC, Delphy, ITAIN, ZSA, SARGA, TBA, SIVE, NILEAS, CVSE, BIORO, RI.NO, AgroDS, NPA, APPAG, APIA, QMULL. Photographs, audio, and visual recordings are formatted as:

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

AgriDataValue provides search keywords in the metadata to optimise the possibility for discovery and potential re-use.

4.2.2 Making data openly accessible

Data that are created in the project are owned by the partners, farmers, or service providers who generate them and are under their responsibility. Raw and processed data will be curated and preserved by project partners following their internal procedures. These data are largely sensor data that are linked with production data, therefore these data will be minimised and anonymised. For this type of data, we use the principle 'as open as possible, as closed as necessary', because of the commercial importance of some of these data. The data of a partner can only be used in agreement with the project partner that owns the data or supervises them. Processed and aggregated datasets will be stored in repositories and made available to the technology developers during the project. Unique identifier numbers are used as indicated above. Where possible, public datasets will be used in the development of the platforms of platforms.

Each partner must disseminate its results as soon as possible unless there is a legitimate interest to protect the results. There will be a clear difference in accessibility of data. Farm data and sensor data are only accessible for this project. Data coming from modelling or platform building will be open access. Pilot data should be deposited in a repository as soon as possible. Project participants who wish to withhold patentable or proprietary data can do so. During embargo periods, information about the restricted data will be published in the data repository, and details of when the data will become available will be included in the metadata. Processed and aggregated data must be made available to AgriDataValue technology partners upon request. Data sharing with third parties will be subject to a data-sharing agreement. The agreement will indicate the conditions of use, criteria for access, and acknowledgments.

Sharing information among project partners and toward a larger community will be based on partner or general repositories. Some research institutions have started to develop in-house strategies to promote data sharing and respect FAIR principles. As an alternative to institutional repositories, the use of Zenodo (https://zenodo.org/) will be promoted. Zenodo will augment the AgriDataValue data Findable and Accessible. An AgriDataValue community will be created on this tool, after which any partner can upload a dataset there by using the project's uid (this is directly connected to ECAS). In any case, the chosen repositories will comply with the FAIR principles and should be open, free, searchable, and structured with flexible licensing allowing for storing all types of data: datasets, images, presentations, publications, and software. All the public data of the project will be openly accessible at the repository. Non-public data will be archived in a repository using "closed access" or "on request" options. Each dataset will get a Digital Object Identifier (DOI) to be easily and uniquely citable. Repositories will provide access to metadata and data files provided over standard protocols such as HTTP and OAI-PMH.

For processed data generated during the project, data will be shared mostly using CSV, TXT, or XLS formats, hence no specific tools will be required to read these data. In specific cases, JSOIN and image files are needed. All central

information will be kept on the OwnCloud managed by Synelixis Solutions S.A., Greece. Data are stored with version control. GITLAB is used for code management. If necessary, there will be a manual created explaining how to read and handle the data. This is part of the method created to increase tool usability.

4.2.3 Interoperable

The data usually needs to be integrated with other data. In addition, the data needs to interoperate with applications or workflows for analysis, storage, and processing. Whenever possible, standard operation procedures, describing data collection, existing or worked out during the time of the project, will be applied. This will be a combined initiative of the pilot and technology partners to allow maximal and optimal data exchange and modelling. Otherwise, standards and formats will be detailed and referenced to in the metadata files associated with each dataset.

The expected type of data to be handled within the AgriDataValue project is research data as reports or spreadsheets, test data (simulated or historical anonymised data) as spreadsheets, and source code that might be subject to license and/or copyright.

- **Reports.** Research data will be consolidated in specific reports, i.e. deliverables, and will be shared with the research community via open repositories and project website. They will be indexed with metadata as described in previous sections.
- **Test data.** Generated and/or measured data together with the resulting data will follow standards or methodologies and will be stored and if appropriate shared in interoperable formats like: JSON, XML, REST, and CSV.
- **Source code/ML models.** Whenever possible, the AgriDataValue project will adhere to standards for formats, as much as possible compliant with available (open) software applications.

4.2.4 Increase data re-use (through clarifying licenses)

Public data will remain available for long after the project grant ends, for the lifetime of the underlying public archives. This will be achieved by accurate recording of metadata, associated protocols, and deposition in public repositories, as mentioned above. These data will therefore be reusable by any party. In some cases, there will be an embargo on the data till the publications are published or, if applicable, till a patent is obtained.

Some basic or raw data may not become publicly available as they may hold sensitive commercial data. Nevertheless, these data remain reusable during the project and thereafter after a signed agreement with the partner has been obtained, as long the data are kept. To make the reuse possible, a procedure will be set up, so that the data can be used by a third party after obtaining a signed agreement of the partner that generated the data.

The **EUDAT B2SHARE** tool [8] will be used to facilitate the selection of an adequate license for research data if applicable. Partners of the project are committed to giving access to all processed data not subject to patent application one year after the end of the project at the latest. For data under embargo, the way data will be made available will be discussed on a case-by-case basis and terms will be detailed in a data-sharing agreement. The central AgriDataValue interface and data registry will be maintained as long as the AgriDataValue website is maintained. Moreover, each partner will set up an **internal control that guarantees the data quality**.

4.3 Allocation of resources

There are no direct identified costs associated to datasets deposited in the public repositories. Thus, the cost for making AgriDataValue open data FAIR is estimated to be zero, which means archiving data in the AgriDataValue repository will be free of charge. The choice of using external and public repositories is to benefit from the involvement of partners or public organizations to cover the costs of their long-term repository infrastructures. However, the cost of sharing some data along with 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. Costs for open access publication of scientific outputs will be carried by the publishing partner.

ILVO is responsible for the AgriDataValue Data Management Plan (DPM), INTRA is responsible for the AgriDataValue Data Protection Officer (DPO) and SYN is responsible for capacity building and for the development of a central registry.

Each partner that generates data is responsible for their data management or the data generated to feed the ML models. They are responsible for the version management, merging is done by the task leaders. The outputs will be distributed through trusted knowledge depositories and the project's knowledge reservoir which will be kept alive at least 5 years after the project 's end. Some partners keep their data for 3-6 years or after publication, others store them permanent on a server. Moreover, as has already been explained, specific partners may choose to offer data and ML models in a prosumer approach, via the AgriDataValue mechanism and tools to get specific incentives and benefits and to be motivated to actually offer the relevant data.

Original data and ML models will be kept, while intermediate stages are not kept to save space and costs.

5 Data Security

For the duration of the project, raw data and processed data will be stored and secured on the responsible partner's storage facility according to their internal procedures. However, as a general rule these data cannot be kept only on temporary disks or personal folders as storage medium. A backup process has been worked out to avoid loss of data. Redundancy is built into the network to cope with failures to some extent. Sufficient backup equipment is in place to provide a minimum level of service.

Additional secure storage among partners will be implemented on a case-by-case basis for high-dimensional data (i.e. visual and multispectral images mainly from drones, and video recordings used for dissemination video material). Secure storage using the OwnCloud platform [9], managed by the project coordinator (Synelixis), is used during the project for document storage and management.

Following data stored on a public repository, all the responsibilities concerning data recovery and secure storage will be with the repository infrastructure. The project does not foresee any sensitive data. Yet, in case capture or transferring of sensitive data is needed, it will not be done via mail but via encryption mail server or stored encrypted and then downloaded using the necessary credentials.

With regards to personal data, the AgriDataValue 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, 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) (please refer to Section 6).

AgriDataValue will take measures to preserve anonymity and appropriately curate the collected data. The aim is to gather data valuable for the 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, intelligible, and usable beyond the original purpose for which it was collected and interoperable to specific quality standards. The AgriDataValue project activities will evaluate and implement issues related to data protection & privacy and evaluate informed consent (to guarantee 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 for ensuring 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.

5.1 Data Protection Impact Assessment

The Data Protection Impact Assessment (DPIA) process aims at providing assurance that controllers adequately address privacy and data protection risks of 'risky' processing operations. By providing a structured way of thinking

about the risks to data subjects and how to mitigate them, DPIAs help organisations to comply with the requirement of 'data protection by design' where it is needed the most, i.e. for 'risky' processing operations [10].

A DPIA is required for:

- systematic and extensive evaluation of personal aspects relating to natural persons based on automated processing, including profiling, that could produce legal effects concerning the natural person or similarly significantly affect the natural person;
- processing on a large scale of special categories of data referred to in Article 10, or of personal data relating to criminal convictions and offenses referred to in Article 11;
- systematic monitoring of a publicly accessible area on a large scale.

As the AgriDataValue project does not include personal data in any format, beyond anonymized data, DPIA methodology is not applicable in this project. Nevertheless, the AgriDataValue project will conduct a data minimisation review for all available datasets to ensure that all data that we intend to process are needed, adequate, relevant, and limited to what is necessary in relation to the purposes of our research.

6 Ethics and GDPR compliance

Data protection and good research ethics are major topics for the consortium of this project. AgriDataValue 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 and that the use of both existing and new data is agreed upon with the data provider and owner. Data records containing personal data will be managed in accordance with the General Data Protection Regulation (GDPR, EU: 2016/679) [11].

The AgriDataValue project activities will evaluate and implement issues related to data protection & privacy and address informed consent procedures for communication with stakeholders to guarantee voluntary participation in research as it is one of the most important procedures to address privacy issues in research.

The AgriDataValue Data Protection Officer (DPO) will be *Dr. Despina Anastasopoulos* (F) from Netcompany-Intrasoft. The DPO will lead the Privacy, Ethical, Legal & Regulatory Compliance monitoring tasks. AgriDataValue DPO actions will be compliant with the GDPR (EU 2016/679, EU 2016/680) [12] [13]. More specifically, he will ensure that the following rules as described in the AgriDataValue project Grant Agreement are followed:

- Personal Data are properly anonymized/pseudo-anonymized and processed legally and fairly
- 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 a Pilot Governance Committee (PGC) as described in section §2.1, which 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.

It should be noted that the Nagoya Protocol on Access and Benefit Sharing [14] is not applicable to this project.

6.1 Additional Ethical aspects

In order to ensure all Ethical Aspects, the project DPO will carry out a questionnaire to ensure that no personal data which do not follow specific rules (e.g. anonymisation) are shared in any way. Data in the delineated section of the survey concerning questions such as sex, age category, farm experience and education level, when combined with information from other resources such as company home pages, may be used to identify a person with some degree of certainty. This data will enable the AgriDataValue project to evaluate the demographic profile

of the respondents in regards to questions such as regional differences, differences by seniority and fields of expertise. This will help the partner to identify key differences between different groups of respondents in regards to the views on FAIR or Open data, fulfilling the criteria of public interest of Article 6(1)(e) in GDPR.

All e-mail addresses are now handled via specialised email lists by Synelixis as project coordinator and will be deleted by the completion of the AgriDataValue project to fully anonymise the dataset in accordance with Article 5(1)(c) and Article 5(1)(e).

Data with the delineated section removed will be treated as anonymised and published as supplementary information when applicable while the complete survey in its full form or anonymized state will be stored by the data controller and data processor(s) involved in AgriDataValue as necessary. The risk of data breaches will be minimised by ensuring storage with GDPR-compliant vendors and appropriate safeguards such as suitable passwords being used. The impact of a data breach is deemed low due to the nature of the questions in the survey.

6.2 Animal experiments

Livestock data are mostly collected from normal farm practices that are not subject to ethical approval. If some data should be needed from animal trials that can be classified as “animal experiments” these were or will be submitted to ethical approval.

7 Conclusions and further steps

This deliverable has been dedicated to the second iteration of the DMP for the AgriDataValue Project. As a second version of the data management plan (DMP) this document has outlined the way in which datasets in the different pilots are being generated, organised, documented, stored, backed up, preserved, and shared, if possible, with other researchers after the publication of the project's main results. This second version also identifies and designates, key responsibilities for data management within the project's consortium. The individual typically designated as the principal investigator is ultimately accountable for managing the data for the project.

Following this approach in this second version of this DMP further refines the structure of each pilot and the captured data and further updates, adjusts, and refines the data gathering and quality management and creation and execution of the DMP of each pilot site.

The DMP is a living document that is continuously shared and updated by the overall consortium project team and reviewed regularly when necessary, hence official updates may be performed when deemed necessary.

8 References

[1] "H2020 IoF2020 (Internet of Fodd & Farm 2020)," [Online]. Available: https://www.iof2020.eu/.

[2] "EIP-AGRI Network," [Online]. Available: https://ec.europa.eu/eip/agriculture/en/node.

[3] "GitLab: the DevSecOps platform," [Online]. Available: https://about.gitlab.com/.

[4] "MPL 2.0," [Online]. Available: https://www.mozilla.org/en-US/MPL/2.0/FAQ/.

[5] "GNU Lesser General Public License," [Online]. Available: https://www.gnu.org/licenses/lgpl-3.0.html.

[6] "Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020,," [Online]. Available: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.html.

[7] [Online]. Available: WWW.researchgate.net.

[8] European Commission, "EC H2020 Programme, Guidelines on FAIR Data Management in Horizon 2020, Version 3.0," [Online]. Available: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

[9] "EUDAT: Store and publish your research data," [Online]. Available: https://b2share.eudat.eu/.

[10] "Owncloud platform," [Online]. Available: https://owncloud.com/.

[11] European Data Protection Supervisor, "Data Protection Impact Assessment (DPIA)," 2023. [Online]. Available: https://edps.europa.eu/data-protection-impact-assessment-dpia_en.

[12] European Parliament and the Council, "Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)," April 2016. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[13] European Parliament and the Council , "Directive (EU) 2016/680 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation," April 2016. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016L0680.

[14] "General data protection regulation (GDPR)," 2022. [Online]. Available: https://eur-lex.europa.eu/EN/legal-content/summary/general-data-protection-regulation-gdpr.html.

[15] "Nagoya Protocol," 2015. [Online]. Available: https://www.cbd.int/abs/about/.