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

D3.1 SCALEAGDATA GENERIC ARCHITECTURE AND DATA GOVERNANCE, SHARING META-ARCHITECTURE AND INTEGRATION OF THE RI LABS, V1

31 January 2024

Authors		
Name	**Organization**	**Draft release date**
Marios Vlachos	ICCS	16/01/2024

Approval Signature on behalf of Technical Steering Committee		
Name	**Organization**	**Date of approval**
Koen Van Rossum	VITO	30/01/2024
Nikos Kalatzis	NP	30/01/2024
Nuno Grosso	DES	30/01/2024

Revision Records			
Version	**Date**	**Changes**	**Authors**
1.0	16/01/2024	Original Document	ICCS
2.0	24/01/2024	Revised based on Reviewers' comments	ICCS

Acronyms and Abbreviations

Acronyms and Abbreviations	
AGRO	The Agronomy Ontology
AGROVOC	Agriculture Vocabulary
AI	Artificial Intelligence
AIM	Agricultural Information Model
ANN	Artificial Neural Network
AoI	Area of Interest
API	Application Programming Interface
APSIM	Agricultural Production Systems sIMulator
ATB	Institut für angewandte Systemtechnik Bremen GmbH
AUTH	Aristotle University of Thessaloniki
AVR	AVR BVBA
AWS	Amazon Web Services
BioPAR	Biogeophysical PARameters
BLE	Bluetooth Low Energy
CA	Consortium Agreement
CGIAR	Consultative Group on International Agricultural Research
CNH	CNH INDUSTRIAL BELGIUM
CPU	Central Processing Unit
CSA	Coordination and Support Action
CSV	Comma-Separated Values
CSW	Catalogue Services for the Web
DA	Data Act
DBMS	Database Management System
DES	Deimos Spain
DGA	Data Governance Act
DHI	DHI A/S
DIAS	Data and Information Access Services
DMA	Digital Markets Act
DME	DEIMOS ENGENHARIA SA
DMK	DMK Deutsches Milchkontor GmbH
DSA	Digital Services Act
DSS	Decision Support System
DSSC	Data Space Support Center
EC	European Commission
EEAB	External Expert Advisory Board

EDIB	European Data Innovation Board
EGM	Easy Global Market SAS
ENVO	The Environment Ontology
EO	Earth Observation
EOD	Earth Observation Data
ETSI	European Telecommunications Standards Institute
EURAC	Accademia Europea di Bolzano (Eurac Research)
EVA	Evapotranspiration
EV ILVO	Eigen Vermogen van het Instituut voor Landbouw en Visserij Onderzoek
ExBo	Executive Board
FAIR	Findability, Accessibility, Interoperability, and Reuse
FEU	Farm Europe
FMIS	Farm Management Information System
fPAR	fraction of absorbed Photosynthetically Active Radiation
GA	General Assembly
GAEC	Good Agricultural and Environmental Conditions
GARDIAN	Global Agricultural Research Data Innovation and Acceleration Network
GDPR	General Data Protection Regulation
GeoJSON	Geographic JSON
GeoTIFF	Geographic Tag Image File Format
GHG	Greenhouse Gas
GIS	Geographic Information System
GNDVI	Green Normalized Difference Vegetation Index
GPP	Gross Primary Productivity
GPR	Gaussian Process Regression
GPU	Graphics Processing Unit
HE	Homomorphic Encryption
HORTA	HORTA SRL
HPC	High Performance Computing
HTTP	Hypertext Transfer Protocol
IA	Innovation Area
ICCS	Institute of Communication and Computer Systems
IDSA	International Data Spaces Association
IFAPA	Instituto Andaluz de Investigación y Formación Agraria, Pesquera y Alimentaria
IoT	Internet of Things
IPR	Intellectual Property Rights
JSON	JavaScript Object Notation

KO	Kick Off
KPI	Key Performance Indicator
KUVA	Kuva Space Oy
LAI	Leaf Area Index
LCA	Life Cycle Assessment
LoRaWAN	Long Range Wide Area Network
LUE	Light Use Efficiency
LUKE	Natural Resources Institute Finland
MAE	Mean Absolute Error
MARS-OP	Monitoring Agricultural ResourceS – Operations
MIGAL	MIGAL Galilee Research Institute
ML	Machine Learning
MLP	Multi-layer Perceptron
MPC	Multi-Party Computation
MS	Microsoft
MST	Management Support Team
NDMI	Normalized Difference Moisture Index
NDRE	Normalized Difference Red Edge
NDVI	Normalized Difference Vegetation Index
NGSI-LD	Next Generation Service Interface with Linked Data
NP	Neuropublic SA
NPK	Nitrogen, Phosphorus, and Potassium
NPP	Net Primary Productivity
OAuth2	Open Authorization
OGC	Open Geospatial Consortium
OHB DS	OHB Digital Services GmbH, Bremen, Germany
OIDC	Open ID Connect
OWL	Web Ontology Language
PATO	Phenotype And Trait Ontology
PBA	Professional Business Analysis
PCSM	Ploutos Common Semantic Model
PDA	Personal Digital Assistant
PEF	Product Environmental Footprint
PET	Privacy Enhancing Technologies
pH	potential of Hydrogen
PMI	Project Management Institute
PO	Project Officer

PSNC	Instytut Chemii Bioorganicznej Polskiej Akademii Nauk
RACI	Responsibility Assignment Matrix
R&D	Research and Development
RDF	Resource Description Framework
REST	REpresentational State Transfer
RF	Random Forest Regressor
RGB	Red, Green and Blue
RH	Relative Humidity
RIE	Research and Innovation Environment
RIL	Research and Innovation Lab
RMSE	Root Mean Squared Error
S3	Simple Storage Service
SAR	Synthetic Aperture Radar
SAREF4AGRI	The SAREF ontology (the Smart Appliance REFerence ontology)
SDK	Software Development Kit
SFTP	Secure File Transfer Protocol
SMART	Specific, Measurable, Achievable, Relevant and Time-bound
SME	Small and Mid-size Enterprise
SOC	Soil Organic Carbon
T	Temperature
TBD	To Be Decided/Determined
TRL	Technology Readiness Level
UAV	Unmanned Aerial Vehicle
UGent	Universiteit Gent
UMA	User Managed Access
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
UTC	Universal Time Coordinated
VI	Vegetation spectral Indices
VITO	Vlaamse Instelling voor Technologische Onderzoek
VP	Vapor Pressure
VPD	Vapor Pressure Deficit
VRA	Variable Rate Application
VRI IES	Foundation "Institute for Environmental Solutions"
VTT	Technical Research Centre of Finland Ltd.
Wi-Fi	Wireless Fidelity
WMS	Web Map Services

WODR	Wielkopolski Osrodek Doradztwa Rolniczego w Poznaniu
WP	Work Package
WPS	OGC Web Processing Service
W3C	World Wide Web Consortium
XML	eXtensible Markup Language

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Project Overview

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

To do so, ScaleAgData has five objectives:

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

During its lifecycle, the project will explore seven innovation areas: innovative sensor technology, edge processing, data sharing architecture and data governance, satellite data augmentation, from data assimilation to service development, privacy-preserving technology, and data integration methodologies.
Six Research and Innovation Labs (RIL) have been identified within the project, across various biogeographical regions of Europe, where different data upscaling and integration models or approaches will be evaluated and demonstrated. The six RILs are: water productivity, crop management, yield monitoring, soil health, grasslands and sustain dairy. Recommendations will be formulated on how such integrated datasets can be capitalized to help national and regional policy making to strengthen both the competitiveness and sustainability of European agriculture.
Recommendations will be formulated on how such integrated datasets can be capitalized to help national and regional policymaking to strengthen both the competitiveness and sustainability of European agriculture.

1.2 Scope

The scope of this deliverable is to provide the specifications of (i) the overall architecture and components and (ii) the architecture defined in each lab based on the common architectural blueprint, data governance schemes and data space connectivity (first iteration).

1.3 Intended Readership

This deliverable is primarily aimed to the Consortium partners including the Commission services as a reference document with the objective of providing the specifications of (i) the overall architecture and components and (ii) the architecture defined in each lab based on the common architectural blueprint, data governance schemes and data space connectivity, within the ScaleAgData project.

1.4 Document Structure

This document is structured as follows:

- Chapter 1 provides a project overview and then goes on to describe the scope, responsibilities, and structure of this deliverable.
- Chapter 2 presents a summary of the design methodology that was applied in the determination of the high-level architecture of the ScaleAgData project. It also includes the identification of the various functional components, the basic building blocks and services of the system along with information regarding data management, data formats, data governance schemes, data sharing strategies and data space connectivity issues.

1.5 Evolution of the Document

This version of the document shows the initial approach to ScaleAgData generic architecture and data governance, sharing meta-architecture and integration of the RILs. This is intended to be a living document in which information will be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur. Therefore, each update of the ScaleAgData generic architecture and data governance, sharing meta-architecture and integration of the RILs will have a clear version number and include a timetable for updates. The project foresees two versions of the document: one at month 13 and one at month 33. Additional updates will take place if necessary due to changed circumstances that require alterations to the approaches presented herein.

2. ScaleAgData System Architecture

2.1 Methodology for Architecture Definition

The methodology followed to derive the ScaleAgData system architecture is, on one hand, based on standard literature definitions and methods for architecture derivation[1,2], which suggest the use of 'views' and 'perspectives' for a holistic and successful description of the system components and functionality. On the other hand, architecture derivation relies on the ScaleAgData user scenarios and system specifications while also taking into consideration the already established processes and functions of the RILs' infrastructure.
A well-established approach is to decompose the architectural description into views. Each view deals with a different aspect of the system. A formal definition of an architectural view is provided below: "A view is a representation of one or more structural aspects of an architecture that illustrates how the architecture addresses one or more concerns held by one or more of its stakeholders"[3].

The views that will be considered by the current architecture document are the following:

- **Functional View**: defines and describes the system's functional components and their related functions and interfaces.
- **Information View**: defines the static information structure and presents dynamic information and data flows; in other words, describes how to "define, structure, store, process, manage and exchange information"[4].
- **Deployment & Operation View**: handles installation, hardware, integration with existing infrastructure, maintenance issues; it proposes selected technologies for the system deployment.

Additionally, perspectives are useful to describe non-functional features (sometimes referred to informally as "illities") of the system. A formal definition may be found in[5]: "An architectural perspective is a collection of activities, checklists, tactics and guidelines to guide the process of ensuring that a system exhibits a particular set of closely related quality properties that require consideration across a number of the system's architectural views".

The perspectives that will be addressed by the current architecture definition are the following:

- **Security**: describes the security features of the ScaleAgData system architecture, which tackle issues such as data integrity and confidentiality, access control, authentication and authorization.
- **Performance & Scalability**: discusses trade-offs between increased performance of system components and scalability issues; for instance, a larger area that is mapped via drone vs increased cost of deployment.

[1] Woods, E. (2005). Software Architecture Using ViewPoints and Perspectives. SET2005. Zurich.
[2] Michael A. Ogush, D. C. (2000). A Template for Documenting Software and Firmware Architectures.
[3] Nick Rozanski, E. W. (2005). Software Systems Architecture: Working with Stakeholders Using Viewpoints and Perspectives.
[4] Magerkurth, C. (2012). IoT-A Deliverable D1.4 Converged architectural reference model for the IoT v2.0. IoT-A Consortium.

2.2 Overall Architecture View

The ScaleAgData system consists of a variety of technological innovations that cover seven innovation areas/approaches: innovative sensor technology, edge processing, data sharing architecture and data governance, satellite data augmentation, from data assimilation to service development, privacy-preserving technology, and data integration methodologies.

The aforementioned technologies and innovations leverage Earth Observation (EO), internet of things (IoT), digital interfaces, advance analytics, machine learning, artificial intelligence and business models.
At the same time, efficient mechanisms will be employed in order to ensure interoperability with existing control systems, as well as improved accessibility and sharing of data through harmonized and standardized means, whilst also demonstrating their uptake by relevant stakeholders for improved decision-making.

The following diagram (Figure 1) provides an overview of the high-level ScaleAgData architecture, while Figure 2 shows the functional view of the ScaleAgData system, summarizing the basic building blocks of the system, while including all of its basic layers.

The overall architecture follows a decentralized approach, where each RIL manages and processes its own data internally, from the initial data collection from the sensors at the edge to the data management and transformation. In this distributed way each RIL operates as an interconnected node inside a bigger network with other RILs. Each RIL can share data with each other directly and also with the Research and Innovation Environment (RIE) or any other external entities.

At the bottom layer is the deployment of the in-situ sensors. This is the primary collection data source for every RIL, together with EO data with which each RIL creates AI models and tools. In this layer the data types as well as the data transmission protocols may vary a lot for each RIL as each may have its own instruments to measure different physical properties that depend to the Lab's specific research focus.

On top of the sensor layer, there is the edge layer whose main focus is to bridge the in-situ data collectors with the data management layer, inside each single RIL. This layer is located near the physical location of the sensors. The edge component can act as a simple gateway for the data management layer supporting the protocols needed to collect data from sensors like Bluetooth Low Energy (BLE), Long Range Wide Area Network (LoRaWAN), Wi-Fi. When necessary, in this layer real-time data processing components can be deployed to reduce the load to RIL internal systems and enforce data quality. It can also be used to deploy federated learning algorithms which allows training of algorithms across multiple edge devices inside the same RIL.

The Data Management Layer is a central component that stores and provide APIs for data consumption of other components within it. Data originating from the in-situ sensors are validated and stored in this layer and managed exclusively by each RIL. Depending on the complexity of the data, simple or more complex querying mechanisms maybe included so that data can be easily provided to other components for further data analysis and research. It should be pointed out that only internal entities of the RIL can have access to the Data Management Layer. External parties will have access only through the Data Sharing Layer.

Using the data from Data Management Layer any RIL internally may create its own internal modules that can also retrieve data, if needed, from external services like EO Data. In this way specialized internal modules tailored to the specific research needs that may be developed or already exist can be integrated with the overall architecture.

The Data Transformation Layer is crucial for the data sharing, as it is responsible to convert the data to a universally accessible format using the Resource Description Framework (RDF), which is the base of semantic web. The primary function is to transform and enrich various datasets to a common format to ensure clarity and consistency. Given that each lab manages data using its unique data model internally, without this layer, an entity seeking data from different RILs would need to perform separate data conversions and may also be subject to misinterpretation.

The Data Sharing Layer is a critical component for the decentralized approach in ScaleAgData architecture. It is noteworthy that this is the only part of the architecture that has the same implementation for every RIL. It ensures exchange of standardized data with common API, making each lab not just an isolated research entity but a node in an interconnected network. Each RIL chooses which internal data can be shared and with whom, through the creation and implementation of specific data sharing policies. This data governance scheme respects the unique needs and privacy concerns of each lab. Despite the decentralized control of the data, the fact that the data is standardized, and the API implementation is common ensures a unified method of access. Regarding the authentication and authorization mechanisms of the data sharing layer, OAuth2 will be implemented as it is a robust and widely accepted standard. It ensures secure, flexible access, allowing only authorized entities to access specific data. It is important to note that this implementation will be used only for sharing RILs' data with external entities. Each RIL will still use its own internal authentication and authorization systems, exclusively for managing internal RIL activities and access. In this way, both independent control within labs and a secure, uniform way for external data sharing is provided.

The modules that are going to be integrated to the ScaleAgData system, and may be common or specific to each RIL’s research focus, are the following:

- Yield estimation tool
- Soil reflectance measurement
- Soil hyperspectral processing module
- Drought prediction from IoT and airborne sensor data
- Drought prediction from satellite sensor data
- Vegetation indices calculator
- Agro-environmental policy indicators monitoring
- DSS precision farming
- Digital twin providing forecasts and decision support
- Reporting module
- Milk quality and quantity forecaster
- Grasslands improved biopars (LAI, fPAR)
- Grasslands primary production
- Edge processing component
- Research and Innovation Environment (RIE)

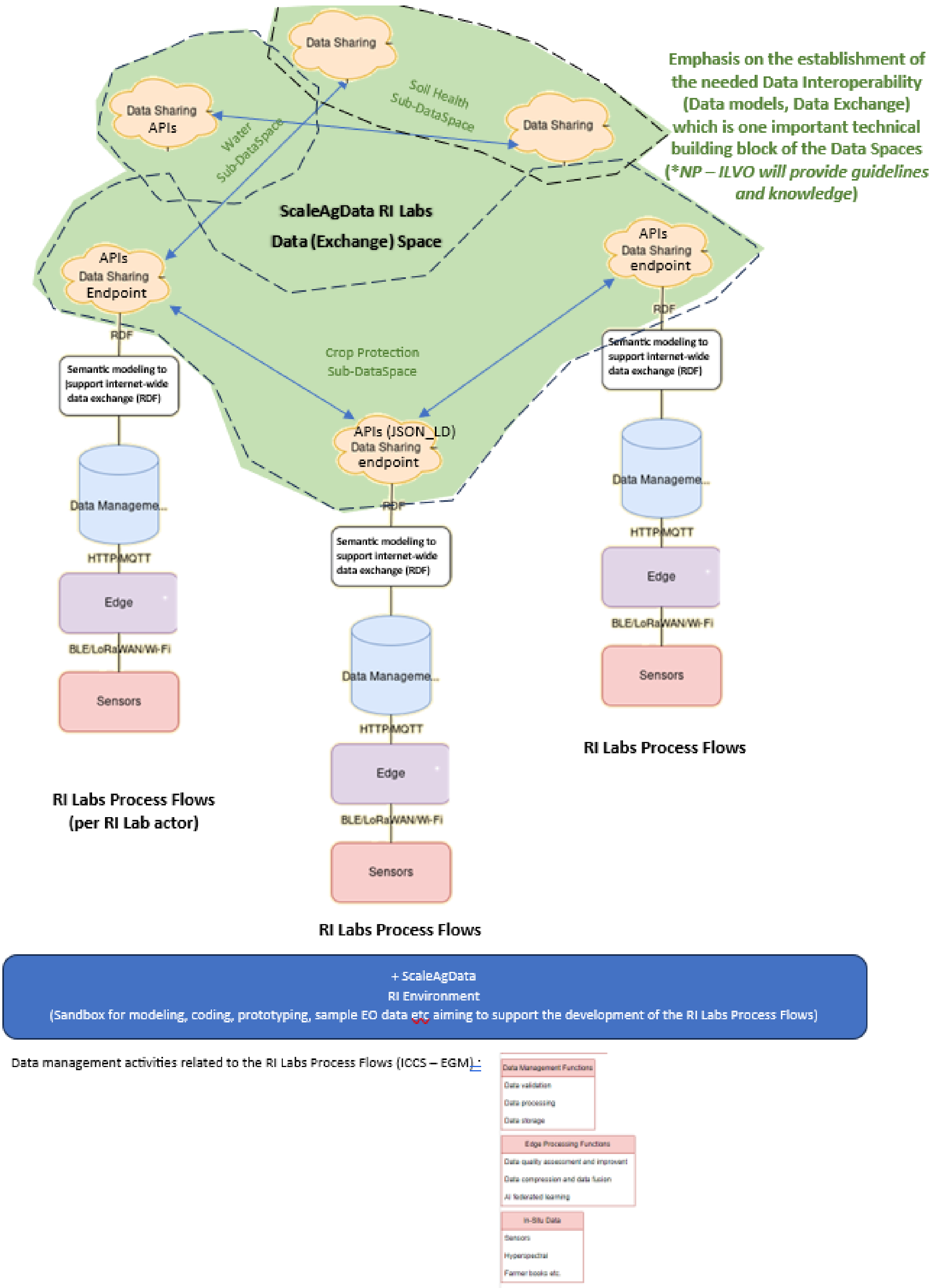


Figure 1 Functional view of the architecture of the ScaleAgData system.

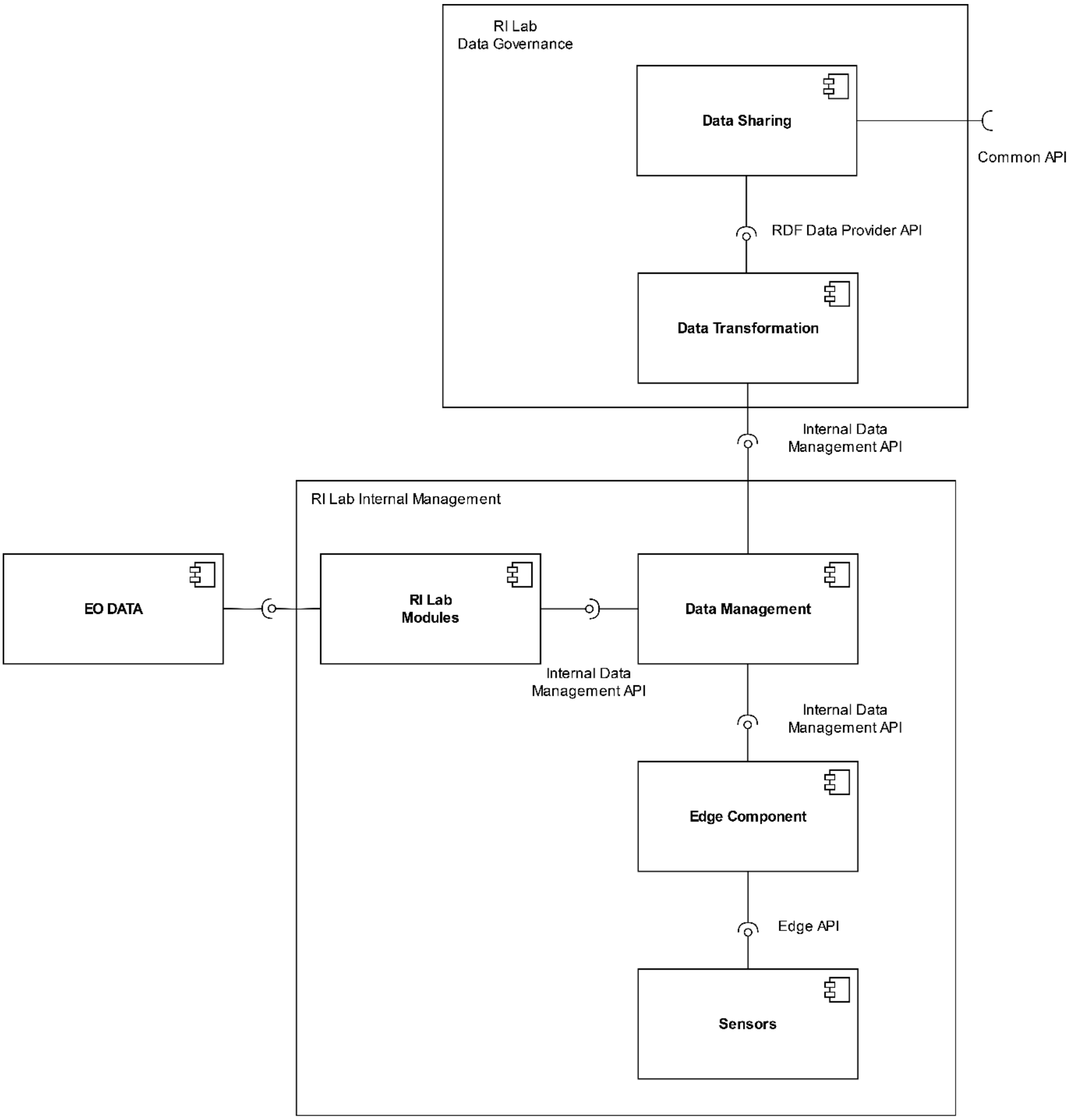


Figure 2 Functional view of the architecture of a node of ScaleAgData system.

2.3 Functional View

The functional view decomposes the system described before into building blocks while also analyzing their related services and interfaces. In component views, focus will be made in provider interfaces to highlight each component's functionality. When analyzing the information flows interactions between consumer components and provider component will become evident. Figure 3 clarifies the nomenclature of the architectural views presented in the following sections.

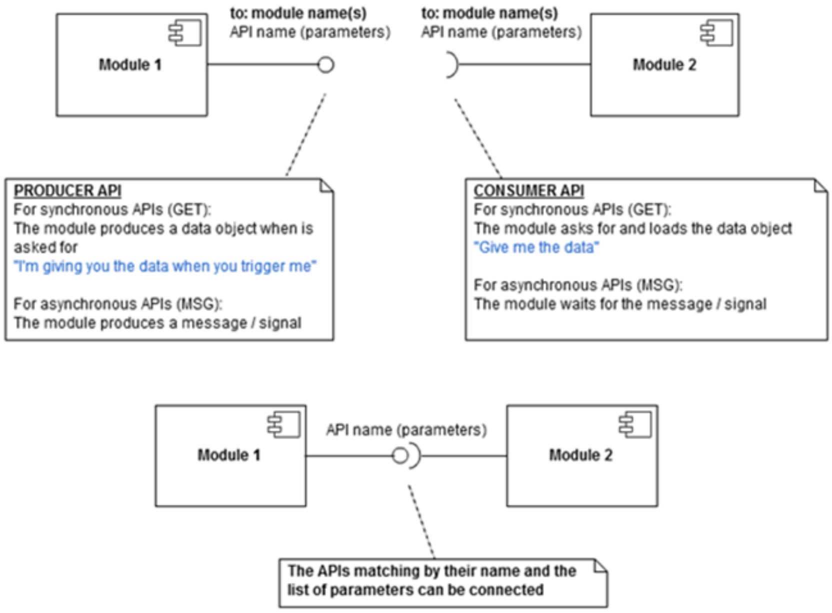


Figure 3 Architectural definition notation.

2.3.1 Yield Estimation Tool

Description	Yield estimates will be derived with an AI model trained with satellite, weather, soil, and historical data and with yield data collected from harvesters. The resulting yield estimates will be made available to external crop monitoring platforms via API. Models can be customized based on application or data availability.
Interfaces	
Requirements association	VITO, AVR, CNH

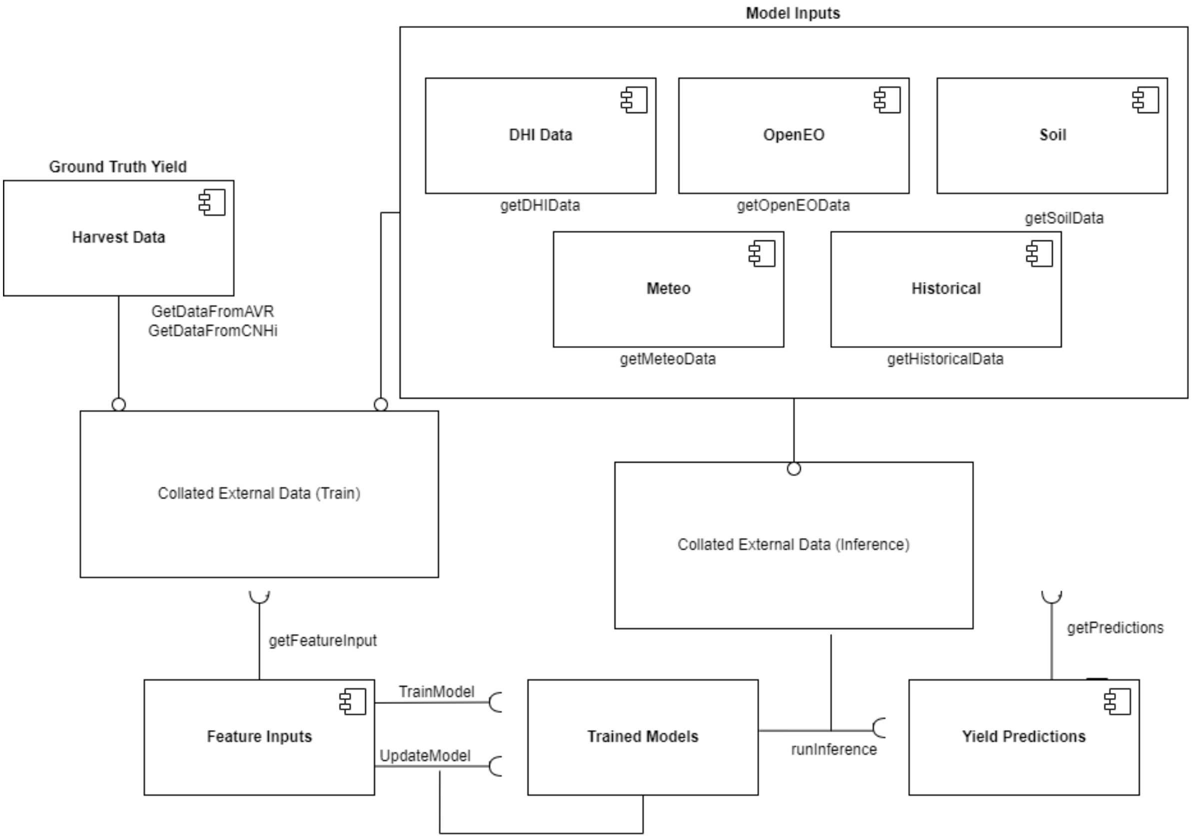


Figure 4 Functional view of yield estimation tool.

2.3.1.1 Input data collection

Yield data collection

Service Name	Service Description
getDataFromAVR getDataFromCNHi	Input parameters: region (field boundaries), time, parameter of interest Given input parameters, return corresponding yield values for the given field.

EO data collection (openEO)

Service Name	Service Description
getOpenEOData	Input parameters: region (field boundaries), time, parameter or index, output format of interest. Given input parameters, return corresponding satellite raster or time series for the given field.

EO data collection (DHI)

Service Name	Service Description
getDHIData	Input parameters: region (field boundaries), time, parameter or index, output format of interest. Given input parameters, return corresponding satellite raster or time series for the given field.

Weather data collection

Service Name	Service Description
getMeteoData	Input parameters: region (field boundaries), time, parameter of interest (temperature, rainfall, solar radiation etc.). Given input parameters, return weather parameter of interest for the given field.

Soil data collection

Service Name	Service Description
getSoilData	Input parameters: region (field boundaries), collection or parameter of interest. Given input parameters, return soil type for the given field.

Historical data collection

Service Name	Service Description
getHistoricalData	Input parameters: region (field boundaries), dates, collection or parameter of interest. Given input parameters, return historical data (climate average, past crops).

2.3.1.2 List Feature Extractors

Service Name	Service Description
getModelArchitectures	Returns possible model architectures and accepted/necessary data inputs.
getTrainedModels	Returns list of trained models. Trained model contains info on regions, input data types, training dates.

getFeatureInput

Service Name	Service Description
getFeatureInput	Input Parameters: Regions, Dates, ModelArchitecture, InputDataTypes Returns collated Inputs for model training for selected regions and time period (Training may be limited to early season). Raise error if input data types not available for regions, or no yield data available.

2.3.1.3 AI Model Training

Service Name	Service Description
TrainModel	Input Parameters: Feature Input(s) Launches training of model. Stores model and produces user metadata for using model.
UpdateModel	Input Parameters: Feature Input(s), Previous Model Launches training of model using weights from previous model. Feature Input dimensions (Date Input Types, length) must match previous model.

2.3.1.4 Yield modelling

Application of the developed yield model

Service Name	Service Description
getModelInputs	Input Parameters: TrainedModel, Regions, Dates Returns Inputs matching trained model inputs. Date range must be same length as date range used in training. Configuration may relate to prediction density.
runInference	Input Parameters: Trained Model, Regions, Inputs Returns yield predictions for each region.

2.3.1.5 Access to yield estimates

Making the results of the yield modelling accessible via API

Service Name	Service Description
getYieldPredictions	Input Parameters: Region Returns all predictions overlapping with region, and metadata on each model used.

1. Tare estimation (UGent in collaboration with AVR)

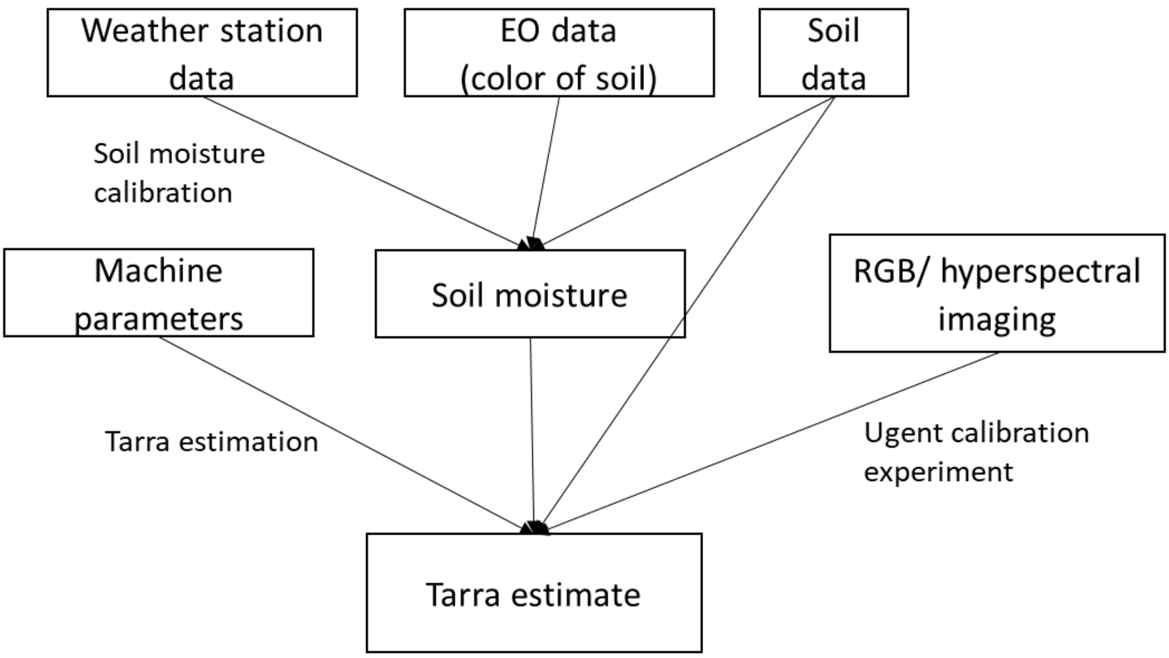


Figure 5 Tare estimation.

Calibration of soil moisture

EO data collection (VITO)

Service Name	Service Description
EO data	In collaboration with VITO. For certain plots (AVR fields) and time (just before harvest), EO data will be used to assess spatial variation in soil moisture.

Weather data collection

Service Name	Service Description
Weather station data	Data from weather station from AVR (?): T°, RH.

Soil data collection

Service Name	Service Description
Soil data	Data from soil scanner from AVR (?): pH, EC, T°, RH, NPK + soil type.

Soil sampling

Service Name	Service Description
Soil moisture	Take samples from different positions on the fields (corresponding to spatial resolution of EO data) to measure soil moisture = calibration data. Alternatively, use data from soil moisture sensors, if available.

Soil moisture calibration

Service Name	Service Description
Soil moisture calibration	Use the input and calibration data to make a regression model to estimate soil moisture. Assess if only weather station data or only EO data can be used for each soil type.

Tare calibration by imaging

Tare calibration

Service Name	Service Description
UGent calibration experiment	Data will be gathered during potato harvest to calibrate tare. Measured weight of the earth stuck on potatoes + stuck on conveyer belt will be linked to RGB images. (DELAYED due to rain) Remark: This calibration experiment should be repeated with the hyperspectral camera of VTT installed in the harvester + with different soil types!

Tare estimation

Machine parameters

Service Name	Service Description
Machine parameters	Data from AVR harvesters (traction, fuel usage etc.).

Tare estimation

Service Name	Service Description
Tare estimation	Use output of soil moisture calibration and machine parameters to make a model (multiple regression analysis) to estimate tare. Calibration data should ideally be measured directly from the field but estimates based on the tare calibration experiment(s) can be used as well. Model should be tested for different soil types.

2. Improved spatial variation of harvester yield maps (UGent in collaboration with CNHi)

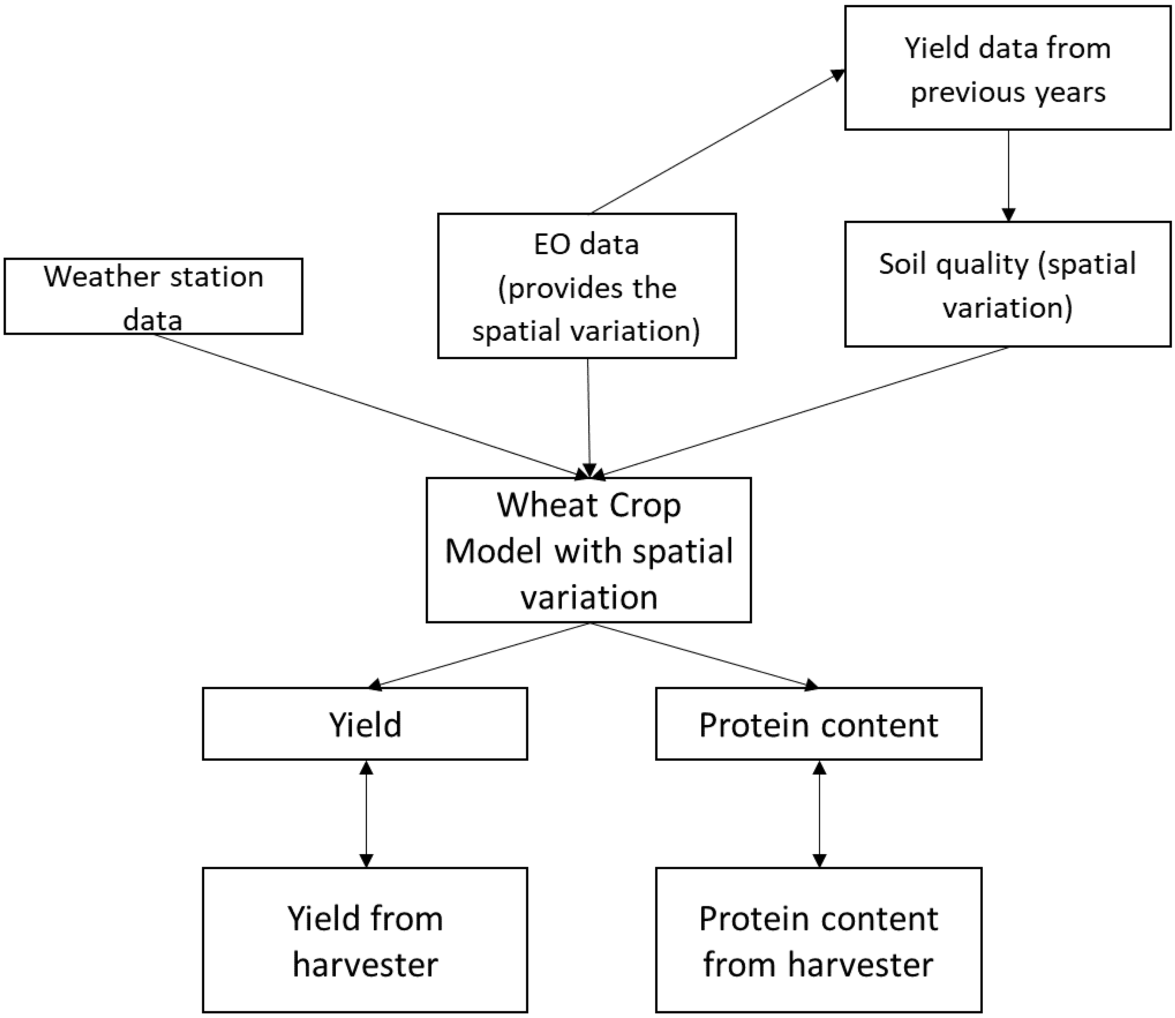


Figure 6 Improved spatial variation of harvester yield maps.

EO data collection (VITO)

Service Name	Service Description
EO data + yield data from previous years	In collaboration with VITO. For certain plots (CNH fields) and time (several time points during growth of wheat), EO data will be used to assess spatial variation in quality of soil/crop. EO data of previous years (around harvest) can provide additional information.

Weather data collection

Service Name	Service Description
Weather station data	Data from weather station from CNH: T°, RH.

Yield data collection

Service Name	Service Description
Yield from harvester + yield data from previous years	Data from CNH harvesters, with spatial variation.

Soil data collection

Service Name	Service Description
Soil quality	Data from soil scanner from CNH: pH, EC, T°, RH, NPK.

Crop model

Service Name	Service Description
Wheat Crop Model	In collaboration with LUKE. Use a digital twin or ML-based approaches to estimate yield (and protein content). The spatial variation in the input data will allow to calculate different yields with the same spatial resolution. This variation in simulation can be compared to the variation from the yield (and protein content) from the harvesters.

3. AVR

Provide yield information

Service Name	Service Description
Yield Information	Based upon x, y positions yield information of fields can be shared.

Field information

Service Name	Service Description
Field information	Input: customer identification. AVR Connect is able to show all the fields of this customer.

4. CNHi

Provide yield information

Service Name	Service Description
Yield Information	Geojson of yield can be shared.

Provide weather information

Service Name	Service Description
Weather stations	Weather data of fields (soil T, air T, rainfall, RH).

Soil data collection

Service Name	Service Description
Soil quality	Data from soil scanner from CNH: pH, EC, T°, RH.

2.3.2 Soil reflectance measurement

2.3.2.1 Hyperspectral sensors

Description, Interfaces and Requirements Association

Description	The sensors are either portable spectrometers that record a single line spectrum, or imaging cameras that capture a hyperspectral data cube (2 spatial and 1 spectral dimension). They are usually off-the-shelf solutions, although we are going to explore with VTT about the possibility of building one such device for our lab.
Interfaces	Depending on the manufacturer, the device may be connected physically (e.g., USB, Ethernet, Camera Link) or wirelessly (e.g., Bluetooth or Wi-Fi to a PDA or other similar device).
Requirements association	None

Main Services

Service Name	Service Description
getSpectralMeasurement	Acquire a spectral measurement of the soil.

2.3.2.2 Processing device

Description, Interfaces and Requirements Association

Description	The processing device is usually purchased together with the spectrometer from the end-user. It contains either a microcontroller or a complete CPU and has all the necessary software and hardware components to connect to the sensor and trigger a spectral acquisition. Depending on the device and manufacturer, some processing of the data takes place here, e.g., converting the signal to reflectance, identifying highly outlying values, etc. This could, e.g., be the Edge Spot device from EGM, or a smartphone device.
Interfaces	• Similar to the hyperspectral sensors in order to interface with them. • [Optional] A wireless connection to the internet (e.g., 4G+, WiFi, etc.) to send the recorded and processed data to the cloud.
Requirements association	Hyperspectral sensor to connect to.

Main Services

Service Name	Service Description

triggerCalibration	Trigger a reference spectral measurement of a reference material and save it in memory.
triggerMeasurement	Trigger a spectral measurement of the soil and save it in memory.
transmitData	Transmit stored data to the cloud. This is optional, and data may be retrieved offline from the memory using standard I/O methods.

2.3.3 Soil hyperspectral processing module

2.3.3.1 Data retrieval

Description, Interfaces and Requirements Association

Description	Data used to generate soil thematic maps include: • Space-borne data (Sentinel-2, hyperspectral from KUVA) • Proprietary soil databases (stored in private userland) This module also processes point spectrometer data.
Interfaces	Python packages that use HTTP connections to query and retrieve satellite data from Copernicus / KUVA catalogues. Ability to upload data via the internet (e.g., HTTP, SFTP, etc.).
Requirements association	None

Main Services

Service Name	Service Description
getSatelliteImages	Given a time period and a specific area, download all satellite images for a given sensor (e.g., Sentinel-2).
getSpectrum	Retrieve a soil spectrum from a central database via HTTP request.
getBareSoilImage	[Maybe] If KUVA generates a L3 product of bare soil reflectance composite, then this service will allow us to automatically download them.
uploadProprietaryDataset	Ability to upload our own proprietary soil databases (e.g., CSV files) to the DEIMOS platform.

2.3.3.2 Bare soil identification and formulation of dataset

Description, Interfaces and Requirements Association

Description	Given a spectral signature, the bare soil identification module classifies the signature as bare soil or other (binary), allowing an apt model to only predict reflectance spectra of soil.
Interfaces	None
Requirements association	Hyperspectral data (images or point data).

Main Services

Service Name	**Service Description**
isBareSoil	Performs binary classification and indicates if this spectrum belongs to a bare soil or not.
generateComposite	Generate a composite map from a multi-temporal product (e.g., multi-temporal satellite images) and store it locally.
extractPoint	Given a composite map, this service returns the composite reflectance at a specific point in space.

2.3.3.3 Model training

Description, Interfaces and Requirements Association

Description	This block develops AI models (using e.g., federated AI) that predict soil properties from soil spectral signatures. The training takes place on DEIMOS platform.
Interfaces	None
Requirements association	Dataset generated from hyperspectral data.

Main Services

Service Name	**Service Description**
trainModel	Given a training dataset as formulated from the bare soil identification block, this process develops an AI model that can infer from the spectra the soil properties and stores it locally.

2.3.3.4

2.3.3.5 Map generation

Description, Interfaces and Requirements Association

Description	This block uses a pre-trained AI model to infer the soil properties from a point spectrum or a hyperspectral image.
Interfaces	HTTP interface to upload/store predictions in an external location.
Requirements association	Pre-trained AI model. New data to infer.

Main Services

Service Name	**Service Description**
predict	Given an input spectrum, this service loads the pre-trained AI model and predicts its soil properties.
upload	This optional service allows the generated predictions (point predictions or maps) to be uploaded to an external location (e.g., geoserver, private data repos, etc.).

2.3.4 Drought prediction from IoT and airborne sensor data

Description	The drought prediction module aims to produce two outputs (water management assessment and yield prediction) using IoT sensor data as a primary data source and airborne sensor data spatial distribution assessment.
Interfaces	Drought prediction module.
Requirements association	IoT sensor data and airborne sensor data.

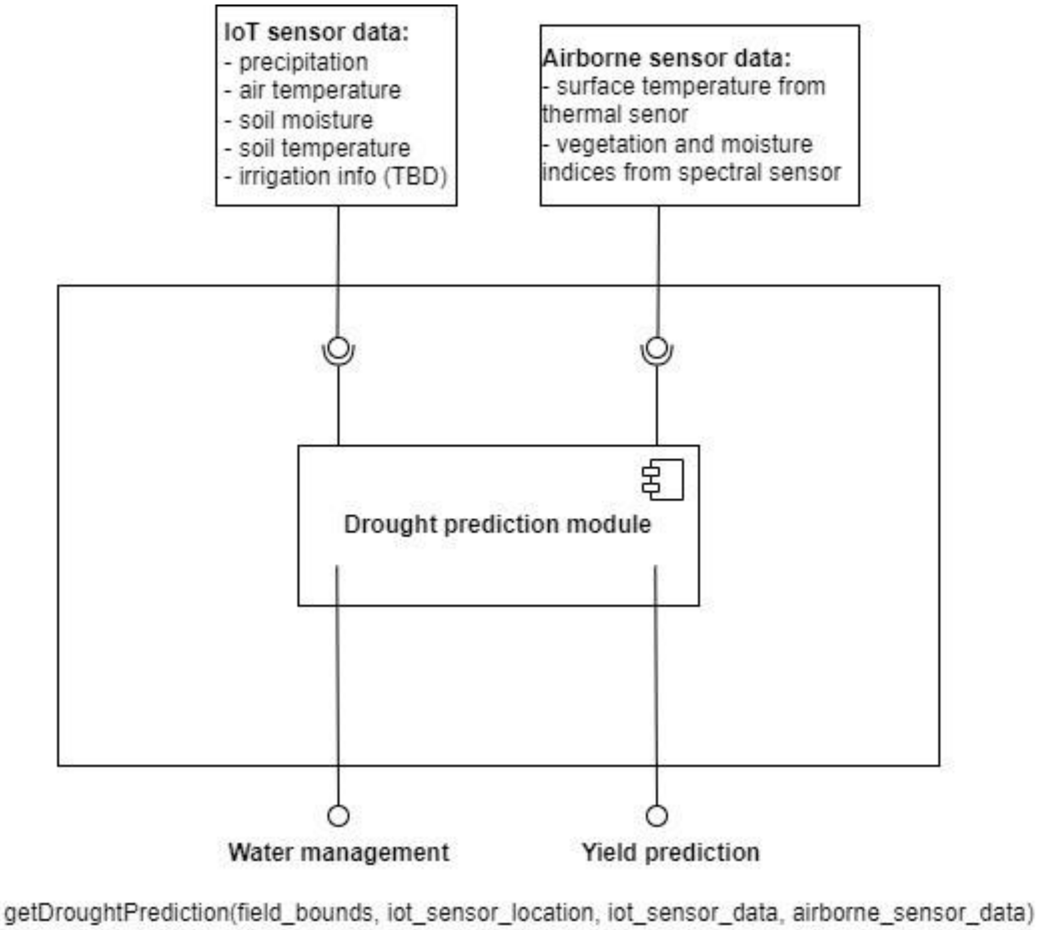


Figure 7 Functional view of drought prediction module for IoT and airborne sensor data.

A particular module could be applied to fields equipped with IoT sensors measuring local weather and soil conditions as well as irrigation regimes. IoT sensor data is updated once every 30 minutes and provided to the module in near real-time. Airborne data acquisition is performed on request and used for the update of parameter spatial distribution.

Service Name	Service Description
getDroughtPrediction	Input parameters: field_bounds, iot_sensor_location, iot_sensor_data, airborne_sensor_data Given input parameters, return water management assessment and yield prediction in the field of interest. Field bounds define the area of interest but IoT sensor location specifies the coordinates of the sensor within the field. IoT sensor data is provided as time series of several parameters: • Precipitation (local meteo station data) • Air temperature (local meteo station data) • Soil moisture (local soil sensor data) • Soil temperature (local soil sensor data) • Irrigation info (irrigation logger data) Airborne sensor data is provided in the form of multiband geospatial rasters and used for the update of output spatial distribution within the field of interest.

2.3.5 Drought prediction from satellite sensor data

Description	The drought prediction module aims to produce two outputs (water management assessment and yield prediction) using satellite sensor data.
Interfaces	Drought prediction module (satellite data-based).
Requirements association	Satellite data.

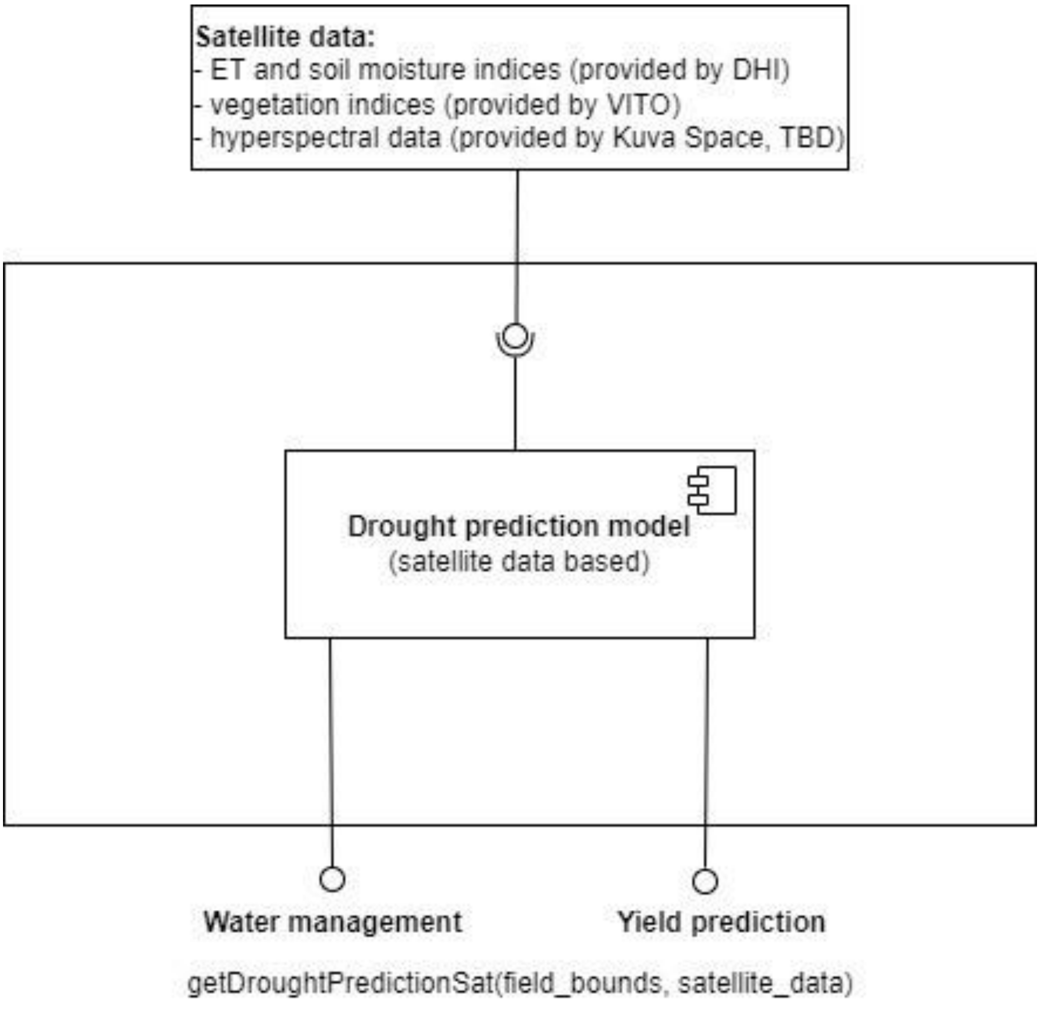


Figure 8 Functional view of drought prediction module for satellite sensor data.

A particular module upscales the local drought prediction module to a regional scale using satellite data as inputs. It is trained using available satellite data products and outcomes of the local drought prediction module.

Service Name	Service Description
getDroughtPrediction	Input parameters: field_bounds, satellite_data Given input parameters, return water management assessment and yield prediction in fields of interest Field bounds define fields of interest. The use of different satellite data products and data should be evaluated but the expected inputs are: • ET and soil moisture data products from DHI • Vegetation indices from VITO • Hyperspectral data from Kuva Space (TBD)

2.3.6 Vegetation indices calculator

Description	Processing API providing access to satellite products from any available data. The component is integrated with image provider to get set of spectral images in specific channels in best quality data to calculate selected vegetation index.
Interfaces	REST
Requirements association	Image provider

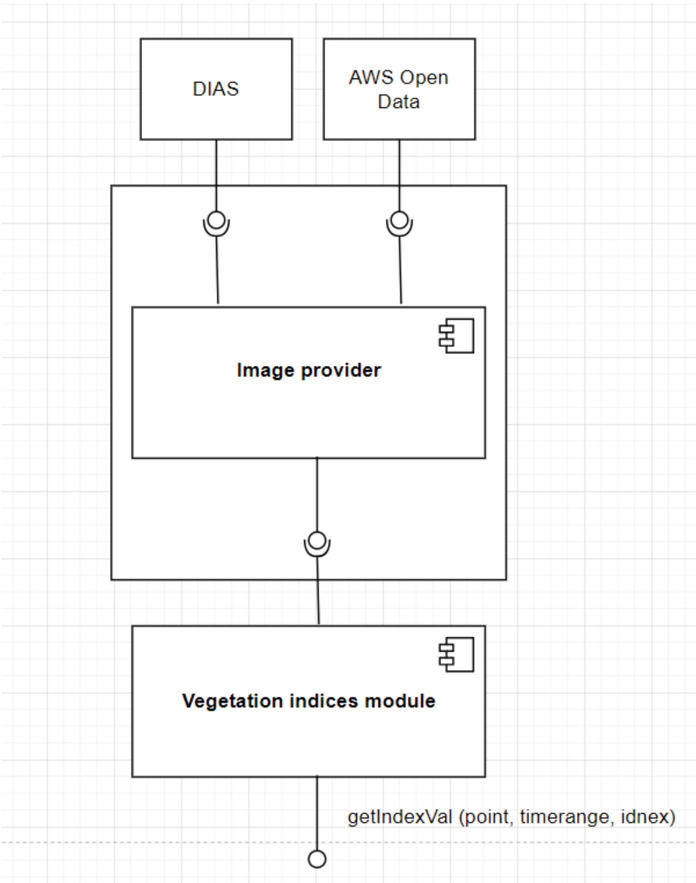


Figure 9 Functional view of DIAS data extraction tool.

The image provider endpoint is a RESTful API interface to various vegetation indexes results (current or calculated on archived data). It provides access to numeric results for single or list of points in defined spatial coordinate system. Service provides available results in time series for normal quality conditions for a given index.

Service Name	**Service Description**
getIndexValue	Input parameters: latitude, longitude, time range, index Given input parameters, return corresponding indexes values for the given point or set od points. Time range is being defined by standardized timestamps in UTC time. Point is defined by two decimal values in defined coordination system Service returns information about available time slots of preferred time range has insufficient data EvalScript enables the user to define additional processing on the returned data with the use of JavaScript scripts.

2.3.7 Agro-environmental policy indicators monitoring

Description	An API that provides calculated outcomes on agricultural policy performance indicators for a specific region. The calculation of indicators is based on aggregation algorithms on datasets collected from various farms in the area (IoT and farm level calendar of cultivation activities). Data assimilation and data fusion methods -utilizing satellite EO data sources- will be utilized for extracting data products reflecting farming activities on regional level. Data and privacy protection mechanisms will be applied in order to comply with ScaleAgData data governance schemes.
Interfaces	Agri-environmental Policy Performance Indicator provider.
Requirements association	• Connection to individual Farm Management Information Systems (FMIS) • Satellite EO data sources

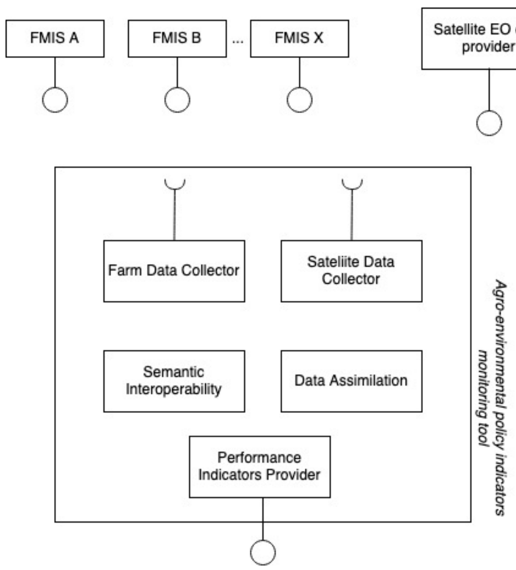


Figure 10 Functional view of Agro-environmental policy indicators monitoring tool.

Performance Indicators Provide

The Agro-environmental policy indicators monitoring tool provides a RESTFul API endpoint that allows service consumers to provide input properties, namely area (coordinates polygon), indicator type (selected from a predefined list of indicators) and time-period. The service will return the respective pre-calculated metrics.

Service Name	Service Description
performanceIndicator	Input parameters: geo_bounds, type_of _indicator Given input parameters, return calculated values and metadata for the requested indicator. Bounds - defined bounding box or geometry (Polygon / Multipolygon)- defines an area for which the retrieved performance indicators are calculated. The service periodically collects datasets from the various FMIS in the area. The datasets include selected parts of the digital farm-books where the various cultivation activities are recorded. Given that FMISs may not be extensively utilized in the area data assimilation mechanisms combined with EO Satellite data will allow to infer additional knowledge on performed farming activities. Given the ground-truth evidence on selected parcels the service then identifies parcels with similar properties in the area (e.g., same crop type, parcel area) and calculates agro-environmental indicators. The inference mechanism will also calculate a confidence level that will escort the calculated indicators denoting the correctness/accuracy of the calculated indicators. The main agri-environmental indicators to be supported are the following: • Cultivation types in the area • Use of digital technologies in the area • Use of pesticides (number and type of active substances, total quantity per active substance, number of interventions) • Use of irrigation (quantity per hectare, total quantity for the selected area, number of interventions) It should be noted that for specific cultivation practices and to ensure the validity of farmers calendar records additional evidenced will be utilized when available. For example, irrigation activities can be escorted by the respective soil moisture sensor recordings (e.g. an increase is expected on the same date/time with the irrigation event). In a similar manner a pesticides application event can be escorted by recordings by the AcID sensor. The service will return both: a) calculated indicators based only on the ground truth evidence (farm calendars) and b) calculated indicators based on data assimilation mechanisms along with the respective annotations.

2.3.8 DSS Precision Farming

HORTA's work plan includes the integration of new indices to the currently implemented library for precision farming services. To date, HORTA is evaluating indexes related to soil moisture, biomass, and yield prediction. The indexes will probably be provided by partners VITO and/or DHI. All the new indexes will be shown in the remote sensing view section of each crop unit. Moreover, the soil moisture indexes will study to improve the current water content balance model adopted in the HORTA's DSSs while the biomass and/or yield production indexes will add information for the yield prediction module and will be studied to improve it.

Indices are expected to be provided in geoTIFF (or similar) format relative to the crop unit polygon. The area concerned may consist of one or more polygonal. Usually, the service provider activates a batch procedure for each submitted polygonal and saves in a data repository the crop unit related geoTIFF file. HORTA's batch procedures query the provider file repository and download the geoTIFF, saving them in their own data structure.

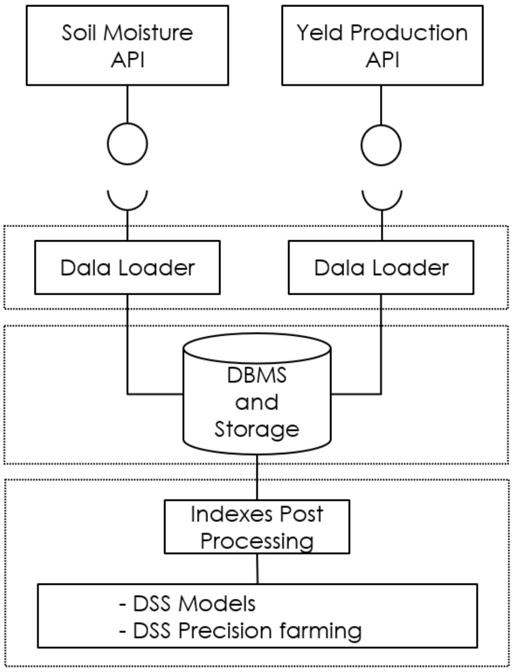


Figure 11 Functional view of the DSS Precision farming.

Soil moisture / Soil water content

Description	An index related to soil moisture or soil water content directly or through a function is required.
Interfaces	DHI APIs services
Requirements association	Are expected APIs that require the crop unit polygons and, if required by the index, the crop type or other attributes.

Biomass / Yield production

Description	An index related to the biomass or yield production directly or through a function is required.
Interfaces	VITO APIs services.
Requirements association	Are expected APIs that require the crop unit polygons and, if required by the index, the crop type or other attributes.

2.3.9 Digital twin providing forecasts and decision support

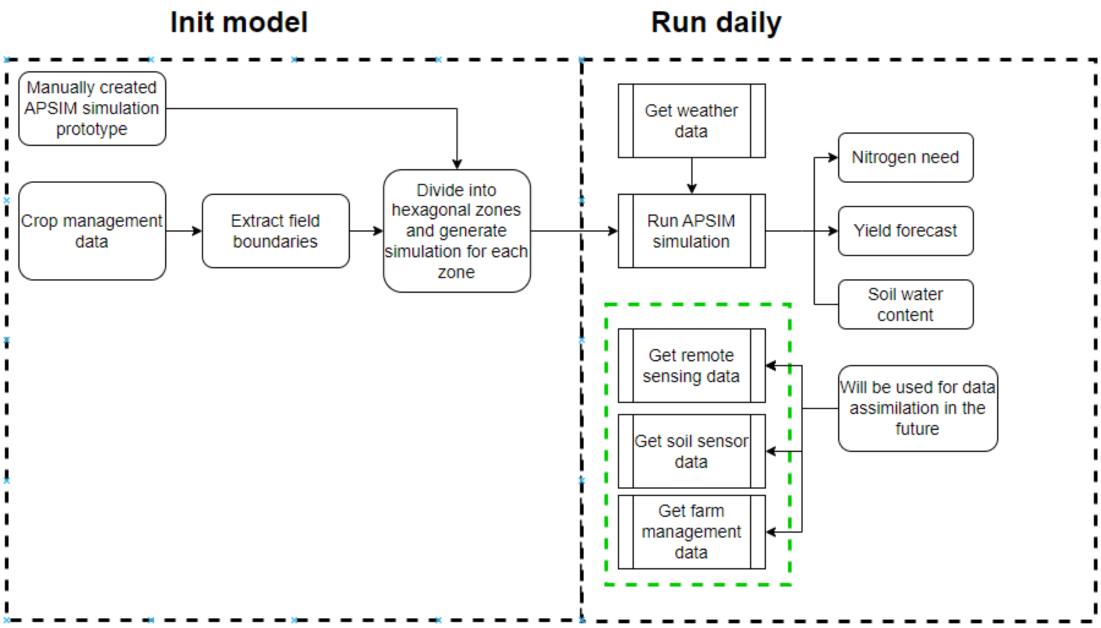


Figure 12 Functional view of Digital twin providing forecasts and decision support.

2.3.9.1 APSIM simulation model

Description, Interfaces and Requirements Association

Description	Model is implemented in C#. dotnet core version > 6.0 is required for running the model. The model provides forecasts and decision to support farming decisions. Separate simulation is run for each field zone. Simulation requires ISOBUS time log data or information from FMIS on field management and field boundaries. Soil properties, cultivar etc. should also be available from e.g. FMIS.
Interfaces	
Requirements association	Crop management data.

2.3.9.2 Python library

Description, Interfaces and Requirements Association

Description	Retrieves model input data for time and area of interest (daily weather; weather forecast, daily remote sensing data; daily soil sensor data; farm management data for each farming action done) etc. Retrieves input data for model calibration (remote sensing data layers, harvester yield data). Carries out simulation at one-day time step and provides estimates of nitrogen need, yield forecast and soil water content. Other forecasts can be added based on RIL needs. Will implement the decision support models developed in T4.1.
Interfaces	Forecasted yield, soil status and nitrogen status. Management recommendations. Currently written to MongoDB database, could be provided through an API e.g. using an NGSI-LD context broker.
Requirements association	APSIM simulation model. Connections to relevant data sources: weather, EO, soil and farm management data.

Main Services

Service Name	Service Description
Get weather data	Retrieves daily weather data and weather forecast (if available).
Get EO data	Retrieves spatial EO data layers for time and area of interest (reflectance, vegetation indices, leaf area index, evapotranspiration estimate, phenological development stages etc.).
Get soil sensor data	Retrieves daily soil sensor data if available from RIL.
Get farm management data	Retrieves farm management data from FMIS or task files from precision agriculture/farm machinery.
Run APSIM simulation	Produces daily estimates of nitrogen need, yield forecast and soil water content.

2.3.10 Reporting module

Description	Automatically generated reports considering the following: • Crop stress (early stress) • NDVI anomalies • Risk events on crops (visible and NDVI) • Recorded temperature anomalies • Recorded soil moisture anomalies • Recorded strong winds anomalies
Interfaces	PDF report generation. Automated mailout to assigned recipients.
Requirements association	Data composition and interpretation, overlaying the various physical and weather parameters to identify crop anomalies and stress.

Main Services

Service Name	**Service Description**
REPORTING	Automatic generation and mailout of report on detection crops stress and vegetation anomalies overlayed with weather data, other sensor data and EO data (optical, NDVI, SAR).

2.3.11 Milk quality and quantity forecaster

The main objective of the dairy lab is to investigate potentials for improving the forecasting of data concerning milk quality and quantity, which is relevant for the production facilities of the involved dairy cooperative. If this will be feasible, an architecture like the one described in this document may be implemented in a later project phase.
An overview of the envisaged architecture is given in the following Figure 13.

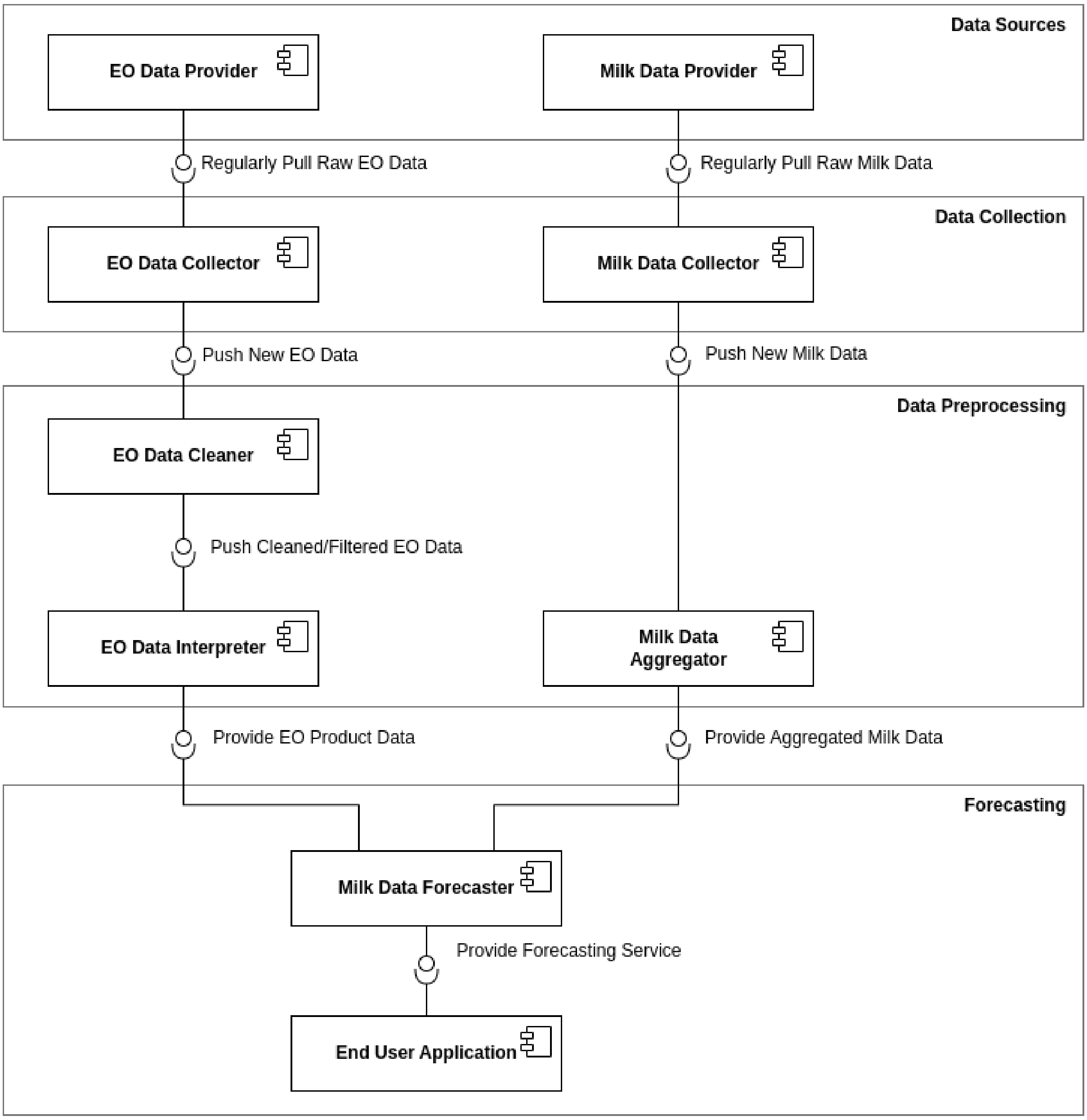


Figure 13 Functional view of Milk quality and quantity forecaster.

Each of the elements from the Figure 13 is further described in the following, grouped by the main components of Data Sources, Data Collection, Data Preprocessing, and Forecasting.

2.3.11.1 Data Sources

EO Data Provider

Description, Interfaces and Requirements Association

Description	Existing public source of satellite image data, e.g. Sentinel Hub.
Interfaces	Provides different APIs to get satellite data, e.g. HTTP/REST API.
Requirements association	Satellite data.

Main Services

Service Name	Service Description
getRawEOData	Provides raw EO data for a given satellite, band, area of interest and temporal period of interest.

Milk Data Provider

Description, Interfaces and Requirements Association

Description	Service that provides data about quality and quantity of milk that was delivered to the dairy processor company.
Interfaces	To be defined, such a service is not yet existing.
Requirements association	Milk data.

Main Services

Service Name	Service Description
getMilkData	For a given geographical area and timeframe, provides data about milk quality (e.g. percentage of fat/protein, based on laboratory checks of the dairy processor company) and milk quantity for all milk deliveries to the dairy processor company.

2.3.11.2 Data Collection

EO Data Collector

Description, Interfaces and Requirements Association

Description	Regularly (based on defined time intervals) pulls new EO data from the EO Data Provider service and pushes that data into the data preprocessing pipeline.
Interfaces	Client to read from EO Data Provider service, client to push data into EO Data Cleaner.
Requirements association	EO data.

Main Services

Service Name	Service Description
updateDataCollectionConfig	Allows to update the configuration for EO data collection, e.g. setting the time interval, the provider API URL, authentication credentials, the satellite bands and the geographical area of interest for which to collect satellite data, etc.

Milk Data Collector

Description, Interfaces and Requirements Association

Description	Regularly, based on a defined time interval (e.g. daily), pulls new milk data from the Milk Data Provider service and pushes that data into the Milk Data Aggregator component.
Interfaces	Client to pull data from Milk Data Provider service, client to push data into Milk Data Aggregator.

Requirements association	Milk data.

Main Services

Service Name	Service Description
updateDataCollectionConfig	Allows to update the configuration for milk data collection, e.g. setting the time interval, the provider service address/URL, authentication credentials, etc.

2.3.11.3 Data Preprocessing

EO Data Cleaner

Description, Interfaces and Requirements Association

Description	Filters/cleans the raw EO data, so that only relevant data are processed further.
Interfaces	Service receiving raw EO data, client to push cleaned data into the EO Data Interpreter.
Requirements association	EO data.

Main Services

Service Name	Service Description
cleanRawEOData	Receives raw satellite image data and filters/cleans it according to predefined rules, e.g. drops image data that is not usable, which could be due to clouds covering the ground, making the actual grassland invisible.

EO Data Interpreter

Description, Interfaces and Requirements Association

Description	Takes cleaned satellite data as input and derives different EO data products.
Interfaces	Service receiving input data, Service providing EO product data.
Requirements association	EO data.

Main Services

Service Name	Service Description
interpretEOData	Receives cleaned/filtered satellite image data and processes it further to derive different EO data products. This may be an index calculation based on formulas (e.g. NDVI or NDMI) or more advanced processing based on ML/AI (e.g. biomass growth rates). Please note that the concrete EO data products that are relevant for the RIL Dairy are still to be defined, depending on the results of the ongoing correlation analysis between historical EO data and historical milk data.

getEOProductData	Returns EO product data (e.g. NDVI, NDMI) for a given geographical area and timeframe.

Milk Data Aggregator

Description, Interfaces and Requirements Association

Description	Aggregates raw milk data over geographical areas.
Interfaces	Service to receive raw data, service to provide aggregated data.
Requirements association	Milk data.

Main Services

Service Name	Service Description
aggregateMilkData	Receives raw milk quality & quantity data and aggregates it over geographical areas (e.g. calculating the average quality & quantity per group of farms or per administrative district). The geographical areas need to be defined according to the requirements by the dairy processor, e.g. the area of interest per processing plant would include all farms that deliver to this processing plant.
getAggregatedMilkData	Returns aggregated milk quality (e.g. average percentage of fat/protein) & quantity data for a given geographical area and timeframe.

2.3.11.4 Modelling and Forecasting

Milk Data Forecaster

Description, Interfaces and Requirements Association

Description	Service that provides milk quality & quantity forecasts, based on latest EO data and milk data. A detailed definition of the required functionality of this service is planned for the second two years' cycle of the project.
Interfaces	Reads latest EO data, reads latest milk data, provides a service to get forecasts of milk quality & quantity.
Requirements association	Eo and milk data.

Main Services

Service Name	Service Description
getMilkDataForecast	Provides a milk quality (e.g. expected average percentage of fat/protein) & quantity forecast for a given geographical area.

2.3.11.5 End User Application

Description, Interfaces and Requirements Association

Description	End-user applications that provide decision support to actors in the dairy processor company, based on forecasts of milk quality and quantity. Please note that the functionality of such an end-user application is still to be defined. This will be subject of the second two years' cycle of ScaleAgData in the dairy lab, provided that the ongoing correlation analysis between historical EO data and historical milk data will be successful.
Interfaces	Reads latest forecasts of milk quality & quantity, provides GUI for end users.
Requirements association	EO and milk data.

2.3.12 Grasslands improved biopars (LAI, fPAR)

2.3.12.1 Data fusion (radar-optical)

Description, Interfaces and Requirements Association

Description	The frequent cloud coverage in central Europe affects the biophysical parameters (biopars) derived from optical remote sensing data, thus causing gaps in the time series. To improve the time series of the grassland biopars, hence fill the cloud-induced gaps, data fusion of different sensors is used. SAR data from Sentinel-1 (S1) can penetrate clouds and is used in a data fusion approach with the biopars Leaf Area Index (LAI) and/or Fraction of Photosynthetically Active Radiation (fPAR) from Sentinel-2 (S2). The S2 biopars are derived using the SNAP Biophysical Processor. S1 preprocessing steps are evaluated, determining if sigma- or gamma-nought preprocessing is suited better for the study area. The data-fusion is done using time-series features of S1, soil moisture content data and precipitation data in non-linear machine learning algorithms. The regressors examined, such as Gaussian Process Regression (GPR) and Random Forest Regressor (RF), are trained using cross-validation and a feature selection is performed to optimize the computation time and prediction accuracy. The predicted biopars are finally fused into the non-gapfilled S2 ones and validated using in-situ measurements over forage-production meadows in the region Trentino-South Tyrol. The accuracy metrics include R^2 (coefficient of determination) and root-mean-squared error (RMSE).
Interfaces	Python, R, SNAP, openEO, GIS Software.
Requirements association	Sentinel-2 LAI and/or fPAR time series, Sentinel-1 SAR data, daily precipitation data, soil moisture content data.

Main Services

Service Name	**Service Description**
S1 LAI model	Machine learning model estimating LAI from S1 derived input features over Alpine grasslands.

Timeseries improvement module	Codes to generate daily timeseries of LAI at parcel scale over Alpine grasslands.

2.3.13 Grasslands primary production

2.3.13.1 Machine Learning integration

Description, Interfaces and Requirements Association

Description	This task outlines a procedure for estimating Gross Primary Productivity (GPP) through the integration of eddy covariance data, Sentinel-2, and Sentinel-1 using machine learning algorithms. The methodology involves comprehensive data preparation, including cleaning and preprocessing of eddy covariance data, normalization and feature extraction from Sentinel-2 and Sentinel-1 data. Two distinct model structures, a Single-layer Artificial Neural Network (ANN) and a Multi-layer Perceptron (MLP), are considered based on data availability. The data undergoes temporal alignment before being split into training and testing sets. Model selection, architecture definition, and subsequent training with appropriate loss functions and optimizers will be considered. Evaluation metrics include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared gauge model performance. The study emphasizes the importance of conditional model structures based on the availability of Sentinel-2 and Sentinel-1 data. Implementation involves utilizing machine learning libraries and visualizing predictions against actual GPP values. The procedure is documented for transparency and replicability, with room for adjustments based on specific data characteristics and desired accuracy levels. The algorithm will be also compared with in in situ measurements of biomass.
Interfaces	Python, GIS Software.
Requirements association	Temporal series of Sentinel-1 and Sentinel-2, Eddy-covariance daily measurements, Weather measurements.

Main Services

Service Name	Service Description
Periodic GPP map estimator from S2	Based on a trained ANN, this service will provide map of GPP based on Sentinel-2 estimation.
Periodic GPP map estimator from a S1 and S2 data fusion	This service will be an update of the previous one introducing the possibility to integrate the lack of data due to S2 limitations with S1.

2.3.13.2 Biophysical integration

Description, Interfaces and Requirements Association

Description	Grasslands net primary productivity (NPP) will be derived following the theory that plant NPP is directly related to absorbed solar energy, that a connection exists between absorbed solar energy and satellite-derived vegetation spectral indices (VI), and the assumption that there are biophysical reasons why the actual conversion efficiency of absorbed solar energy (LUE) may be reduced below the theoretical potential value.
Interfaces	N/A
Requirements association	Sentinel-2 NDVI series, daily weather data, maximum LUE and environmental limits derived for grasslands.

Main Services

Service Name	Service Description
get_EO_VI	Input parameters: selected region/field, period of time, cloud cover percentage. Given input parameters, return the corresponding satellite time series for the given field and compute the vegetation index.
gapfill_day_interpol	Input parameters: VI times series. Given input parameters, return a daily VI time series with gap-filled cloudy areas.
woody_veg_contribution	Input parameters: daily VI time series. Given input parameters, return an annual time series of woody vegetation contributions to the vegetation index signal for each pixel.
LUE_NPP	Input parameters: the period of time and area of interest, time series of gap filled daily VI and annual woody contribution images, weather data including daily values of global solar radiation (Rd, MJ), minimum temperature (Tmin, °C), relative humidity (HR, %), vapor pressure (VP, kPa), and vapor pressure deficit (VPD, kPa), and max_efficiency parameter for each component of the scene. The function returns the time series of daily and accumulated biomass produced by the grassland, based on the given input parameters.

2.3.14 Edge processing component

The detailed Edge Spot specifications are in I_APN_ESPT_04_2.0 - EdgeSpot Specification.pdf.
As a summary, it is a versatile and extensible IoT box, that can connect to any sensor or actuator and can send data to any network. This is possible because of its 3 mikroBUS™ slot which can receive extension cards. More than 1000 of them are available and it is quite easy to design some others, depending on the needs of the data collection. It also has a very adaptable power management system: it can harvest energy from solar panels and manage the load of a battery, but it can also, in case of deployments with scarce energy sources, put itself in very low power mode, consuming a few

tens of microamps, while still being able to react to external stimulations or waking up regularly to take some measurements.
This box is built around a STM32L4 microcontroller, which has enough memory and processing power to embed small AI models (TinyML) or complex applications.
The Edge Spot is thus more a *far edge sensing and computing platform* rather than a simple IoT box. These edge computing capabilities can be leveraged for different usages, depending on the sensor type:

- Data quality assessment, outliers' detection
- Data quality improvement, interpolation of missing points
- Data compression (basically: send only measurements if they have changed by a certain amount)
- Early warning with unsupervised learning
- Data interpretations of complex sensors (like spectrometry, images, sounds etc.): use an AI model to infer some characteristic of the sensed (plant disease, soil characteristics etc.). Send the result of this inference rather than the raw data. This is particularly important in a data massification approach while maintaining a reasonable network bandwidth.
- In a same way, edge processing can be used for sensor fusion: use incomplete but overlapping information coming from various sensors to infer the value of a parameter of interest. This can be a qualitative result like a classification (type of smell from different gas sensors for example, or water quality)
- Edge computing can also be used to train an AI model in common with others but without sharing data. This is called Federated Learning and can be of interest when data cannot be shared among parties (for privacy or security reasons), but a shared AI model is wanted. In that case, the parties use a protocol to exchange some data about the model being trained but not the data that are used to train the model.

At this stage, Edge Computing has been considered for the Soil RIL, especially to embed a model to interpret the raw data coming from a single pixel spectrometer (VTT).
The plan is to start building a versatile sensing and edge computing platform for agriculture, with functionalities like sensors *plug and play*, dynamic edge computing capabilities (i.e., with means to upload processing algorithms or AI models from the cloud), mesh networking etc. In a first step, some simple sensors will be usable (like soil moisture, temperature, or redox potential for example). Then, it will be easy to integrate the VTT sensor and the associated AI models, to get additional data.

2.4 Research and Innovation Environment (RIE)

The RIE made available within ScaleAgData is based on the Virtual Lab solution which is part of the Data Exploitation Platform developed by Deimos in the last years to: a) support EO service providers to build, deploy and operationalize their algorithms/applications; b) provide users with user friendly interfaces to access those applications.

The services4EO Virtual Lab provides an interactive development environment that allows the users to develop algorithms online coded in Python and execute them over discovered data. This functionality can be accessed via Jupyter Notebooks with a subset of SDKs that helps the user during the data discovery, data access, data visualization and data execution operations.

Once the development is finished it is possible to deploy the new/updated script so that it is available to be triggered by the rest of users through the ScaleAgData RIE as a standard toolbox.

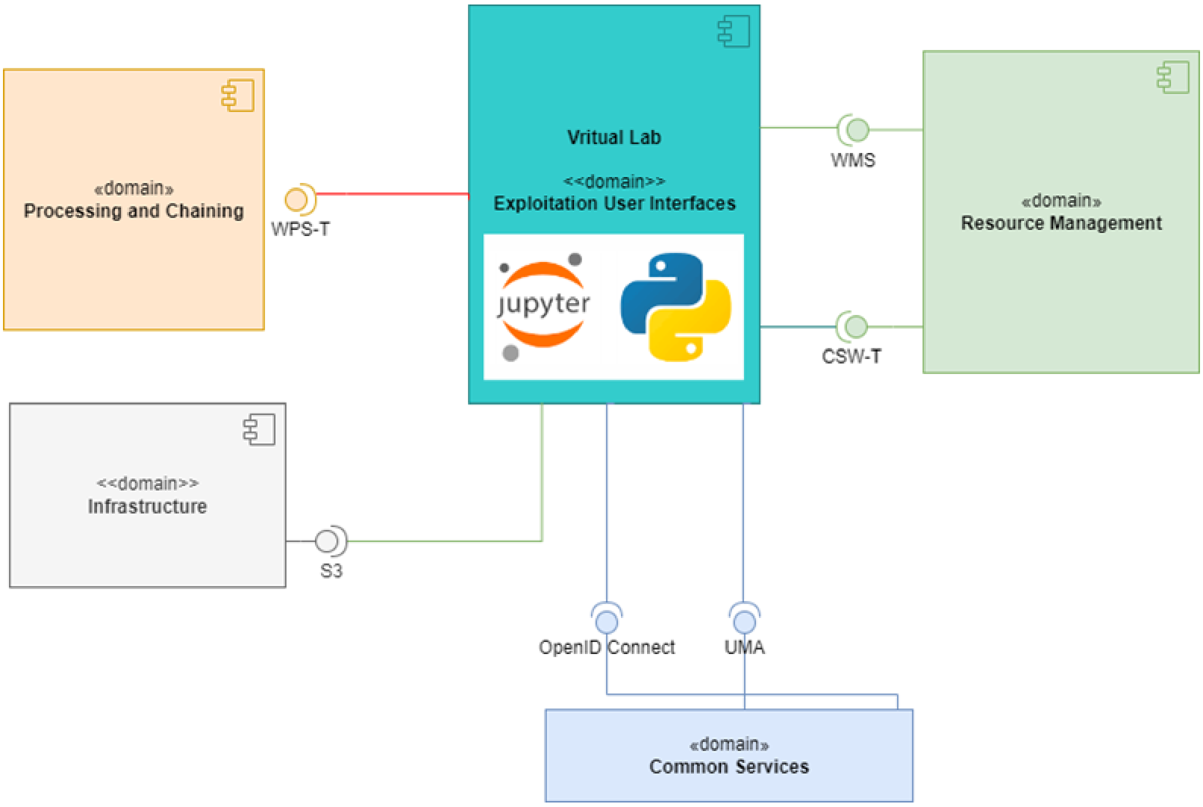


Figure 14 RI Environment.

The Virtual Lab exposes a Jupyter Notebook interface and consumes the following main interfaces provided by the services4EO Platform:

- Open ID Connect (OIDC) and User Managed Access (UMA) for **Identity and Access Management** functionalities provided by the Common Services Domain of services4EO.

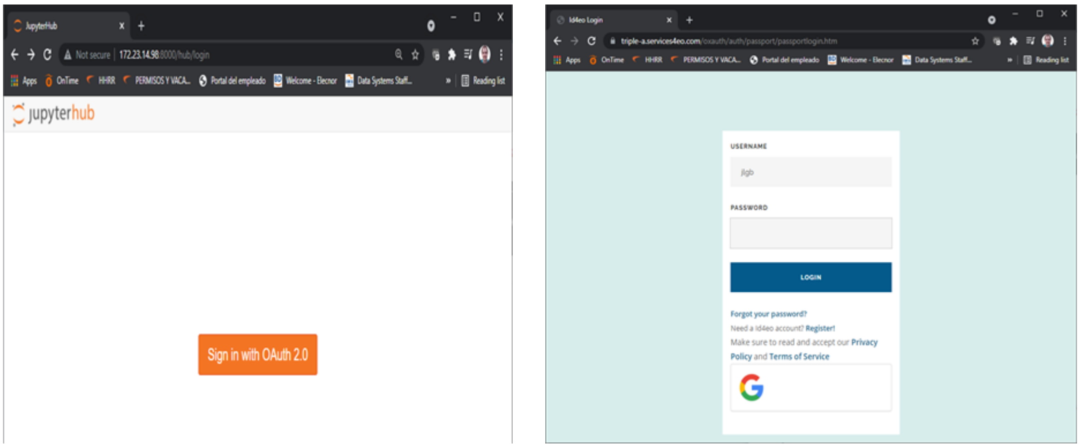


Figure 15 Identity and Access Management.

- OGC Catalogue Services for the Web (CSW) for **Data Discovery** provided by the Resource Management Domain of services4EO.

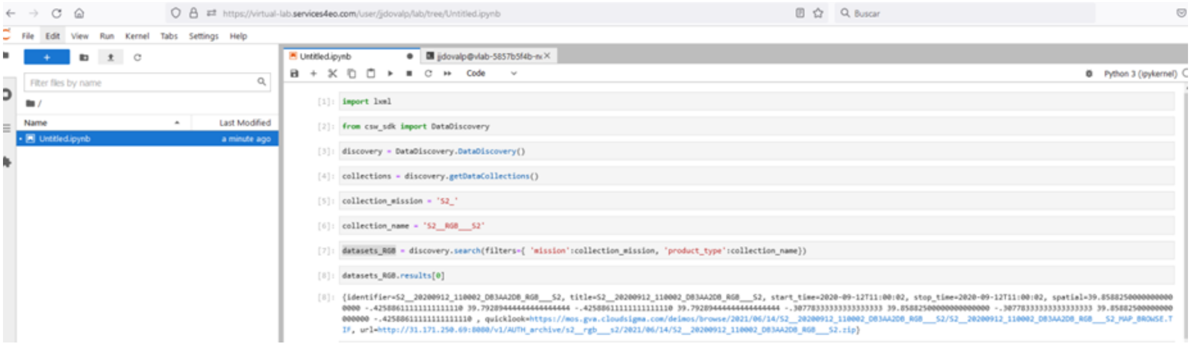


Figure 16 Data Discovery.

- OGC Web Map Services (WMS) for **Data Visualization** provided by the Resource Management Domain of services4EO.

```
array_list=[]
for image in images:
    img=io.imread(image)
    array_list.append(skimage.util.img_as_ubyte(img))
```

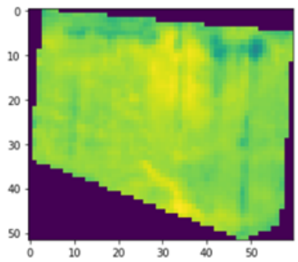
```
array_out=np.mean(array_list, axis=0)
```

```
np.mean(array_list)
```

```
60.988247863247864
```

```
plt.imshow(array_out)
```

```
<matplotlib.image.AxesImage at 0x7fc13a1718d0>
```




```
io.imsave(os.path.join(outputs_dir,"test.tif"), arr=array_out)
```

Figure 17 Data Visualization.

- Simple Storage Service (S3) for **Data Access and Download** provided by the Infrastructure Domain of services4EO.

Download products and store them in a folder inside virtual lab

```
for product in subset:
    url = product.url
    product_download_path = os.path.join(os.path.join(download_dir, url.split("/")[-1]))
    print(product_download_path)

    # check if not already downloaded
    if not os.path.isfile(product_download_path):
        print("Downloading from Archive product " + product.title + '... ')
        urllib.request.urlretrieve(product.url, product_download_path)

    else:
        print('Product: '+ product.title + ' already downloaded in path: ' + product_download_path)

print("Download has finished")
```

```
./Downloads/zips/S2__20220430_000000_FE1D9F21_VITONDVI
Downloading from Archive product S2__20220430_000000_FE1D9F21_VITONDVI...
./Downloads/zips/S2__20220515_000000_1CBC5008_VITONDVI
Downloading from Archive product S2__20220515_000000_1CBC5008_VITONDVI...
./Downloads/zips/S2__20220510_000000_70955634_VITONDVI
Downloading from Archive product S2__20220510_000000_70955634_VITONDVI...
Download has finished
```

Figure 18 Data Access and Download.

The Virtual Lab can be extended for generating an SDK that allows the users to create online an Application Package (composed of the algorithm artefact and an Application Metadata file) that can be deployed as a data processing service and executed, monitored, and controlled via OGC Web Processing Service (WPS) provided Processing and Chaining Domain of services4EO.

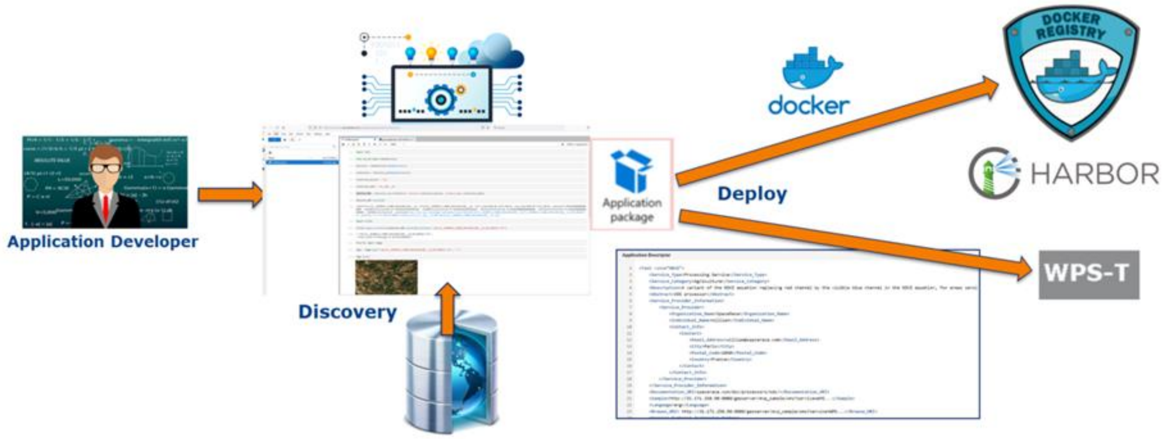


Figure 19 Data Exploitation and Marketplace interfaces.

Those processing services could be also consumed from end users via Data Exploitation or Marketplace interfaces.

More information on how to use the RIE is present in deliverable “D4.1 The Research and Innovation Environment”.

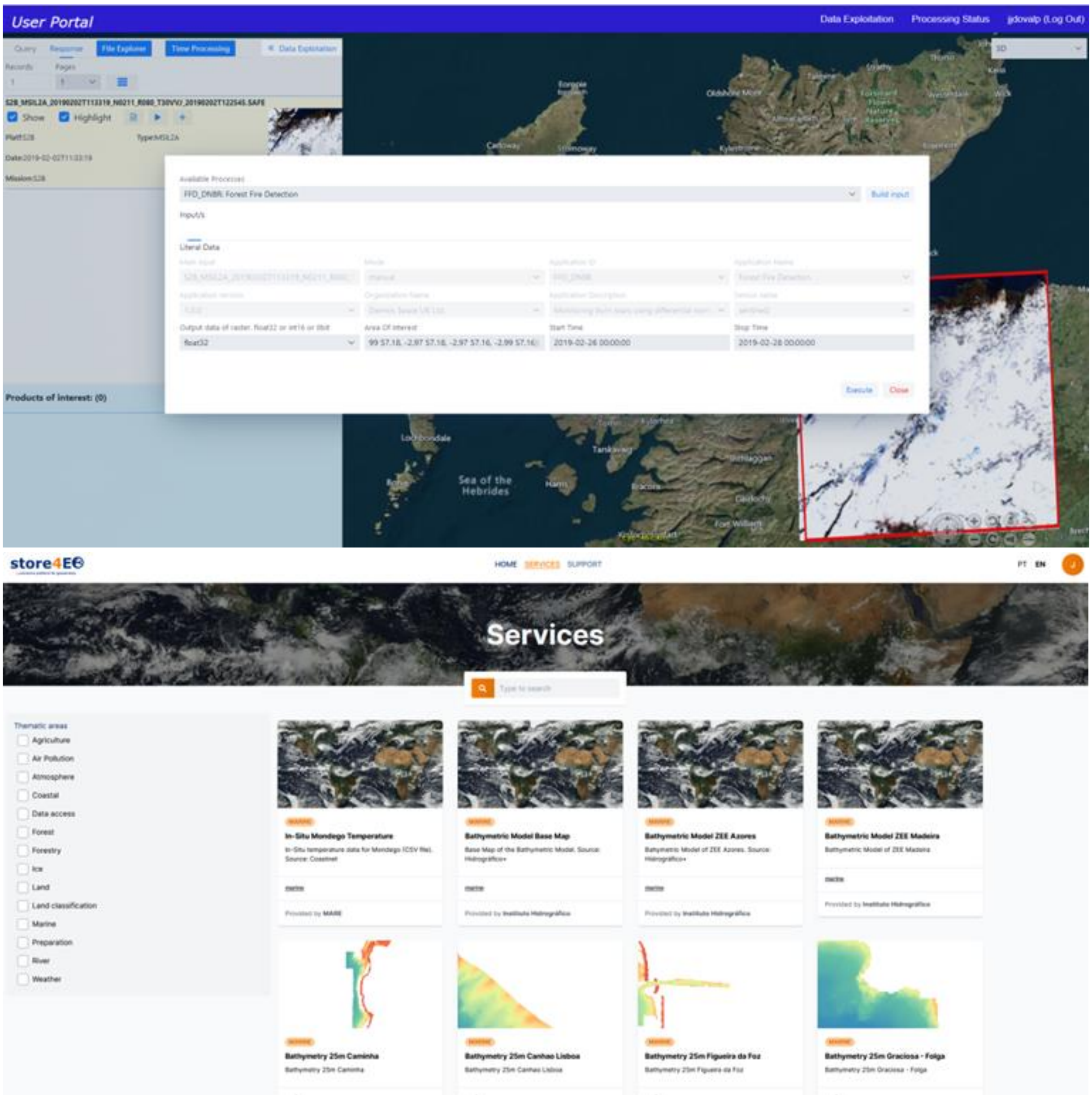


Figure 20 User Portal.

2.5 Data Interoperability and Data Governance

As many different systems become integrated, IoT ecosystems face complex interoperability challenges before creating real datasets and services at higher levels. This is also the case with ScaleAgData where in the various RILs heterogeneous technologies are utilized generating different types of datasets modeled with the use of various formats. Semantic technologies[5] are key in the development of IoT when dealing with, e.g., interoperability, governance, thing's heterogeneity, discovery of devices and services, discovery and management of knowledge, etc. In the recent research and standardization areas there are various IoT architectures (e.g. ETSI ISG-CIM, oneM2M),

[5] Brewster, Christopher et al. 'Data Sharing in Agricultural Supply Chains: Using Semantics to Enable Sustainable Food Systems'. 1 Jan. 2023 : 1 – 31.

open source communities (e.g. FIWARE), and IT vendors (e.g. CISCO IoT) that aim to develop solutions that could be broadly used in the agri sector. However, various research outcomes have shown how difficult it is to avoid the “silo effect”[6] and have a common architecture. On the other hand, there are already significant efforts and outcomes towards a data-driven and interoperable ecosystem for agricultural data.

ScaleAgData will address data interoperability challenges through the reuse and adaptation of existing best practices in the use of semantic technologies aiming to enable data harmonization on semantic level. Standardized data models will be reused as much possible and where necessary will be extended accordingly.

With regards to data interoperability the overall objective is to provide the required data translation mechanisms that will allow to harmonize data derived from RILs and make them available for further processing within the RIE. As it is illustrated in Figure 1 “High-level architecture of the ScaleAgData system” this will be realized with the use of the “Data Transformation” components. These components will act as data translators that will get as input data generated by the RIL, translate them to a common data model and provide them for further analysis.

This section initially performs a short analysis on the challenge of data interoperability for the agri sector and then elaborates on existing best practices on data harmonization particularly focusing on the use of ontologies. As it will be analyzed, the DEMETER Agriculture Information Model -developed by the H2020-DEMETER project- is selected as the most appropriate data model for reuse. This section also provides examples on how the primary data forms -collected by the technologies deployed in RILs- will be translated with the use of AIM semantics. The section closes with a short analysis on the key data governance principles that will dictate the use of ScaleAgData’s RIL datasets.

2.5.1 Data Interoperability

There have been various efforts for defining a methodology to address the interoperability challenge in IoT. In the related literature, there exist various classifications of the interoperability aspects, which are also called levels of interoperability. For example, the European Interoperability Framework designed “to support the delivery of pan-European eGovernment services to citizens and enterprises” [1] defines the following three levels: technical, semantic, and organizational interoperability. In the report presented in [2] and illustrated in Figure 21 an IoT-specific classification split in four levels is presented:

- Technical Interoperability: usually associated with communication protocols and the infrastructure needed for those protocols to operate.
- Syntactic Interoperability: usually associated with data formats and encodings, e.g., XML, JSON and RDF.
- Semantic Interoperability: associated with a common understanding of the underlying meaning of the exchanged content (information).
- Organizational Interoperability: associated with the ability of organizations to effectively communicate and transfer information even across different information systems, infrastructures or geographic regions and cultures.

[6] C. Brewster, I. Roussaki, N. Kalatzis, K. Doolin and K. Ellis, "IoT in Agriculture: Designing a Europe-Wide Large-Scale Pilot," in IEEE Communications Magazine, vol. 55, no. 9, pp. 26-33, Sept. 2017, doi: 10.1109/MCOM.2017.1600528.

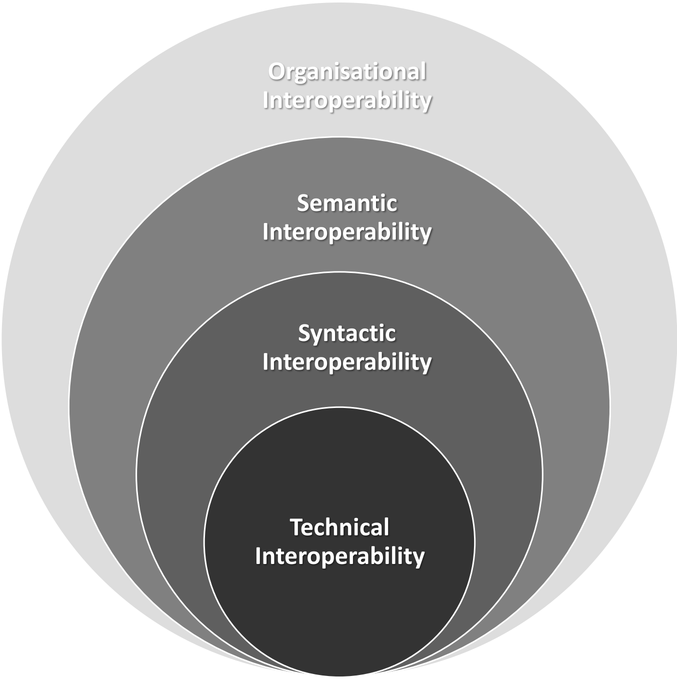


Figure 21 The dimensions of interoperability [2].

Among the most challenging tasks -but with significant outcomes by EC funded projects the recent years- is the issue of semantic interoperability. There is a strong move towards the use of semantic standards and technologies (ontologies) where the data and data models are explicitly specified, where URIs are widely used, and where data integration is consequently made far easier. A short review of the most important data models focusing in the agriculture sector follows. The main work of this state-of-the-art review was realized in the context of the recently completed Innovation Action H2020-Ploutos (D4.2 - A Sustainable Innovation Framework to rebalance agri-food value chains) [3]. However, the conducted analysis is extended with additional outcomes relevant with the ScaleAgData application domains and objectives.

2.5.2 Ontologies

Since the development of ontologies as tools for knowledge representation, in the late 90s, early 00s, there have been a succession of attempts to create formal ontologies expressed in OWL. More recently a series of more specialized ontology or ontology frameworks) have been developed partly under the influence of parallel activities in the Life Sciences/Bioinformatics.

ENVO: The Environment Ontology covers environmental processes, anthropogenic environments, and entities relevant to environmental health initiatives and the global Sustainable Development Agenda for 2030. The ENVO has approximately 6500 concepts and has a large proportion of terms are relevant to the agrifood sector and the ontology has been quite successful in being widely adopted by different projects and its terms reused in other ontologies. It is maintained as a community effort largely from the bioinformatics community.

AGRO: The Agronomy Ontology, follows principles concerning formal ontology design and its focus is the agronomy domain, providing terms to describe agricultural practices, cropping systems, field management, soil, weather and crop phenotypes, building on a number of existing ontologies including ENVO and PATO[7]. The ontology was built and is maintained by the CGIAR network of

[7] https://www.ebi.ac.uk/ols/ontologies/pato

research centers, and forms a core part of the Agronomy Field Information Management System, a CGIAR system to capture field books describing crop management practices[8], as well the GARDIAN data management system[9]. It is a relatively small ontology with just over 2000 classes and is extensively used in a number of crop specific ontologies built as part of the "Crop Ontology Curation Tool" initiative[10].

SAREF4AGRI: The SAREF ontology (the Smart Appliance REFerence ontology) aims to provide a standardized reference data model for the smart appliances domain which is seen as part of the wider Internet of Things ecosystem. The main objective here was to reduce the mappings needed between different types of appliances in the creation of smart homes and other IoT integrated systems. It is a standard sponsored by ETSI[11] and as such has considerable buy-in from the wider IoT and telecommunications industry. SAREF is designed to be modular and has mappings to the widely used W3C SSN ontology[12] designed for sensor data. As part of this modularity, an extension for agriculture has been created – SAREF4AGRI with the intention, as noted in the documentation "to connect SAREF with existing ontologies (such as W3C SSN, W3C SOSA, GeoSPARQL, etc.) and important standardization initiatives and ontologies in the Smart Agriculture and Food Chain domain, including ICAR for livestock data, AEF for agricultural equipment, Plant Ontology Consortium for plants, AgGateway for IT support for arable farming" (ETSI, 2020) [4]. The initial focus of SAREF4AGRI has been livestock farming and irrigation where IoT sensors both play a significant role in generating data. SAREF and SAREF4AGRI are relatively light-weight ontologies mostly used for data sharing (via APIs) rather than formal inference scenarios. The backing of ETSI ensures the longevity and likely maintenance of SAREF4AGRI while the growing role to telecommunications in precision agriculture motivates its further use and adoption. https://saref.etsi.org/saref4agri/v1.1.2/

FOODIE: The FOODIE ontology was built as part of the FOOD EC funded project[13]. This ontology was based on the INSPIRE data standard (which is in XML) for geospatial data, and was "lifted" from the UML version of this standard into RDF/OWL. The main characteristic of the FOODIE ontology is its specialization of the INSPIRE concepts allowing a variety of attributes concerning geospatial attributes, types of agricultural activity and crops at these levels of granularity. The main challenge in reusing this ontology is that it does not seem to reuse concepts from other agrifood ontologies although claims to map (or plan to map) to AGROVOC. In addition, it does not appear to be under development or maintenance. However, the Geospatial focus of this ontology is an important consideration that is sometimes explicitly lacking or not prioritized in other vocabularies.

Demeter AIM: The Demeter project[14] developed an "Agricultural Information Model" (AIM) partly based on the FOODIE ontology. The project is focused on data sharing for precision agriculture and smart farming, so this is the prime target of application. A key difference from other standards mentioned here is that AIM is based on a "property graph" model rather than the more usual RDF/OWL approach. The core foundation is the NGSI-LD Information Model developed by the telecommunication sector (ETSI) and based on the FIWARE NGSI information model. This means a set of cross domain concepts, combined with domain specific modules for various aspects. The main vocabularies reused are NGSI-LD, FOODIE and SAREF4Agri. The proposed set of ontologies are quite comprehensive but standard apart from much of the rest of the sector's activities in standardization.

8 https://agrofims.org/about
9 https://bigdata.cgiar.org/resources/gardian/
10 http://www.cropontology.org/
11 https://www.etsi.org/
12 https://www.w3.org/TR/vocab-ssn/
13 https://foodie-cloud.org
14 https://h2020-demeter.eu/

In compensation, the Demeter AIM does propose a set of very thorough mappings to a variety of ontologies including FOODON, SAREF4AGRI and AGROVOC. The AIM is available here: https://agroportal.lirmm.fr/ontologies/DEMETER-AIM/

Ploutos: The H2020-Ploutos Innovation Action specified the "Ploutos Core Semantic Model" which defines its own Ploutos namespace for the common concepts but reuses existing ontologies as much as possible and defines new concepts and relations only where needed. The PCSM is based on semantic technologies, like RDF and OWL, because it is currently the best way of intuitively defining formal semantics (OWL) and provides the flexibility for modular reuse of existing data models or extend them. To this end, the PCSM is a small, core model that covers the main common concepts in the agrifood domain ranging from farm calendars to agro-environmental parameters. Existing ontologies that already define the required concepts are reused by the PCSM as much as possible.
A thorough analysis on agri-food data sharing needs, the use of ontologies and specifications about the PCSM are provided in [5]. The PCSM is available here: https://gitlab.com/Ploutos-project/ploutos-common-semantic-model/-/blob/master/ploutos.ttl?ref_type=heads

Beyond the standards, vocabularies and ontologies mentioned above, there are great many others that are either specific to the domain or are generic but relevant because they cover concepts that are widely reused across multiple domains (examples include SSN/SOSA[15], QUDT[16], OM[17], W3C Time[18], various weather ontologies[19], etc.).

2.5.3 ScaleAgData RILs - data types

Having analyzed the most dominant, ontology-based data modeling approaches an analysis on the data modeling needs of ScaleAgData RILs follows. For each RIL a categorization of the key data types that are expected to be generated and need to be modeled has been realized. It should be noted that at this stage it is not feasible to perform a one-to-one mapping of the data to be generated in RILs with the respective concepts of the selected ontology. The scope of this exercise is to have a first high-level analysis on the modeling needs that the data model (ontology) that will be selected by ScaleAgData for reuse should address. One key characteristic of the data model to be reused in ScaleAgData is to be versatile enough and able to model additional data types through the implementation of the necessary extensions.

The following tables provide a summary of data types and their properties for each RIL that are extracted from the respective RIL descriptions in section "2.3 Functional View". The data type descriptions of RILs data in "D1.2 Open Science and Data Management Plan" are also considered within our analysis.

Table 1: Yield Estimation Tool Data Types.

RIL title	Data type	Properties
Yield Estimation Tool	Yield data	Harvest weight, grain moisture, number of grains per ear, data from harvesters (ton/ha)
	Field information	Coordinates, size, crop type

[15] https://www.w3.org/TR/vocab-ssn/
[16] https://www.qudt.org/
[17] https://github.com/HajoRijgersberg/OM
[18] https://www.w3.org/TR/owl-time/
[19] https://bimerr.iot.linkeddata.es/def/weather/ connected to SAREF

	EO data	Satellite data, Reflectance data (Red, Green, Blue, Near-Infrared, Shortwave Infrared), Modeled in GeoTiff
	Vegetation indices, Leaf area index (LAI), Evapotranspiration (EVA)	
	Phenological development stages GS_i	GS1: Seedling stage GS2: Tillering stage GS3: Leaf expansion stage GS4: Booting stage GS5: Flowering stage GS6: Grain filling stage GS7: Maturation stage
	Weather data	Temperature (air and soil), precipitation, relative humidity, solar radiation
	Soil data	pH, EC, temperature, moisture, nutrient content (NPK), soil type (e.g., clay loam, sandy loam, silt loam), quality
	Historical data	Yield data, weather data and soil data from previous years
	RGB images	Pixel intensity values (0-255), texture features (e.g., mean, standard deviation, entropy), color features (e.g., red, green, blue, hue, saturation, value)
	Machine parameters	Tractor speed, traction, fuel usage

Table 2: Soil Reflectance Measurement Data Types.

RIL title	Data type	Properties
Soil Reflectance Measurement (Hyperspectral sensors)	Soil data	Moisture, pH, salinity, organic matter

Table 3: Soil Hyperspectral Processing Module Data Types.

RIL title	Data type	Properties
Soil hyperspectral processing module	Soil data	Type (bare soil, eroded soil, topsoil), texture, pH, organic matter
	Weather data	Air temperature, air humidity, wind direction and speed, solar radiation, precipitation, atmospheric pressure
	Soil Proprietary Database	Soil composition, soil texture, soil moisture content, soil pH levels, soil nutrient levels, soil organic matter, soil microbial activity

Table 4: Drought Prediction from IoT and Airborne Sensor Data Types.

RIL title	Data type	Properties
Drought prediction from IoT and airborne sensor data	Spatial data	Field boundaries (polygon data), land cover/land use (raster data)
	IoT sensor data	Time series, data of several parameters including precipitation, air temperature, soil moisture, soil temperature, and irrigation info
	Airborne sensor data	Surface temperature, vegetation and moisture indices, multiband geospatial raster's

Table 5: Drought Prediction from Satellite Sensor Data.

RIL title	Data type	Properties
Drought prediction from satellite sensor data	Spatial data	Field boundaries (polygon data), land cover/land use (raster data)
	ET (evapotranspiration) and soil moisture data products (from DHI)	
	Vegetation indices from VITO	NDVI - normalized difference vegetation index, EVI - enhanced vegetation index, GNDVI - green normalized difference vegetation index, NDRE - normalized difference red edge, LAI - leaf area index
	Hyperspectral data (from Kuva Space)	Surface temperature, normalized difference vegetation index (NDVI), other vegetation indices, such as EVI, GNDVI, and NDRE

Table 6: Vegetation Indices Calculator Data Types.

RIL title	Data type	Properties
Vegetation indices calculator	Geospatial data	Field boundaries (polygon data), (latitude, longitude).
	Time Series data	Reflectance data, leaf index area (LAI), evaporation-transportation
	Index data (NDVI, EVI, GNDVI, NDRE)	NDVI - normalized difference vegetation index, EVI - enhanced vegetation index,

		GNDVI - green normalized difference vegetation index, NDRE - normalized difference red edge, LAI - leaf area index
	Dias and AWS open Data	Satellite imagery, soil data, GIS data, weather data

Table 7: Agro-environmental Policy Indicators Monitoring Data Types.

RIL title	Data type	Properties
Agro-environmental policy indicators monitoring	Farm Management Information Systems (FMIS) data	Digital farm books, crop type, yield, dates of management practices, yield quality
	Satellite EO data	Imagery, land cover classification, vegetation indices, land use, soil moisture
	IoT sensor data	Air temperature, humidity, wind speed, precipitation, soil moisture, soil temperature, soil salinity.
	Geospatial data	Topographical maps, land cover maps, lidar data

Table 8: DSS Precision Farming Data Types.

RIL title	Data type	Properties
DSS Precision Farming	Crop data	Unit polygon, type (User input)
	Other attributes	Attributes, such as soil type or weather data, may also be required for some indices for each polygon
	GeoTIFF data	A geoTIFF file for each polygon, which will contain the calculated indices

Table 9: Digital Twin Providing Features and Decision Support Data Types.

RIL title	Data type	Properties
Digital twin providing forecasts and decision support	ISOBUS time log data	This data contains information about the various field management activities that have been performed, such as planting, fertilizing, and harvesting. It can be used to track the progression of crop growth and identify potential problems.

	FMIS data	Structured data. This data includes information about the field boundaries, soil properties, and cultivar
	Weather data	Time series data like air temperature, air humidity, wind direction speed, solar radiation, precipitation, atmospheric pressure
	Remote sensing data	Raster data like satellite imagery, ground-based sensors, unmanned aerial vehicles (UAVs)
	Soil sensor data	Time series data. i.e., soil moisture, soil temperature etc.
	Farm management data	Structured data i.e., as fertilizer applications and irrigation schedules
	Forecasted yield	Tons/hectare
	Soil status	Healthy, eroded
	Nitrogen status	Parts per million
	Management recommendations	Apply nitrogen fertilizer at x kg/hectare

2.5.4 Reuse of AIM in ScaleAgData

As it was described in the introduction of this section the DEMETER Agriculture Information Model is selected as the best approach for addressing data modeling needs of the ScaleAgData. According to our analysis and given that AIM is tailored to the needs of the agricultural sector already contains most of the domain specific concepts that need to be modeled in the context of ScaleAgData. In addition, it is an extendable ontology, and any additional needed concepts are feasible to be introduced.

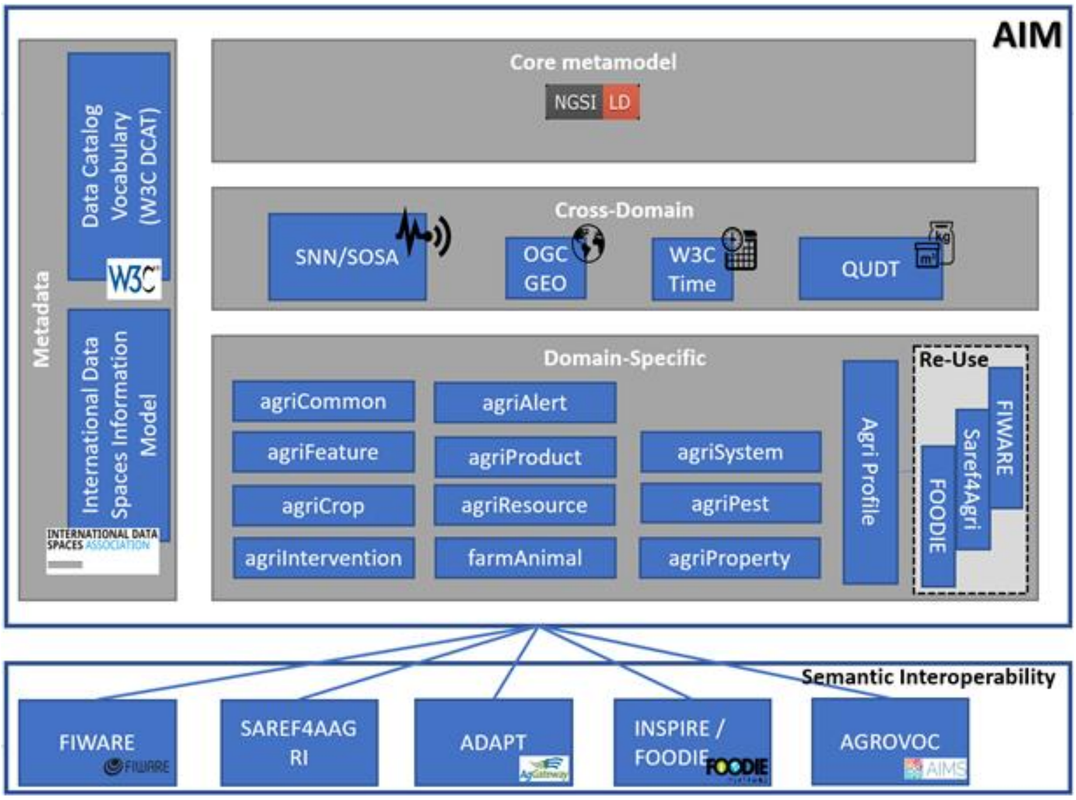


Figure 22 AIM ontology [1].

Figure 22 provides an illustration of AIM's architecture and the application domains that can currently be modeled. In addition, the AIM is currently under standardization by OGC with an emphasis on the re-use of generic OGC standards as appropriate. (see here for more details: https://www.ogc.org/press-release/ogc-forms-new-agriculture-information-model-standards-working-group/)

As a first step towards the reuse of AIM in the context of RILs various data translation exercises are realized. Sample data generated from RILs have been translated with the use of AIM. For example, the following AIM-JSON-LD data object corresponds to soil temperature measurements recorded by an IoT sensor deployed in a parcel of RIL "Agro-environmental policy indicators monitoring".


```
{
 "@context": [
  "https://w3id.org/demeter/agri-context.jsonld",
  {
   "qudt-unit": "http://qudt.org/vocab/unit/"
  }
 ],
 "@graph": [
  {
   "@id": "urn:scaleAgData:plot:RI_lab_policymonitoring:71219362-f8da-4ae7-b85a-a083eea4a493",
   "@type": "Plot",
   "hasGeometry": {
    "@id": "urn:scaleAgData:plot:RI_lab_policymonitoring:71219362-f8da-4ae7-b85a-a083eea4a493",
    "@type": "Polygon",

    "asWKT": "POLYGON (33.17372175137724 34.84488553253772,33.17399994030271 34.84498379420684, 33.1737956962324 34.84540862440489, 33.17351398466683 34.84531325229517,33.17372175137724 34.84488553253772)"
   },
   "area": 0.14,
   "description": "tomato cultivation in Thessaly",
```

"category": "vegetables",
"crop": {
"@id": "urn:scaleAgData:crop:df72dc57-1eb9-42a3-88a9-8647ecc954b4",
"@type": "Crop",
"cropSpecies": "urn:scaleAgData:croptype:df72dc57-1eb9-42a3-88a9-8647ecc954b4",
"cropStatus": "seeded",
"lastPlantedAt": "2022-08-23T10:18:16Z"
}
},
{
"@id": "urn:scaleAgData:croptype:df72dc57-1eb9-42a3-88a9-8647ecc954b4",
"@type": "CropType",
"name": "Vegetables",
"alternateName": "Solanum lycopersicum",
"agroVocConcept": "https://agrovoc.fao.org/browse/agrovoc/en/page/c_4475",
"eppoConcept": "https://gd.eppo.int/taxon/LYPES",
"description": "Solanum lycopersicum.Cultivated throughout the world as a vegetable, mostly under glass or plastic, sometimes naturalizing."
},
{
"@id": "urn:scaleAgData:RI_lab_policymonitoring:observation-20221203",
"@type": "ObservationCollection",
"observedProperty": "https://uri.fiware.org/ns/data-models#soilTemperature",
"hasMember": ["urn:demeter:RI_lab_policymonitoring:soilTemperature:observation-20221203:1","urn:demeter:RI_lab_policymonitoring:soilTemperature:observation-20221203:2","urn:demeter:RI_lab_policymonitoring:soilTemperature:observation-20221203:3"]
},
{
"@id": "urn:scaleAgData:RI_lab_policymonitoring:soilTemperature:observation-20221203:1",
"@type": "Observation",
"resultTime": "2022-12-03T04:17:12Z" ,
"hasResult": {
"@id": "urn:scaleAgData:RI_lab_policymonitoring:observation-20221203:1/result",
"@type": "QuantityValue",
"numericValue": "14.2",
"unit" : "qudt-unit:DEG_C"
}
},
{
"@id": "urn:scaleAgData:RI_lab_policymonitoring:soilTemperature:observation-20221203:2",
"@type": "Observation",
"resultTime": "2022-12-03T04:17:12Z",
"hasResult": {
"@id": "urn:scaleAgData:RI_lab_policymonitoring:observation-20221203:2/result",
"@type": "QuantityValue",
"numericValue": "14.1",
"unit" : "qudt-unit:DEG_C"
}
},
{
"@id": "urn:scaleAgData:RI_lab_policymonitoring:soilTemperature:observation-20221203:3",
"@type": "Observation",
"resultTime": "2022-12-03T06:18:12Z" ,
"hasResult": {
"@id": "urn:scaleAgData:RI_lab_policymonitoring:observation-20221203:3/result",
"@type": "QuantityValue",
"numericValue": "14.6",
"unit" : "qudt-unit:DEG_C"
}
}
]
}

At this stage validation of the generated schemas are performed with the tools like https://json-ld.org/playground/ and visualization of the semantic graphs with tools:
https://issemantic.net/rdf-visualizer
For example, the AIM graphs visualized in Figures 23 and 24 correspond to the AIM JSON-LD sample data object above.

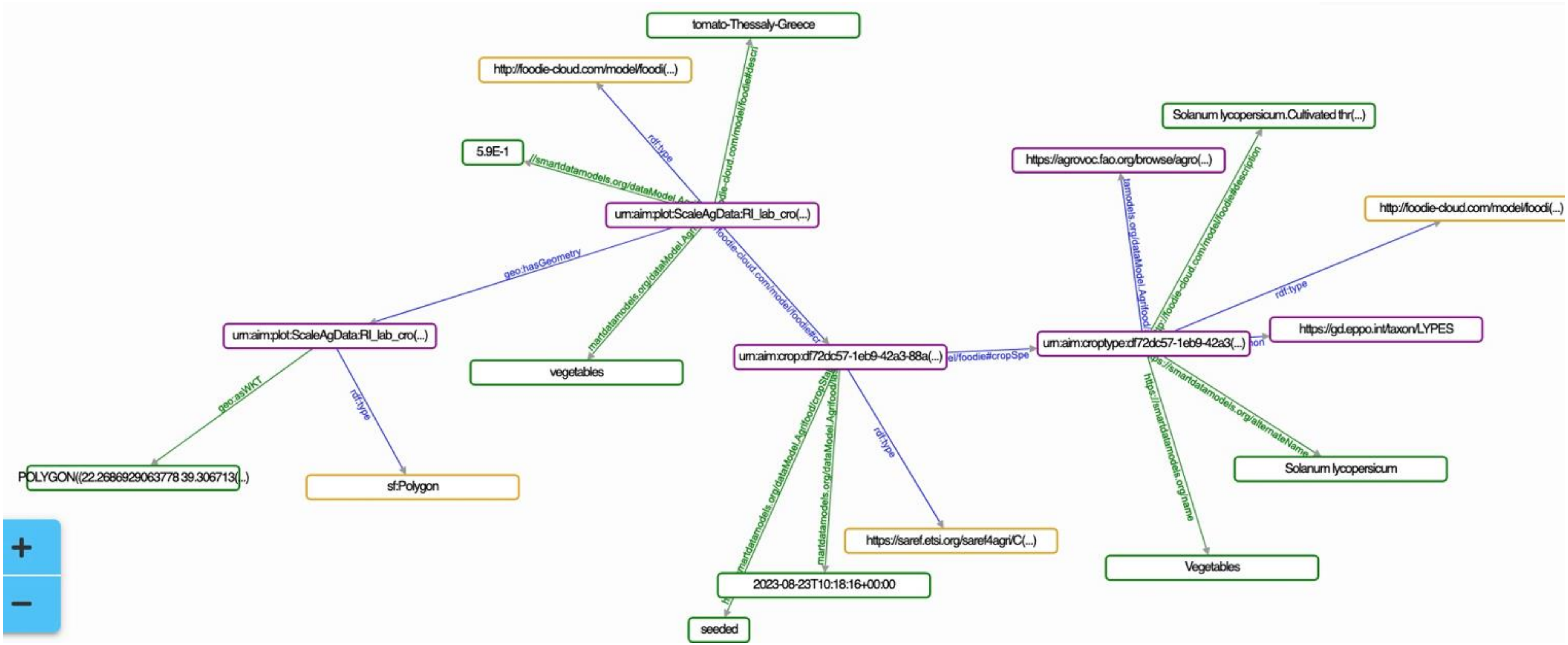


Figure 23 Parcel properties described with the use of AIM.

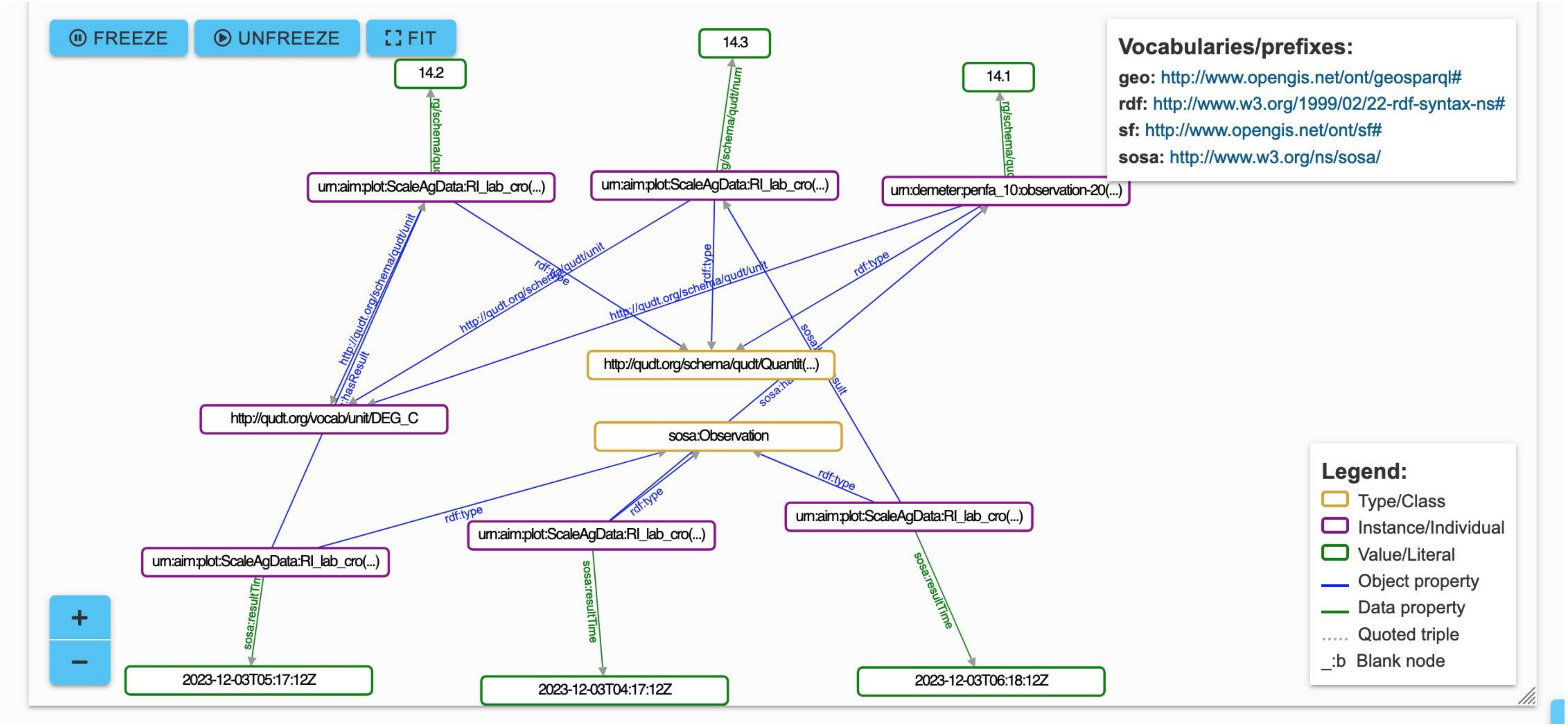


Figure 24 Soil temperature recordings described with the use of AIM.

Next steps include the development of software translators that will get as input RIL datasets modeled in custom formats and translate according to AIM specifications.

2.5.5 Data governance schemes

Among the dominant data governance principles that ScaleAgData aims to comply with are the Findability, Accessibility, Interoperability, and Reuse of digital assets (FAIR). From a practical perspective data sharing has meant the application of semantic technologies, and to a greater or lesser extent the adoption of the FAIR data principles. The use of ontologies supports such governance schemes:

- The Findability of RIL data is ensured using APIs.
- The Accessibility of data is ensured using standardized protocols integrated with access control.
- Interoperability of data is ensured by the (re)use of widely used ontologies/vocabularies that are accessible online.
- Reusability of data is ensured by using community standards and by ensuring data provenance is a consequence of data ownership and control.

The AIM has already been evaluated against these principles and proved that it is capable to address them to a large extend especially with regards to Interoperability and Reusability. Figure 25 provides the assessment questions outcomes of AIM's FAIRness. (source: https://agroportal.lirmm.fr/ontologies/DEMETER-AIM).

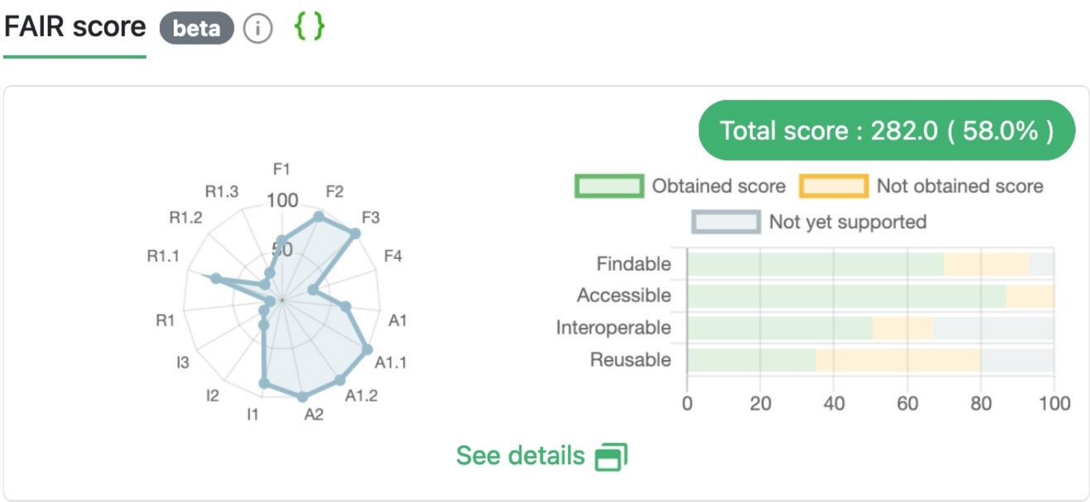


Figure 25 AIM's score against FAIR principles.

Overall, the details of the data governance models will be specified in collaboration with T2.4 (Governance models for the vertical domains of the RILs**)** also considering legal, operational and functional agreements as well as technical standards. T2.4 will develop, elaborate and validate new governance models within the second iteration, considering the openDEI design principles for data spaces and the preparatory actions for the data space for Agriculture as way to support the interoperability between the vertical domains.

2.6 External Dataspaces

2.6.1 ScaleAgData Key points

ScaleAgData is involved in developing data products and services, particularly focusing on various data types, including IoT data. The project is working on transactions within data ecosystems, which may vary regarding organization and governance. The key objectives revolve around addressing the issue of fair access to and use of data, with a specific emphasis on facilitating data sharing and creating value within the vertical defined by RILs.

Here's a breakdown of the key points mentioned:

- **Types of Data:** ScaleAgData deals with different data types, including IoT data. This suggests a broad scope regarding the data they handle, including agricultural and rural data.
- **Data Products and Services:** The primary focus is on the development of data products and services. This could involve creating solutions that leverage data for specific applications, potentially in the agricultural sector.
- **Transactions within Data Ecosystems:** ScaleAgData is involved in transactions within data ecosystems. This could refer to activities such as data exchanges, collaborations, or partnerships within the broader context of data management.
- **Organization and Governance:** The data ecosystems may vary in terms of organization and governance. This implies that ScaleAgData may operate in diverse environments with different levels of structure and control over data.
- **Fair Access to and Use of Data:** Ensuring fair access to and use of data is a key concern. This suggests a commitment to equitable data practices, addressing issues related to data ownership, privacy, and sharing.
- **Technical and Non-Technical Building Blocks:** ScaleAgData considers both technical and non-technical building blocks of data spaces. This indicates a holistic approach involving both the technological aspects of data management and the broader organizational and procedural components.
- **Facilitating Data Sharing:** A key goal is to facilitate data sharing. This aligns with broader industry trends where collaboration and data sharing contribute to innovation and value creation.
- **Creation of Value from Data:** ScaleAgData aims to enable the creation of value from data. This underscores the belief that data, when properly managed and utilized, can yield significant value, potentially benefiting stakeholders within the defined vertical of Rural Innovation Labs.

In summary, ScaleAgData is positioned at the intersection of data management, technology, and agricultural innovation, focusing on creating value through fair and effective use of diverse types of data.

2.6.2 Common European Data Spaces

The European strategy for data aims to create a "single European data space," a unified market for data open to global sources. This space secures personal and non-personal data, including sensitive business data, providing businesses easy access to high-quality industrial data for growth and value creation. The strategy emphasizes the need for both horizontal actions and sector-specific data spaces in areas like Agriculture and Green Deal. It seeks a secure, unified data space with effective governance, advanced tools, and improved data quality, promoting collaboration and value across sectors.

In practice, a data space is an infrastructure that enables data transactions between different data ecosystem parties based on the governance framework of that data space. According to the DSSC, a data space should be generic enough to support the implementation of multiple use cases, which is the case for the ScaleAgData RILs.

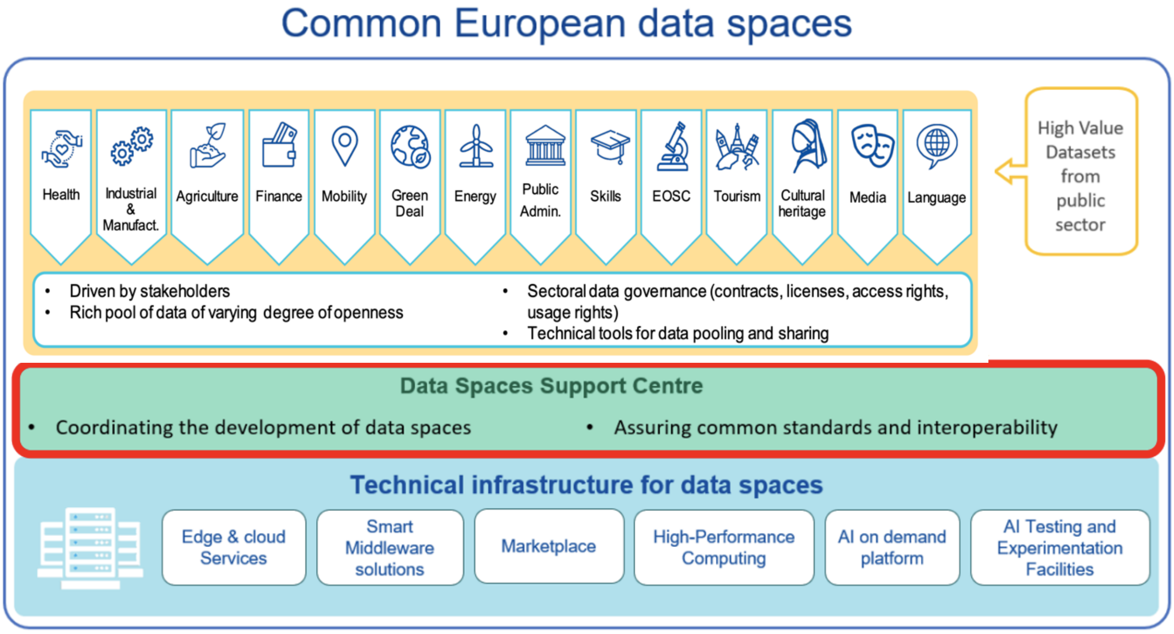


Figure 26 The common European data spaces, updated version. The ScaleAgData project has links with the Agriculture and the Green Deal data space.

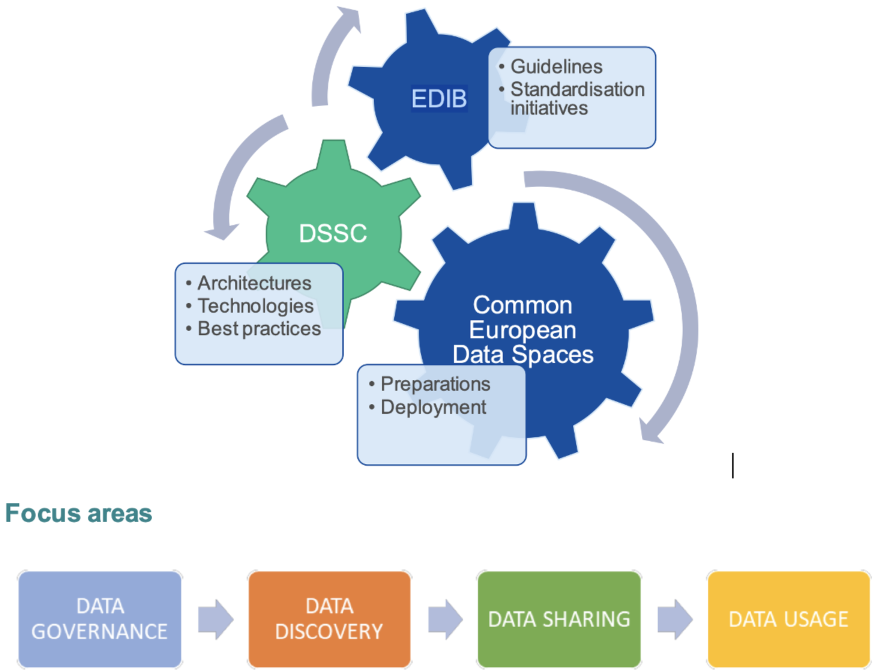


Figure 27 For developing the Common European Data Spaces, the European Data Innovation Board (EDIB), the Data Space Support Center (DSSC) and the preparation and deployment projects like AgDataSpace have complementary roles. The standardization needs to support interoperability, and data governance is a focus area.

2.6.2.1 Agricultural data space and its challenges

According to the Demeter HE project, the key challenges in establishing effective Agriculture Data Spaces are related to the following.

- **Diverse Nature of Agricultural Data**: Agricultural data is varied, including livestock, land, climate, financial, compliance, and food-related information. Data modelling standards are needed for effective collection, sharing, and comparison.
- **Trustful Environment:** Building trust between agrifood sector players and farmers is crucial. Ensuring data sovereignty for farmers and promoting the benefits of data sharing is essential for creating a trustworthy environment.
- **Data Interoperability and Portability:** Low interoperability and portability between data tools in the agrifood ecosystem result in farmers using multiple platforms. Semantic descriptions of data formats can facilitate decentralized and interoperable data management.
- **Access to Digital Single Market:** A single market of technologies, services, and data sources is needed for SMEs to reach farmers effectively. Agriculture Data Spaces can contribute by reducing market fragmentation and enhancing competition.
- **Sustainable Business Models:** The implementation of European agrifood Data Spaces requires clear and sustainable business models. Putting data owners in control, incentivizing data sharing, and defining business models for intermediation services are crucial.
- **Data Quality:** Data quality is essential for informed decision-making in agriculture. Data Spaces can contribute by addressing data quality assessment for structured and linked data to ensure its usefulness.
- **Involving the Entire Food Chain:** To maximize data-sharing benefits, connecting the farm world with the entire food system is vital. Consideration of technical, legal, ethical, socio-economic, and business aspects is necessary for successful implementation.

2.6.2.2 Data Space Building Blocks

To break down data spaces into manageable pieces, the DSSC adopted the concept of building blocks: a basic unit or component that can be implemented and combined with other building blocks to achieve the functionality of a data space. We should mention that the building blocks are not isolated. Most functionalities the data space participants need result from an interplay between multiple building blocks.

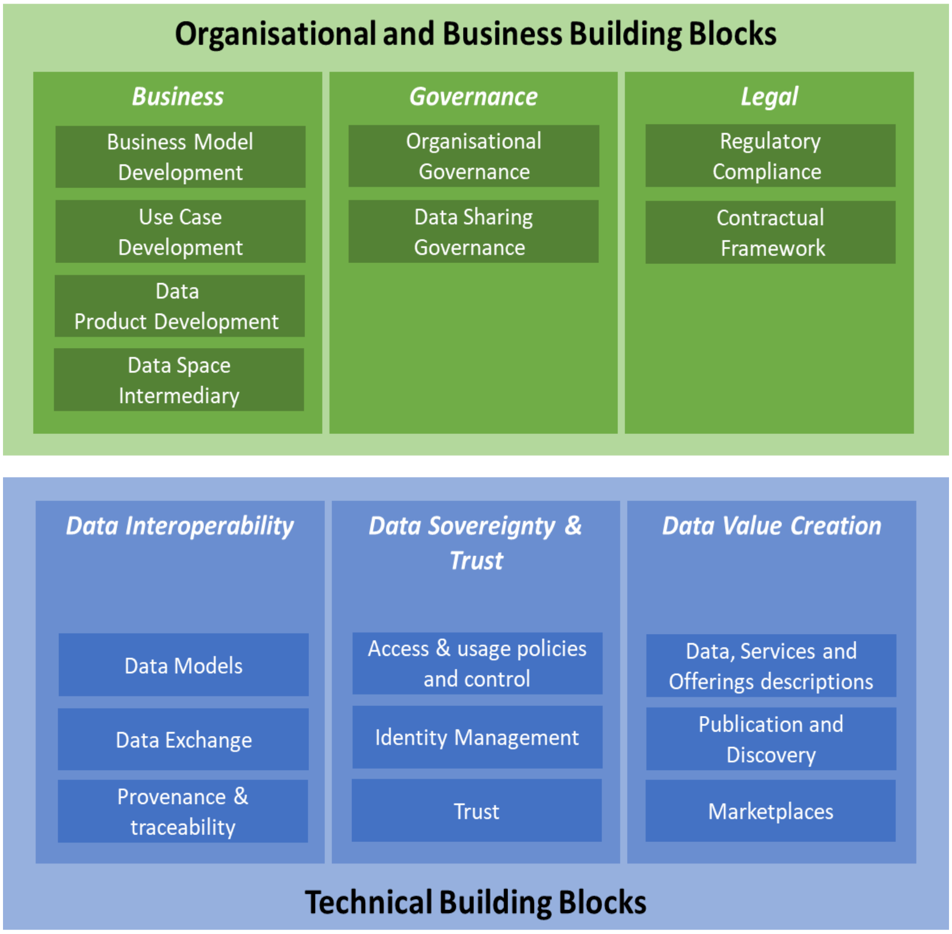


Figure 28 Building Block Taxonomy.

<u>Organizational and Business Building Blocks</u>

Are organized into three categories:

Business Category: *provides the essential concepts necessary in the business modelling of a data space.*
Building blocks:
1. Business Model Development:
   - Supports a data space in developing its business model.
   - Identifies key elements and considerations for the governance authority.
2. Use Case Development:
   - A strategic approach to amplify the value of a data space.
   - Fosters the creation, support, and scaling of use cases.
3. Data Product Development:
   - Considers data product templates, governance rules, and network effects.
   - Focuses on enhancing synergy between data providers and users.
4. Data Space Intermediaries:
   - Supports business and governance decisions related to data space intermediaries.

Governance Category: Focuses primarily on data space-level governance. Emphasizes the dynamic nature of governance that needs to adapt as the data space evolves.
Building blocks:
1. Organizational Governance:
   - Guides setting up the data space governance authority.

- Identifies key decision points and options for establishing inclusive governance.

2. Data Sharing Governance:
    - Supports the governance authority in establishing common rules.
    - Promotes effective and reliable data-sharing processes and introduces ways to organize data transactions.

Legal Category: Provides guidance and resources for data space initiatives to ensure compliance and establish a robust contractual framework.
Building blocks:

1. Regulatory Compliance:
    - Raises awareness of the legal landscape for data space initiatives.
    - Aids in assessing applicable regulatory requirements to ensure compliance and alignment with EU values.
2. Contractual Framework:
    - Supports a data space by establishing clear and enforceable rights and obligations.
    - Provides contractual resources for data space participants to regulate their data transactions.

<u>The Technical Building Blocks</u>
Are organized into three categories:

Data Interoperability Category: Capabilities: The exchange of data requires (semantic) models, data formats, and interfaces (APIs). Includes Functionalities for provenance and traceability.
Building blocks:

1. Data Models:
    - Capabilities: Define and use shared semantics in a data space.
    - Purpose: Enhance understanding and consistency in how data is represented and used.
2. Data Exchange:
    - Capabilities: Facilitate the actual exchange and sharing of data.
    - Purpose: Enable seamless sharing and transfer of data within the data space.
3. Provenance and Traceability:
    - Capabilities: Track the process of data sharing.
    - Purpose: Ensure traceability and compliance by making the data-sharing process transparent and auditable.

Data Sovereignty and Trust Category: Capabilities: Identification of participants and assets in a data space, establishment of trust, and the ability to define/enforce access and usage control policies.
Building blocks:

1. Access and Usage Policies and Control:
    - Purpose: Enables the specification and enforcement of policies within a data space.
    - Responsibility: Handled by both the data space authority and individual participants.
2. Identity Management:
    - Purpose: Manages identities within a data space.
    - Responsibility: Ensures effective management of participant identities.
3. Trust:
    - Purpose: Verifies that a participant in a data space follows specific rules.
    - Enables: Confidence in the reliability and adherence of participants to established guidelines.

Data Value Creation Category: Capabilities: Enable value creation in a data space. Examples: Registering and discovering data offerings or services, providing marketplace functionality, and enabling monetization of data sharing.
Building blocks:

1. Data, Services, and Offering Descriptions:
    - Purpose: Provides tools for data providers to describe data products comprehensively and understandably.
    - Includes: Information on data policies and ways to obtain the data product.
2. Publication and Discovery:
    - Purpose: Enables data providers to publish descriptions of their data, services, and offerings.
    - Follows: FAIR principles (Findable, Accessible, Interoperable, Reusable) to enhance discoverability by potential users.
3. Marketplace:
    - Purpose: Offers marketplace capabilities for providers and users to engage in relationships.
    - Enables: Access, provision, and use of data products previously published and discovered in the data space.

2.6.2.3 ScaleAgData approach towards the Common European Data Spaces

As presented in Figure 29, within the ScaleAgData project, the RILs are performing transactions aiming at the development and upscaling of data products and services. Adopting Data Space's best practices for data sharing, value creation, interoperability, trust, and digital sovereignty can support the RILs participants not only in performing the same transactions more easily but also in generating the required conditions for upscaling. There are several key challenges in establishing effective Agriculture Data Spaces, and to support this transition methodologically, ScaleAgData will follow the Building Blocks Taxonomy suggested by the Data Space Support Center (DSSC) (Figure 28), elaborating and targeting specific building blocks related to governance, interoperability, data sovereignty and trust considering the relevant legal frameworks.

As already described within D2.1, ScaleAgData will deal with:
- The organizational governance and Regulatory Compliance within Work Package 2-Task 4 (Governance models for the vertical domains of the RILs).
- The data-sharing governance is mainly within Work Package 2-Task3/Task4 and Work Package 3-Task 4 (Data Governance, Sharing Meta architecture, and Integration**)**. Similar for the Data, Services and Offerings description because it is related to the catalog services and the metadata organization.
- The data models and data exchange within Work Package 3-Task 4.
- The access and usage policy, together with the identity management within Work Package 3-Task 4 and Work Package 4-Task 3 (Research and Innovation Environment).
- Business Model Development and Data Product Development within Work Package 6-Task 4 (IPR Management, Definition of Business Models and Policy brief).

Within the current deliverable, we will provide the needed info related to the data models, data exchange, identity management, access, and usage policy, providing an architectural design that incorporates elements or building blocks that support the establishment and operation of data space (architecture for a data sharing governance).

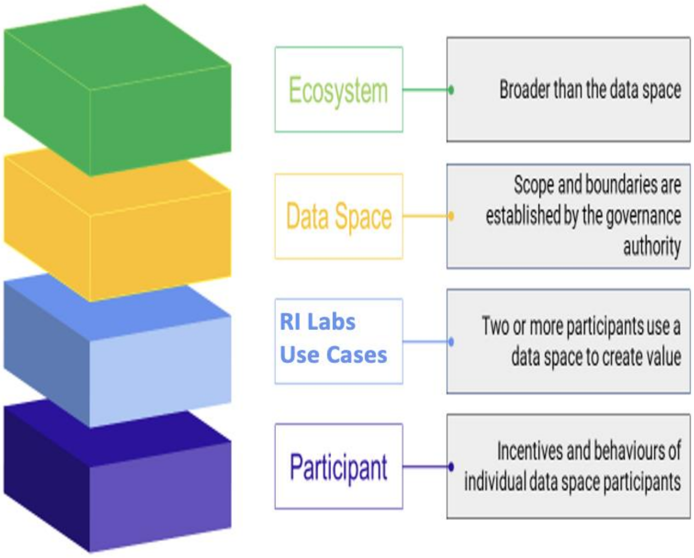


Figure 29 The ScaleAgData RILs use cases targeting different innovation areas and materialize expected transactions of a Data space.

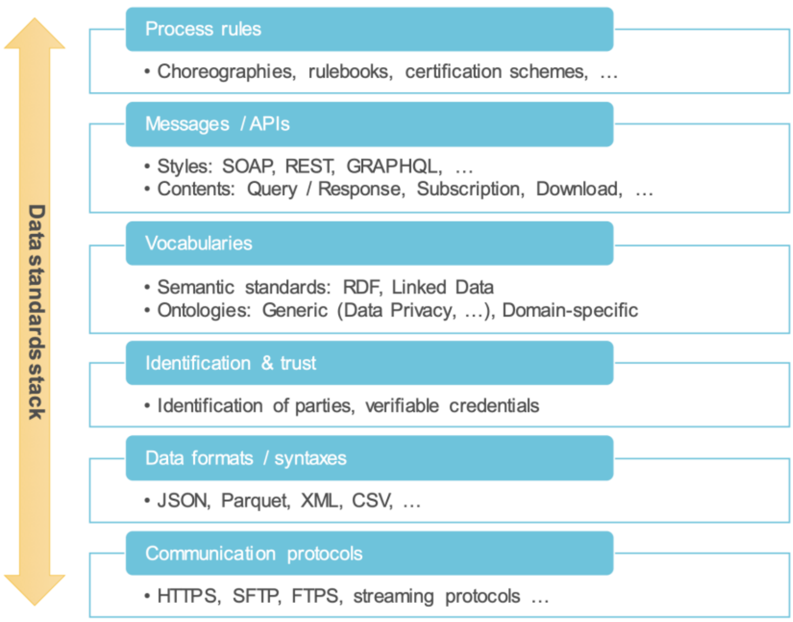


Figure 30 A visualization of the data standard stack. ScaleAgData will try to support RILs participants on the building blocks related to data models (vocabulary), data exchange (Messages/APIs) and Identification and Trust.

Additionally, although ScaleAgData recognizes that the Data space functionality is implementation-agnostic, having common standards and software components is important to improve efficiency and transparency and achieve growth. For data space designers, it is important to know not only what functionality needs to be provided (Architecture of What) but also which standards and software implementations are available to support their implementation (Architecture of How, Figure 31). For this reason, the ScaleAgData architecture model will provide info related not only to the business, functional information, and process but also to the system layer (Figure 32), which means for the technical components that support the materialization of a data space.

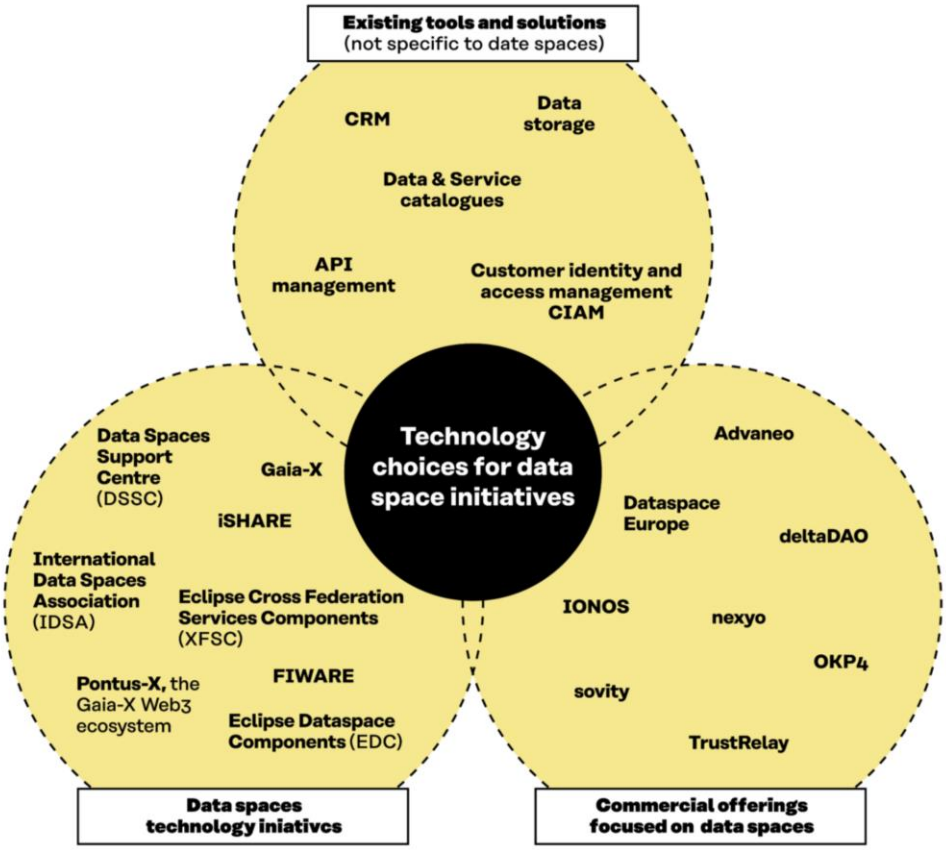


Figure 31 The data space technology landscape allows multiple choices or offers different directions.

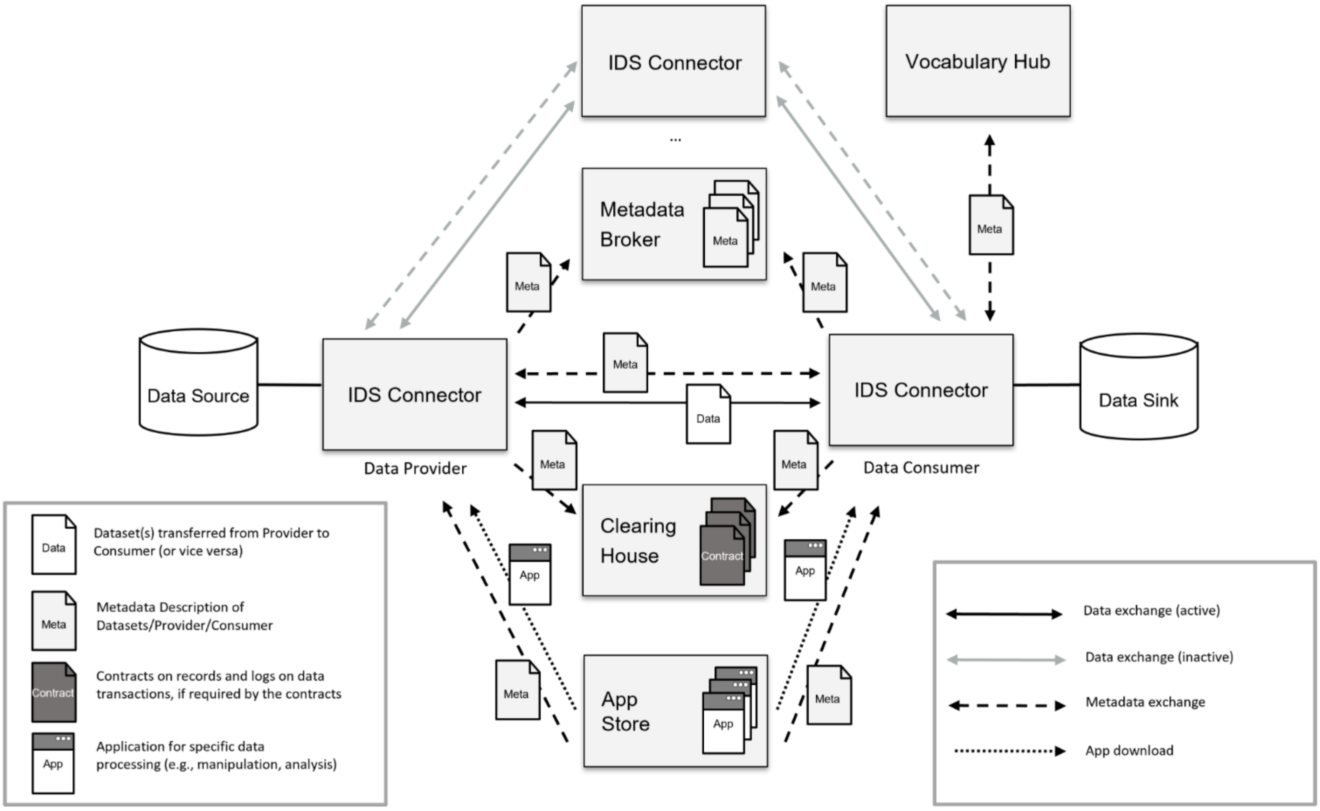


Figure 32 Interaction of the technical components (system layer) that implement a data space using the IDSA Reference Architecture Model.

Data Models

In the realm of a data space, addressing the challenges posed by diverse data interpretation necessitates achieving semantic interoperability through a shared language among participants. This has been realized by implementing key practices:

- **Providing Metadata**: Crucial metadata for attributes, including descriptions and data types, is presented in machine-readable formats using Semantic Web standards like RDFS, OWL, and RIF.
- **Establishing References:** Connecting to external sources through linked data URIs enhances the comprehensive understanding of data.
- **Vocabulary Metadata:** Information such as licensing, contact details, modification dates, and ontology language within vocabularies is available in a machine-readable format, following open standards like DCAT-AP for consistency and accessibility.
-

The data model building block's key capabilities encompass vocabulary, vocabulary management processes, and a vocabulary hub. In the context of ScaleAgData, a particular emphasis is placed on vocabularies. RILs participants, whether acting as data providers or receivers, must reference vocabularies to prevent miscommunication, misinterpretation, errors, and suboptimal decisions. Vocabularies, including ontologies, data models, schema specifications, mappings, and API specifications, are often domain specific. They can be either provided by the data space as shared or standardized or created ad-hoc when no applicable vocabulary exists, and the data space allows for such flexibility.

Effectively implementing the data model building block involves defining the scope and purpose of shared vocabularies and specifying included aspects and entities. Once established, the focus shifts to creating or adopting standardized vocabularies adhering to open metamodel standards like RDFS, OWL, JSON, XML schema, SKOS, etc., tailored to domain and application requirements.

Table 10: Cross-domain standards.

Candidate Specification	Related Building Block	Publisher	Data Spaces endorsing
NGSI-LD	Data Exchange	ETSI	DS4SSCC, int:net, DATES, DS4.0
OpenAPI Spec	Data Exchange	OpenAPI initiative	DS4SSCC, LDS, DATES
JSON-LD	Data Models	W3C	DS4SSCC, int:net, LDS, DS4SKILLS, GREAT
Smart Data Models	Data Models	SDM initiative	DS4SSCC, DATES, DS4.0
XACML	Access and usage policies &control	OASIS	DS4SSCC, LDS, DATES, GREAT
eIDAS/eIDAS2	Identity Management	EU	DS4SSCC, int:net, LDS, DS4SKILLS, GREAT
ODRL	Access and usage policies &control	W3C	DS4SSCC, LDS, DATES
EUDI RAF	Identity Management	EU	DS4SSCC, LDS, DS4.0
Oauth2	Identity Management	IETF	DS4SSCC, int:net, LDS, DATES, DS4SKILLS, GREAT
OpenID	Identity Management	OpenID Foundation	DS4SSCC, LDS, DATES, GREAT
ICAM	Identity Management	GAIA-X	int:net, LDS, DATES
SAML 2.0	Identity Management Access and usage policies &control	IETF	DS4SSCC, LDS, DATES, DS4SKILLS, GREAT
DCAT-AP	Publication & Discovery	W3C	DS4SSCC, int:net, LDS, DATES, GREAT
DID	Identity Management	W3C	LDS, DATES, DS4SKILLS
SOLID	Identity Management	MIT	DS4SSCC, LDS, DS4SKILLS
Verifiable Credentials Data Model	Identity Management Trust	W3C	LDS, DATES, DS4SKILLS
Dublin Core	Data Models	ISO	DATES, DS4SSCC, GREAT
Electronic Signatures and infrastructures	Identity Management	ETSI	DS4SSCC, LDS, GREAT
Federated catalogue	Marketplaces	GAIA-X	LDS, DS4.0, GREAT
IDS information model	Publication and discovery	IDSA	Int:net, LDS, GREAT
OGC	Data models	Open Geospatial Consortium	DS4SSCC, LDS, GREAT
XML	Data Models	W3C	Int:net, LDS, GREAT

<u>Data exchange</u>

The data space participants must use a common protocol to exchange the data which includes the syntax, the semantics, and the order of the interaction. This protocol is used to define an Application Programming Interface (API), which allows applications to communicate with each other. APIs can be different per domain, and standardization is often needed within a data space to enable interoperability.

The data exchange building block provides the following capabilities:

- Meta specifications and best practices for the adoption of existing data exchange APIs
- Generic purpose data exchange API, including methods to query, update and delete data.
- Tooling to use and maintain data exchange APIs.

This building block has a strong relation with the data models building block. It provides the ground of data models in technical formats (payload of the API), e.g. JSON.

<u>Implementation of data exchange</u>

The data exchange process in a data space has two phases. We need to distinguish between the control plane and data plane. In the control plane phase, the data space participants must agree on the data assets and data exchange agreements. Once agreed, during the data plane phase, the actual data exchange starts.

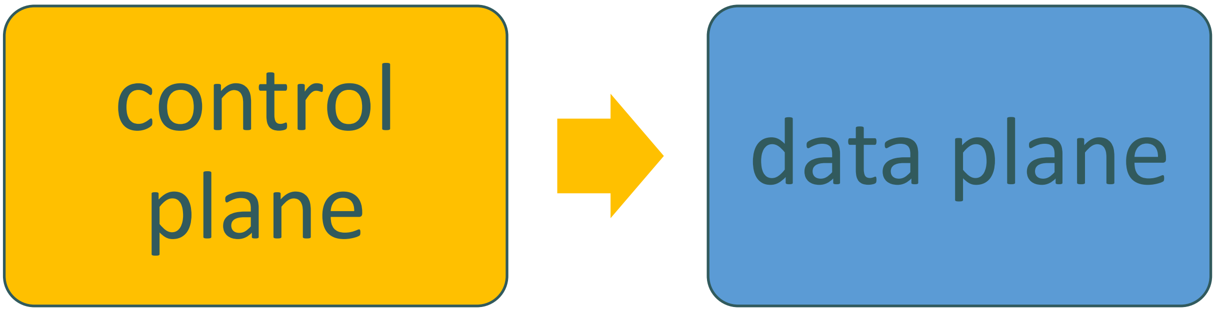


Figure 33 The data exchange process has two phases. First, the participants agree (control plane), and then the actual data exchange starts (data plane).

The coordination of the actual transmission of data can be performed by the Data Space Connectors (not the only technical option; see Figure 31). A data space connector is a collection of technical modules that any participant has to deploy and run to connect to a data space to support transactions with other participants.

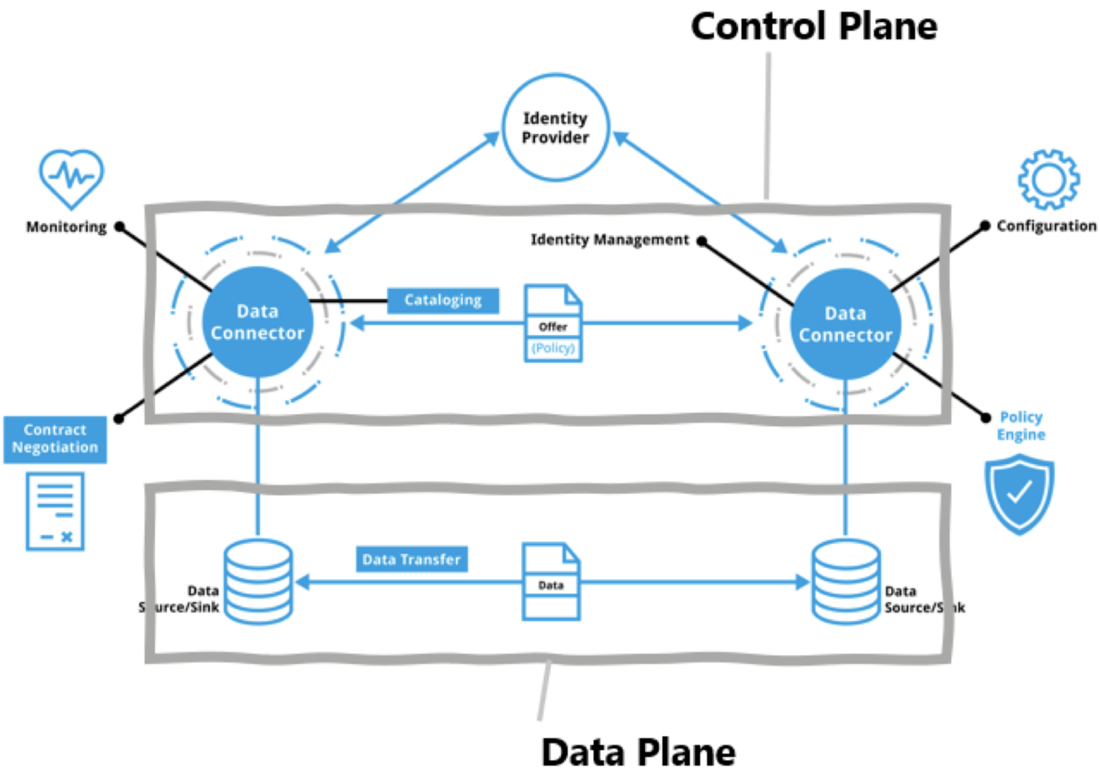


Figure 34 The control plane provides interoperability at the data space level. The data plane manages interoperability for data exchange.

Data within a data space can follow two paths: it can flow through the data connector ('in-band') or be exchanged through a separate channel ('out-of-band'). Regardless of the approach, using a standardized, machine-readable specification of the data exchange protocol, commonly known as an API, is crucial to ensure that systems on both ends can effectively send and receive data. These protocols, designed for interactions between participants, play a pivotal role in achieving interoperability.

In the realm of the data plane, data exchange protocols adopt a layered approach, building on standard internet protocols such as HTTP/HTTPS. Additional agreements, like API Protocols (e.g., REST), sit on top of these to enhance interoperability. Further up is the specific interface specification for sharing data (e.g., NGSI-LD, recommended by the Demeter HE project), outlining API methods, operations, calls, and payload formats. The Data Exchange Building Block operates in this third layer, defining a data space-specific API.

API Protocols come in various types (e.g., REST, SOAP, RCP) and can be either open for public use or private for internal or partnership use. The choice of API Protocols depends on the user and purpose, and an API can employ multiple protocols. REST, utilizing HTTP/HTTPS communication protocols, stands out as one of the most widely used web APIs.

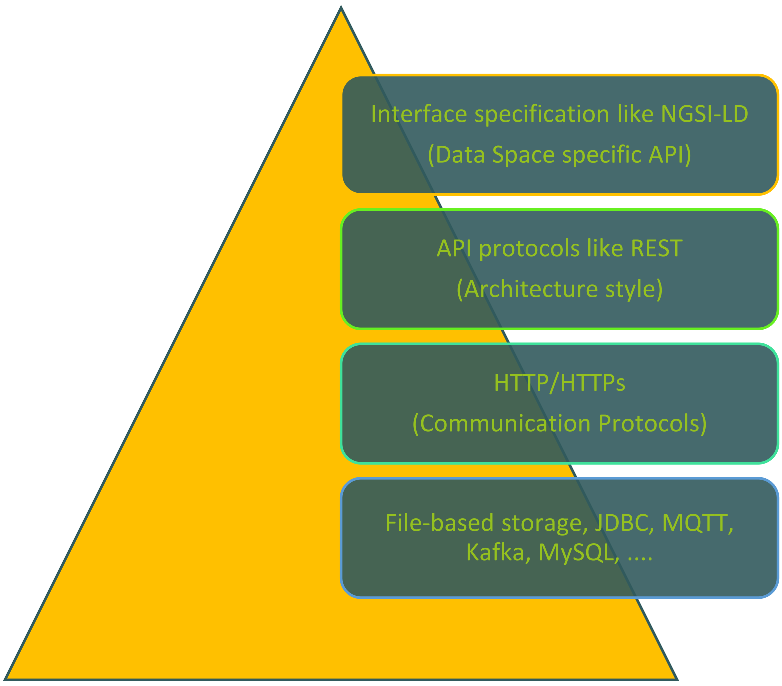


Figure 35 The data exchange protocols rely on a layered approach.

The data exchange block mainly deals with the upper layer, which means the domain-specific and the domain agnostic interface specifications. Together with the Data Models building block, this building block is essential to implement the required interoperability within the data space and, furthermore, across data spaces. The absence of common data exchange APIs makes data transfer, access or exchange between participants in a data space difficult. It is also essential for the Data Models building block to ground the data models into specific formats to be included in the API calls.

To ensure effective data exchange within ScaleAgData, the data exchange APIs need to be formalized in a machine-readable format, like OpenAPI or AsyncAPI specifications. This involves providing best practices for API maintainers to ensure:

- Explicit Link with Semantics: Establish a clear connection with semantics, making it easy for users to understand the meaning and purpose of the data exchange.
- Utilization of Vocabularies: Encourage the use of vocabularies from the data models building block, ensuring consistency and shared understanding.
- Ease of Implementation: Facilitate straightforward implementation for software engineers, ensuring the APIs are user-friendly and accessible.
- Best Practices: Provide guidelines, including design rules such as REST API design rules or naming conventions like OpenAPI Naming and Design Rules | UNECE.

Moreover, the data exchange protocols should be published and made available through user-friendly documentation platforms, such as Swagger UI, a plugin for a static site generator, or a comprehensive API specification management platform like Swagger Hub. This documentation should include detailed information on each API endpoint, supported methods, request/response data formats, examples, and other relevant details to enhance accessibility and usability for developers.

2.7 High-level architecture per RIL

The general high-level architecture of the ScaleAgData system is depicted in Figure 2.

From this general high-level architecture, the architecture of each RIL can be derived by selecting only the specific components that match the RIL needs, based on the innovative areas and the deployment scenarios that must be supported per case.

The seven innovative areas are:
IA1 - Innovative sensor technology,
IA2 - Edge processing
IA3 - Data sharing architecture and data governance
IA4 - Satellite data augmentation
IA5 - From data assimilation to service development
IA6 - Privacy-preserving technology
IA7 - Data integration methodologies

2.7.1 High-level architecture of RIL 1 – Water Productivity

The innovations areas that must be supported on the RIL 1 – Water Productivity based on the deployment scenarios are:
IA3 - Data sharing architecture and data governance
IA4 - Satellite data augmentation
IA5 - From data assimilation to service development

The general high-level architecture of the RIL 1 – Water Productivity is depicted in Figure 36.

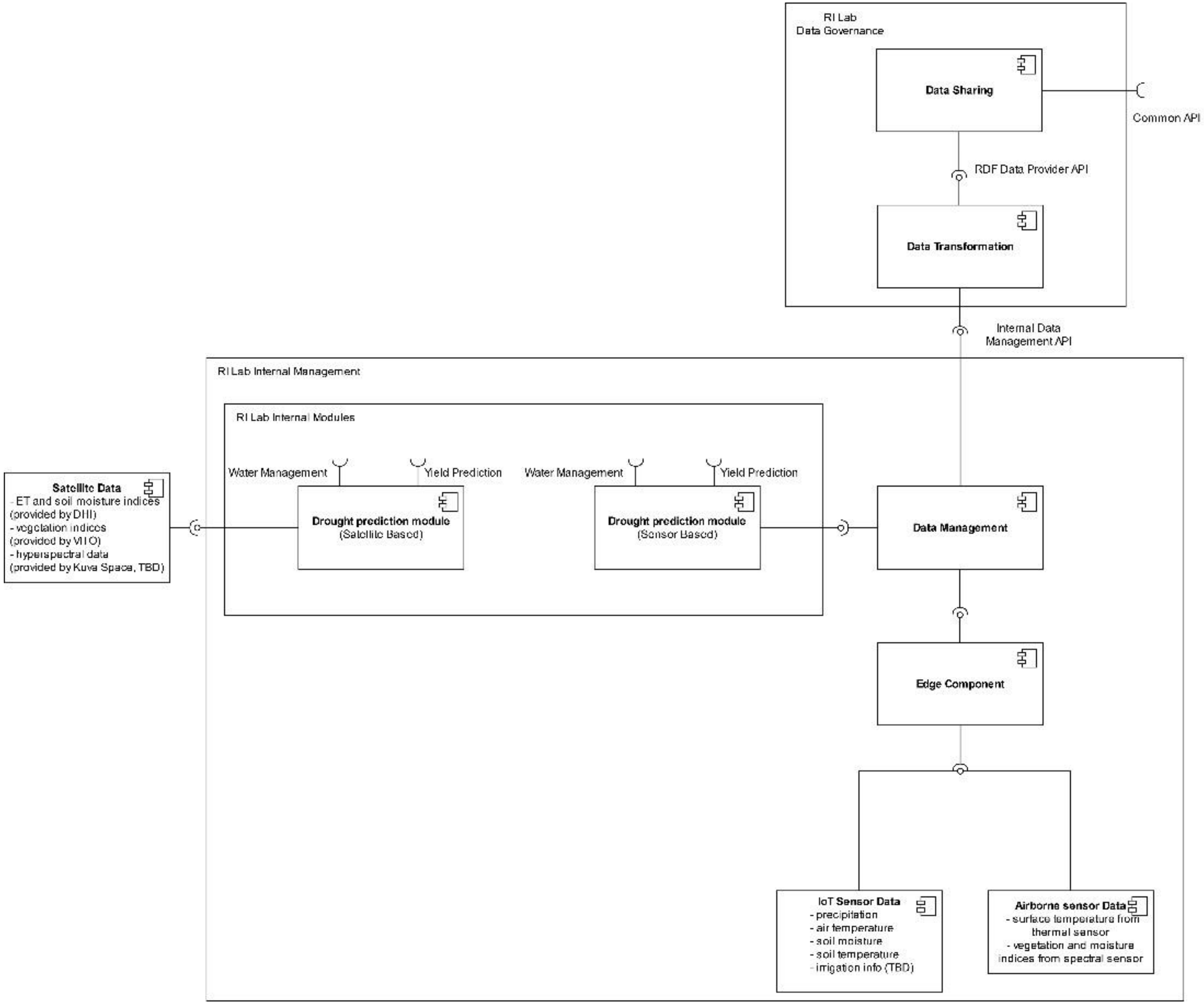


Figure 36 High level architecture of RIL 1 – Water Productivity.

2.7.2 High-level architecture of RIL 2 – Crop Management

The innovations areas that must be supported on the RIL 2 – Crop Management based on the deployment scenarios are:
IA1 - Innovative sensor technology,
IA2 - Edge processing
IA3 - Data sharing architecture and data governance
IA5 - From data assimilation to service development
IA6 - Privacy-preserving technology
IA7 - Data integration methodologies

The general high-level architecture of the RIL 2 – Crop Management is depicted in Figure 37.

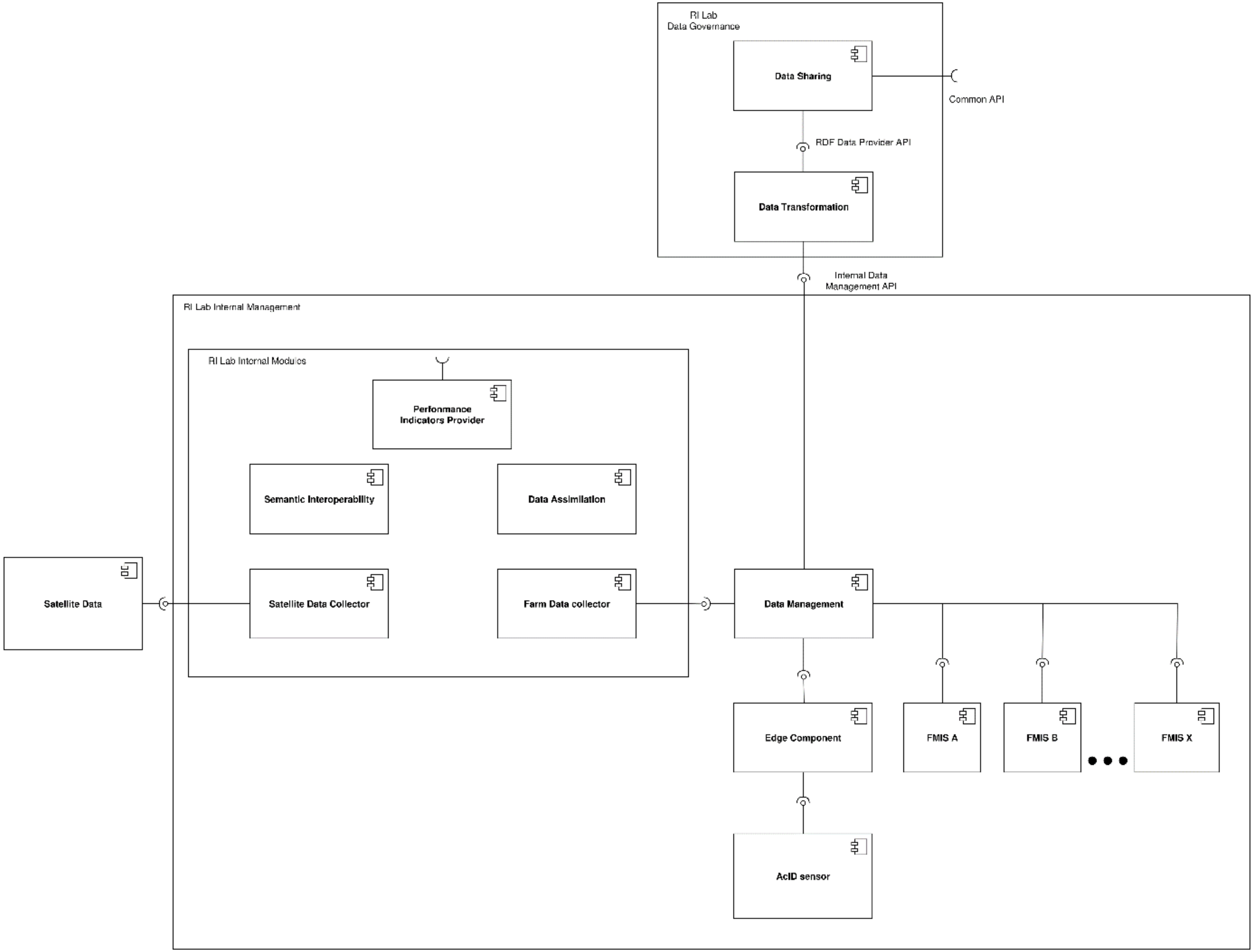


Figure 37 High level architecture of RIL 2 – Crop Management.

2.7.3 High-level architecture of RIL 3 – Yield Monitoring

The innovations areas that must be supported on the RIL 3 – Yield Monitoring based on the deployment scenarios are:
IA3 - Data sharing architecture and data governance
IA5 - From data assimilation to service development
IA6 - Privacy-preserving technology
IA7 - Data integration methodologies

The general high-level architecture of the RIL 3 – RIL 3 – Yield Monitoring is depicted in Figure 38.

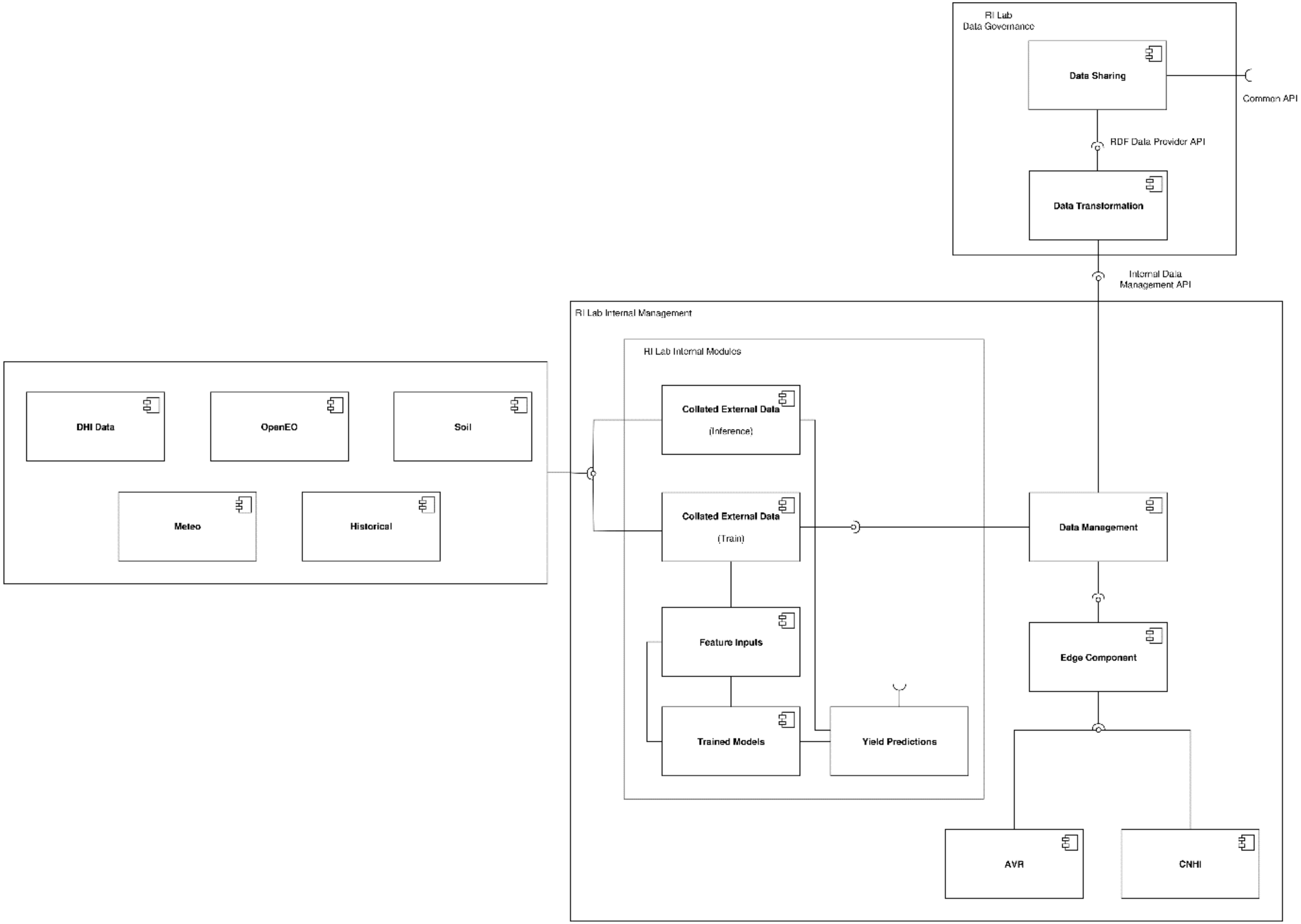


Figure 38 High level architecture of RIL 3 – Yield Monitoring.

2.7.4 High-level architecture of RIL 4 – Soil Health

The innovations areas that must be supported on the RIL 4 – Soil Health based on the deployment scenarios are:

IA1 - Innovative sensor technology,
IA2 - Edge processing
IA3 - Data sharing architecture and data governance
IA6 - Privacy-preserving technology
IA7 - Data integration methodologies

The general high-level architecture of the RIL 4 – Soil Health is depicted in Figure 39.

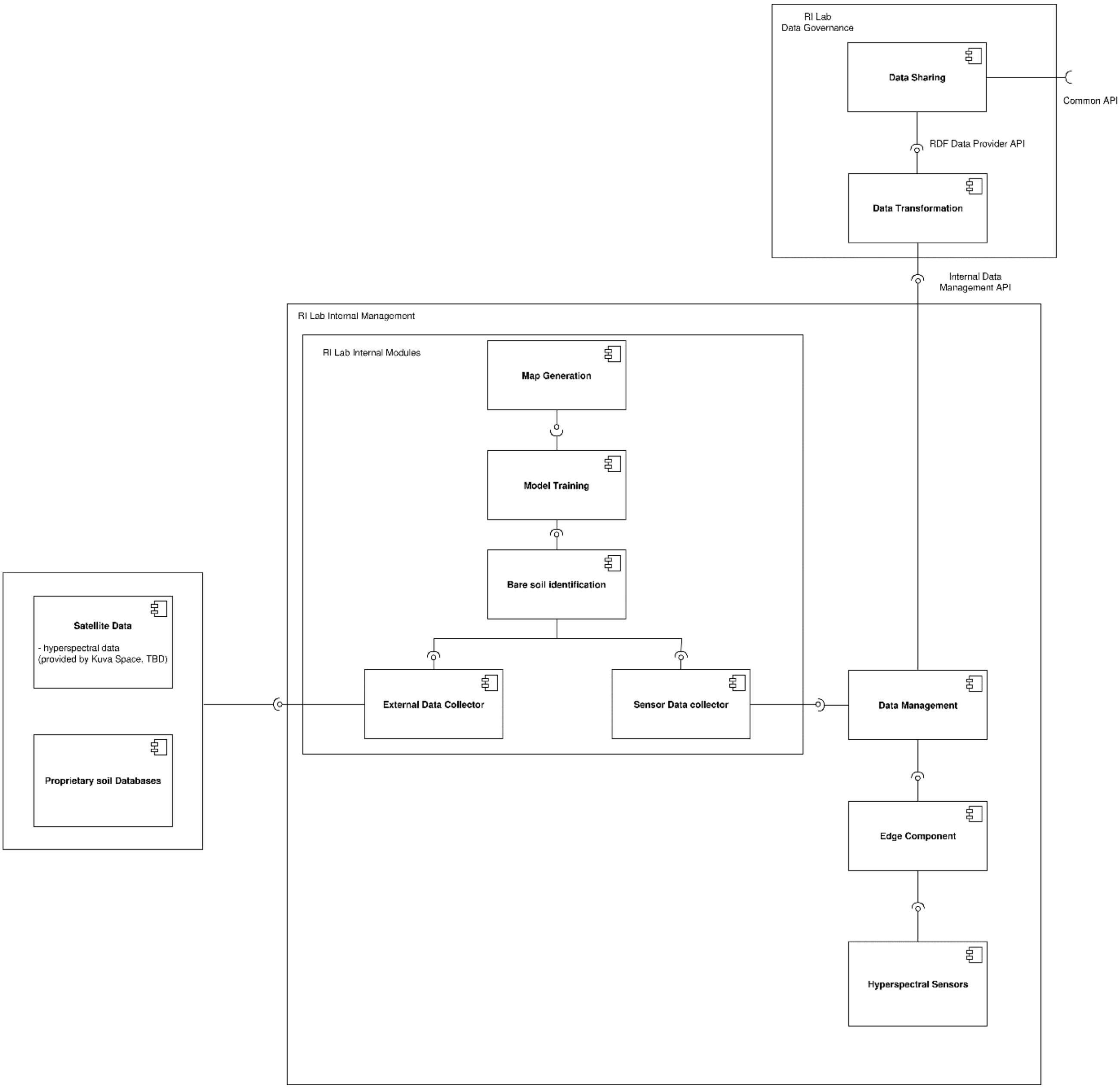


Figure 39 High level architecture of RIL 4 – Soil Health.

2.7.5 High-level architecture of RIL 5 - Grasslands

The innovations areas that must be supported on the RIL 5 – Grasslands based on the deployment scenarios are:
IA3 - Data sharing architecture and data governance
IA4 - Satellite data augmentation
IA7 - Data integration methodologies

The general high-level architecture of the RIL 5 – Grasslands is depicted in Figure 40.

Figure 40 High level architecture of RIL 5 – Grasslands.

2.7.6 High-level architecture of RIL 6 – Sustain Dairy

The innovations areas that must be supported on the RIL 6 – Sustain Dairy based on the deployment scenarios are:

IA1 - Innovative sensor technology,
IA2 - Edge processing
IA3 - Data sharing architecture and data governance
IA4 - Satellite data augmentation
IA5 - From data assimilation to service development
IA6 - Privacy-preserving technology
IA7 - Data integration methodologies

The general high-level architecture of the RIL 6 – Sustain Dairy is depicted in Figure 41.

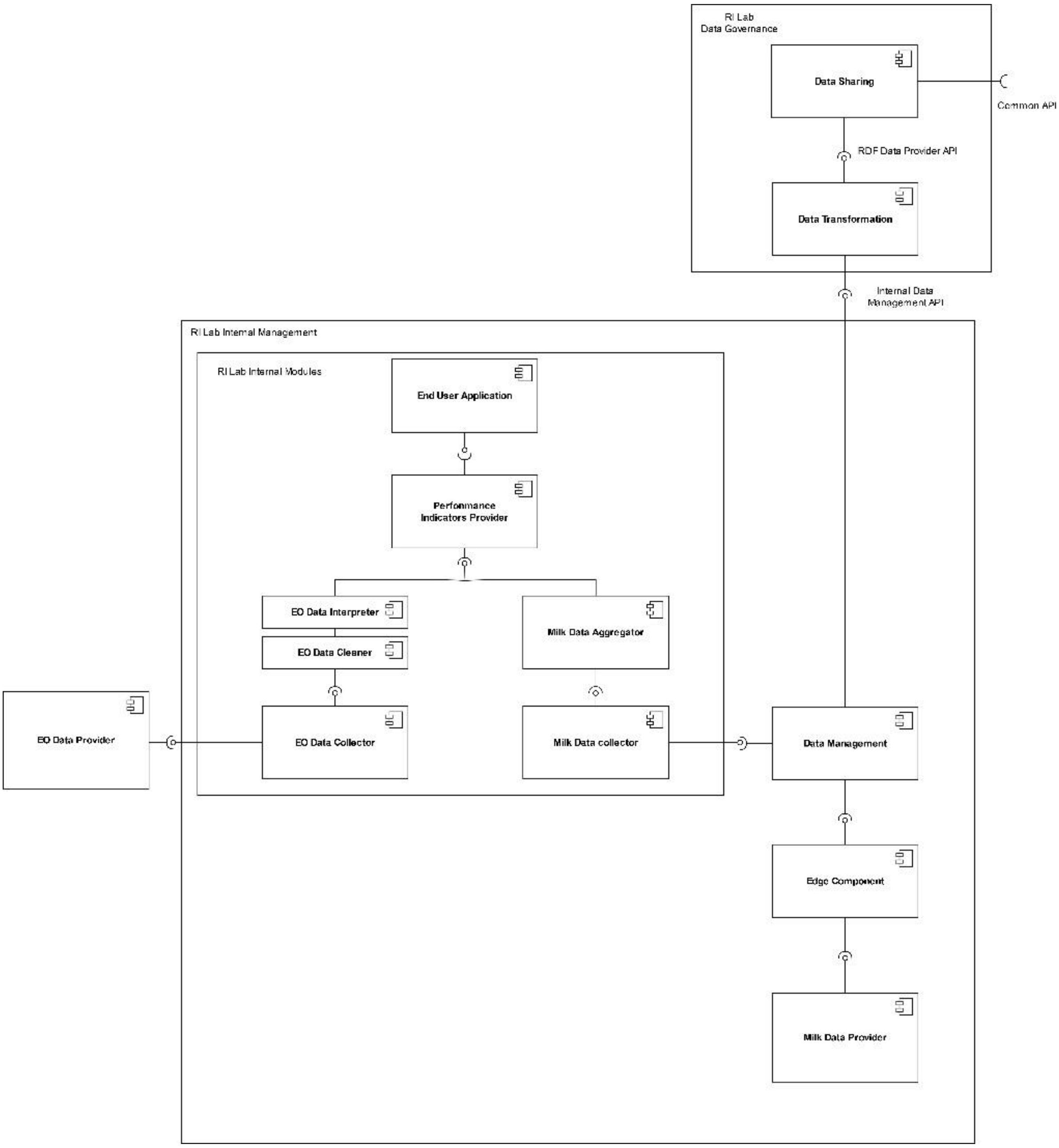


Figure 41 High level architecture of RIL 6 – Sustain Dairy.

2.8 Information View

The information view defines the static information structure and presents dynamic information and data flows. In more detail:

- Regarding the static information structure of the components of the ScaleAgData System, it is important to note that it is not ready yet. The components are currently in ongoing development and this information is not fully complete at this stage. There are still elements that are to be determined (TBD) in the development process. However, these details will be fully developed and included in the next iteration of this deliverable (expected around month 33).
- Information flow patterns and specific data flows (using Message Sequence Charts, MSCs), for all the components that have already been described in 2.3, of specific indicative scenarios are provided.

<u>Overall Architecture Information Flow</u>
The Information Flow within ScaleAgData shows the communication pathways that facilitate the data collection and sharing. As shown by the previous analysis there are various components and enough diverse in the technological foundations and functionalities that are integrated, so the Information Flow will be a high level one trying mostly to show some generalized common data flows focusing on data sharing. Not all security requests are shown as it may depend on the specific protocols used internally by each RIL, which may depend for example on the sensor supported protocols, and their security practices may vary.

<u>Sensor to Data Management Information Flow</u>

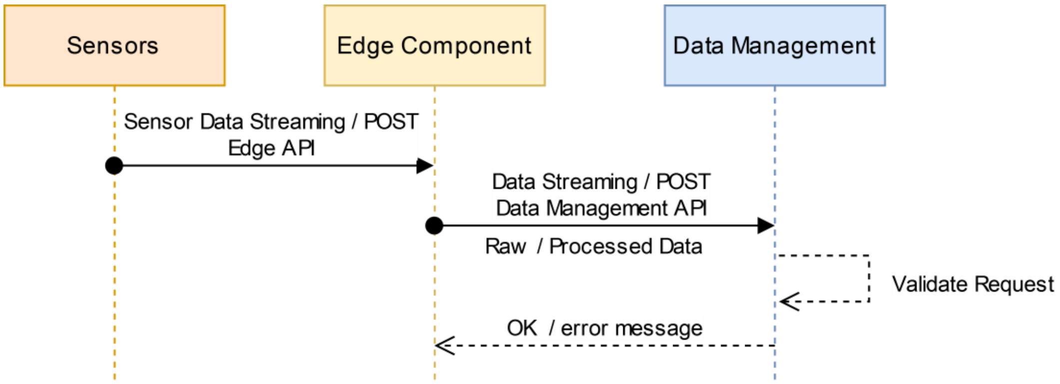


Figure 42 Sensor to Data Management Information Flow.

The sequence diagram shown in the Figure 42 shows a simple flow from the sensors to the Data Management layer. The sensors may send data first to the Edge Component using their supported transfer protocol. The Edge Component afterwards, may transmit the data directly to the data management layer or do a local processing and transmit processed Data depending on the use case. Security best practices should be followed during this data transfer that are not shown, as it is internal to each RIL, but a validation of any request, or data streaming interface must be applied.

Data Management to RIL Module Information Flow

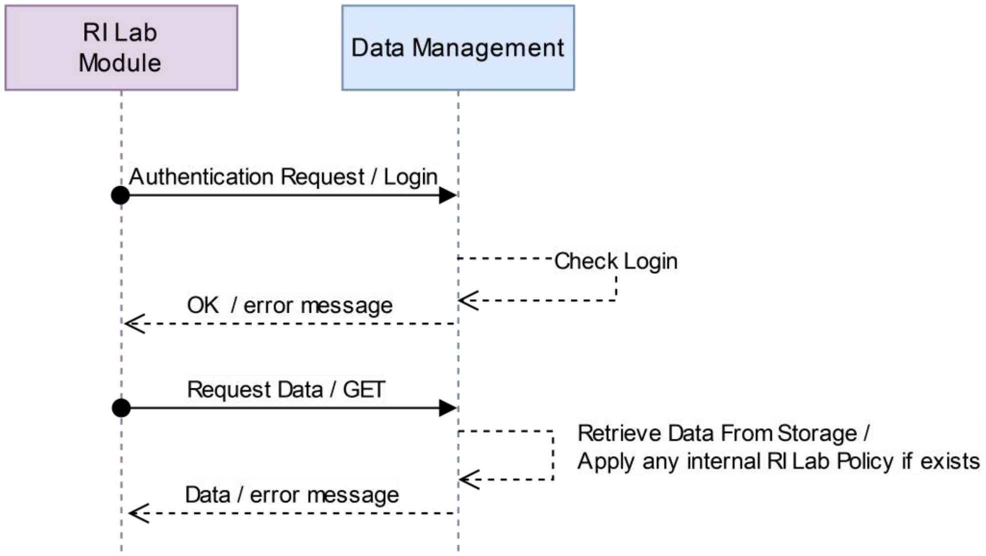


Figure 43 Data Management to RIL Module Information Flow.

Each RIL develops some internal modules guided by research interests. These modules can get data from the Data Management platform to perform their analysis. The sequence flow is shown in Figure 43. The internal module must make a login request to gain access to the platform and then request the data that it needs. Any internal to the RIL Authorization policy upon the data may be applied before retrieving and sending the data back to the internal module.

Data sharing Sequence Flow

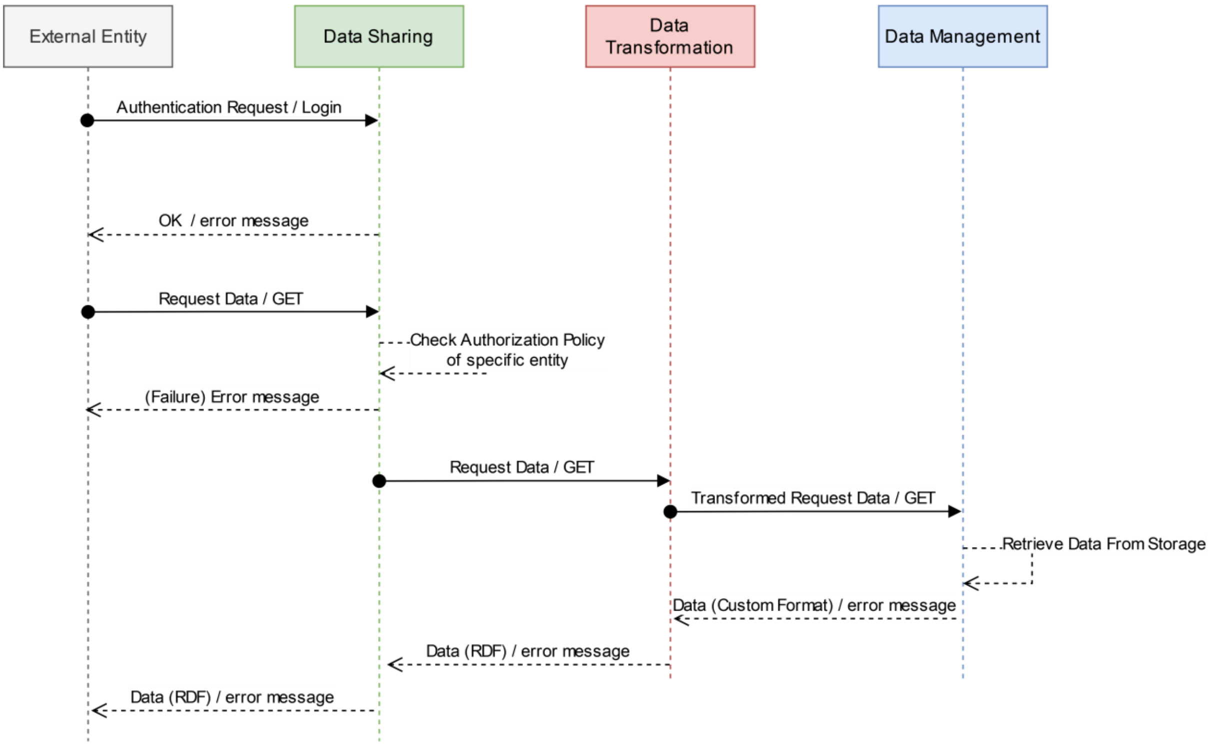


Figure 44 Data Sharing Sequence Flow.

Figure 44 shows the data sharing sequence flow. Any external entity, for example another RIL, the RIE, can request data using the public Common API exposed by each RIL. An authentication request must be validated first to identify the external entity. Afterward the external entity can request data that it may be interested in retrieving from the RIL. Afterwards any policy is checked to verify whether this external entity is authorized to access the requested data. If there is such policy then the data are retrieved from the data management layer, are transformed to the common data model based on RDF vocabularies and then responded back to the external entity.

3. References

[1] European Commission. European Interoperability Framework for Pan-European Egovernment Services; EU Publications: Luxembourg 2011; ISSN 92-894-8389-X.

[2] Serrano, M.; Barnaghi, P.; Carrez, F.; Cousin, P.; Vermesan, O.; Friess, P. IoT Semantic Interoperability: Research Challenges, Best Practices, Recommendations and Next Steps; IERC: European Research Cluster on the Internet of Things Technical Report, March 2015. Available online: http://www.internet-of-things-research.eu/pdf/IERC_Position_Paper_IoT_Semantic_Interoperability_Final.pdf (accessed on 15 January 2019).

[3] "D4.2: Data interoperability for the agri-food sector", June 2021. https://ploutos-h2020.eu/wp-content/uploads/2022/11/Ploutos_D4.2-Data-interoperability-for-the-agri-food-sector_v1.0_2021-06-11.pdf

[4] ETSI. (2020). SmartM2M; Extension to SAREF; Part 6: Smart Agriculture and Food Chain Domain (Technical Specification 103 410-6; p. 26). ETSI. https://www.etsi.org/deliver/etsi_ts/103400_103499/10341006/01.01.02_60/ts_10341006v010102p.pdf

[5] Brewster, Christopher et al. 'Data Sharing in Agricultural Supply Chains: Using Semantics to Enable Sustainable Food Systems'. 1 Jan. 2023: 1 – 31.