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

EO DATA COLLECTION, SPATIO-TEMPORAL PREPARATION SERVICES

Revised version, 13 May 2026

Authors			
Name	**Organization**	**First release date**	**Update**
Isabelle Piccard	VITO	25/9/2025	12/5/2026
Rado Guzinski	DHI	25/9/2025	7/5/2026
Paolo Cosmo Silvestro	Deimos	25/9/2025	12/5/2026
Nuno Grosso	Deimos	29/9/2025	30/4/2026
Mariapina Castelli, Giulia Graldi	EURAC	25/9/2025	12/5/2026
Maria Pat Gonzalez-Dugo	IFAPA	25/9/2025	12/5/2026
Valentina Manstretta	HORTA	25/9/2025	7/5/2026
Thomas Papakosmas	NP	25/9/2025	11/5/2026
Nick Berkvens	ILVO	25/9/2025	7/5/2026
Haris Ampas	AUTH	25/9/2025	7/5/2026
Joris Jaruschewski	OHB	25/9/2025	7/5/2026
Hanna-Maria Tarjamo	Kuva Space	25/9/2025	7/5/2026
Giorgos Tsilimanis	ICCS	25/9/2025	7/5/2026

Approval Signature on behalf of the Executive Board		
Name	**Organization**	**Date of approval**
Koen Van Rossum	VITO	29/9/2025; 13/05/2026
Nick Berkvens	ILVO	29/9/2025
Giorgos Tsilimanis	ICCS	29/9/2025
Thomas Papakosmas	NP	29/9/2025

Revision records			
Version	**Date**	**Changes**	**Authors**
0.1	31/01/2024	First draft version	VITO et al.
1.0	30/09/2025	First version	VITO et al.
1.1	13/05/2026	Revised version	VITO et al.

Acronyms and Abbreviations

Acronyms and Abbreviations	
AI	Artificial Intelligence
API	Application Programming Interface
AOI	Area of Interest
ATB	Institut für angewandte Systemtechnik Bremen GmbH
AUTH	Aristotle University of Thessaloniki
AWS	Amazon Web Services
BIOPAR	Biophysical Parameters
BOA	Bottom of Atmosphere
CCC	Canopy Chlorophyll Content
CDS	Climate Data Store
CDSE	Copernicus Data Space Ecosystem
CWC	Canopy Water Content
DEM	Digital Elevation Model
DES	Deimos Spain
DME	DEIMOS ENGENHARIA SA
DMK	DMK Deutsches Milchkontor GmbH
DOY	Day of the Year
EC	European Commission
ECMWF	European Centre for Medium-Range Weather Forecasts
EO	Earth Observation
ERA5	ECMWF Re-analysis version 5
ET	Evapotranspiration
EURAC	Accademia Europea di Bolzano (Eurac Research)
EV ILVO	Eigen Vermogen van het Instituut voor Landbouw en Visserij Onderzoek
F(A)PAR	Fraction of (Absorbed) Photosynthetically Active Radiation
FCOVER	Fraction of Vegetation Cover
GCP	Ground Control Point
ICCS	Institute of Communication and Computer Systems
GDPR	General Data Protection Regulation
GIS	Geographic Information System
GRD	Ground Range Detected products
HF-1	Hyperfield-1
HOBE	Hydrological Open Air Laboratory
HPC	High Performance Computing

HSI	Hierarchical Storage Interface
IFAPA	Instituto Andaluz de Investigación y Formación Agraria, Pesquera, Alimentaria y de la Producción Ecológica
IFS	Integrated Forecasting System
IPR	Intellectual Property Rights
ISMN	International Soil Moisture Network
LAFIS	Informationssystem der Autonomen Provinz Bozen für Land- und Forstwirtschaft
LAI	Leaf Area Index
LAT	Latitude
LAU	Local Administrative Unit
LON	Longitude
LST	Land Surface Temperature
LUE	Light Use Efficiency
LUKE	Natural Resources Institute Finland
MIGAL	MIGAL Galilee Research Institute
ML	Machine Learning
MPC	Multi-Party Computation
MSI	Multispectral Instrument
NBR2	Normalized Burn Ratio 2
NDII	Normalized Difference Infrared Index
NDRE	Normalized Difference Red-Edge Index
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NP	Neuropublic SA
NPP	Net Primary Production
NRT	Near Real Time
OHB DS	OHB Digital Services GmbH, Bremen, Germany
OLCI	Ocean and Land Colour Instrument
PET	Privacy Enhancing Technologies
PSNC	Instytut Chemii Bioorganicznej Polskiej Akademii Nauk
PSRI	Plant Senescence Reflectance Index
R&D	Research and Development
RBF	Radial Basis Function
ResNet	Residual Neural Network
RGB	Red Green Blue
RIE	Research and Innovation Environment
RIL	Research and Innovation Lab

RTC	Radiometrically Terrain Correct
SAR	Synthetic Aperture Radar
SCL	Scene Classification Layer
SDK	Software Development Kit
SLSTR	Sea and Land Surface Temperature Radiometer
SM	Soil Moisture
SME	Small and Mid-size Enterprise
SNAP	Sentinel Application Platform
SNR	Signal to Noise Ratio
SOC	Soil Organic Carbon
SR	Super Resolution
STAC	SpatioTemporal Asset Catalog
SVM	Support Vector Machine
TOA	Top of Atmosphere
TOC	Top of Canopy
TSEB	Two-Source Energy Balance
UGent	Universiteit Gent
UTM	Universal Transverse Mercator
VIS	Visual
VITO	Vlaamse Instelling voor Technologisch Onderzoek
VRI IES	Foundation "Institute for Environmental Solutions"
VTT	Technical Research Centre of Finland Ltd.
WODR	Wielkopolski Osrodek Doradztwa Rolniczego w Poznaniu
WMS	Web Map Service
WP	Work Package
XGBoost	eXtreme Gradient Boosting

Table of Contents

List of tables

List of figures

1. Introduction

1.1 Project overview

ScaleAgData is a response to the call HORIZON-CL6-2022-GOVERNANCE-01-11 Upscaling (real-time) sensor data for EU-wide monitoring of production and agri-environmental conditions. The ScaleAgData project will run from January 2023 until 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 into how the complex data streams should be governed and organised (governance call). On the other hand, it aims to develop the data technology needed to scale data collected at the farm level to regional datasets, agri-environmental monitoring and the management of agricultural production.

To do so, ScaleAgData has five objectives:

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

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

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

1.2 Scope of the document

This document provides a description of the EO data sets that have been collected, processed and delivered to the RILs for data product development. It makes a distinction between (i) existing datasets and (ii) new or improved datasets developed in ScaleAgData (WP3).

The document should be considered as “supporting documentation” for deliverable D3.3 EO data collection, spatio-temporal preparation services (DATA).

More information on how exactly these EO data are used for data product development in the various RILs can be found in deliverables D4.3 and D4.5 (Sensor-integrated data products, v1 and v2).

1.3 Document structure

This document is structured as follows:

- Chapter 1 provides a project overview and then goes on to describe the scope, responsibilities, and structure of this deliverable.
- Chapter 2 provides an overview of the raw EO data and derived EO data products that have been collected / processed and delivered to the RILs for data product development as well as the different delivery mechanisms
- Chapter 3 focuses on the new or improved EO data products that have been developed and are made available to the RILs.
- Chapter 4 describes where the data can be found and how they are made available to the RILs

1.4 Evolution of the document

Version 1.0 of this document provided an overview of the situation on 30 September 2025.

This version (v1.1) includes updates on the collection of hyperspectral data for the project (introduction of chapter 2). It also provides more details on the timing and functional role of the EO products in the RILs (Tables 1 and 4). For a more detailed description of the use of EO data and the integration of EO and sensor data, reference is made to deliverable D4.5 (Sensor-integrated data products, v2). Finally, in chapter 4, the role of the Research and Innovation Environment (RIE) related to EO data collection and sharing is clarified and an update is given on EO product ingestion in the RIE.

History of changes	
Section	**Nature of change and reason (if applicable)**
2. Standard EO data	Table 1 is updated to provide more details on the timing and functional role of the EO products and includes minor updates reflecting recent changes in the use of these EO products in the RILs. Updates are provided on the use of hyperspectral EO data within the project.
3. New or improved EO data	Introduction is updated to explain the delay in EO product development. Table 4 is updated to provide more details on the timing and functional role of the new/improved EO products and includes minor updates reflecting recent changes in the use of these EO products in the RILs. Section 3.4 now provides updates on the framework that has been implemented for sensor data integration and that is illustrated for generating sensor-integrated soil moisture products.
4. Overview of collected EO data on the RIE	This section is updated with a clarification on the role and set up of the RIE and includes a timeline for EO product ingestion into the RIE.

2. Standard EO data

ScaleAgData Task 3.3 aims at providing analysis-ready EO data to the RILs and supporting partners (Task 5.2) via the Research and Innovation environment (RIE) (Task 4.3).

EO data from the Copernicus initiative, as well as other sources of (EO) data, are available on a wide array of platforms, in different stages of usability (e.g. atmospherically corrected or not, analysis-ready or not).

Table 1 provides an overview of the commonly available EO data (both raw data and derived products) and weather data used by the ScaleAgData partners for developing new products and services in the RILs. It also specifies the EO data providers, the thematic use case, the role of the EO data, the AOIs where the data are used as well as the time period for which the EO data are collected.

Table 1: Derived EO products used in the RILs, the EO data provider, RIL reference, the thematic use case, the role of the EO data (E=experiments, C=calibration, V=validation, I= input to analytical workflows), the AOI (area of interest) / test sites and time period of EO data collection. RIL1: Water productivity, RIL2: Crop management, RIL3: Yield monitoring, RIL4: Soil health; RIL5: Grasslands, RIL6: Dairy

EO products	EO data provider	RIL (partner)	Thematic use case	Role of EO data	AOI/test sites	Time period
Sentinel-2 products (NDVI, LAI, fAPAR, Fcover, RGB)	VITO	RIL1 (IES, MIGAL)	Vegetation health assessment	I	Latvia, Israel	2024-2025
		RIL2b (HORTA)	Water stresses assessment	I	N-Italy	2024-2025
		RIL3 (UGent, VITO, LUKE)	Crop yield estimation	C, V, I	Belgium, Finland	2022-2025
Sentinel-1/2 fused products (CropSAR1D fAPAR)	VITO	RIL3 (UGent, VITO, LUKE)	Crop yield estimation	C, V, I	Belgium, Finland	2022-2025
		RIL5 (EURAC, IFAPA, DEIMOS)	Grassland biomass estimation (input for decision support system for farmers)	C, V	S-Spain	2024-2025
		RIL6 (OHB, ATB)	Monitoring development of available biomass on grasslands	E	N-NE-Germany, parts of the Netherlands	2018-2023
Sentinel-2 Yield potential maps	VITO	RIL2b (HORTA)	Yield potential assessment	I	N-Italy	2024-2025
Sentinel-1/2 Biomass production	VITO	RIL6 (OHB, ATB)	Monitoring development of available biomass on grasslands	E	N-NE-Germany, parts of the Netherlands	2018-2023
Sentinel-2/3 Evapotranspiration	DHI	RIL1 (IES, MIGAL)	Water availability status assessment	C, V, I	Latvia, Israel	2023-2025

EO products	EO data provider	RIL (partner)	Thematic use case	Role of EO data	AOI/test sites	Time period
		RIL2b (HORTA)	Water availability status assessment	I	N-Italy	2023
		RIL3 (UGent, VITO, LUKE)	Crop yield estimation	C, V, I	Belgium, Finland	2022-2023
Sentinel-2/3 Soil moisture	DHI	RIL1 (IES, MIGAL)	Water availability status assessment	C, V, I	Latvia, Israel	2023-2025
		RIL2a (NP)	Input for spatial planning of SM sensors installation	E	Central Macedonia, Greece	2026
		RIL2b (HORTA)	Water availability status assessment	I	N-Italy	2023
		RIL3 (UGent, VITO, LUKE)	Crop yield estimation and tare estimation of potato harvest	C, V, I	Belgium, Finland	2022-2023
Sentinel-2/3 products (NDVI)	OHB	RIL6 (ATB, OHB)	Monitoring development of available biomass on grasslands	I, E	N-NE-Germany, parts of the Netherlands	2018-2024
Sentinel-1/2 products	DEIMOS	RIL5 (EURAC, IFAPA, DEIMOS)	Grassland gross primary Production	V, I	S-Spain,	2017-2024
Sentinel-2 MSI L2A reflectances	EURAC	RIL5 (EURAC, IFAPA, DEIMOS)	Estimation of grassland biophysical parameters	I (S2 LAI)	NE-Italy	2023
Sentinel-1 backscattering	EURAC	RIL5 (EURAC, IFAPA, DEIMOS)	Development of a data fusion algorithm between Sentinel-1 and Sentinel-2	I (S1 LAI)	NE-Italy	2023
Sentinel-2 / 1 LAI	EURAC	RIL5 (EURAC, IFAPA, DEIMOS)	Grassland monitoring	C, V (S1 LAI) / I (biomass)	NE-Italy, S-Spain	2023-2025
Sentinel-2 vegetation indices (NDVI, NDRE, LAI, NDWI, PSRI)	IFAPA	RIL5 (EURAC, IFAPA, DEIMOS)	Estimation of grassland biophysical parameters	C, V	S-Spain	2023-2025
Sentinel-2 derived crop classification product	NP	RIL2a (NP)	Input for parcel data regional aggregation for policy makers	I	Crete, Thessaly, Central & Western Macedonia	2022-2025
Sentinel-2 NDVI	HORTA	RIL2b (HORTA)	Input for decision support models	I	N-Italy	2023-2025
Sentinel-2 bare soil indices	ILVO	RIL4 (ILVO, AUTH)	input for estimating soil parameters	I	Flanders	2023-2025

EO products	EO data provider	RIL (partner)	Thematic use case	Role of EO data	AOI/test sites	Time period
Sentinel-2 bare soil indices	AUTH	RIL4 (ILVO, AUTH)	input for estimating soil parameters	I	Central Macedonia	2023-2025
ERA5 Reanalysis	OHB	RIL6 (OHB, ATB)	Milk quality estimation	I	Germany	2018-2026
ECMWF IFS	OHB	RIL6 (OHB, ATB)	Milk quality estimation	I	Germany	2018-2026

Most EO data listed in Table 1 are derived from Sentinel-1/2/3 multispectral or SAR sensors. During the first iteration phase of ScaleAgData, there was limited interest among the RILs in using hyperspectral EO data, such as those from the EnMAP or PRISMA satellite sensors. Only the Dairy and Soil RILs expressed some interest, with the latter considering hyperspectral data merely as a backup option in case Hyperfield data would be delayed. In addition, EnMAP faced significant challenges, including delays in satellite tasking and a low temporal resolution of the images. Universities and other public institutions also appeared to have higher priority for ordering EnMAP data compared to SMEs. As a result, none of the data requested for ScaleAgData could ultimately be acquired. During the second iteration phase, however, Hyperfield data became available for use within the RILs as described in chapter 3.

As can be seen from Table 1, the EO data used in the RILs for data product development have been collected and processed by a variety of "providers". On one side, ScaleAgData "EO and technology providers" (WP3 partners) such as Deimos, VITO and DHI have either supplied EO data directly to the RILs (DHI) or enabled RIL partners to order data through their platforms (VITO, Deimos). On the other side, RIL partners themselves—such as NP, EURAC, IFAPA, ILVO, and AUTH—have also collected and processed EO data for use within the RIL.

The sections below (2.1 to 2.12) give an overview, for each "EO data provider", of where the 'standard' EO products that are delivered to/collected by the RILs come from and how they are processed.

Chapter 4 will describe how the analysis-ready EO data are made available via the Research and Innovation Environment (RIE). When EO data providers (ScaleAgData partners) have direct access portals that can be used by the RIL partners, this is also described in the sections below.
More details on the use of the EO data and more specifically on the integration of EO and sensor data to develop data products for agri-environmental monitoring, can be found in deliverable D4.5 "Sensor-integrated data products, v2".

2.1 Sentinel-1/2 vegetation indices, biophysical parameters, phenology, biomass products provided by VITO

2.1.1 Via Terrascope

The EO data provided by VITO to the RILs are generated from Sentinel-2 Level 1C (TOA) or 2A (TOC) products and/or Sentinel-1 GRD products that are retrieved from ESA's Science Hub.

Depending on the service, data processing may include

- For optical data: atmospheric correction, cloud and shadow detection, calculation of derived indices or parameters for agricultural monitoring
- For SAR data: Orbital correction, radiometric and thermal noise removal, radiometric calibration
- For fused products: specific algorithms to derive parameters for agricultural monitoring

The RIL partners can order VITO EO products via the Terrascope EOplaza (https://portal.terrascope.be/catalogue) after registration.

Registration includes the following steps:

- create an account and sign in: https://docs.terrascope.be/#/Developers/EOplaza/Start?id=are-you-interested-in-using-or-publishing-services
- share your username and email address with VITO. With this information, VITO will set up an umbrella organization for you. Following that, we will send you an invitation to join this organization. Afterwards, you can add other colleagues to this organization, if needed. We will assign a number of credits to this umbrella organization that can be used to order EO products. This means that the products will be free for ScaleAgData users.

After registration, ScaleAgData partners can use all services available in the EOplaza catalogue page https://portal.terrascope.be/catalogue. (see Figure 1)

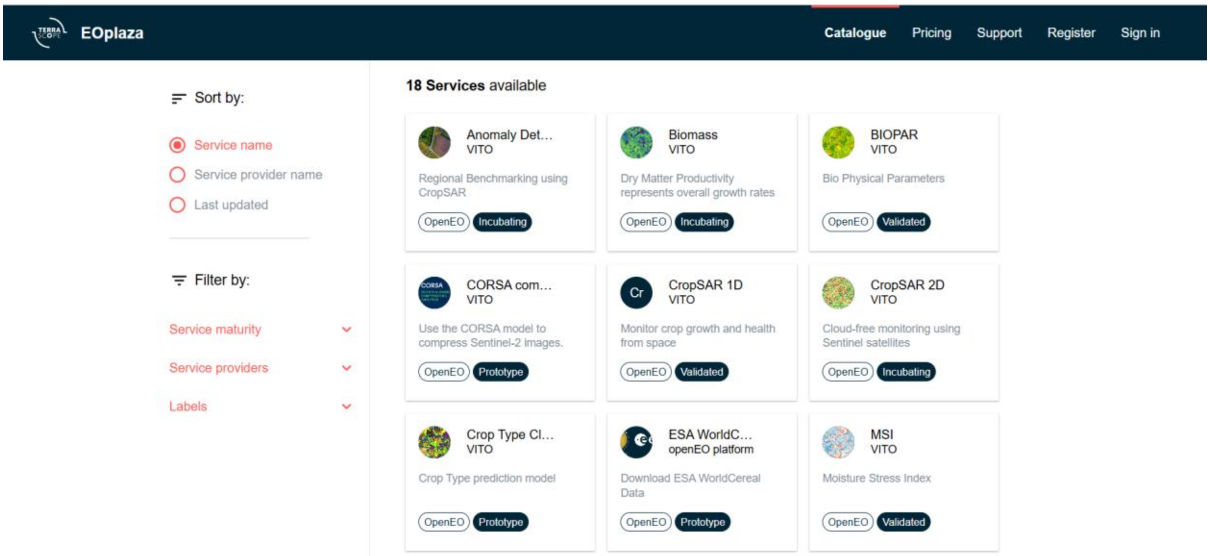
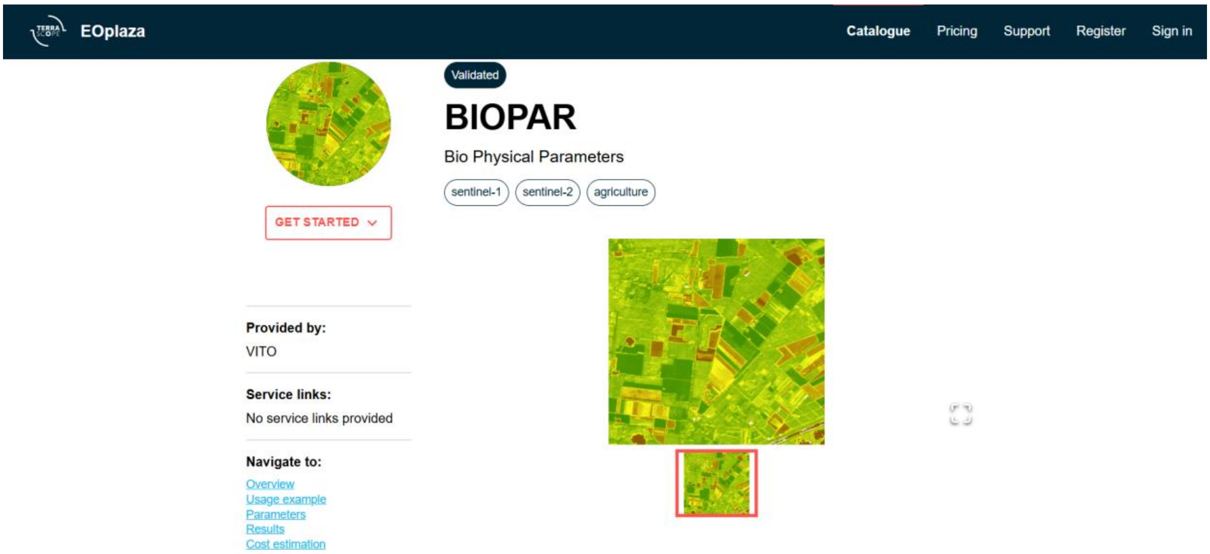


Figure 1: VITO's EOPlaza catalogue

When clicking on any of these services, you will be redirected to the service details page where you can find information about the service such as a general description and instructions on how to execute the service. Figure 2 shows the example of the BIOPAR service page.



BIOPAR

Bio Physical Parameters

Provided by:
VITO

Service links:
No service links provided

Navigate to:
Overview
Usage example
Parameters
Results
Cost estimation

Figure 2: BIOPAR service details page

More information on the services that are used by the ScaleAgData partners in the RILs can be found here:

- Sentinel-2 NDVI: https://portal.terrascope.be/catalogue/app-details/14
- Sentinel-2 biophysical parameters LAI, fAPAR, Fcover, CCC, CWC: https://portal.terrascope.be/catalogue/app-details/21
- Sentinel-1 and Sentinel-2 fused products - CropSAR Fapar: https://portal.terrascope.be/catalogue/app-details/10
- Sentinel-2 biomass estimates: https://portal.terrascope.be/catalogue/app-details/17
- Sentinel-2 vegetation water indices - MSI: https://portal.terrascope.be/catalogue/app-details/11; NDWI: https://portal.terrascope.be/catalogue/app-details/13; NDII: https://portal.terrascope.be/catalogue/app-details/12
- Yield potential maps: https://portal.terrascope.be/catalogue/app-details/33
- Phenology - Start, peak, end of season: https://portal.terrascope.be/catalogue/app-details/80

2.1.2 Via OpenEO and CDSE

The RILs that are using the Presto framework for few-shot learning that has been set up in WP4 can get access to the expected Sentinel-1 and -2 input EO data via the OpenEO platform.

In the following notebook on the project's GitHub repository, the EO data extraction and the requirements for running the extractions are described:
https://github.com/ScaleAGData/scaleag-vito/blob/main/notebooks/few_shot_learning_dummy_datasets.ipynb

It requires setting up an account in the Copernicus Data Space Ecosystem (CDSE). After signing up for free, you will have a monthly budget available of 10,000 credits.

The EO data that are extracted to be used by Presto are decadal/monthly time series of:

- Sentinel-2 L2A data (all bands)

- Sentinel-1 VH and VV

In addition, the following datasets are also extracted:
- Average air temperature and precipitation sum derived from AgERA5
- Slope and elevation from Copernicus DEM

2.2 Sentinel-2/3 evapotranspiration and soil moisture provided by DHI

DHI uses the following input data:

- Data acquired by the MSI sensor on Sentinel-2 satellites and SLSTR and OLCI sensors on Sentinel-3 satellites:
  - Sentinel-2: Downloaded as L1C S2 product, used as input to Sen2Cor atmospheric correction model to obtain 20-m top-of-canopy (TOC) reflectance.
  - Sentinel-3: The Sentinel-3 Synergy SY_2_SYN product combines the surface reflectance from shortwave optical bands on OLCI and SLSTR instruments. The Sentinel-3 SLSTR L2 product contains Land Surface Temperature (LST), which is required to characterize energy fluxes at the land surface, including evapotranspiration (ET).
- World Cover Data (ESA) ESA WorldCover 10 m 2021 v200 (zenodo.org)
- DEM - Copernicus 30 m DEM (COP-DEM GLO-30 European Space Agency and Airbus, 2022, Copernicus Digital Elevation Model - Copernicus Contributing Missions Online). The DEM is used during atmospheric correction of Sentinel-2 reflectance, during sharpening of Sentinel-3 LST and in terrain correction of meteorological parameters. The change of DEM is expected to have minimal impact on the final ET maps but fits with the aim of using mainly Copernicus data in the study
- ERA5 Reanalysis - ERA5 meteorological data (Hersbach et al., 2020, The ERA5 global reanalysis - Hersbach - 2020 - Quarterly Journal of the Royal Meteorological Society - Wiley Online Library) from Copernicus Climate Change Service

The following EO products are derived from these data:

Evapotranspiration:
Atmospherically corrected and cloud masked shortwave multispectral observations from Sentinel-2 satellites are used to determine surface biophysical properties (green and total leaf area index, albedo, etc.) at 20 m spatial resolution. The thermal infrared Sensor on board Sentinel-3 satellites provides LST observations that are sharpened from the native resolution of around 1 km to 20 m using Data Mining Sharpener (Gao et al., 2012; Guzinski and Nieto, 2019) with Sentinel-2 reflectance and digital elevation model (DEM) derived parameters used as explanatory variables.

Meteorological forcing comes from the ERA-5 reanalysis (Hersbach et al., 2020) as provided by Copernicus Climate Change Service and is subsequently corrected for topographical effects using a DEM at 300 m resolution before being resampled to 20 m resolution using bilinear interpolation. Finally, surface parameters that cannot be directly determined from Sentinel-2 data (e.g. vegetation height, canopy clumpiness or leaf angle orientation) are set using a land-cover map and a look-up table.

A biophysical model (TSEB) using the above input layers is then used to derive ET.

Soil moisture:

A water balance model driven by the EO ET-derived layers is used to compute the DHI soil moisture products.

ET and soil moisture products are distributed as GeoTiff files, accessible through Amazon S3 buckets or downloadable from Azure Blob Storage. DHI provides the user with the relevant keys to access Azure Blob Storage and Amazon S3 buckets.

2.3 Sentinel-2/3 products provided by OHB

The EO data provided by OHB DS to the RILs, are derived from Sentinel-2 Level L2A and Sentinel-3 SLSTR data. These are gathered from different sources such as CDSE (Copernicus Data Space Ecosystem) but also the Planetary Computer platform.

The processing may encompass the following options:

- diverse masking options of clouds, scenery, static scenery, land use (Copernicus data, regional datasets (non-EO))
- calculating of derived indices or parameters for agricultural monitoring
- calculating statistics for masked EO data
- potential downscaling algorithms for LST data

2.4 Sentinel-1/2 products provided by Deimos

The EO data acquired by DEIMOS for the RIL – Grasslands are Sentinel-2 Level 2A or Sentinel-1 GRD. They were retrieved from the Copernicus Browser (https://browser.dataspace.copernicus.eu/).

EO data processing:

- Sentinel-2 Level 2A: Image resizing to the area of interest, pixel resampling, adaptation of the data to the footprint of the ground sensor
- Sentinel-1 GRD: correction, radiometric and thermal noise removal, radiometric calibration, resizing to the area of interest, pixel resampling, and adaptation of the data to the footprint of the ground sensor.

2.5 Sentinel-1/2 LAI product collected and processed by EURAC

The EO data provided by EURAC are Sentinel-2 MSI Level 2A reflectances retrieved from the Copernicus Data Space Ecosystem (https://dataspace.copernicus.eu/). The Sentinel-1 RTC (radiometrically terrain corrected) gamma-nought backscatter coefficients from ascending and descending orbits are retrieved from the Microsoft Planetary Computer (https://planetarycomputer.microsoft.com/dataset/sentinel-1-rtc).

EO data processing:

- Sentinel-2 Level 2A: tiles T32TPR, -TPS, -TPT, -TQS and -TQT covering the study area are resampled to 10 m, the LAI is calculated using SNAP Biophysical Processor using the 20 m neural network, cloud-masking is performed.
- Sentinel-1 RTC: is processed to RTC by Catalyst from GRD backscatter in IW mode (radiometrically calibrated, radiometrically terrain corrected, orthorectified), Lee Sigma speckle filtering is applied, logarithmic scaling to dB and index calculation are performed, data are extracted over the polygons.

- Sentinel-1 (predicted) LAI:
    - Model Training/Validation/Test: a domain specific transformer-based convolutional network was developed to fill temporal gaps using Sentinel-1 RTC with ancillary information (Soil Moisture, Topographical Classes, DOY) and S2 LAI as a target.
      We have selected 10% of the polygons classified as permanent meadows based on the agricultural information system of the province of Bolzano (LAFIS, Informationssystem der Autonomen Provinz Bozen für Land- und Forstwirtschaft) stratified across eight types of grasslands meadows (altitudinal) for the training/validation/test. Out of which 55% for training, 5% for validation and 40 % for the test.
    - Model Prediction: The S1 LAI was predicted over all the 33601 LAFIS polygons of the province of Bolzano between April 2023 and October 2023.We have also predicted the S1 LAI between April 2023 and October 2023 over the eight field sites in the provinces of Bolzano and Trento where field measurements of grassland biophysical parameters were performed during the first two project years.l .

2.6 Sentinel-2 biophysical parameters collected and processed by IFAPA

The EO data acquired by IFAPA for their research activities in the RIL – Grasslands are Sentinel-2 MSI Level 2A, which are retrieved from the image collection 'COPERNICUS/S2_HR_HARMONIZED'. Dataset provided by the European Union/ESA/Copernicus.

EO data processing:
Sentinel-2 Level 2A reflectances with 10 days and 10 meters of temporal and spatial resolutions, respectively, were used on a field scale. The data were first filtered by the area of interest, date range, and cloud cover (<30%). During cloud masking, we filtered the original imagery to remove any pixels obscured by clouds. The data were resampled and adapted to field-level AOIs, including *dehesa* pilot farms and flux tower footprints in the northern area of Córdoba province (southern Spain).

Calculation of biophysical parameters:
This entails the calculation of various indices such as NDVI, which is derived from Sentinel-2A imagery after cloud masking and resampling. NDVI time series over the *dehesa* are used to estimate the fraction of photosynthetically absorbed radiation (fPAR), after subtracting the contribution of tree canopy to isolate the herbaceous layer. In addition, externally derived products of LAI and fPAR from Terrascope and the Land Monitoring Service were integrated for comparison and validation. The combination of NDVI, fPAR and local meteorological data (radiation, temperature, vapor pressure deficit) allows the implementation of an adapted Light Use Efficiency (LUE) model, which relates incident solar radiation to the photosynthetic activity of the vegetation. This adaptation produces daily estimates of net primary production (NPP) and its accumulated version (NPPa), which are provided as raster time series (GeoTIFF). These products are used both for grassland monitoring and for validation against flux tower and field measurements.

2.7 Sentinel-2 vegetation indices and crop classification product collected and processed by NP

2.7.1 Vegetation indices

The Earth Observation (EO) data used by NP for the research activities in the RIL – Crop monitoring are retrieved and processed through the Sentinel Hub platform using custom Evalscript-based workflows. The data source is Sentinel-2 Level-2A products, which provide atmospherically corrected surface reflectance. These products are accessed via the Copernicus Data Space Ecosystem, targeting specific tiles that cover the RIL2a study areas: Thessaly and Crete. The data preparation pipeline is implemented through Sentinel Hub's Process API, utilizing Evalscripts to define band selection, masking logic, and index computation. The processing script performs the following steps :

- Retrieves relevant bands (e.g., B04 [Red], B08 [NIR], SCL [Scene Classification Layer], and dataMask).
- Computes vegetation indices such as NDVI: `NDVI = (B08 - B04) / (B08 + B04)`
- Applies conditional masking to exclude invalid pixels:
    - Pixels where the denominator (B08 + B04) is zero are masked out.
    - Pixels classified as open water (SCL = 6) are excluded.
    - The native Sentinel Hub `dataMask` is used to ensure cloud-free and valid observations.

The Evalscript outputs both the index values and an associated binary `dataMask`, ensuring only valid, cloud-free, and non-water pixels are used in downstream analysis. The temporal coverage corresponds to the crop growing season, and data are aggregated at a daily interval (P1D). The resulting datasets are resampled to a defined spatial resolution and can be further used for classification and phenological profiling.

In addition to NDVI, similar scripts will be applied to compute other spectral indices such as:

- NDRE (Normalized Difference Red Edge)
- NDWI (Normalized Difference Water Index)
- PSRI (Plant Senescence Reflectance Index)

2.7.2 Sentinel-2 derived crop classification product

The data used to produce the crop classification product included both Earth Observation (EO) imagery and ancillary vector data. Specifically, a supervised Support Vector Machine (SVM) classifier with a Radial Basis Function (RBF) kernel was trained using parcel-level data, which contained polygon geometries, crop type labels, and information on the start and end of the growing season. The EO data comprised filtered and cloud-masked Level 2A, Sentinel-2 imagery collected during the crop growth period, utilising selected spectral bands: B2 (blue), B3 (green), B4 (red), B5, and B6 (red-edge). Additionally, local administrative boundaries (LAU) and land use datasets were used to set the mask area.

2.8 Sentinel-2 NDVI collected by HORTA

The EO data retrieved by Horta for its research activities in the RIL 2b – Crop management are Sentinel-2 data, which are retrieved from a commercial provider (Airbus). The provider delivers biophysical parameters, such as NDVI (Normalised Difference Vegetation Index) automatically processed, de-clouded, clipped to the field and provided via API.

2.9 Sentinel-2 indices collected and processed by ILVO

The EO data acquired by EV ILVO for their research activities in the RIL – Soil Health is Sentinel-2 MSI Level 2A, which is retrieved from Google Earth Engine. In this way we can use the ample computer processing capabilities of the Google Earth Engine platform to preprocess large volumes of EO data before further processing in our data pipelines. The extraction and preprocessing data pipeline can be found in the ScaleAgData GitHub at https://github.com/ScaleAGData/SOIL-RILAB/blob/main/Googe_earth_engin_extract_sentinel2_data.py Sentinel-2 tiles have been acquired for the whole region of Flanders during multiple years.

Processing contains the following steps: filtering the original imagery to remove any pixels obscured by clouds (cloud masking), applying computational indices (e.g. NDVI, NBR2, VNSIR, …) to identify and exclude pixels containing green vegetation, dry vegetation, and areas with high soil moisture content, as these factors can alter the soil's spectral signature (vegetation and moisture masking), employing the ESA Worldwide land cover map to further refine the collection by selecting only pixels classified as cropland or grassland (land cover classification). A ML model is then developed correlating the "reflection signatures" (unique light reflection patterns) extracted from the masked EO imagery and the corresponding Soil Organic Carbon (SOC) measurements obtained from the field. In a final step the deployed model is applied to all remaining EO pre-processed pixels for the Flemish region after which we aggregate the SOC from pixel level to parcel level, on presenting the information to the service consumers or the end-users; the final product is a SOC map for Flanders.

We will equally attempt to use EO data provided by Kuva Space and described in section 3.1. A similar methodology will be applied for the Kuva Space EO data. Pre-processing of the Kuva Space data is planned to be done in the Research and Innovation Environment of DEIMOS.

2.10 Sentinel-2 bare soil indices collected and processed by AUTH

Sentinel-2 L2A (i.e., reflectance) data have been downloaded from the Copernicus Data Space Ecosystem, or alternate sources (e.g., Google Cloud) for tiles which are not feasible to retrieve from the long term storage. Tiles from the region of Central Macedonia (indicatively, T34TFL, T34TGL, T34TGK, T34TFK, T34TEL, T34TEK) have been downloaded across multiple years and aggregated in order to form the bare soil reflectance composites.

The data are processed to identify bare soil on each scene, and are subsequently aggregated (i.e., mean across the temporal dimension) to form bare soil mean reflectance composites which present the spectral reflectance of the soil at each pixel. The processing entails the calculation of various indices (e.g., NDVI and NBR2), the use of the built-in scene classification layer of L2 products provided by ESA, and other layers of information (e.g., CORINE Land Cover, or others).

Following the generation of the bare soil mean reflectance composites, selected points are extracted whose soil properties are known, and AI models are subsequently developed to correlate the spectra with the soil properties. The models will be applied on the whole area to derive soil property maps of the region.

2.11 ERA5 Reanalysis meteorological data from GCP / CDS provided by OHB

ERA5 reanalysis data is produced by ECMWF and made available through platforms such as the Copernicus Climate Data Store (CDS) and Google Cloud Platform (GCP). To enable simple access to the RIL partners and efficient extraction of comprehensive time series, the data is provided in one unified store in cloud-optimised Zarr format with enhanced chunking especially in time dimension. Further, spatial and pressure level chunking is optimised for the usage in individual RILs. To enable NRT applications, such as milk quality forecasting in RIL6 Dairy, the store is continuously updated.

Table 2: ERA5 Reanalysis meteorological data from GCP/CDS provided by OHB

Format	Parameters	Dimensions / Extend	Chunking	Location
.zarr (v2)	On request from: Temperature Humidity Precipitation Wind speeds Soil temperature ...	**Spatial: full Europe** Latitude: [-10°,36°] Longitude: [34°,70°] Resolution: 0.25° **Time** 1st Jan 2018 to 25th Aug 2025 Hourly **Pressure Levels** On request, 1-1000hpa **SoilLayer** On request, 1-4	Lat: 64 Lon: 64 Time:1024 PLevel: 1 (or other)	RIE Cloud Storage OHB

2.12 ECMWF IFS weather forecast data from ECMWF / AWS provided by OHB

ECMWF IFS forecasts are produced by ECMWF and are initially made available through the ECMWF Data Store and Open Data portal, recurrently covering data of forecast runs of the last 4 days. However, the AWS Marketplace redistributes the same forecast data with access to a much longer historical range. All providers offer the datasets in GRIB2 format for each forecast run and step. GRIB2 does not offer any chunking, range requests or download of subsets in spatial dimensions and is therefore, inefficient for cloud-native applications, as the full dataset with worldwide coverage needs to be processed. Further, the forecast horizon, available parameters and dimensions vary across the forecast dates. Addressing this lack, within ScaleAgData, IFS data is collected, assimilated, and merged into one common Zarr store with optimised chunking in all dimensions of the dataset. For interoperability, it shares as close as possible specifications as the ERA5 Zarr store from the previous section.

IFS forecast data is made accessible from November 2024 onwards, but can be extended further into the past if requested. For products or applications based on meteorological forecast data, such as milk quality forecasting in RIL6 Dairy, historical forecasts enable a retrospective analysis of the performance. Additionally, historical ECMWF IFS time series can be used in NRT applications to complement ERA5 time series, which are subject to a 5-day publication delay.

Table 3: ECMWF IFS weather forecast data from ECMWF / AWS provided by OHB

Format	Parameters	Dimensions / Extend	Chunking	Location
.zarr (v2)	Typical params as: Temperature Humidity Precipitation Wind speeds Soil temperature ...	**Spatial: full Europe** Latitude: [-10°,36°] Longitude: [34°,70°] Resolution: 0.25° **Time** 13th Nov 2024 to 31th Aug 2025 Twice daily 00z and 12z **Step: full 360h forecasts** Step: 0-85 (0h-360h) **Pressure Levels** On request, 1-1000hpa **SoilLayer** On request, 1-4	Lat: 64 Lon: 64 Time:1 Step:85 PLevel: 1 (or other)	RIE Cloud Storage OHB

3. New or improved EO data

In addition to the commonly used EO data, described in Chapter 2, ScaleAgData partners in the RILs have also explored new sources of EO data for product development such as hyperspectral data from Hyperfield, provided by Kuva Space. More details can be found in section 3.1.

Based on the users' needs, innovative methods have been developed for improving the EO products that were used during the first iteration phase for product development in the RILs. The requirements only became apparent at the end of the first iteration, following initial testing with standard EO products. Consequently, the development of improved EO products was initiated only during the second iteration.

Methods have been developed for improving the spatial and temporal resolution of the data or for cloud-filling. APIs have been developed to provide the RILs and partners with EO data with a resolution of less than 10m and/or at daily timesteps, independent of cloud cover. These new methods and products are presented in sections 3.2 to 3.4.

Table 4 below provides an overview of the new and/or improved EO products (raw data and derived products) developed by the ScaleAgData technology providers and made available to the RILs for evaluation and/or development of new products & services in the Labs.

Table 4: New or improved EO products that are used in the RILs, the EO data provider, the thematic use case, the role of the EO data (C=calibration, V=validation, I=model input) and the AOI (area of interest) / test sites and time period for which the EO data are collected. RIL1: Water productivity, RIL2: Crop management, RIL3: Yield monitoring, RIL4: Soil health, RIL5: Grasslands, RIL6: Dairy

EO products	EO data provider	RIL (partner)	Thematic use case	Role of EO data	AOI/test sites	Time period
Hyperfield data	Kuva Space	RIL4 (ILVO, AUTH)	input for estimating soil parameters	I	Flanders, Central Macedonia	2025 - 2026
Sentinel-2/3 Super Resolution products	ICCS	RIL3 (UGent)	Potato tare estimation	V	Belgium	2024
	ICCS	RIL3 (VITO, LUKE)	Crop yield estimation	V	Belgium, Finland	2024-2025
Sentinel-1/2 fused products (CropSAR 2D Fapar)	VITO	RIL3 (UGent, VITO, LUKE)	Crop yield estimation	C, V, I	Belgium, Finland	2022-2025
		RIL5 (EURAC, IFAPA, DEIMOS)	Grassland biomass estimation (input for decision support system for farmers)	C, V	S-Spain	2024-2025
		RIL5 (EURAC, IFAPA, DEIMOS)	Grassland gross primary Production estimation	V, I	S-Spain	2017-2024
Sentinel-2/3 Sensor-	DHI	RIL1 (IES, MIGAL)	Water availability status assessment	C, V, I	Latvia, Israel	2024-2025

integrated soil moisture	DHI	RIL2b (HORTA)	Water availability status assessment	C, V	N-Italy	2023-2024

The sections below (3.1 to 3.4) provide for each EO data provider a description of the new EO products, the methodology used to generate these products and how the RILs can access them. These new / improved EO products are also made available via the Research and Innovation Environment (RIE) as described in Chapter 4.

More details on the evaluation of these improved EO data for potential use within the RILs can be found in deliverable D4.5 "Sensor-integrated data products, v2".

3.1 Hyperfield data provided by Kuva Space

Since the start of the ScaleAgData project, Kuva Space has diligently worked to develop the capability to provide Earth observation services as part of ScaleAgData. The company's first satellite mission Hyperfield-1A was launched successfully with SpaceX in August 2024, after having been postponed by more than half a year by SpaceX due to technical issues on their side. Several months of careful commissioning, calibration and data validation activities followed the launch. A rigorous commissioning phase is critical for ensuring a satellite performs optimally and provides high-quality data for the intended purposes.

The satellite entered into operation in January 2025. It collects hyperspectral data in the VIS and NIR wavelength ranges (475-950 nm), from an altitude of about 500 km providing 50x50 km 2D-snapshot images. The satellite performs reliably, consistently acquiring approximately 50 GB of hyperspectral data per day.

Kuva Space's second hyperspectral satellite, Hyperfield-1B, also covering the VIS-NIR spectral range, was successfully launched in late June 2025. Thanks to the previous experience, the commissioning period - initially focused on hardware and communications validation, followed by full lunar cycles dedicated to spectral calibration - was finalised in just a couple of months. The satellite is fully operational at the time of submission of this report.

Product description and delivery

Hyperfield satellites are capable of providing two major observation modes. In scanning mode, they continuously acquire images in a systematic manner. These images are then stitched together to produce the final product. The second mode available is the staring mode which on demand can take pictures of specific locations with the camera pointing towards the target for a more prolonged period of time. This comes with two benefits. In the first place, the system can be tasked to point off-nadir to a target it might not be exactly passing over. This reduces the time between the request and the image acquisition. Secondly, in staring mode, there is more time available for the imager to acquire additional frames, at each wavelength, that can be used to increase either the signal-to-noise-ratio (SNR) or the resolution of the images.

The Kuva Space processing pipeline is capable of producing all higher-level products in less than 2 minutes after they become available at the ground station. The company strives to provide newly acquired images within 6 hours of their acquisition in the first six months after each satellite's commissioning. The company's goal is to reduce this delivery time to less than 2 hours by the end of 2025.

Considering HF-1 orbit, nadir-pointing, and a swath of 50km, the following revisit times can be achieved with one satellite, while the revisit times vary greatly:

- Between 0°N and 25°N (e.g. between the equator and e.g. Dubai, UAE): 23 days
- At 50°N (e.g. Frankfurt, Germany): 7 to 8 days
- At 75°N (e.g. Kullorsuaq, Greenland): 5 to 6 days

Off-nadir pointing allows for a higher revisit time. As always in earth observation, weather conditions impact the frequency of acquisitions.

Kuva Space offers the following HF-1A data level products:

Table 5: Hyperfield 1A product description

Product level	Product Description
1B	Fully stitched hypercube containing all the VIS and NIR camera bands. Images are provided as radiance or Top Of Atmosphere (TOA) reflectance.
1C	Orthorectified hypercube. Orthorectified and re-projected to a coordinate system associated with the closest Universal Transverse Mercator (UTM) zone. Georeferencing is available via Ground Control Point (GCP) on the GeoTIFF metadata. Images are provided as radiance or Top Of Atmosphere (TOA) reflectance.
2A	Atmospherically corrected hypercube resulting from applying atmospheric correction to the Level 1C product. Images are provided as Bottom Of Atmosphere (BOA) surface reflectance.
2B	Index product - Level 2A product with application-specific arrays to be defined by the user. Examples include biochemical and geophysical indices.

All HF-1 products have a width of 50 km and a length that is one or more multiples of 50 km. The amount of detail in each pixel is expressed in units of bits (12 bit). The data products are available as a zip file containing the GeoTIFF together with a custom sidecar metadata file adapted to the needs of the mission. The use of GeoTIFFs makes it possible to use and visualise the hyperspectral data straightforwardly, without having to preprocess or repack the image into a usable format. It also allows the images to be dropped into common GIS tools. The HSI file will be accompanied by additional TIFF files containing cloud and bad pixel masks.

Kuva Space has designed and implemented a data access platform. After registration and account validation by an administrator, the user is invited to search the data catalogue according to three distinct filters: (1) spatial, (2) temporal, and (3) cloud coverage-based. Data products can be downloaded directly via the platform in compressed ZIP format.

Users can also access, explore and fetch data products in the same manner via an API.

ScaleAgData project's requirements

The requirements of the ScaleAgData project have been carefully managed throughout the entire preparation period and will continue to be diligently overseen. Project partners have been consistently informed about the specifics of Kuva Space's data offerings and delivery plans.

Partners have entered their requested areas of interest into a table provided on the project SharePoint, along with descriptions of their intended use and accompanying GeoJSON files. A meeting was organised with ILVO and AUTh to discuss their needs for SOC maps over Flanders and Central Macedonia. Similarly, VTT's requirements for vegetation health monitoring in Finland were discussed.

The project partners' Areas of Interest (AOIs) are included in Kuva Space's acquisition programme and the automatic satellite scheduler, within the tasking limitations of the early operational period. Data collection commenced in these areas in spring and for specific partner needs based on their planned use cases. To date, the satellite has successfully captured 50 cubes of hyperspectral data for the project.

The partners have been given access to Kuva Space's service platform Kuva Sense™ and they have been able to download the data acquired so far for them. Their AOIs remain in the satellite scheduler until the end of their given period of interest.

3.2 Sentinel-2/3 Super Resolution products provided by ICCS

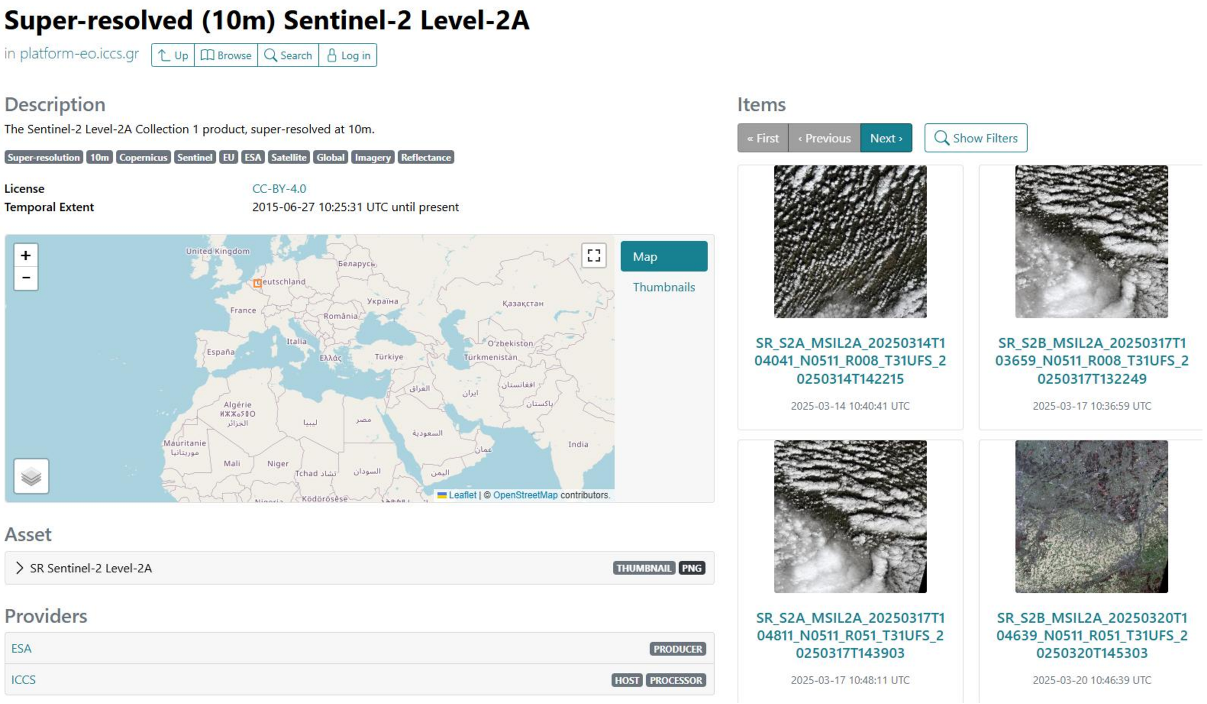
A suite of Super-Resolution (SR) tools has been developed to enhance the spatial detail and analytical value of satellite-based EO products. This suite comprises two distinct tools, each designed to address specific EO data sources and application domains:

- The **Sentinel-2 Super Resolution** tool is designed to enhance the spatial resolution of the coarser 20 m and 60 m spectral bands by integrating them with the finer 10 m bands. This approach improves spatial detail while preserving the original spectral fidelity of the data, enabling more accurate monitoring of vegetation, land cover, and agricultural features.
- The **Sentinel-3 SLSTR Thermal Sharpening** tool applies advanced sharpening techniques to integrate thermal data from the Sentinel-3 SLSTR instrument with super-resolved Sentinel-2 imagery. The result is high-resolution land surface temperature maps at 10 m resolution, supporting more precise environmental monitoring and heat-related analyses such as evapotranspiration modelling.

The Sentinel-2 and Sentinel-3 SR tools have been made available to interested RILs via APIs. For both services, users are required to define an AOI and a temporal window. The principal limitation across services remains the availability of input EO data from the respective satellite imagery data providers.

Comprehensive details on the methodology and outputs of these SR tools is provided in Deliverable D4.5.

A dedicated platform (https://platform-eo.iccs.gr/stac-browser/?.language=en) has been developed to give the users simple and efficient access to the enhanced EO products generated within the project. Through this platform, users can easily search for and download products by applying spatial and temporal filters that match their needs.



Super-resolved (10m) Sentinel-2 Level-2A

in platform-eo.iccs.gr

Description

The Sentinel-2 Level-2A Collection 1 product, super-resolved at 10m.

Super-resolution 10m Copernicus Sentinel EU ESA Satellite Global Imagery Reflectance

License	CC-BY-4.0
Temporal Extent	2015-06-27 10:25:31 UTC until present

Asset

SR Sentinel-2 Level-2A — THUMBNAIL PNG

Providers

ESA — PRODUCER

ICCS — HOST PROCESSOR

Items

SR_S2A_MSIL2A_20250314T104041_N0511_R008_T31UFS_20250314T142215
2025-03-14 10:40:41 UTC

SR_S2B_MSIL2A_20250317T103659_N0511_R008_T31UFS_20250317T132249
2025-03-17 10:36:59 UTC

SR_S2A_MSIL2A_20250317T104811_N0511_R051_T31UFS_20250317T143903
2025-03-17 10:48:11 UTC

SR_S2B_MSIL2A_20250320T104639_N0511_R051_T31UFS_20250320T145303
2025-03-20 10:46:39 UTC

Figure 3: ICCS EO Platform

The platform is built on SpatioTemporal Asset Catalog (STAC) specifications, a widely recognized standard for structuring geospatial data. In practice, this means that a STAC catalog acts as the entry point, a STAC collection represents a dataset (e.g., Sentinel-2 SR), and a STAC item corresponds to an individual EO product. This clear and consistent structure makes it possible to explore multiple datasets, refine searches by AOI and timeframe, and directly retrieve the products of interest.

To make the functionality more tangible, a demo Jupyter Notebook has also been created. It walks users through the entire process: from defining the STAC API endpoint and specifying the AOI geometry (with the support of a map visualization), to querying the API for available products, and finally downloading the selected assets. The Jupyter Notebook serves both as a hand-on example and a practical guide for users who wish to integrate the platform into their workflows.

3.3 Sentinel-1/2 fused products (CropSAR2D) provided by VITO

In 2019, VITO released CropSAR, an innovative technology that fuses Sentinel-1 radar and Sentinel-2 optical satellite data to generate cloud-free time series of key biophysical parameters such as NDVI, fAPAR and fCover at field level. These CropSAR1D products are available for ScaleAgData partners as described in section 2.2.1.

In the meantime, the CropSAR technology has been upscaled from field level to pixel level so that we can now also generate cloud-free image time series at 10 m resolution every 5 days (see Figure 4).

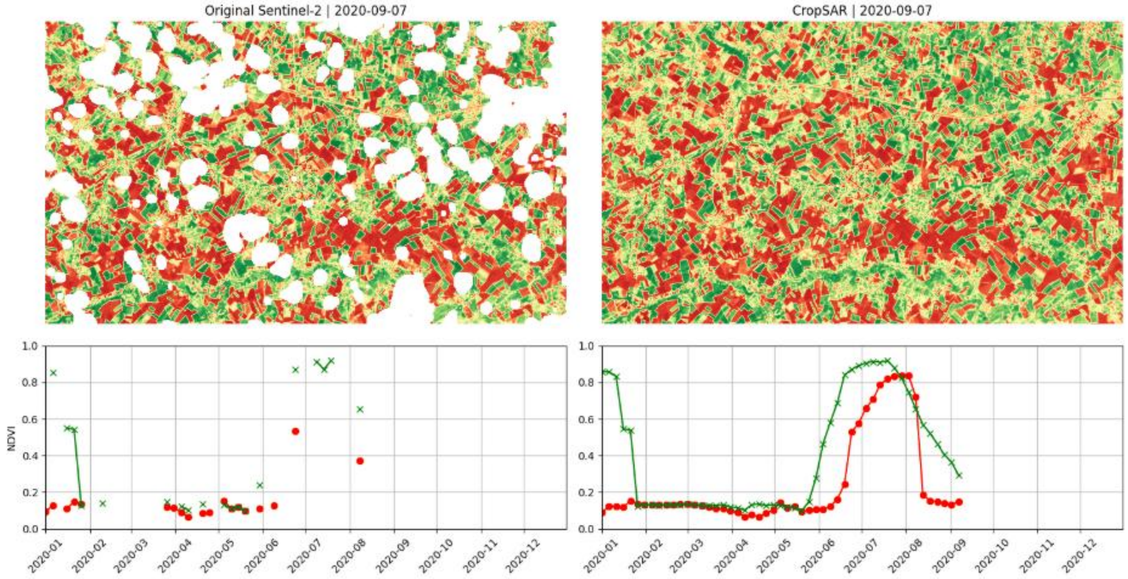


Figure 4: CropSAR 2D service: comparison of native Sentinel-2 with frequent cloud cover (left panel) and CropSAR generated equivalent without clouds (right panel).

The CropSAR2D algorithm uses a deep learning model that combines ResNet blocks (per-image spatial encoding/decoding) and transformers (per-pixel temporal gap-filling) to generate a spatially and temporally consistent time series of the requested output (Figure 5). It uses collocated Sentinel-1 and Sentinel-2 data cubes as input and returns a completely cloud-free Sentinel-2-like output.

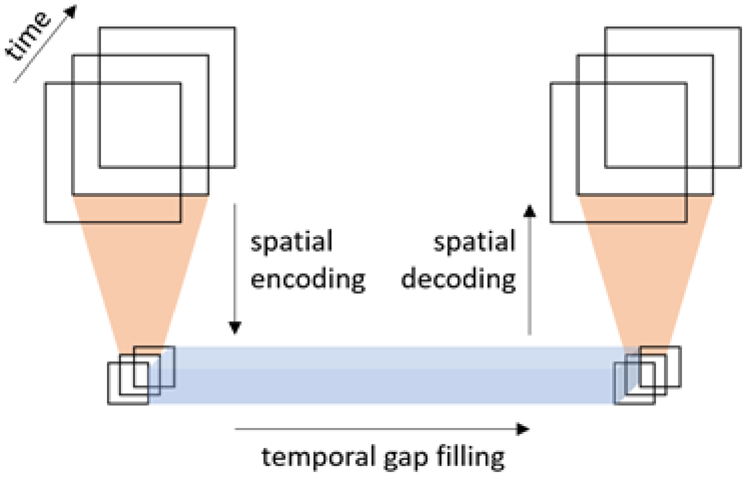


Figure 5: Conceptual CropSAR 2D architecture decoupling spatial and temporal encoding and decoding for optimal reconstruction of clouded images.

The upscaled "CropSAR2D" service enables continuous, high-resolution crop performance monitoring at sub-field level. Next to the focus on biophysical parameters such as fAPAR, fCover and NDVI, it also generates cloud-free RGB-NIR image time series. Beyond agricultural applications, the CropSAR2D service is also well-suited for a wide range of use cases that require consistent vegetation time series across large areas.

The CropSAR2D service is available for ScaleAgData partners through Terrascope (https://portal.terrascope.be/catalogue/app-details/86).

3.4 Sentinel-2/3 sensor integrated soil moisture provided by DHI

In ScaleAgData (WP4) DHI has developed a methodological framework that can be used to enhance soil moisture estimation by integrating different data sources, including in situ soil moisture sensors and Earth Observation (EO) datasets, using machine learning (ML) methods. A reproducible codebase is provided to facilitate data integration pipelines for different downstream data integration tasks and applications.

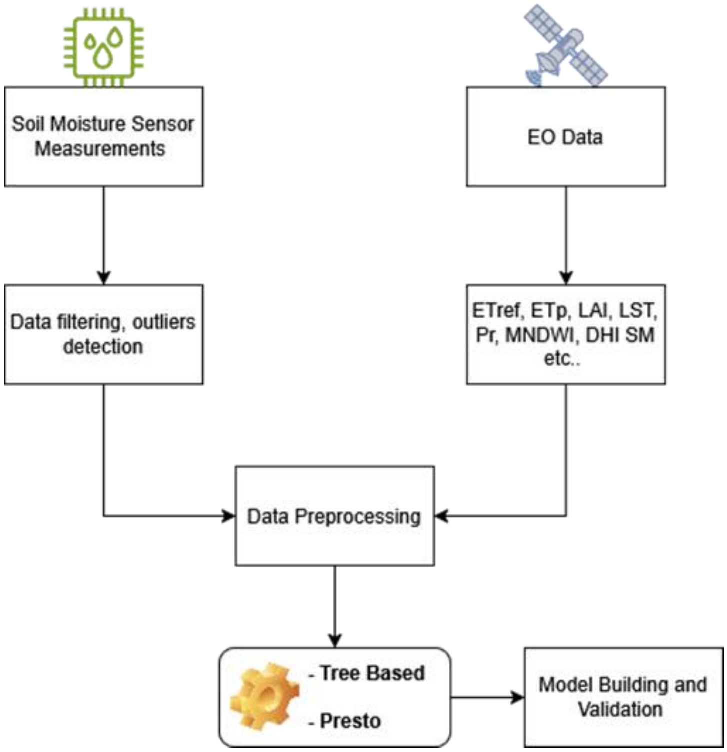


Figure 6: Workflow representing the different stages implemented for the in situ-EO-ML integration.

Figure 6 represents the workflow including the following stages:

1. **Acquisition of Earth Observation Datasets**: Collection of multi-source EO data from Sentinel-2, ERA5, along with DHI EO-derived products (E.g; ETref, ETp, LST, DHI SM data).
2. **Data Preprocessing**: Spatial and temporal harmonisation of EO datasets, georeferencing, resampling, and gap-filling procedures. Sensor data filtering and outlier detection.
3. **In Situ and EO Data Integration**: Integration of sensor-based soil moisture data (e.g., from HORTA and the International Soil Moisture Network) with EO-derived variables.
4. **Machine Learning Data Preparation**: Feature engineering, data normalisation, and selection of most significant explanatory variables.
5. **Model Development and Validation**: Application of ML models like tree-based algorithms (Random Forest, XGBoost) and performance evaluation.

Sensor Data Sources:
Following in-situ data sources were used for building and validating the integrated in situ–EO-ML pipeline:

- **HORTA soil moisture data**: data from ScaleAgData partner.
- **International Soil Moisture Network (ISMN)**: Provides a distributed network of soil moisture sensor data. HOBE network sensors from ISMN over Denmark are also used and tested.

Deployment of the EO-in situ fusion pipelines:

The methodological framework was deployed in October 2025. A demonstration of the framework was given on the 17th of October 2025, after which the Jupyter Notebooks and Python scripts were handed over via GitHub to the RILs for testing.

Evolution of the EO-in situ fusion pipelines
Following the deployment of the framework, and guided by feedback from potential users, a second pipeline is being prepared in which the requirement for DHI-derived higher level EO products is reduced through the use of learned representation (i.e. embeddings) from Presto foundation model. Presto is a transformer-based geospatial model pre-trained on 21.5 million multi-sensor pixel time series using Sentinel-1, Sentinel-2, ERA5, Dynamic World, SRTM, and metadata to extract transferable representations. This framework will be deployed and evaluated in May 2026.

4. Overview of collected EO data on the Research and Innovation Environment

The analysis-ready EO products described in Chapters 2 and 3 are made available through the Research and Innovation Environment (RIE) that has been set up for ScaleAgData.

The Research and Innovation Environment (RIE) has been established as a cloud-based internal environment to support the consortium's methodology development, integration, and validation activities during the project. In line with the project objectives, the RIE provides a shared technical space in which relevant datasets, intermediate products, analytical components, and development resources can be organised and made available to partners involved in research and prototyping activities. Its primary function is therefore to support the development and testing phase of the project, rather than to operate as a public-facing or production-level service environment.

The RIE has been conceived as an environment where consortium partners can access relevant project data, develop and test analytical workflows, integrate EO and in-situ information, and validate prototype solutions in an iterative manner. In this respect, it supports the technical work associated with the refinement of methods, the testing of data integration approaches, and the assessment of prototype suitability for the RI Labs. Dedicated development resources, including notebook-based environments and supporting software components, are made available to facilitate experimentation and technical validation by partners during implementation.

It is important to clarify that the RIE is not intended to be an operational platform for external users, nor a public environment where datasets and applications would be openly accessible for execution by third parties. Rather, it functions as a consortium-facing research and development workspace, used internally during the implementation period while components, methods, and services are still under active development. For this reason, access to the development environment is restricted to project partners, as this is consistent with its role as a controlled setting for technical development, testing, and integration.

Additionally, the use of the RIE as a development and testing environment is not considered mandatory for all partners and all developments. Instead, the RIE remains available as an optional internal environment, to be used where a shared project setting can provide added value for integration, testing, or validation purposes. In this way, the RIE functions as a flexible support environment within the project, while allowing partners to continue working within their existing technical ecosystems where appropriate.

The role of the RIE should be understood primarily as an internal research, integration, and testing infrastructure that supports the consortium during the iterative development and validation stages of the project. It enables partners to explore technical options, test selected solutions, and assess their suitability for RI Lab needs, but it is not designed to become a permanent operational environment hosting all project services for external use. This distinction is important for understanding both the restricted accessibility of the environment and the level of technical maturity of some of its components during the project period.

At project completion, the elements intended to remain publicly available are more limited and focus on dissemination and reuse of results rather than operational service provision. In particular, the project will maintain a catalogue of input and output datasets and a public GitHub repository containing the applications and software components developed within the project, where these are intended for sharing and reuse. The RIE itself should therefore be considered as the project's internal technical environment for development and validation, while the public legacy is reflected in the documented datasets, software assets, and associated project outputs.

On the ‘input side’, the RIE provides access to a catalogue and a set of scripts to access the EO data that have been collected for product development in the RILs. These EO products may originate from ScaleAgData “EO and technology providers” (WP3 partners) or may be collected and processed by the RIL partners themselves. To facilitate the exchange and reuse of the collected EO data, the partners are encouraged to share the data on the RIE. If needed, they may submit a request to Deimos for the data to be added.

Currently there are two main access points to the ScaleAgData EO data collections, where RIL users could search for EO products:

- the RIE catalogue
- via scripts in the RIE Jupyter notebook environment using the dedicated SDK

In those endpoints you will find three different data collection types

- **EO data collections searchable and directly accessible from the RIE**
- **EO data collections hosted on third-party platforms (Terrascope, Kuva Sense,…).** RILs typically access these platforms to request and download data directly into their personal space within the RIE, without storing it in the central repository. In these cases, the RIE catalogue includes entries for the datasets, with metadata profiles containing links to the corresponding third-party platforms.
- **EO data collections that are accessed and processed in third-party platforms (Google Earth Engine, OpenEO) via scripts being developed in the scope of WP4 technologies/methods** (e.g. Few-shot Learning, Sensor integrated soil moisture, Federated Learning). Similarly to the second case, the RIE catalogue includes entries for the datasets, with metadata profiles containing links to the corresponding scripts that are available in the project GitHub repository and where these datasets are accessed.

In the next sections an overview is provided of how a user can access the different ScaleAgData datasets in the RIE catalogue or Jupyter notebook environment.

4.1 RIE catalogue

The EO data collections metadata is available in the ScaleAgData catalogue and can be accessed following this procedure:

1. Access the ScaleAgData catalogue: https://scaleagdata.nextgeoss.eu . The figure below shows the expected view when you enter the catalogue.

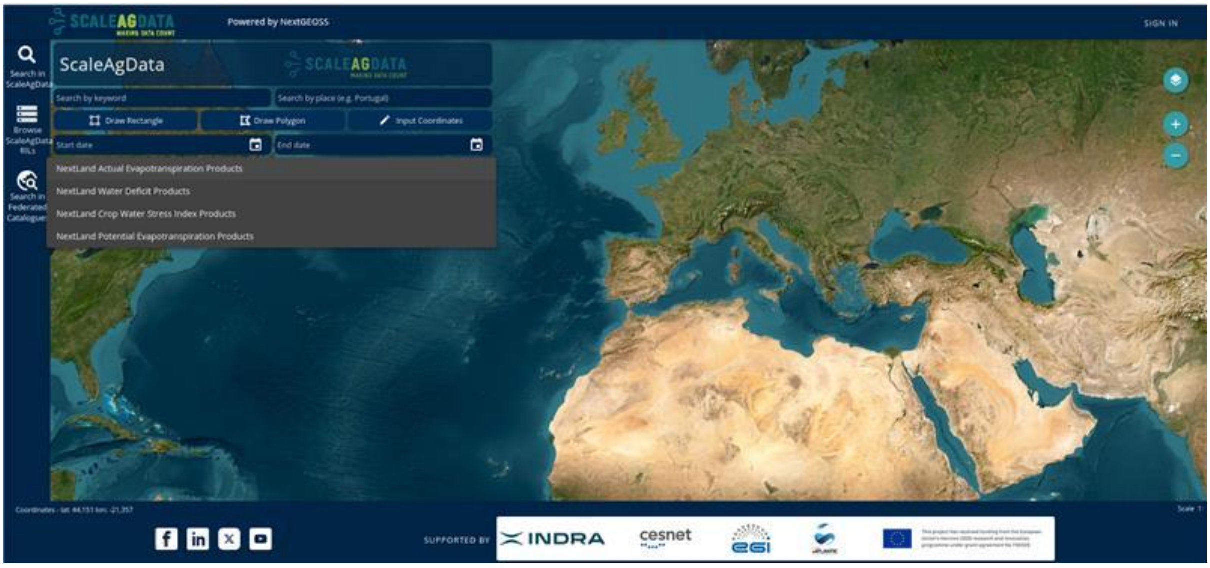
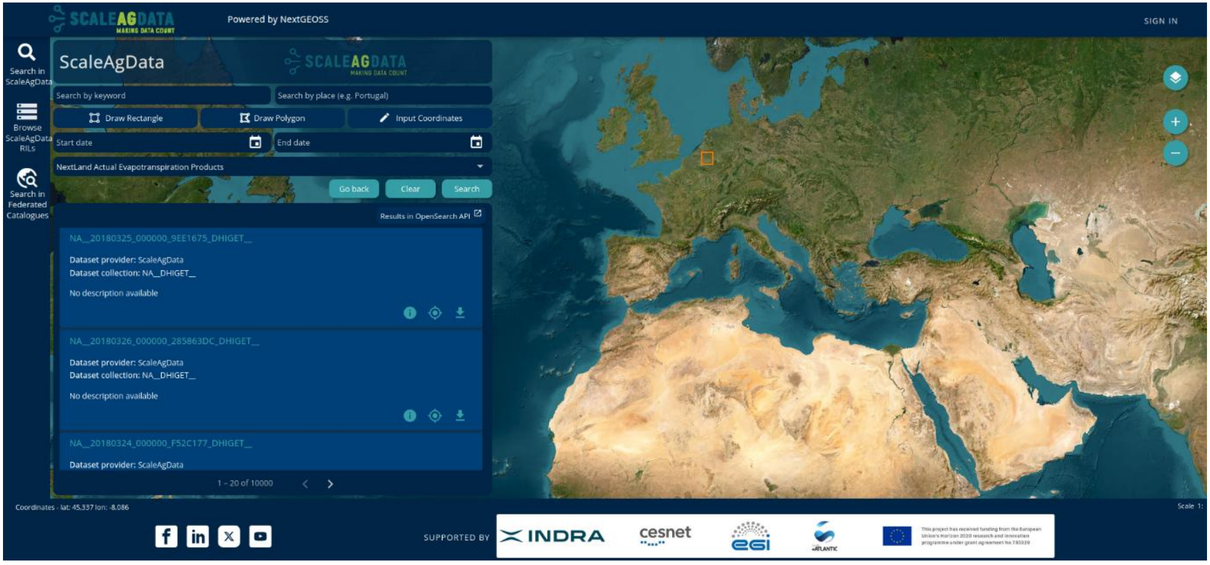


Figure 7: Entry page of the ScaleAgData catalogue

2. In the search interface the user is able to search directly for the different data collections by inputting the appropriate search parameters



NA_20180325_000000_9EE1675_DHIGET_

Dataset provider: ScaleAgData
Dataset collection: NA_DHIGET_

No description available

NA_20180326_000000_285863DC_DHIGET_

Dataset provider: ScaleAgData
Dataset collection: NA_DHIGET_

No description available

NA_20180324_000000_F52C177_DHIGET_

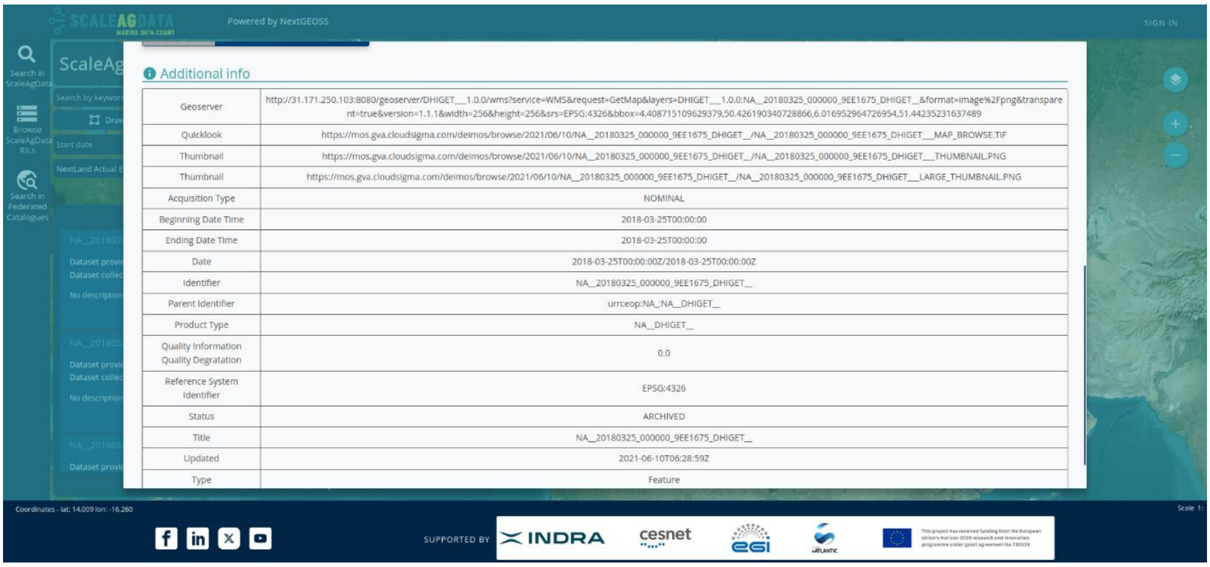
Dataset provider: ScaleAgData

Figure 8: Data Collection search result

3. The user is also able to browse over the different RILs to get an overall description of each lab and the associated data collections

Figure 9: Example of one RIL and its associated data collections

4. Independently of the way the user searches (general search or by browsing the different RILs) the user can click over the data collections/items to access their metadata. The metadata include a map interface to show the spatial footprint of the data collection, the different metadata fields and the link to the associated resources, which can either be a link to the RIE archive, a link to third party platforms where some data collections are stored, or a link to a ScaleAgData script present in the project GitHub repository, in the cases where the EO data collection is accessed and processed in a third party platform via this script.



Geoserver	http://31.171.250.103:8080/geoserver/DHIGET___1.0.0/wms?service=WMS&request=GetMap&layers=DHIGET___1.0.0:NA__20180325_000000_9EE1675_DHIGET__&format=image%2Fpng&transparent=true&version=1.1.1&width=256&height=256&srs=EPSG:4326&bbox=4.408715109629379,50.426190340728866,6.016952964726954,51.44235231637489
Quicklook	https://mos.gva.cloudsigma.com/deimos/browse/2021/06/10/NA__20180325_000000_9EE1675_DHIGET__/NA__20180325_000000_9EE1675_DHIGET___MAP_BROWSE.TIF
Thumbnail	https://mos.gva.cloudsigma.com/deimos/browse/2021/06/10/NA__20180325_000000_9EE1675_DHIGET__/NA__20180325_000000_9EE1675_DHIGET___THUMBNAIL.PNG
Thumbnail	https://mos.gva.cloudsigma.com/deimos/browse/2021/06/10/NA__20180325_000000_9EE1675_DHIGET__/NA__20180325_000000_9EE1675_DHIGET___LARGE_THUMBNAIL.PNG
Acquisition Type	NOMINAL
Beginning Date Time	2018-03-25T00:00:00
Ending Date Time	2018-03-25T00:00:00
Date	2018-03-25T00:00:00Z/2018-03-25T00:00:00Z
Identifier	NA__20180325_000000_9EE1675_DHIGET__
Parent Identifier	urn:eop:NA_:NA__DHIGET__
Product Type	NA__DHIGET__
Quality Information Quality Degratation	0.0
Reference System Identifier	EPSG:4326
Status	ARCHIVED
Title	NA__20180325_000000_9EE1675_DHIGET__
Updated	2021-06-10T06:28:59Z
Type	Feature

Figure 10: Example of the metadata of one RIL data collection

The ingestion of the project's input and output EO data collections has been delayed but will advance at a faster pace during the next month and it is expected that it will be near completion by the end of September 2026. During the remaining months of the project, the activities will focus on the ingestion of final products coming out of the RI Labs, following the close out of their development and validation activities. The table below defines a calendar for the ingestion of the collected EO input products into the catalogue. The ingestion of the output products is discussed in the revised version of deliverable D4.5 (sensor-integrated data products, v2).

Table 6: Calendar for data ingestion of ScaleAgData input EO datasets into project catalogue

Standard Input EO products	EO data provider	RI Lab (partner)	Date to conclude ingestion into catalogue
Sentinel-2 products (NDVI, LAI, fAPAR, Fcover, RGB)	VITO	RIL1 (IES, MIGAL)	29/05/2026
		RIL2b (HORTA)	29/05/2026
		RIL3 (UGent, VITO, LUKE)	29/05/2026
Sentinel-1/2 fused products (CropSAR1D fAPAR)	VITO	RIL3 (UGent, VITO, LUKE)	29/05/2026
		RIL5 (EURAC, IFAPA, DEIMOS)	
		RIL6 (OHB, ATB)	29/05/2026
Sentinel-2 Yield potential maps	VITO	RIL2b (HORTA)	29/05/2026
Sentinel-1/2 Biomass production	VITO	RIL6 (OHB, ATB)	29/05/2026
Sentinel-2/3 Evapotranspiration	DHI	RIL1 (IES, MIGAL)	01/04/2026
		RIL2b (HORTA)	05/06/2026
		RIL3 (UGent, VITO, LUKE)	01/04/2026
Sentinel-2/3 Soil moisture	DHI	RIL1 (IES, MIGAL)	01/04/2026
		RIL2a (NP)	05/06/2026
		RIL2b (HORTA)	05/06/2026
		RIL3 (UGent, VITO, LUKE)	01/04/2026
Sentinel-2/3 products (NDVI)	OHB	RIL6 (ATB, OHB)	05/06/2026
Sentinel-1/2 products	DEIMOS	RIL5 (EURAC, IFAPA, DEIMOS)	05/06/2026
Sentinel-2 MSI L2A reflectances	EURAC	RIL5 (EURAC, IFAPA, DEIMOS)	22/05/2026
Sentinel-1 backscattering	EURAC	RIL5 (EURAC, IFAPA, DEIMOS)	22/05/2026
Sentinel-1/2 LAI	EURAC	RIL5 (EURAC, IFAPA, DEIMOS)	22/05/2026
Sentinel-2 vegetation indices (NDVI, NDWI, PSRI)	IFAPA	RIL5 (EURAC, IFAPA, DEIMOS)	05/06/2026
Sentinel-2 derived crop classification product	NP	RIL2a (NP)	05/06/2026
Sentinel-2 NDVI	HORTA	RIL2b (HORTA)	12/06/2026
Sentinel-2 bare soil indices	ILVO	RIL4 (ILVO, AUTH)	12/06/2026
Sentinel-2 bare soil indices	AUTH	RIL4 (ILVO, AUTH)	12/06/2026
ERA5 Reanalysis	OHB	RIL6 (OHB, ATB)	29/05/2026
ECMWF IFS	OHB	RIL6 (OHB, ATB)	29/05/2026

New / improved EO products	EO data provider	RI Lab (partner)	Date to conclude ingestion into catalogue
Hyperfield data	Kuva Space	RIL4 (ILVO, AUTH)	12/06/2026
Sentinel-2/3 Super Resolution products	ICCS	RIL3 (UGent)	12/06/2026
	ICCS	RIL3 (VITO, LUKE)	12/06/2026
Sentinel-1/2 fused products (CropSAR 2D Fapar)	VITO	RIL3 (UGent, VITO, LUKE)	19/06/2026
Sentinel-1/2 fused products (CropSAR 2D Fapar)	VITO	RIL5 Biomass (EURAC, IFAPA, DEIMOS)	19/06/2026
Sentinel-1/2 fused products (CropSAR 2D Fapar)	VITO	RIL5 GPP (EURAC, IFAPA, DEIMOS)	19/06/2026
Sentinel-2/3 Sensor-integrated soil moisture	DHI	RIL1 (IES, MIGAL)	19/06/2026
	DHI	RIL2b (HORTA)	19/06/2026

4.2 RIE Jupyter Notebook environment

An SDK has been implemented with the functionality to allow users to discover collections and datasets in the archive4EO product by DEIMOS. More information can be found in the project deliverable D4.1 “The Research and Innovation Environment”.

The SDK includes different classes to search and access the data over the RIE catalogue. This is the class structure provided to the operator:

- **DataDiscovery:** contains discovery methods
    - Methods:
        - **getDataCollections():** Returns an object of the class DataCollections with the list attribute filled with the collections available in the catalogue. Each collection in the list is an object of the class DataCollection.
        - **search(filters={}, sortby=’<parameter>’, sort=’<asc|desc>’) :** returns an object of the class EODatasets with the results attribute filled with the list of datasets that meet the searching parameters.
        - **getSearchParameters():** returns a list with the available search parameters.
        - **getSortParameters():** returns a list with the available sort parameters.
- **DataCollection:** containing attributes about a collection
        - Attributes
            - **mission:** collection’s mission
            - **name:** collection’s name
    - **DataCollections:** containing a list of data collections and methods over that list
        - Attributes:
            - **list**: list of objects of class DataCollection
        - Methods:
            - **searchCollections(mission=’<substring>’, name=’<substring>’):** returns a list of collections whose mission and name contain the substrings passed as arguments. Each collection on the list is an object of the class DataCollection.
    - EODataset: containing the attributes with the metadata values of a specific dataset
        - Attributes:

        - **startTime**
        - **stopTime**
        - **spatial**
        - **url**
        - **quicklook**
        - **title**
        - **identifier**
- EODatasets: containing a list of datasets (results) and methods over that list
    - Attributes:
        - **results:** list of objects of class EODataset
    - Methods:
        - **convertResultsToDict():** returns a list of dictionaries containing results information.
        - **getDataset(id):** returns an object of class EODataset with the id passed as argument (if it exists).

The steps below provide examples of how a user can apply the SDK to search and access the different ScaleAgData datasets using the SDK classes and methods.

Search for data collections available in the archive:

```
from csw_sdk import DataDiscovery
discovery = DataDiscovery.DataDiscovery()
collections = discovery.getDataCollections()
```

Each collection (object from class DataCollection) has mission and name as attributes, to access them:

```
collection = collections.list[x]
collection_mission = collection.mission
collection_name = collection.name
```

Filter collections by name and/or mission (substrings are allowed)

```
collections = collections.searchCollections(mission=’S1B’,name=’S1PL1_SAFE’)
collection = collections[x]
collection_mission = collection.mission
collection_name = collection.name
```

Search for datasets

```
datasets = discovery.search(filters={'file_name':'%ED0E%', ’mission’: ‘S2_’, ’product_Type’:’
S2__VITONDVI’, start_time:’2022-05-01T00:00:00’, stop_time:’2022-05-15T00:00:00’, bbox:’1.45
49.418 6.95 52.0’, ...}, sortby=’creation_date’, sort=’asc,desc’);
```

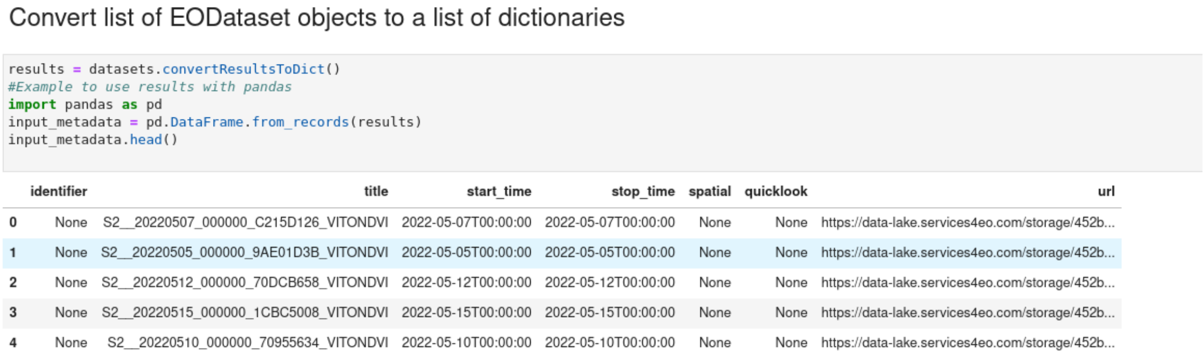
The allowed parameters for searching and sorting are located on a configuration file. To check which parameters can be used, the user may use methods getSearchParameters() and getSortParameters() of DataDiscovery class:

- valid_filters = discovery.getSearchParameters()
- valid_sort_parameters = discovery.getSortParameters()

Convert list of EODataset objects to a list of dictionaries

```
results = datasets.convertResultsToDict()
#Example to use results with pandas
```

import pandas as pd
input_metadata = pd.DataFrame.from_records(results)
input_metadata.head()



Convert list of EODataset objects to a list of dictionaries

```
results = datasets.convertResultsToDict()
#Example to use results with pandas
import pandas as pd
input_metadata = pd.DataFrame.from_records(results)
input_metadata.head()
```

	identifier	title	start_time	stop_time	spatial	quicklook	url
0	None	S2__20220507_000000_C215D126_VITONDVI	2022-05-07T00:00:00	2022-05-07T00:00:00	None	None	https://data-lake.services4eo.com/storage/452b...
1	None	S2__20220505_000000_9AE01D3B_VITONDVI	2022-05-05T00:00:00	2022-05-05T00:00:00	None	None	https://data-lake.services4eo.com/storage/452b...
2	None	S2__20220512_000000_70DCB658_VITONDVI	2022-05-12T00:00:00	2022-05-12T00:00:00	None	None	https://data-lake.services4eo.com/storage/452b...
3	None	S2__20220515_000000_1CBC5008_VITONDVI	2022-05-15T00:00:00	2022-05-15T00:00:00	None	None	https://data-lake.services4eo.com/storage/452b...
4	None	S2__20220510_000000_70955634_VITONDVI	2022-05-10T00:00:00	2022-05-10T00:00:00	None	None	https://data-lake.services4eo.com/storage/452b...

Figure 11: Script example and output on how to convert a list of EODataset objects to a list of dictionaries

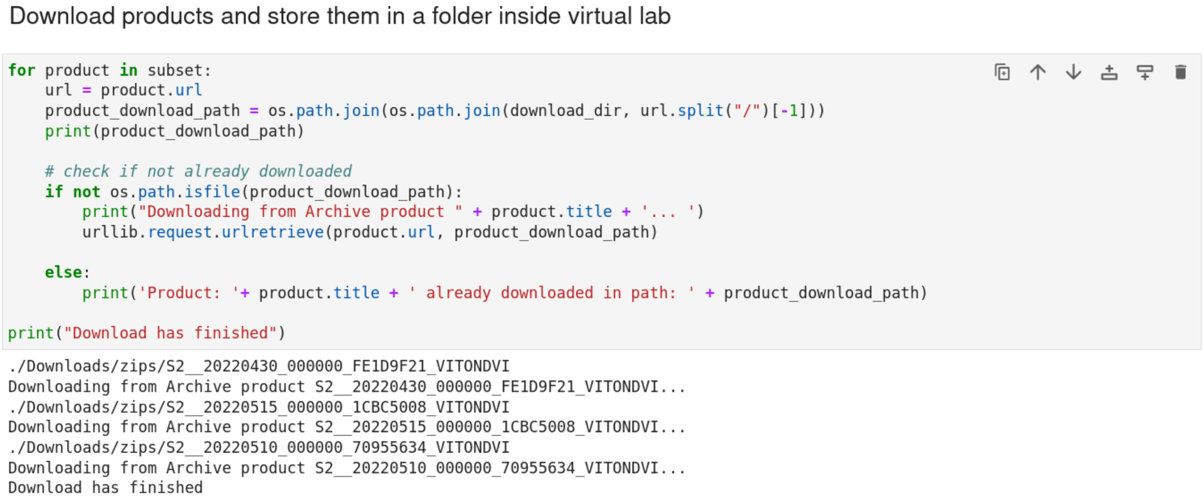
Select a dataset from the list

product= datasets.results[x]
product= datasets.getDataset(id)

Access to dataset

link_to_storage = product.url

Download products and store them in a folder



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 12: Script example on how to Download products and store them in a folder

After the download of product finishes the user can take a look into the products inside the Download Directory:

List the products inside the Download Directory

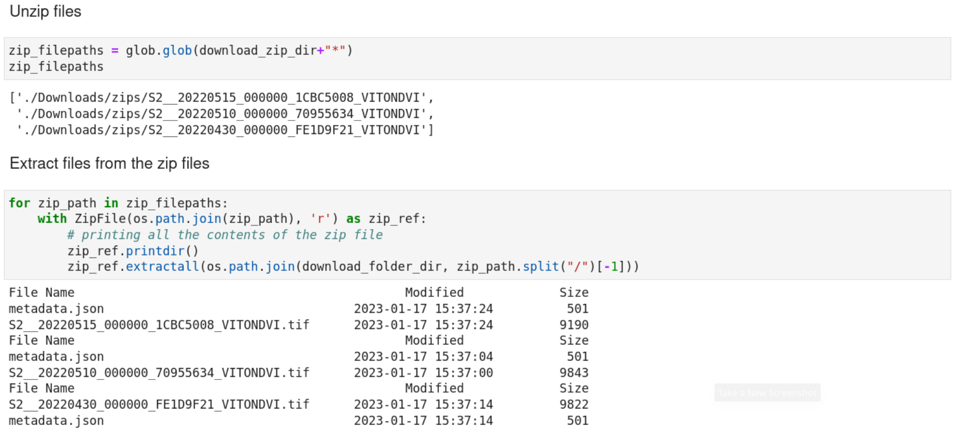
```
os.listdir(download_dir)
```

```
['S2__20220515_000000_1CBC5008_VITONDVI',
 'S2__20220510_000000_70955634_VITONDVI',
 'S2__20220430_000000_FE1D9F21_VITONDVI']
```

Figure 13: Script example on how to list the products inside the Download directory

Extract Files from Zip downloaded

The downloaded products will always be zip files so the next steps explain how to unzip and extract all files and store them in different folders.



Unzip files

```
zip_filepaths = glob.glob(download_zip_dir+"*")
zip_filepaths
```

```
['./Downloads/zips/S2__20220515_000000_1CBC5008_VITONDVI',
 './Downloads/zips/S2__20220510_000000_70955634_VITONDVI',
 './Downloads/zips/S2__20220430_000000_FE1D9F21_VITONDVI']
```

Extract files from the zip files

```
for zip_path in zip_filepaths:
    with ZipFile(os.path.join(zip_path), 'r') as zip_ref:
        # printing all the contents of the zip file
        zip_ref.printdir()
        zip_ref.extractall(os.path.join(download_folder_dir, zip_path.split("/")[-1]))
```

```
File Name                                             Modified             Size
metadata.json                                  2023-01-17 15:37:24          501
S2__20220515_000000_1CBC5008_VITONDVI.tif      2023-01-17 15:37:24         9190
File Name                                             Modified             Size
metadata.json                                  2023-01-17 15:37:04          501
S2__20220510_000000_70955634_VITONDVI.tif      2023-01-17 15:37:00         9843
File Name                                             Modified             Size
S2__20220430_000000_FE1D9F21_VITONDVI.tif      2023-01-17 15:37:14         9822
metadata.json                                  2023-01-17 15:37:14          501
```

Figure 14: Script example and output on how to Extract Files from Zip downloaded

Following these steps, the user can access all the datasets that are stored in the RIE archive.

5. Conclusions

This document provides an overview of the standard EO data products used by the RILs for data product development, as well as the new and improved EO data products developed by the ScaleAgData technology partners and provided to the RILs for evaluation and/or implementation. It also describes how the analysis-ready EO data are made available to the RILs through the Research and Innovation Environment (RIE).