D4.1 ARCHITECTURE AND SERVICES SPECIFICATIONS REPORT

Project: Monitoring of Environmental Practices for Sustainable Agriculture Supported by Earth Observation

Acronym: ENVISION

This project has received funding from the European Union`s Horizon 2020 research and innovation programme under grant agreement No. 869366.

Document Information

Grant Agreement Number	869366	**Acronym**	ENVISION
Full Title	Monitoring of Environmental Practices for Sustainable Agriculture Supported by Earth Observation		
Start Date	1st September 2020	**Duration**	36 months
Project URL	https://envision-h2020.eu/		
Deliverable	D4.1 Architecture and Services Specifications report		
Work Package	WP4 – ENVISION services		
Date of Delivery	**Contractual** M7	**Actual**	M7
Nature	Report	**Dissemination Level**	Public
Lead Beneficiary	DRAXIS		
Responsible Author	Stavros Tekes, George Efthimiou, Niki Papadatou		
Contributions from	NOA, EV ILVO, AgroApps		

Document History

Version	Issue Date	Stage	Description	Contributor
D0.1	20/3/2021	Draft	Draft for review	DRAXIS
D0.2	26/3/2021	Draft	Review feedback	NOA
F1.0	31/3/2021	Final	Final version	DRAXIS

Disclaimer

This document and its content reflect only the author's view, therefore the EASME is not responsible for any use that may be made of the information it contains!

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

CONTENT

The ENVISION project has received funding from the European Union's Horizon 2020 research
and innovation programme under grant agreement No 869366

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

LIST OF TABLES

LIST OF FIGURES

The ENVISION project has received funding from the European Union's Horizon 2020 research
and innovation programme under grant agreement No 869366

Abbreviations

Acronym	Full term
ANN	Artificial Neural Network
API	Application Programming Interface
AOI	Area Of Interest
AVIST	Analytics on Vegetation and Soil Index Time-series
BFAST	Breaks For Additive Season and Trend
BOA	Bottom-Of-Atmosphere
BSM	Burnt Scar Mapping
CAP	Common Agricultural Policy
CC	Cross-Compliance
CCTM	Cultivated Crop type Maps
Chl	Chlorophyll content
CSS	Cascading Style Sheets
cGANs	Conditional Generative Adversarial Networks
CNNs	Convolutional Neural Networks
DFS	Data Fusion Service
DL	Deep Learning
DNN	Deep Neural Network
EFA	Ecological Focus Area
EO	Earth Observation
FAPAR	Fraction of Absorbed Photosynthetically Active Radiation
fCover	Fraction of Vegetation Cover
FTP	File Transfer Protocol
GAEC	Good Agricultural and Environmental Conditions
GDAL	Geospatial Data Abstraction Library
GEE	Google Earth Engine
GIS	Geographical Information System
GMED	Grassland Mowing Events Detection Service
gml	Geography Markup Language
GRD	Ground Range Detected
GRU	Gated Recurrent Units
HTTP	Hypertext Transfer Protocol
InSAR	Interferometric Synthetic Aperture Radar
json	JavaScript Object Notation
LAI	Leaf Area Index
LPIS	Land Parcel Identification System
LULC	Land Use Land Cover
MVP	Minimum Value Product
ML	Machine Learning
MSI	Multispectral Instrument

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

MVT	Model-View-Template
NBR	Normalized Burn Ratio
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NIR	Near Infrared
NOA	National Observatory of Athens
NVZ	Nitrate Vulnerable Zone
OGC	Open Geospatial Consortium
OTSC	On-The-Spot-Check
PA	Paying Agency
ReCAP	Reinforcing CAP
RESTful API	Representational state transfer Application Programming Interface
RNN	Recurrent Neural Networks
RON	Relative Orbit Number
RS	Remote Sensing
RUSLE	Revised Universal Soil Index Equation
S1	Sentinel-1
S2	Sentinel-2
SAVI	Soil-Adjusted Vegetation Index
SAR	Synthetic Aperture Radar
SCL	Scene Classification Map
SD	Standard deviation
SDI	Spatial Data Infrastructure
SEN4CAP	The Sentinels for Common Agricultural Policy
SLC	Single Look Complex
SMR	Statuary Management Requirements
SMTP	Simple Mail Transfer Protocol
SOA	Service-oriented Architecture
SOAP	Simple Object Access Protocol
SOC	Soil Organic Carbo
SWIR	Short-wave infrared
TCP	Transmission Control Protocol
TOC	Top of Canopy
TMS	Tile map service
UAT	User Acceptance Testing
UDP	User Datagram Protocol
UI	User Interface
US	User Stories
USLE	Universal Soil Index Equation
VH	vertical transmit, horizontal receive
VI	Vegetation Index
VV	vertical transmit, vertical receive

The ENVISION project has received funding from the European Union's Horizon 2020 research
and innovation programme under grant agreement No 869366

WCS	Web Coverage Service
WFS	Web Feature Service
WMTS	Web Map Tile Service
WMS	Web Map Service

Executive Summary

The aim of this deliverable is to define the system architecture, the software and hardware requirements, the integration of the various components as well as the services specifications.
This deliverable is based on the Deliverable 2.2 Report of customer requirements from ENVISION services which contributed to the specification of the main points of the current deliverable and acts as the basis for its development.
This document defines the set of techniques and technologies that they are considered to be the most appropriate in order to produce the system architecture and design as well as provides the specifications (data, visualisation) of the services that will be developed during the project. Specifically, the sections of this deliverable are:

Section 1 – ENVISION Architecture: Contains a description of the ENVISION architecture, the design inputs, the system components, the dataflow diagram and the system monitoring and security.

Section 2 – ENVISION Services Specifications: Contains a description of each service that will be developed including the internal and external dependencies as well as dependencies from other services.

Section 3 – ENVISION's development pipeline: Contains the software development methodology that will be used during the development of the project.

Section 4 – Conclusions: Contains the main conclusions of this deliverable.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

1 ENVISION Architecture

The overall objective of ENVISION is to develop a robust and cost-efficient system fulfilling the need for continuous and systematic monitoring of agricultural land, shifting the focus from fragmented monitoring limited to specific fields and dates to territory-wide and all-year-round monitoring.
ENVISION fully exploits the wealth of data made available through GEOSS, Copernicus and DIAS and its synergetic use with other data to develop data products such as: Cultivated crop type maps; Soil Organic Carbon (SOC); Distinction of organic – conventional farming; Grassland mowing/ ploughing; Soil erosion. It makes use of heterogeneous types of available data (EO-based, in situ, open data, and historical on-field check data) and state-of-the-art technologies and methodologies (automatic pixel/texture/object-oriented change detection and classification methods, machine learning, data fusion, multi-source and multi-temporal data management) for providing a fully-automated and scalable toolbox of services, built in close interaction with its future customers.

1.1 Architecture Description

The architecture strategy follows practices of a Spatial Data Infrastructure (SDI) implementing a framework of geographic data, metadata, users and tools that are interactively connected in order to use spatial and remote sensing data in an effective and flexible manner.
An SDI would enable the discovery and delivery of data from a data repository, via a spatial service provider to the end user. Specifically:

- **Catalogue service –** discovering, browsing, and querying of metadata or spatial services and databases as well as other resources.
- **Spatial data service –** allowing the delivery of the data via the internet
- **Processing services**
- **(Spatial) data repository –** storing data form processing products
- **Web Geographic Information System (GIS) User Interface (UI) capabilities**
- **User access permissions management**

1.2 Design Inputs

The development of the system is not a monolithic solution but requires the interconnection of different components with each other responsible to implement specific requirements of the application.
Service-oriented Architecture (SOA) defines a way to make software components re-usable via services interfaces, which utilise common communication standards in such a way that they can be rapidly incorporated into new applications without having to perform deep integration each time.
Each service in a SOA embodies the code and data integrations required to execute a complete, discrete business function. The service interfaces provide loose coupling, meaning they can be called with little or no knowledge of how the integration is implemented underneath. The services are exposed using standard network protocols – such as Simple Object Access Protocol (SOAP)/ Hypertext Transfer Protocol (HTTP) or JSON/ HTTP – Web Map Service (WMS)/ Web Coverage Service (WCS) to

The ENVISION project has received funding from the European Union's Horizon 2020 research
and innovation programme under grant agreement No 869366

send requests to read or change data. The services are published in a way that enables developers to quickly find them and re-use them to assemble new applications.

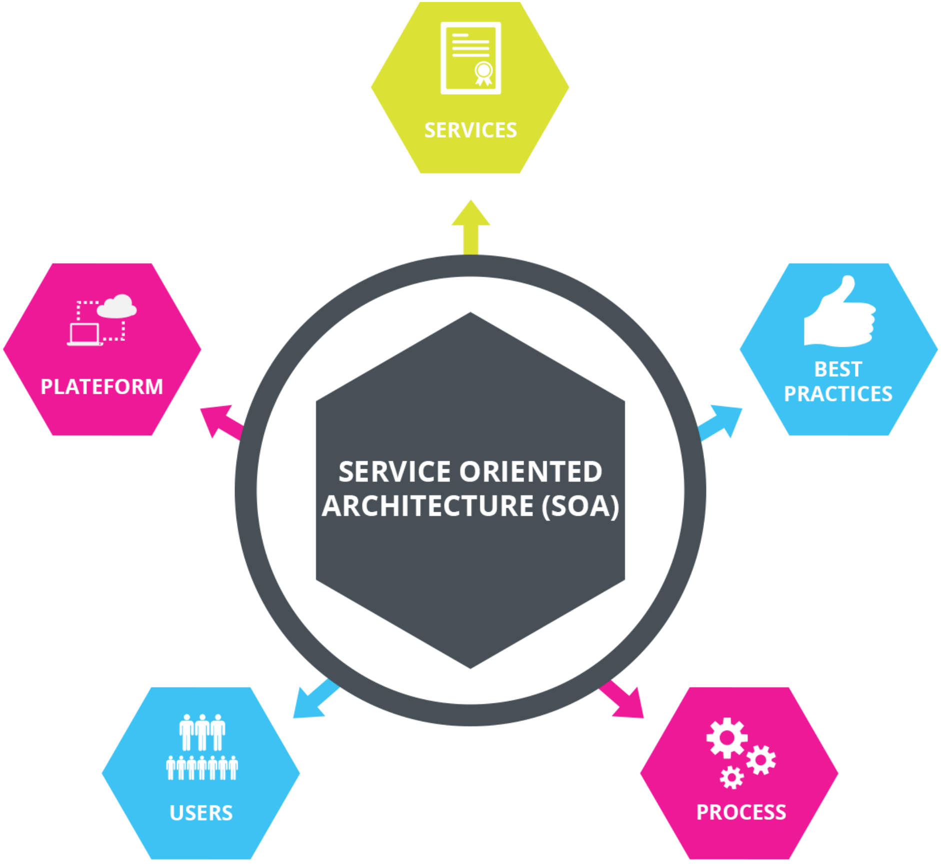

Figure 1: Service-oriented Architecture diagram

Furthermore, the software development will focus on core parts of the system to properly orchestrate the SOA components so as to meet the various business needs. The Core codebase will be based on the python programming language implemented with the Django framework.

Django is a high-level python web framework that enables rapid development of secure and maintainable web systems. Built by experienced developers, Django takes care of much of the hassle of web development, so as shifting the focus to the application development without the need of reinventing the wheel. It is free and open source and it is complemented with a thriving and active community, great documentation and many options for free and paid-for support.

Django is based on Model-View-Template (MVT) architecture, which is a software design pattern for developing a web application.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

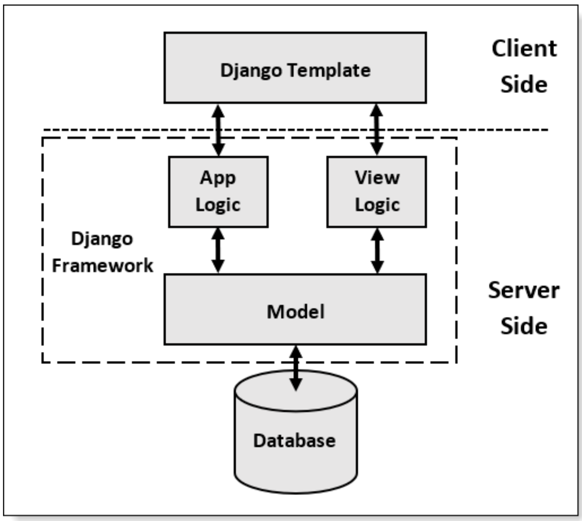

Figure 2: Model-View-Template Structure

MVT structure has the following parts:

- **Model:** Model is going to act as the interface of your data. It is responsible for maintaining data and is the logical data structure behind the entire application represented by a database.
- **View:** The View is the user interface – what you see in your browser when you render a website.
- **Template:** A template consists of static parts of the desired Hypertext Markup Language (HTML) output as well as some special syntax describing how dynamic content will be inserted.

1.3 System Components

The following SOA approach is framed by technologies that are going to handle the application development business logic and the subsystems should be able to meet the needs for GIS, Web mapping and GIS caching capabilities.

The data will be stored in a geospatial database (for the case of vector files) and in the file system for the cases of raster files. Central Authentication system is going to control the rules of user access to the application data. Alerts' production subsystem will be triggered when required client applications (mobile or web) may act as a point of interaction with users.

More specifically, the components and technologies to be applied to the SOA approach are:

- **Geospatial Server:** GeoServer allows you to display your spatial information to the world. Implementing the WMS, WMS-T standard, GeoServer can create maps in a variety of output formats. GeoServer is built on GeoTools, an open-source Java GIS toolkit. There is much more to GeoServer than nicely styled maps. GeoServer conforms to the Web Feature Service (WFS) standard, and WCS standard which permits the sharing and editing of the data that is used to generate the maps. GeoServer also uses the Web Map Tile Service (WMTS) standard to split your published maps into tiles for ease of use by web mapping and mobile applications. GeoServer is a modular application with additional functionality added via extensions. These extensions for Web Processing Service open up a wealth of processing options, you can even write your own!

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

- **Web Server:** NGINX is open-source software for web serving, reverse proxying, caching, load balancing, media streaming, and more. It started out as a web server designed for maximum performance and stability. In addition to its HTTP server capabilities, NGINX can also function as a proxy server for email (IMAP, POP3, and SMTP) and a reverse proxy and load balancer for HTTP, Transmission Control Protocol (TCP), and User Datagram Protocol (UDP) servers.
- **Authentication Server:** Standalone authentication server that undertakes to maintain and apply user access rules to all systems centrally, based on Keycloak that is an open-source identity and access management solution for modern applications and services.
- **Geospatial Database:** PostgreSQL is a powerful, open-source object-relational database system that uses and extends the SQL language combined with many features that safely store and scale the most complicated data workloads.

  PostgreSQL has earned a strong reputation for its proven architecture, reliability, data integrity, robust feature set, extensibility, and the dedication of the open-source community behind the software to consistently deliver performant and innovative solutions. PostgreSQL runs on all major operating systems, has been ACID-compliant since 2001, and has powerful add-ons such as the popular PostGIS geospatial database extender.

  PostGIS is an Open Geospatial Consortium (OGC) compliant software used as an extender for PostgreSQL, which is a form of object-relational database. While PostGIS is free and open source, it is used in both commercial (e.g., ArcGIS) and open-source software (e.g., QGIS). PostGIS extends capabilities of PostgreSQL to increase its management capabilities by adding geospatial types and functions to enhance spatial data handled within a relational database structure. The language of PostGIS is similar to SQL and allows spatial analysis and typical queries to be performed on spatial data with relative ease. This makes it a relatively powerful backend for databases within larger software, helping projects to use SQL-like functionality to do more complex spatial analysis and query.
- **Authentication Proxy:** The use of authentication server also requires the development of an authentication proxy that aims to follow the operational requirements of the application in terms of user accessibility rules in the system and data. For the needs of the authentication proxy, a python Django application embedded in the core code base of the application will be implemented, which will maintain a connection with the Authentication server.
- **Web UI Clients:** The Web application will be a Client wrapper that will enable action with the system to the end user. It will contain all the necessary Web elements and Tools to provide a friendly user experience while at the same time it will be able to integrate external components with iframes whenever necessary such as Web Maps. The development technology that will be used for the Frontend part of the system is Vue.js Vue.js is a library for building interactive web interfaces. The goal of Vue.js is to provide the benefits of reactive data binding and composable view components with an API that is as simple as possible.Vue.js itself is not a full-blown framework - it is focused on the view layer only. It is therefore very easy to pick up and to integrate with other libraries or existing projects. On the other hand, when used in combination with proper tooling and supporting libraries, Vue.js is also perfectly capable of powering sophisticated Single-Page Applications.
- **Mobile Client:** The development of the application will be based on hybrid technologies such as Vue.js together with Capacitor. Capacitor is an open-source native runtime for building Web

The ENVISION project has received funding from the European Union's Horizon 2020 research
and innovation programme under grant agreement No 869366

Native apps. Create cross-platform iOS, Android, and Progressive Web Apps with JavaScript, HTML, and Cascading Style Sheets (CSS).

- **Alerting component:** Custom python based alerting component will be used for alerts sending needs whenever necessary.
- **Web map interface:** For the needs of the visualisation on a map and general consumption of OGC services that will be produced by Geoserver, TerriaJS will be used as it is considered an innovative open-source solution enabling publishers to efficiently get their spatial data on the web, including 3D and 4D data.
- **HTTP Data Injection Interface**
  Component which will provide via secure HTTP protocol (s) the Injection of data from providers into the system. The Interface will provide additional data control rules that meet the conditions for entry into the system and will inform the client with the appropriate status code. It could provide the possibility for bulk insertion after consultations with the providers.

All components are to be developed as Docker Containers. A container is a standard unit of software that packages up code and all its dependencies so the application runs quickly and reliably from one computing environment to another. A Docker container image is a lightweight, standalone, executable package of software that includes everything needed to run an application: code, runtime, system tools, system libraries and settings.

Due to the variety of subsystems, technologies and dependencies the whole system has been split into Docker Containers and therefore, environments as described below:

Table 1: Docker Containers along the respective environments

Container A	Web Server , Python Django Code base , Alerting stack, Authentication Proxy, HTTP interface Component
Container B	Geoserver/Geowebcache
Container C	Postgres/PostGIS Geospatial Database
Container D	Vue.js Wrapper codebase
Container E	Monitoring tool (Grafana, Prometheus)
Container F	TerriaJS Web Mapping UI
Container G	Keycloack Authentication server

By splitting the system as described before, the systems' dependencies are isolated avoiding conflicts between libraries, services and different programming languages. In addition, using the Docker-Compose tool, the appropriate orchestration between the containers can be added in order to keep

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

the whole system running as a unique service. Docker – Compose is a tool for defining and running multi-container Docker applications.

1.4 Dataflow Diagram

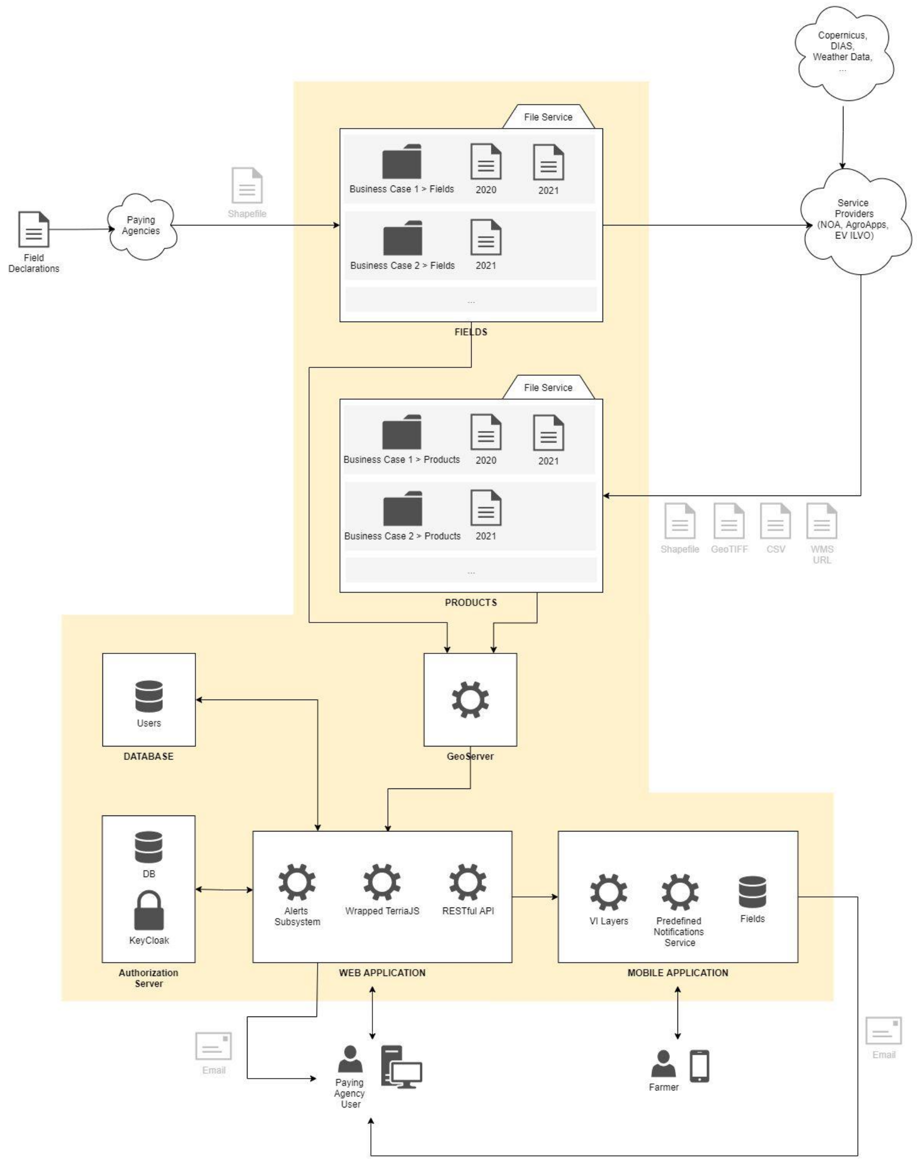

Figure 3: Dataflow Diagram

The above data flowchart presents a draft communication approach of the different components with each other. The data interconnection between Paying Agencies/ Certification Bodies and service providers will be done via HTTP Data injection Interface.

The products produced by the execution of the models will be a data source for the geospatial server which in turn will offer OGC web services (WMS/ WCS/ WMS-T) at the web level of the application.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

OGC services will be consumed by the TerriaJS Web Mapping subsystem for end-user viewing, as well as internally by alerting and Restful Services for the production of notifications to users and the business use of geospatial data respectively.

All data will go through the Authentication server Keycloack to manage permissions which will also be responsible for authenticating the Iframes contained in the Wrapped Web Page built by Vue.js .

Finally, the Restful API will be the backend system that will drive the Mobile client applications that will be used by farmers.

1.5 Component Interface

The table below is presented an analysis of the internal interfaces used by the components as well as how they are interconnected.

Table 2: Analysis of the internal interfaces

Geospatial server	Integrated with Postgres/PostGIS via Postgres Driver Integrated with Web via OGC services Integrated with Model's output via OS file system
Geowebcache	Integrated with Geoserver out of the box (Embedded)
Postgres	Integrated with Django via ORM Integrated with Geospatial Server via Postgres/PostGIS Driver
Keycloack	Integrated with Python Proxy via Rest API (json)
Mobile	Integrated with System's backend via Rest API (json)
TerriaJS	Integrated via Geospatial Server via WMS/WFS/WCS/WPS and time series protocols
Wrapper Page	Integrated with external components (Web Mapping) via Iframes and backend system via Rest API (json)
Alerts	Integrated with Mail server via SMTP
Web Server	Integrated with Interface via HTTP(s) protocol

1.6 Data Inputs/ Outputs

The following table reports the input and output data that the system is going to receive and generate:

Table 3: Data inputs/ outputs

Vector files	shp files provided by the Paying Agencies/ Certification Bodies
Raster files	Geotiff files from service providers as service products
Rest Payloads	JSON data for communication between components and external clients

1.7 Application Interface

The system will provide capabilities for communication with external systems (third parties) through the following protocols:

Table 4: Protocols for communication with external systems

Application Programming Interface (API)	Via Restfull API third party systems or developers can consume systems functionalities
OGC Services	Via Common OGC services third party SDI or Desktop GIS software can consume Geospatial Products
Postgres Connector	Via Postgres connection third party Postgres Clients can gain access to system's database
HTTP	Via HTTP protocol service providers can push service's products into system

1.8 System Monitoring and Security

System monitoring will be based on checking the status of each Component separately. Containers monitoring tools will be used to ensure their proper operation. The Monitoring tool that will be used will be a combination of System Dashboards Grafana with the Prometheus system.

Grafana equips users to query, visualize, and monitor metrics, no matter where the underlying data is stored. With Grafana, one can also set alerts for metrics that require attention, apart from creating, exploring, and sharing dashboards with their team and fostering a data-driven culture. In addition, Prometheus is an open-source system monitoring and alerting toolkit originally built at SoundCloud. Since its inception in 2012, many companies and organizations have adopted Prometheus, and the project has a very active developer and user community. It is now a standalone open-source project and maintained independently of any company.

The system will communicate with the outside world permanently via secure HTTP connection. The access to the systems that maintain the source code of the application, the data of the database will be done through SSH and specifically with the use of public keys.

Internal data exchange communication if needed will be based on SFTP interfaces.

The installation of Software Firewall will ensure secure communication in order to manage open ports as well as prevent attacks such as SQL injection, XSS attachment etc.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

2 ENVISION Services Specifications

This chapter provides a description of each service that will be developed including the internal and external dependencies as well as dependencies from other services.

2.1 Cultivated Crop type Maps

2.1.1 Name of the services – provider

This service will deliver a number of EO derived products for cultivated crop type maps consisting of:

- Crop type maps
- Crop compliance with Greening-1 rule
- Alert mechanism for smart sampling

The service will make multiple classification of the crops throughout the year to ensure the confidence of the classification process. To ensure high accuracy, the system will utilize both Sentinel-1 and Sentinel-2 data. The provider of the service is NOA, whereas the results of the service will be provided either via a RESTful API or as shape files.

2.1.2 Description of the service

The Cultivated Crop Type Maps (CCTM) is an Earth Observation (EO) crop classification module that exploits satellite data along with the usage of Machine Learning techniques in order to provide products related to the validation of the declared crop type by a farmer. In addition, it provides the knowledge of the compliance with certain environmental rules such as Greening requirements. Thus, it can be used from the Paying Agencies as a tool to enhance the process of checking the declarations of the farmers at the time of declarations, but also to assist via smart sampling of parcel to be checked during the validation process (OTSCs). This can be achieved as the service that informs the PAs about the parcels that have a high probability of being wrongly declared.
The Crop Diversification Service exploits the Land Parcel Information System (LPIS) and the declarations of the farmers. Thus, it will monitor crop types included in the aforementioned files and it will make a merge of crop types if the exhaustive process cannot distinguish between two or more of them. As a result, the predicted crop type of each declared parcel will be generated along with a percentage of prediction's confidence and it will be used as indicator for the declaration process. Last but not least, taking into account last year's farmers declarations, this service will be able to point out possible Land Use/Land Cover (LULC) changes, if they exist.

2.1.3 Service: Crop classification, alert mechanism and smart sampling

The service will provide:

1. Alerts based on significant differences between classification prediction and declarations on a higher level early in the year;
2. Alerts based on significant differences on crop type between classification predictions and declarations in the middle of the year;
3. Crop type maps via a shape file which will be exported so to be transferred via FTP or HTTP in an automated way.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

The provided service addresses to different user segments:

- **Paying agencies.** The service will provide them with continuous information about the crop type of each parcel. The first classification process will take place early in the year in a higher level giving the potential for the agencies to check the validity of the declaration. Then, multiple executions of the service will produce results in a lower level for each declared crop type. Thus, it can assist the paying agencies in decision-making as it allows them to make targeted inspections of parcels in shorter time periods so to validate the declared crop type.

2.1.4 Service: Crop diversification

The service will provide:

1. Alerts based on the Greening-1 compliance rules for each parcel
2. Compliance maps via a shape file which will be exported so to be transferred via FTP or HTTP in an automated way

The provided service addresses to different user segments:

- **Paying agencies.** The service will highlight the probability of parcels to follow or not the crop diversification rules or to be exempted. Therefore, it can assist the agencies on the faster and better validation of the parcels' compliance.
- **Policy makers.** The service will assist policy makers in taking the best decision on planning the crop suitability and to analyse the potential of expanding the crops following the necessities of crops diversification (various crops on adjacent area) and crops rotation (different crops throughout the consecutive cultivation periods). Moreover, it will be a significant tool, assisting on the better and most efficient design of new Common Agricultural Policy (CAP) post 2020 policies, adjusted on the local factors and needs.

2.1.5 Dependencies of the service from other services

The technological implementation of the satellite-based CCCD Service of NOA, needs to find a compromise between the following elements:

1. Availability of Sentinel-1 and Sentinel-2 satellite data coming from the data cube service

2.1.6 Literature on the service/ product

The CAP includes a series of rules and practices to monitor the land cover and control the subsidy payments. Paying Agencies are responsible for the distribution of CAP funds based on the farmers' declarations and their compliance to the CAP[1]. The crop classification enhances the procedure of validation in terms of accuracy and time. Remote sensing data contribute to the crop type classification process[2] along with the Machine Learning techniques such as support vector machines**Error! B**

---

[1] Sitokonstantinou V, Papoutsis I, Kontoes C, Lafarga Arnal A, Armesto Andrés AP, Garraza Zurbano JA. Scalable Parcel-Based Crop Identification Scheme Using Sentinel-2 Data Time-Series for the Monitoring of the Common Agricultural Policy. Remote Sensing. 2018; 10(6):911. https://doi.org/10.3390/rs10060911

[2] Rousi, Maria & Sitokonstantinou, Vasileios & Meditskos, Georgios & Papoutsis, Ioannis & Gialampoukidis, Ilias & Koukos, Alkis & Karathanassi, Vassilia & Drivas, Thanassis & Vrochidis, Stefanos & Charalabos, Kontoes & Kompatsiaris, Ioannis. (2020). Semantically Enriched Crop Type Classification and Linked Earth Observation Data to Support the Common Agricultural Policy Monitoring. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. PP. 1-1. 10.1109/JSTARS.2020.3038152.

The ENVISION project has received funding from the European Union's Horizon 2020 research
and innovation programme under grant agreement No 869366

ookmark not defined.,[3,4] and random forest**Error! Bookmark not defined.**,**Error! Bookmark not defined.**,[5]. The usage of both optical and SAR data enhances the feature space required for the crop classification service. Sentinel-2A and Sentinel-2B constellation results 5-day revisiting frequency and is widely used in crop classification**Error! Bookmark not defined.**,[6,7]. Nonetheless, the existence of clouds, especially i n certain areas, become a drawback as the optical radiation cannot penetrate them. This drawback is strongly related to the classification performance[8]. To solve this problem, integrated SAR and InSAR observations have been widely used in order to deal with the problem of frequent clouds and improve accuracy. To that direction, research[9] has proposes a more careful feature selection and proper combination of them that will drive to the better interpretability of the respective predictive models and more efficient management of computation expenses. In addition, taking into consideration the significant impact of parcels' size, shape and boundaries will assist on the design of more robust ML training routines.

2.1.7 Input data

The input data required are EO data and Paying agencies' data to derive crop information.

I. Satellite: Sentinel-2 L2A, Sentinel-1
II. Products: Spectral bands, backscattering coefficients, coherence (if needed), Vegetation Indices (VIs)
III. Paying agencies: Declared Crop Type, Validated Crop type, Polygon data

2.1.8 Output data

The service will provide:

I. Crop type maps as a shape file over the registered parcels early in the year grouped in a higher class.
II. Crop type maps as a shapefile over the registered parcels in a predefined frequency.
III. Compliance information as a shape file over the registered parcels in a predefined frequency.

[3] Duro, D.C.; Franklin, S.E.; Dubé, M.G. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. Remote Sens. Environ. 2012, 118, 259–272.
[4] Ok, A.O.; Akar, O.; Gungor, O. Evaluation of random forest method for agricultural crop classification. Eur. J. Remote Sens. 2017, 7254.
[5] Raczko, E.; Zagajewski, B. Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. Eur. J. Remote Sens. 2017, 50, 144–154.
[6] Immitzer, M.; Vuolo, F.; Atzberger, C. First experience with Sentinel-2 data for crop and tree species classifications in central Europe. Remote Sens. 2016, 8, 166.
[7] Sonobe, R.; Yamaya, Y.; Tani, H.; Wang, X.; Kobayashi, N.; Mochizuki, K. Crop classification from Sentinel-2-derived vegetation indices using ensemble learning. J. Appl. Remote Sens. 2018, 12, 026019.
[8] Whitcraft, A.K.; Vermote, E.F.; Becker-Reshef, I.; Justice, C.O. Cloud cover throughout the agricultural growing season: Impacts on passive optical earth observations. Remote Sens. Environ. 2015, 156, 438–447.1
[9] Orynbaikyzy, Aiym & Gessner, Ursula & Mack, Benjamin & Conrad, Christopher. (2020). Crop Type Classification Using Fusion of Sentinel-1 and Sentinel-2 Data: Assessing the Impact of Feature Selection, Optical Data Availability, and Parcel Sizes on the Accuracies. Remote Sensing. 12. 10.3390/rs12172779.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

2.2 Data fusion

2.2.1 Name of the service – provider

This service will provide a fully-automated Machine Learning module, aiming to assist in the enrichment of the Sentinel feature space for the rest of ENVISION services providing full time series of observations (no gaps due to clouds).

The main goal of this service is to tackle the problem of cloudy observations of Sentinel-2 images, taking into consideration information coming from Sentinel-1 and the rest of cloudless Sentinel-2 cases. Given that, the rest of the services will be able to provide results of higher accuracy and performance, since more complete and dense measurements of the respective indicators and vegetation indices will be at their disposal.

The provider of the Data Fusion Service (DFS) is NOA, whereas the service is provided as a backend module, easy to be exploited by other modules on top of DataCube.

2.2.2 Description of the service

The Data Fusion (DFS) workflow combines Sentinel-1 and Sentinel-2 data in order to "increase" the number of cloud-free observations. Optical sensors are sensitive to clouds resulting in gaps in the time series. As the complete time series is crucial for crop monitoring, data fusion comes as a solution to this problem. Sentinel-1 data are weather independent and not affected by clouds, therefore they can assist in predicting the Sentinel-2 values in the case of cloud obstructions.

The combination of both Sentinel-1 and Sentinel-2 using Deep Neural Networks (DNN) or other common Machine Learning (ML) algorithms will generate cloud-free time series ready to be ingested in the data cube and used as input to machine learning and/or other processes or be directly visualized. The cloud free products will enhance not only the size of the feature space but also the discrimination among the crop types.

The service will provide complete Sentinel-2 images (no gaps) for certain Optical Bands and Vegetation Indices such as NDVI, NDWI, etc. (which bands and indices TBC), acting as a assistive interpolation routine in order to fill and replace the respective rasters where cloudy image pixels have been detected and masked out.

2.2.3 Dependencies of the service from other services

3.2.3.1 Internal dependencies

Dependencies on outputs of other services/ products:

1. Sentinel-1 GRD
2. Sentinel-2 L2A
3. Cloud Masks

3.2.3.2 External dependencies/ Data acquisition plan

Databases, Hubs, Libraries

1. Python
2. Tensorflow or Pytorch
3. OSGEO/ GDAL Library
4. NumPy, Scikit-learn and Pandas Libraries

The ENVISION project has received funding from the European Union's Horizon 2020 research
and innovation programme under grant agreement No 869366

2.2.4 Literature on the service/ product

Monitoring of agricultural crops and grasslands is necessary for decision-making. In order to address the problem of frequent cloud cover that multiple regions present, especially in Northern geographical latitudes, systematic routines needs to be developed exploiting SAR and InSAR data that are not affected from these obstacles, but also the already available cloud-free observations. To that direction multiple Deep Learning (DL) techniques have been tried out during the recent past years.
From a computer-vision perceptive based on the benchmark study of[10] and cGANs (conditional Generative Adversarial Neural Networks) and pix2pix architecture, multiple attempts and highlighted works in low-level fusion task have been published. In the field of Earth Observation (EO), authors[11,12,13] are proposing several Deep Learning approaches of SAR-to-Optical image translations using cGANS and CNN networks using SAR, InSAR and past cloud-free Optical images available as inputs.
However, SAR and InSAR data are vulnerable not only to climatic conditions such as atmospheric humidity and precipitation, but also to topographical factors of the AOI like slope and exposition. Taking into consideration the aforementioned, in their illuminated work[14,15] are suggesting Recurrent Neural Networks and GRUs in order to take advantage of the temporal phenology characteristic that specific grasslands type presents, in an effort to ameliorate the performance of grassland monitoring and mowing events detection task.
Finally, others not that sophisticated models and methodologies have been advised using more common ML methodologies (e.g. Random Forests, Support Vector Machines, Boosting Algorithms, etc.) exploiting SAR time-series within the scope of Data Fusion for crop monitoring and crop type mapping**Error! Bookmark not defined.**[,Error! Bookmark not defined.,16,17].

2.2.5 Input data

The input data required to derive fused Sentinel-2 data is:

I. Satellite: Sentinel-1 GRD Backscatter, Sentinel-1 SLC Coherences, Sentinel-2 L2A, Cloud Masks
II. Products: Spectral bands.

---

[10] Isola, Phillip & Zhu, Jun-Yan & Zhou, Tinghui & Efros, Alexei. (2017). Image-to-Image Translation with Conditional Adversarial Networks. 5967-5976. 10.1109/CVPR.2017.632.
[11] Haas, Jan & Ban, Yifang. (2017). Sentinel-1A SAR and Sentinel-2A MSI data fusion for urban ecosystem service mapping. Remote Sensing Applications: Society and Environment. 8. 10.1016/j.rsase.2017.07.006.
[12] Fuentes Reyes, Mario & Auer, Stefan & Merkle, Nina & Henry, Corentin & Schmitt, Michael. (2019). SAR-to-Optical Image Translation Based on Conditional Generative Adversarial Networks—Optimization, Opportunities and Limits. Remote Sensing. 11. 2067. 10.3390/rs11172067.
[13] Zhang, Qian & Liu, Xiangnan & Liu, Meiling & Zou, Xinyu & Zhu, Lihong & Ruan, Xiaohao. (2021). Comparative Analysis of Edge Information and Polarization on SAR-to-Optical Translation Based on Conditional Generative Adversarial Networks. Remote Sensing. 13. 128. 10.3390/rs13010128.
[14] Garioud, Anatol & Giordano, Sébastien & Valero, Silvia & Mallet, Clément. (2019). Challenges in grasslands mowing event detection with multimodal Sentinel images.
[15] Garioud, Anatol & Valero, Silvia & Giordano, Sébastien & Mallet, Clément. (2020). On the joint exploitation of optical and SAR satellite imagery for grassland monitoring.. 10.5194/isprs-archives-XLIII-B3-2020-591-202.
[16] Filgueiras, Roberto & Mantovani, Everardo & Althoff, Daniel & Fernandes-Filho, Elpídio & Cunha, Fernando. (2019). Crop NDVI Monitoring Based on Sentinel 1. Remote Sensing. 11. 1441. 10.3390/rs11121441.
[17] Filgueiras, Roberto & Mantovani, Everardo & Fernandes-Filho, Elpídio & Cunha, Fernando & Althoff, Daniel & Dias, Santos H.. (2020). Fusion of MODIS and Landsat-Like Images for Daily High Spatial Resolution NDVI. Remote Sensing. 12. 1297. 10.3390/rs12081297.

The ENVISION project has received funding from the European Union's Horizon 2020 research
and innovation programme under grant agreement No 869366

2.2.6 Output data

The service will provide:

I. Sentinel-2A interpolated and cloud free rasters

Table 5: Output data for the cultivated crop type maps

Short description	File type	Expected size	Frequency
Crop Classification Maps (including classification confidence, possible alerts, smart sampling alerts etc)	Shapefile	100-500 MB per file	*max 3/month (Lithuania) *~3/growing season (Cyprus)

2.3 Sentinel data Preprocess

2.3.1 Name of the service – provider

Sentinel data Preprocess (CCD)
This micro-service aims at the generation of spectral, spatial and temporal features of Synthetic-aperture radar (SAR) data derived from Sentinel-1 (S1) satellites. The provider of the Preprocess Service is National Observatory of Athens (NOA). The service is provided as a backend module, easy to be exploited by other modules.

2.3.2 Description of the service

S1 preprocess workflow consists of a series of steps in order to make standard corrections to the initial S1 Ground Range Detected (GRD) data. These steps include, among others, radiometric calibration, terrain correction and noise removal. Moreover, this micro-service gives the potential for spatial adjustment of S1 data to Sentinel-2 (S2) data grids so to enhance the data fusion techniques. Finally, it provides also coherence images calculated from 6-day Sentinel-1 image pairs and in specific S1 Single Look Complex (SLC) products.

3.3.2.1 Service: Sentinel-1 GRD preprocess

The service will provide backscatter coefficients in dB units.
The provided service generates raster files of two polarization Vertical transmit - Horizontal receive (VH) and vertical transmit - vertical receive (VV) with standards corrections enabling the potential for the development of new products. At the same time, the processed GRD data can be co-registered to S2 data grids paving the way for data fusion algorithms.

3.3.2.2 Service: Sentinel-1 Coherence Generation

The service will provide average VV and VH coherence from pairs of images of the same relative orbit and direction.
The provided service addresses the need of providing temporal interferometric coherence for detecting grassland mowing events.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

2.3.3 Dependencies of the service from other services

The technological implementation of the satellite-based CCD Service of NOA, needs to find a compromise between the following elements:

1. Availability of satellite data coming from the DIAS catalogue

3.3.3.1 Internal dependencies

1. System/ server crash

3.3.3.2 External dependencies/ Data acquisition plan

1. SNAP software[18]
2. Snappy Python library

2.3.4 Algorithm Outline

In this section, two methodologies are presented for the generation of the VV/VH backscatters and the VV/VH coherence via SNAP software and snappy Python library. These two initial approaches, based on several studies[19,20], can be slightly updated as a result of a fine-tuning process.

VV/VH Backscatter generation process:

1. Apply Orbit File. Due to the fact that the included to the metadata orbit state vectors are not so accurate, the precise orbits of the satellites have to be automatically downloaded and updated so to determine a more accurate satellite position.
2. Radiometric Calibration. This process converts the values of pixels to radiometrically calibrated SAR backscatter.
3. Speckle Filtering. This process reduces speckle which is appeared as granular noise in the products. The refined Lee filter has proved to be better for visual interpretation[21] compared to other speckle filters that SNAP offers such as Media, IDAN et cetera.
4. Terrain Correction. Range Doppler Terrain Correction exploits a digital elevation model for correcting the location of each pixel so to reduce any geometric distortions.
5. Conversion to dB. Backscatter coefficients come without units. Thus, this process converts unit less products to dB using a logarithmic transformation.

VV/VH Coherence generation process:

1. TOPS Split. The split process isolates each sub-swath so to be processed separately. This separation allows the reduction of loaded data in the memory.
2. Apply Orbit File. As described before.
3. Back Geocoding. The procedure co-registers the pair of images taking into consideration the applied orbit information.
4. Coherence generation. In this step a calculation of pixels' similarity between the pair of images is calculated resulting a raster with values in range 0 to 1.

[18] SNAP Software, Help Document 2019. Available online: https://step.esa.int/main/toolboxes/snap

[19] Tamm, T.; Zalite, K.; Voormansik, K.; Talgre, L. Relating Sentinel-1 Interferometric Coherence to Mowing Events on Grasslands. Remote Sens. 2016, 8, 802. https://doi.org/10.3390/rs8100802

[20] Filipponi, F. Sentinel-1 GRD Preprocessing Workflow. Proceedings 2019, 18, 11. https://doi.org/10.3390/ECRS-3-06201

[21] Lee, J.S.; Jurkevich, L.; Dewaele, P.; Wambacq, P.; Oosterlinck, A. Speckle filtering of synthetic aperture radar images: A review. Remote Sens. Rev. 1994, 8, 313-340.

The ENVISION project has received funding from the European Union's Horizon 2020 research
and innovation programme under grant agreement No 869366

5. TOPS Merge. Merge of the isolated sub-swaths.

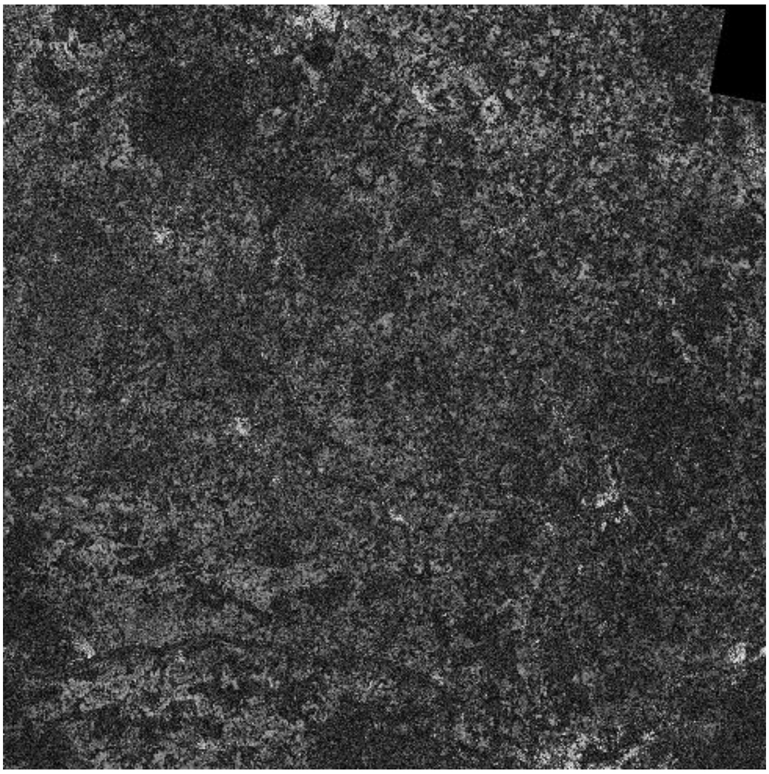

Figure 4: Sentinel 1A 6-day VH coherence for part of Lithuania in the middle of September, 2019.

2.3.5 Input data

The input data required are EO data.

I. Satellite: Sentinel-1
II. Products: SLC and GRD

2.3.6 Output data

The service will provide:

I. Sentinel-1 backscatter coefficient at VV and VH polarizations
II. Sentinel-1 coherence at VV and VH polarizations

2.4 Grassland Mowing Events Detection

2.4.1 Name of the service - provider

This service will provide a fully automated identification of Grassland Events module, with a view to assist in the valid and on-time identification of main events taking place in grasslands, such as mowing and grazing (if possible).

The service will contribute into the direct supervision of the Paying Agencies (PAs) of the compliancy of grasslands farmers to the respective regulation of pilot countries regulations and indication of possible declination from them. Given that, PAs will be able to organize and realize more accurate field visit campaigns to more specific locations pinpointed from that service and as a result will drive into the reduction of the inspections cost.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

This service will take advantage of EO derived indicators of agriculture monitoring, assisting to track the aforementioned grassland events. These indicators will be constructed using the available Sentinel-1 and Sentinel-2 images.
The provider of the Grassland Mowing Events Detection (GMED) service is NOA, whereas the results of the service will be provided either via a RESTful API or as shapefiles.

2.4.2 Description of the service

The Grassland Mowing Events Detection micro-service is an Earth Observation (EO) change detection module that exploits satellite data along with the usage of Decision Trees enhanced with Machine Learning (ML) and Artificial Neural Networks (ANN) algorithms (TBD). Based on the reproduction (partially or entirely) and enhancement of other similar projects' routines (e.g. SEN4CAP) pipelines, the main scope is to efficiently monitor grassland activity and precisely track the key dates of those cultivation events taking place. More specifically, it combines Sentinel-1 data (VV, VH and VV-VH ratio backscatter polarization coefficients and VV, VH coherences) and Sentinel-2 (fused if needed) NDVI time series, incorporating texture features such as homogeneity, entropy, contrast and dissimilarity (if needed), along with FAPAR, fCover and LAI indices and potentially VHR data (if needed), as provided by the PAs.
The product will be updated with every new image acquisition. Continuous change detection products will issue alerts for the detection of a cultivation event, such as mowing, and if it is possible discriminate those from grazing activity. Finally, the grassland events detection product will be provided dynamically accompanied with the respective confidence level, delivering updated versions to the user with every new acquisitions. The grassland mowing detection processing chain will be built on top of the Data Cube, allowing for its large scale and timely application.
Finally, this information can be used from the PAs and in parallel with the grassland regulations provided into the system encapsulating the maximum number of possible events and the exact period these can take place, this service will export an estimation regarding the compliancy from the farmers.
This service will provide:

1. Grassland Events Maps via a shapefile which will be exported in order to be transferred via FTP or HTTP in an automated way.

The provided service addresses to different user segments:

- **Paying Agencies:** The service will provide them with continuous information regarding grassland activity of each parcel. This will give PAs the ability to monitor abrupt changes into the field's canopy though the entire cultivation period and track the main events taking place and the respective time-periods. Moreover, given the specific regulations applied from each country, grassland mowing events detection micro-service can assist the PAs in the faster and better identification of farmers' compliance.
- **Policy Makers:** The service will assist policy makers in taking the best decision on planning the more suitable number of grassland events allowed during the entire cultivation period and to analyze the potential of grassland maintenance.

2.4.3 Dependencies of the service from other services

The technological implementation of the satellite-based GMED service of NOA will use:

1. Available satellite data coming from the Data Cube service;

2. Available detected/declared parcels (masked layers) coming from the Cultivated crop type maps (CCTMS) micro-service.

3.4.3.1 Internal dependencies

1. System/ Server Crash

3.4.3.2 External dependencies/ Data acquisition plan

Mowing data:

- The Paying Agencies must provide shapefiles containing the grassland parcels to be checked. These files must contain, among others, the geometry of each parcel and the declared grassland crop type.
- The Paying Agencies must provide the exact regulations characterize the grassland mowing or grazing policies. These files must contain, among others, the maximum number of allowed mowing events and the exact period, during these events can take place.
- Past decisions of OTSCs for training and validation of models

2.4.4 Literature on the service/ product

In order to address the challenge of mowing detection, data of HR and VHR optical and SAR imagery will be provided for the computation of vegetation parameters indicative for the current grassland condition, managing to detect the various anomalies and changes on index profiles.

Researchers using BFAST toolbox[22] are able to detect potential changes as well as the magnitude of these on vicinal regions in northwestern Slovenia. Exploiting data of Sentinel-2 and constructing the respective NDVI, they applied a pixel-based BFAST monitor time-series analysis and a semi-automated methodology. The above has been examined for meadows and crop fields respectively. BFAST performs a decomposition into the signal trends, seasonal and remained components in order to pinpoint anomalies in the total behavior of the investigated index. However, since it is susceptible to overestimation on the number of anomalies of permanent meadows candidates due to the short time series and trend overestimation, it is followed and supplemented by a parcel-based graph analysis (descriptive statistics) and a pixel-based time-series standard deviation computing in order to verify these results.

Researchers[23] constructed a grassland mowing index (GMI) based on the NDVI and its sudden drops from consecutive clear observations in their disposal extracted from areas of Switzerland as it displays a more representative image of the mowing events and characterizes mowing schedule and frequency according to the Swiss regulations. Moreover, they performed different training scenarios for parcel and shrunken-parcel cases, as well as optimized situations (after various level of cloud masking - blue-band masking approach) in matter of time consumption against common ready-to-use algorithms. It is worthy to be mentioned the pixels allocation step into few grass/non-grass identified clusters (characterization was aided by careful visual inspection) in order to masked out from the NDVI time series before into-parcel aggregation level. All things considered significant findings showed that

[22] Kanjir, Ursa & Duric, Natasa & Veljanovski, Tatjana. Sentinel-2 based temporal detection of agricultural land use anomalies in support of common agricultural policy monitoring. ISPRS International Journal of Geo-Information, 7:405, 10 2018.

[23] Kolecka, Natalia & Ginzler, Christian & Pazur, Robert & Price, Bronwyn & Verburg, Peter. Regional scale mapping of grassland mowing frequency with sentinel-2 time series. Remote Sensing, 10:1221, 08 2018.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

grassland mapping is not limited on the number of cloudless images as long as they are carefully selected and acquired in crucial moments of mowing season.

On the contrary, in a not sophisticated methodology, a time-series algorithm[24] has been presented based on the estimated NDVI from Sentinel-2 images for different regimes across Germany on 10-day composites on an effort to characterize mowing behavior and evaluate the acquired results based on in-situ data taken from farmers. Essentially, a 3rd order polynomial is fitted based on time-series vertices for every pixel available. Lastly, a mowing event is characterized when the respective residual between the NDVI points and the fitted polynomial overcomes a specific threshold.

Alternatively, Sentinel-1 data can be used for the estimation of mowing events. Research work has been conducted**Error! Bookmark not defined.** for the area of Rannu in Estonia where researchers have c orrelated polarization coherence of VV and VH transmission/receival for different relative orbit numbers (RON) and image resolutions (window-size) since the median values of those measures can present remarkable soaring trends even after 24 to 36 days of an actual mowing event compared to pre-event values.

All in all, other researchers[25,26] are using Sentinel-1 SAR acquisitions in order to tackle the problem of scarce optical imagery and low temporal resolution since it is highly affected from the cloud coverage. For this reason, they formulate an estimated data-driven NDVI based on a RNN architecture and Sentinel-1 originated input data accompanied with an extra ancillary temporal information (day of year, month, season, etc.). However, issues have been arisen here since SAR signals characterized by high sensitivity and as a result S1 images are also affected by shadow, layover, and foreshortening (occurring differently between ascending and descending orbits). Auxiliary information provided aims to mitigate such predicament, acting as an attention vector using multiple supplementary data based on the region and time. Ultimately, the reconstructed NDVI obtained could be used as a starting point for the subsequent grassland analysis.

2.4.5 Input data

The input data required are EO data and Paying Agencies data to derive grassland mowing events detection:

I. Satellite: Sentinel-1, Sentinel-2 L2A, VHR
II. Products: Spectral bands, backscattering coefficients, coherence coefficients, Vegetation Indices (Vis), FAPAR, LAI
III. Paying Agencies: masked grassland crop type maps, polygon data, mowing regulations for the specific AOI.

---

[24] Griffiths, Patrick & Nendel, Claas & Pickert, Jurgen & Hostert, Patrick Towards national-scale characterization of grassland use intensity from integrated sentinel-2 and Landsat time series. Remote Sensing of Environment, 238:111124, 04 2019.

[25] Garioud, Anatol & Valero, Silvia & Giordano, Sébastien & Mallet, Clément. (2020). On the joint exploitation of optical and SAR satellite imagery for grassland monitoring. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. XLIII-B3-2020. 591-598. 10.5194/isprs-archives-XLIII-B3-2020-591-2020.

[26] Garioud, Anatol & Giordano, Sébastien & Valero, Silvia & Mallet, Clément. (2019). Challenges in grasslands mowing event detection with multimodal Sentinel images.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

2.4.6 Output data

The service will provide:

- Events Map (shapefile) of grassland mowing detection per parcel encapsulating all the extracted information regarding the detected events, their confidence levels and their compliance into the respective mowing regulations.

Table 6: Output data for the grassland mowing events detection

Short description	File type	Expected size	Frequency
Grassland mowing detection	Shapefile	100-500 MB per file	max 3/month (Lithuania)

2.5 Analytics on Vegetation and Soil Index Time-series

2.5.1 Name of the service – provider

Analytics on Vegetation and Soil Index Time-series (AVIST) aims at providing vegetation and soil indices, along with geospatial analytics such as growth trends, change detection, phenological metrics, soil specific indicators, static indicators with respect to rainfall erosivity and soil erodibility, cover management factor for soil erosion, Natura2000 areas hotspot detection, Burnt Scar Mapping and Runoff Risk assessment for the reduction of water pollution in Nitrate Vulnerable Areas. The final set of subservices remains to be defined after the processing of user requirements. Some and not all of the aforementioned services will be developed.

This micro-service will deliver a number of EO derived products for crop monitoring and data analytics consisting of:

- Vegetation status map and agricultural monitoring of Ecologically Focused Areas (EFA) practices such as Catch-Crops, Nitrogen-Fixing crops and Fallow lands.
- Runoff Risk assessment for the reduction of water pollution in Nitrate Vulnerable Areas.
- Buffer Strips for the Proximity to water-ways nearby.
- Minimum soil cover for Soil Erosion
- Automated Burnt Scar Mapping (BSM) for the maintenance of organic matter in soil and identification of stubbles burning.
- Natura2000 regions activity hotspot detection (clearing) (T.B.D)
- Other GIS querying functionalities (T.B.D)
- The service will provide multiple Analytics reports throughout the year taking advantage of both Sentinel-1 and Sentinel-2 data available along with additional products generated from the raw data. The provider of the service is NOA, whereas the results of the service will be provided either via a RESTful API or as shape file.

The service will provide multiple Analytics reports throughout the year taking advantage of both Sentinel-1 and Sentinel-2 data available along with additional products generated from the raw data. The provider of the service is NOA, whereas the results of the service will be provided either via a RESTful API or as shape file.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

2.5.2 Description of the service

The Analytics on Vegetation and Soil Index Time-series (AVSIT) is an Earth Observation (EO) monitoring module that exploits satellite data along with the usage of Machine Learning (T.B.D) algorithms on top of Datacube platform in order to provide into users' realistic indications related to the field activities and the control of cross-compliance policies. More specifically, this module will take advantage of numerous potentials of Datacube introducing an innovative Big Data framework in the field of CAP monitoring from Paying Agencies (PAs).

Initial assessment and visualizations from the derived satellite signals can be performed in order to check the degree of compliance of agricultural parcels and the advisable cultivations periods of the respective Catch crops and Nitrogen Fixing crops, as well as the maintenance of no crops bare lands or spontaneous cultivated areas for the production of green manure such as black and green fallows respectively. Monitoring of activity in the above cases is considered as a necessity from the perspective of GAEC 4 and similar soil-erosion regulations, reinforcing EFA practices. Furthermore, CAP Cross Compliance conditions such SMR 1 and GAEC 1 are tackled by the proposal of a risk assessment for water pollution in the Nitrate Vulnerable Zones (NVZs). The risk assessment employs RUSLE products and the proximity to surface waters. The SAVI index is used for the identification of parcels with high soil coverage, providing a monitoring solution to the GAEC 4 conditions. Finally, Burnt Scar Mapping algorithms are developed to fit the ENVISION needs and specifically answer the GAEC 6 requirement. Last but not least, detection of activity inside Natura2000 regions will be a critical challenge remained to be solved since Sentinels spatial resolution as well as the scarcity of such events as ground truth poses significant deterrents on that venture.

This micro-service combines Sentinel-1 and fused Sentinel-2 data so to generate vegetation and soil indices at national scale. At a next level, the service aims to provide geospatial analysis. This analysis comprises of charts, statistics and data visualizations for understanding complex relationships, finding trends and reveal changes throughout the time. Moreover, it can help on predicting what is going to happen next by pattern recognition. Finally, several algorithms that will be implemented (ideally or modified) based on the legacy and the rich methodology of ReCAP project that NOA developed, since similar problems tackled with success in the general goal of CAP monitoring.

3.5.2.1 Service: Agricultural Monitoring

The service will provide:

1. A general monitoring and visualization tool of parcels vegetation status.
2. Identification of illegal clearing in Natura2000 zones (TBD)
3. Indications of main events such as Harvest of Main Crops, onset of catch crops (TBD)
4. Depiction of allowable periods for EFA practices application. (TBD)
5. Provision of the respective confidence levels regarding CAP rules conformity.

3.5.2.2 Service: Runoff Risk Assessment – SMR1

The service will provide:

1. A set of maps according to the needs of users and data acquired indicating the runoff risk level assessment of the parcels.

The ENVISION project has received funding from the European Union's Horizon 2020 research
and innovation programme under grant agreement No 869366

3.5.2.3 Service: Burnt Scar Mapping (BSM) -GAEC 6

Remote sensing has been extensively used for the robust, accurate and timely assessment of forest wildfire damages, focusing on their extent, severity and other crucial indicators. This type of information has proved to be of paramount importance for public environmental agencies and relevant stakeholders, responsible for the mitigation policy making in the aftermath of crisis. Therefore, substantial research has been undertaken in the accurate mapping of burnt areas, with multiple methodologies available in literature.
The mapping of burnt areas is also of primary importance in the monitoring of the agricultural sector. The need for Burnt Scar Mapping (BSM) is expressed by the stakeholders in stubble burning monitoring requirement and is shaped into policies and practices through GAEC 6: Maintaining the level of organic matter in soil. The identification of burnt crop parcels would directly assist the monitoring of GAEC 6 compliance, as clearly indicated in the rule's specifications.
Data of variable spatial resolution will be utilized for burnt scar analysis, starting from moderate resolution MODIS, AVHRR for the real-time monitoring due to their high temporal resolution. Moreover, high spatial resolution data of Landsat TM and Sentinel-2 MSI along with VHR imagery of Formosat 2, WorldView, Ikonos and Quickbird could also be employed for more accurate, event-driven assessment of fire impact.

3.5.2.4 Minimum Soil Cover – GAEC 4

The remote sensing solution for GAEC 4 cannot be of deterministic nature as clear distinction between complying and noncomplying parcels is a difficult task. However, the SAVI index was found to be a satisfactory proxy for the indication of land parcels of increased soil percentage.

$$SAVI = \frac{(1+L)(NIRband8 - REDband4)}{NIRband8 - REDband4 + L}$$

The SAVI index is essentially a modification of the NDVI index with the addition of factor L[27]; in this case its value is set at L=0.5. In areas covered by low vegetation, the soil surface is dominant and therefore the reflectance in the NIR and Red spectra can significantly impact the vegetation index values. This poses a problem when comparisons are made between land areas of different soil types that have differing reflectance in the Red and NIR spectra. The SAVI index modifies NDVI in a way to limit the varying soil brightness problem in regions of little vegetation cover. The value of the factor L is chosen based on the amount of vegetation in the area inspected. Low values of L would refer to areas of high vegetation, whereas high values would refer to areas of minimal vegetation. The optimal choice of L for the present scenario was found to be at 0.5, as the land parcels of interest are described by a mixture of vegetation and soil cover. The index is produced with the concept of being used as ancillary product to indicate parcels of very low vegetation cover and thereafter be further inspected for the GAEC 4 compliance decision. The SAVI product shall be given at a regular basis (i.e. monthly) in order to monitor the temporal evolution of parcels identified to be of significantly high soil coverage. For example, a summer crop that will be found to be of minimal vegetation coverage during the periods of seeding and harvesting but not during the rest of the year, it would not be considered a problematic parcel. The parcels that would require attention are the ones that do not take any measures to respect the GAEC 4 requirements after the harvest period and up until the next sowing period. The minimum

---

[27] Huete A.R. , "A soil-adjusted vegetation index (SAVI)," Remote Sens. Environ., 25 295 –309 (1988). https://doi.org/10.1016/0034-4257(88)90106-X Google Scholar

The ENVISION project has received funding from the European Union's Horizon 2020 research
and innovation programme under grant agreement No 869366

soil cover condition is estimated by employing a modified NDVI index for land cover of increased soil abundance, SAVI. SAVI was shown to be adequately responsive to small changes in the soil percentage within a parcel. The provided information gives at the very least a continuous indication of parcels in the risk zone of not complying with GAEC 4. Threshold SAVI values shall be identified to give a binary classification of parcels requiring attention and parcels of adequate vegetation cover. This threshold is expected to geographically vary and therefore proper evaluation of the method's performance shall be done in the pilot phase of the project.

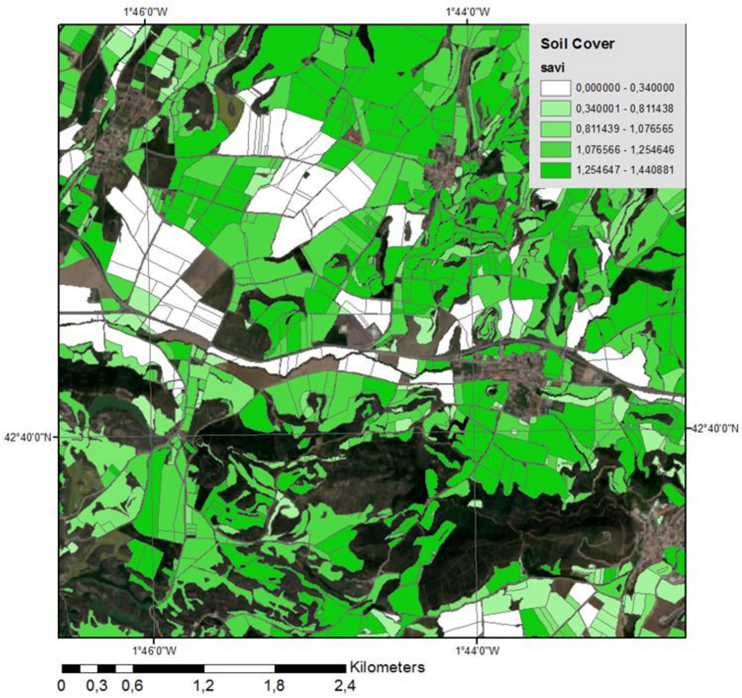

Figure 5: SAVI index, depicting with white the below threshold parcels.

The provided services address to different user segments:

1. Paying agencies. The service will provide them with continuous information regarding the parcels cultivation phases and the respective compliance of the current CAP policies. Multiple executions during the entire cultivation period and visualizations of the service will give them a clearer picture of the current farmers' activity. In parallel, the provision of the respective confidence levels will assist them in decision-making as it allows them to make more accurate field inspections (through RS or OTSC) and reduce the cost of field campaigns.
2. Policy Makers. In the dawn of a new POST 2020 CAP, this service can be a valuable supportive tool on the design of new area-specific cross-compliance policies. Policy makers can use this service as a second hand in order to define new regulations focusing on the topical characteristics and specifications of the inspected regions.

2.5.3 Dependencies of the service from other services

The technological implementation of the satellite-based GMED service of NOA will use the

1. Generated Sentinel-1 VV/VH Backscatter indexed in the Data cube
2. Generated Fused Sentinel-2 L2A indexed in the Data cube, if needed

3.5.3.1 Internal dependencies

1. System/ Server Crash

3.5.3.2 External dependencies/ Data acquisition plan

Databases, Hubs, Libraries

1. Python
2. OSGEO/ GDAL Library
3. NumPy and Pandas Libraries
4. Datacube API for connecting to the cube

2.5.4 Literature on the service/ product

The CAP requirements responsible for reducing the water pollution in Nitrate Vulnerable Zones (NVZs) (SMR 1 rule), among other things, requests the farmers to perform themselves risk assessment on the susceptibility of a parcel to contribute nitrate-rich soil to nearby surface water. The assessment shall involve the following considerations:

1. The slope of the land, especially if over 12°
2. Ground cover
3. Proximity to surface water
4. Weather conditions
5. Soil type and condition
6. The presence of land drains.

Based on the requirements of the risk assessment, it can be easily seen that emphasis will be given on the soil erosion and runoff susceptibility. Since the monitoring and estimation of the above conditions, only via means of Sentinel-2 remote sensing will not be possible; an alternative approach will be devised. The Universal Soil Loss Equation (USLE) and the revised version RUSLE are very common soil erosion modeling techniques[28,29]. The soil loss equation is considered to be an appropriate proxy for the estimation of risk in terms of the SMR 1 requirements.

The combined use of GIS and erosion models, such as USLE/RUSLE, has been proved to be an effective approach for estimating the magnitude and spatial distribution of erosion and soil loss[30,31,32]. Five major factors such as rainfall pattern, soil type, topography, crop system, and management practices are to be used in USLE/RUSLE for computing the expected average soil loss through the following equation[33].

---

[28] Panagos, Panos, Pasquale Borrelli, and Katrin Meusburger. "A New European Slope Length and Steepness Factor (LS-Factor) for Modeling Soil Erosion by Water." Geosciences 5.2 (2015): 117-26.

[29] Panagos, Panos, Pasquale Borrelli, Katrin Meusburger, Christine Alewell, Emanuele Lugato, and Luca Montanarella. "Estimating the Soil Erosion Cover-management Factor at the European Scale." Land Use Policy 48 (2015): 38-50.

[30] Yitayew, M., Pokrzywka, S. J., Renard, K.G. (1999). Using GIS for facilitating erosion estimation. Applied Engineering in Agriculture 07/1999; 15(4).

[31] Fernandez C, Wu JQ, McCool DK, Stockle CO (2003) Estimating water erosion and sediment yield with GIS, RUSLE,and SEDD. J Soil Water Conserv 58:128–136

[32] Wu Q., Wang M. (2007).A framework for risk assessment on soil erosion by water using an integrated and systematic approach. Journal of Hydrology, 337: 11–21

[33] Renard, K.G., Foster, G. R., Weesies, G.A., Porter, J.P. (1991b). RUSLE, Revised Universal Soil Loss Equation. J. Soil Wat. Conserv. 46(1), 30-33.

The ENVISION project has received funding from the European Union's Horizon 2020 research
and innovation programme under grant agreement No 869366

Figure 6: Overall RUSLE values in parcel scale

$$A = R \times K \times L \times S \times C \times P$$

where A is the computed spatial average soil loss and temporal average soil loss per unit area per year (t/ha year-1), R is the rainfall-runoff erosivity factor [MJ mm/(ha h year-1)], K is the soil erodibility factor [t ha h/(ha MJ mm)], L is the slope length factor, S is the slope steepness factor, C is the cover management factor, and P is the conservation support practice factor. L, S, C, and P are all dimensionless. Individual GIS layers will be generated for each factor in the RUSLE equation and they will be combined at parcel-scale to spatially predict the soil loss.

The proximity to surface water accounts for the parcel's slope orientation and its relation to the proximity orientation. Therefore, multiple parcels can be completely exempted from the proximity risk assessment as their slope orientation is significantly deviating from the proximity orientation. As a results specific threshold of acceptable orientation angle differences will be set in order to indicate perilous cases.

Figure 7:Special Case of an exempted parcel according to water proximity method

The SMR 1: NVZ Water Pollution clearly identifies the requirement for the risk assessment of runoff of polluted water and soil into surface water. Since the risk assessment requirement heavily depend on soil erosivity and erodibility, the RUSLE method of soil loss estimation was considered appropriate. The RUSLE product functions as a proxy to the susceptibility of each parcel to lose soil outside its boundaries.
On the other end, the proximity to surface water function will be the second quantitative output of the methodology; where together with RUSLE form a look-up table of combinations between the two values, estimating the overall risk. The proximity-to-surface-water attribute accounts for the orientation overlap between the proximity orientation and the overall parcel slope orientation. This way, only parcels that could actually contribute polluted soil and water to nearby watercourses will be recorded.
Unfortunately, for the validation of the proposed method, actual SMR 1 compliance records should be available. At this point the algorithm attempts to understand in depth the requirements of the risk assessment of SMR and propose solid and verified GIS and RS techniques to tackle the respective elements of the rule. Finally, the product in question could also be exploited for the identification of parcels that ought to comply with the buffer zones defined in the GAEC 1: Buffer Strips CC requirement.

2.5.5 Input data

The input data required to derive fused Sentinel-2 data is:

(i) Satellite: Sentinel-1, Sentinel-2 L2A
(ii) Products: Spectral bands, backscattering coefficients, coherence coefficients, Vegetation Indices (Vis)
(iii) Auxiliary Shapefiles: e.g LPIS, Agricultural Practices Descriptions, hydrographic networks, Natura2000 regions, etc.

2.5.6 Output data

The service will provide either in file format (GeoTIFF, Shapefile) or via a RESTful API the following

(i) Vegetation Indices
(ii) Soil Indices
(iii) Maps and Analytics (Phenological Metrics, Growth trends, Change Detection et cetera)

Table 7: Output data for the analytics on vegetation and soil index time-series

Short description	File type	Expected size	Frequency
Aggregated satellite or vegetation indices values per polygon request	PreGenerated Layers (raster)	-	*max 3/month between July-October (Lithuania)
CC indicators of compliance (tbd - traffic light system) (SMR 1, GAEC 6, GAEC 4, etc..)	Shapefile	100-500 MB per file	-

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

Automated Burnt Scar Mapping (BSM) and Stubble Burning for the maintenance of organic matter in soil (GAEC 6.)	Raster (GeoTIFF) or/and shapefile	a few MB per file	*max 5/month between March-May (Lithuania)
Runoff Risk assessment for the reduction of water pollution in Nitrate Vulnerable Areas (SMR 1.)	Raster (GeoTIFF) or/and shapefile	up to 100MB per file	1/growing season (Cyprus) 1/growing season (Lithuania)
Minimum soil cover for Soil Erosion (GAEC 4.)	Raster (GeoTIFF) or/and shapefile	up to 100MB per file	*max 5/month between July-October (Lithuania) *max 5/month between January-February (Cyprus)
Natura2000 regions activity hotspot detection	Raster (GeoTIFF) or/and shapefile	up to 100MB per file	max 3/month (Cyrpus)

2.6 Identification of organic farming practices

2.6.1 Name of the service – provider

This service will provide a fully-automated Organic crop identification service, which aims at identifying whether a particular crop type declared as organic is classified as such, based on a traffic light system. The service will contribute to replace direct and guide on-field checks for priority control and will result in the reduction of inspections costs and of the Certification Bodies (CBs) administrative burden, thus ensuring targeted and efficient controls and faster delivery of payments/organic certifications to farmers.

The service will exploit a number of EO derived indicators and tools to ensure effective monitoring of the crop condition variability and phenological stages, in both space and time. To ensure high temporal coverage of the data, the system will utilize data from different spaceborne remote sensors, namely the Sentinel-2 and Sentinel-1 missions.

The provider of the Identification of organic farming practices service is AgroApps. The service will be integrated and delivered as an earth observation component of the ENVISION platform.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

2.6.2 Description of the service

Plants cultivated under organic and conventional farming principles present bio-chemico-physical differences that can be detectable by satellite imagery[34], especially during the vegetative and reproductive growth stages. The Identification of organic farming practices service will benefit from these differences to discriminate between organic and non-organic (conventional) crops. The logic behind the service is to identify distinct patterns characterizing the growth and evolution of organic and conventional crops during the growing season, through the use of both high resolution optical and radar satellite images depicting the phenological status of the cultivated parcels.

Machine learning classifiers (MLC) will be trained to understand the temporal and spectral signatures of conventional and organic crops. The accuracy to discriminate between the two agricultural practices (organic, conventional) of both supervised Support Vector Machines (SVM), and Random Forests classifier schemes will be investigated.

The Predictor Layers will include:

(i) Sentinel-2 MSI's 10 m and 20 m resolution bands in the visible, Near Infrared (NIR), and Short-Wave Infrared (SWIR) parts of the spectrum to exploit Texture Analysis Features,
(ii) multispectral VIs such as NDVI, Normalized Difference Water Index (NDWI), Plant Senescence Reflectance Index (PSRI) and REIP (Red Edge Inflection Point),
(iii) crop biophysical parameters such as leaf area index (LAI), FCOVER and FAPAR, at 20 m spatial resolution),
(iv) radar VIs, such as MPDI (or Polarization Ratio) that represents a normalized polarization, calculated from VV and VH (co-polarized, vertical-vertical and cross-polarized, vertical-horizontal polarization) backscatter signal images captured by Sentinel-1 satellites, PRI (Polarization Ratio Index), which is the VH/VV ratio, as well as the VH and the VV backscatter signals,
(v) Orfeo ToolBox phenological analysis for image time series. In this case, complementary products resulting from fitting parametric models (double logistic function, asymmetric Gaussians) to VIs time profiles, will be used. These are: a. the date of the maximum positive gradient, b. length of the plateau and c. senescence slope.

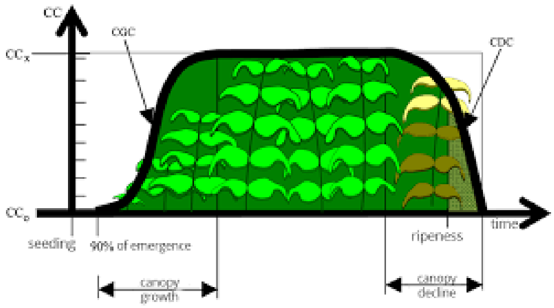

Figure 8: Phenological analysis for image time series.

Once developed, the most accurate ML classifiers will be incorporated in the ENVISION Platform. The service will automatically access and download satellite imagery and provide the infrastructure for the storage of remote sensing data, their preprocessing and calculation of VIs and additional parameters. Regarding the predictor features that will provide the best accuracy for the ML classifiers, either

[34] https://doi.org/10.1007/s13593-015-0313-2

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

Sentinel-1 or cloud-free Sentinel-2 images, or both, will be downloaded every time they become available, to feed the ML scheme and yield the results. It is estimated that at least 9 images per crop type per growing season (crop dependant) will be acquired to clearly depict the timeline of the crops' phenology stages and growth through time.

The Identification of organic farming practices service will provide information to the CBs in parcel level, and upscale their capacity throughout their control and certification processes, by capitalizing on EO data and ML techniques. The identification of organic parcels will be crop specific for soybean, wheat, sunflower and maize for the country of Serbia.

The procedure for organic parcel identification, is described in the following steps:

1. The system communicates with the Copernicus Open Access Data Hub through an API in order to retrieve the corresponding Sentinel product ID and saves the results in an .xml file. Based on the area of interest (AOI), date and cloudiness level thresholds, the CDS-API responds with a unique product ID that matches the search criteria and is the identifier for the targeted tile. The products are then located in the respective satellite image provider.
2. Once the satellite product is identified, a request is issued to the respective provider to acquire it, specifying optionally the bands to download. Data are acquired through an API gateway in a programmatic way. A Python script is used for this task, modified to suit the needs of the service. Possessing the unique ID, the system proceeds to collect the necessary information layers from the EO provider.
3. The primary raw satellite data are corrected for atmospheric and topographic effects, before masking. The analyzed following procedures are included in the workflow of the system.
    - Resampling and Layer Stack. After the data have been corrected, all spectral bands that will be imported into the database are Stacked together into one multiband image using GDAL.
    - Regarding Sentinel-2 optical satellite layers come in different resolutions across the different spectral channels. This is a result of the different spectral range each band exhibits, and the reciprocal relation between band range and pixel resolution. However, a multi sized product cannot be further processed by raster algebra operators since pixels do not align one below the other. Therefore, the multi sized stack has to be resampled appropriately, before any other processing takes place. In this step, all the 10 m (02-04 and 08) and 20 m (05-07 and 8A, 11, 12) of the Sentinel-2 bands are merged together – after resampling. The Bottom-Of-Atmosphere (BOA) reflectance Sentinel-2 L-2A dataset will be used as baseline for the derivation of a number of vegetation indices. The Sentinel-2 L-2A products are atmospherically, corrected. The resulting Sentinel-2 L-2A datasets contain information over the cloud coverage, water presence, thin cirrus presence, etc. A Scene Classification Map (SCL) is included in the L-2A datasets together with Quality Indicators for cloud and snow probabilities.
    - Regarding Sentinel-1 the IW swath acquisition mode will be used, which captures three sub-swaths by employing the Terrain Observation with Progressive Scans SAR (TOPSAR). IW is the main acquisition mode for the systematic monitoring of surface deformation and land changes, providing data with a 250 km swath. Each image has a spatial resolution of 5×20 m with double polarization (VV and VH). Also, the GRDH

The ENVISION project has received funding from the European Union's Horizon 2020 research
and innovation programme under grant agreement No 869366

(Ground Range Detected in High resolution) format will be used. Level 1 GRD products concern SAR data detected, multi-looked, and projected to ground range using an earth Ellipsoid Model, with an approximate square pixel resolution. The ESA SNAP software is used for geocoding and pre-processing SAR data and converting the pixel values from digital numbers (DNs) into measurements of the sigma naught values. To this end, the 1-arcsec SRTM DEMs are being used. SAR data pre-processing follows specific steps like (a) calibration, (b) implementation of a speckle filter (Lee Sigma filter with window size 3 by 3), (c) geometrical correction and registration to a UTM 34N WGS84 ellipsoid with standard terrain correction applied (Range-Doppler Terrain Correction), (d) (Calibration).

- Data Quality Masking. Before the data can be imported into the database, it's very important to remove the abnormal values from it. Pixel reflectance values can be easily corrupted by various types of clouds or cloud shadows. In some cases, it is very easy to detect these pixels, such as those that are covered by thick clouds. But it's very hard to distinguish pixels covered by cirrus clouds or thin clouds, even with visual inspection. To this aim, the system applies a cloud masking technique.
- Cropping the Area of Interest (AOI) and rejecting cloudy tiles. This step is primarily introduced to save memory if needed– for the cases where area of interest only falls into a small portion of the tile. That is why an additional step is performed – cropping the image to a specified polygon using GDAL utilities (gdalwarp with -cutline parameter).

4. Once the desired information is downloaded and after the processes of cloud masking and clipping (based on the requested area of interest), the system applies the ML classifiers to discriminate conventional and organic crops. The polygons of interest (parcels) vector and cropping data are used to partition the time-series of imagery into parcel objects and thus provide parcel specific thematic information, regarding the identification of organic and conventional crops.
5. The processed EO data will be stored in Rasdaman and Post GIS. Geoserver will serve WMS for the purpose of data visualization, and Rasdaman's API will be used for querying the data. Data is inserted into Rasdaman using an import script and a .json file with necessary parameters, such as input files paths, name of the collection where to insert the data, network address of service endpoint, storing options like tiling, band names and date-time extraction etc.

3.6.2.1 Service: Traffic Light System – Organic vs Conventional Identification

The service will provide a traffic light system with the cultivation method classification at parcel level (vector data). It will be set up operationally on the ENVISION Platform to identify the cultivation practices by the end of the growing season. The traffic light system will enable a smart sampling technique for the inspections. Each parcel will be characterized with the confidence of its classification decision (red, green, blue). These smart inspections methodology will identify potential breaches of compliance and assign the appropriate color to suspicious parcels declared as organic, based on their deviation from the classification decision, as provided below.

The ENVISION project has received funding from the European Union's Horizon 2020 research
and innovation programme under grant agreement No 869366

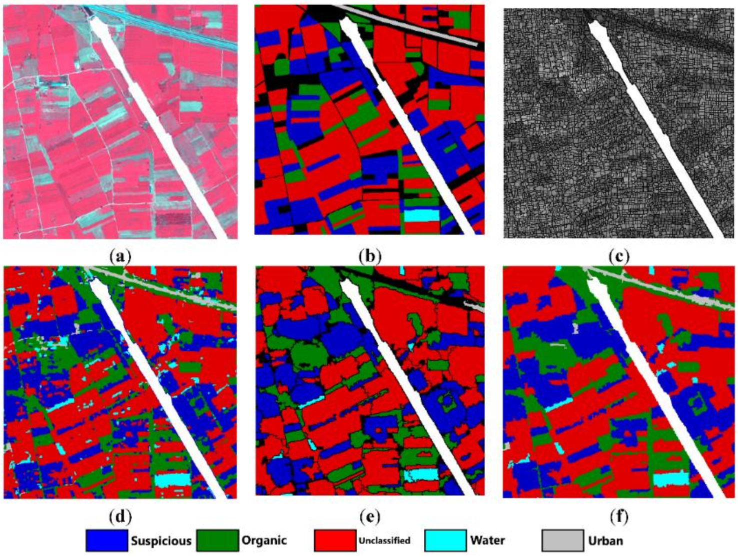

Figure 9: Classification decision

The provided service addresses the needs of:
Certification Bodies (CBs). The service will showcase how systematic and timely EO-based crop monitoring can assist controls, suggesting an effective and efficient service towards the required shift from randomly sampled to targeted inspections. The capacity of the CBs is increased in terms of number and extend of controls, and reduces the involved administrative and operational costs through actionable EO-derived services. It allows CBs to deploy more targeted inspections in shorter time periods and mitigate any difficulties that would otherwise arise in the management of inspectors' workload; it enables the performance of more timely risk-based inspections.

2.6.3 Dependencies of the service from other services

The technological implementation of the satellite-based Identification of organic farming practices service needs to find a compromise between the following elements:

1. availability of input data, with minimal contribution from end-users;
2. elaboration and processing time, with minimum possible timelag between E.O. data acquisition date and information delivery to final users;

3.6.3.1 Internal dependencies

1. System/server crash
2. Maintenance and time for the re-initialization or re-instating of the service at different servers

3.6.3.2 External dependencies/ Data Acquisition plan

1. Crop data.
    - The CBs must upload shp file of farm area, or can draw the farm area on the base map, or can provide coordinates of the farm in the reference World Geodetic System (WGS84). The farm must be considered to comply with organic farming practices and will be monitored throughout the growing season to verify its eligibility and compliance;
    - The external user must specify crop type;
    - The external user must specify sowing date;

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

2. Spectral data. The service can either acquire Sentinel-2 or Sentinel-1, or both, satellite images from any available service provider. Based on the imagery used, the appropriate bands and products will be assimilated for the calculation of the indices that will feed the ML classifier, the results will be crop specific, notifications will be produced based on the decisions of the object-based (parcel) analysis, and visualization (graphs, reports, widgets) will be populated based on the results.

The table below presents the required data sources that will be used for the development and operation of the Identification of organic farming practices service, as well as the spatial resolution of the data, the derived parameters and the update frequency. The polygon – parcel data, that will be made available by the CBs, will contribute the geospatial input for crop delineation and farmers' crop declarations, and will be employed i. for the object partitioning of the images, ii. the supervised classifier's training, and iii. to provide the classification decisions.

Table 8: The data required for the development and operation of the service

Source	Required Data	Spatial resolution	Derived Parameters	Update Frequency
Sentinel-2 mission	Sentinel-2 L-2A L-1C, optical multispectral	10 m, 20 m	Spectral Bands, VIs, biophysical parameters	4-6 days
Sentinel-1 mission	Sentinel-1 SAR, IW, GRD	5 x 20 m	Backscatter signal and VIs	6 days
LPIS (Land-Parcel Identification System)	Parcels vector data acquired	Polygon Data Crop Type	Parcel Geometry	Yearly
CBs	Parcels cropping data	Polygon attributes Farmer's declaration of the cultivation method	Parcel Crop Type	Yearly

2.6.4 Literature on the service/ product

The logic of organic crop identification relies on the fact that plants grown under organic farming principles tend to be less robust due to organic crop management techniques that apply less fertilizer and pesticide, do not use growth regulators and their sowing density is often smaller[35]. The machine learning model will use EO derived features such as vegetation indices, biophysical parameters, SAR backscatter, GLCM textural indices, Phenology metrics, as predictor variables to discriminate between organic and non-organic practices. To this aim, spectral indicators related to the plants spectral signatures will be fully exploited in the multi-temporal training datasets, to calculate the phenological metrics that will enable accurate discrimination. Zonal Statistics of the EO indicators per parcel will form the multi-temporal training datasets to feed the ML classifiers. SVM or Random Forest classifiers will be investigated in their accuracy to provide accurate results and form the final models to be incorporated in the platform.

[35] https://doi.org/10.1007/s13593-015-0313-2

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

2.6.5 Input data

The input data required are EO data to derive VIs and CBs data to derive crop information.

(i) Satellite: Sentinel-2 L2A/Sentinel-1 (EO data)

(ii) Products: Spectral bands, backscattering coefficients, VIs, Phenology Analytics (Predictor Features)

(iii) CBs: Crop Type, Sowing Date, Polygon Data (In Situ Data)

Table 9: Sentinel-2 bands for the calculation of vegetation indices and texture Analysis Features

Band number	Central wavelength (nm)	Spatial resolution (m)
2	490	10
4	665	10
5	705	20
6	740	20
7	783	20
8	842	10
11	1610	20
12	2190	20

Table 10: Sentinel-1 backscattering coefficients

Coefficient	Bandwidth	Polarization	Spatial Resolution (m)
σ°	IW	VV	5x20
σ°	IW	VH	5x20

Table 11: Orfeo ToolBox Analytics applied at the end of crop cycle

Analytics	Parameter
Starting date	Date
End Date	Date
Growth Slope	Number
Senescence Slope	Number
Length of Plateau	Number

Table 12: Data from CBs

Type of Data	Parameter	Source	Units
Crop Data	Crop Type Sowing Date Polygon Data	Farmer	selection date

2.6.6 Output data

The service will provide maps of decision on the cultivated practices and whether these are organic or conventional over a registered parcel by the end of the growing period or within the growing period,

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

updated every time satellite images are available (Sentinel-2 or Sentinel-1). The product is accompanied with a legend showing the values of “organic”, “non-organic”, “not classified” (when the decision’s accuracy is lower than an acceptable value).

Table 13: Table of variables

Type of Data	Parameter	Spatial Resolution	Temporal Resolution
Vector	Decision (Y/N/NC) Organic – Non Organic Not classified	Object-Based (Parcel-Based)	Annually – at harvest (or within the growing period)

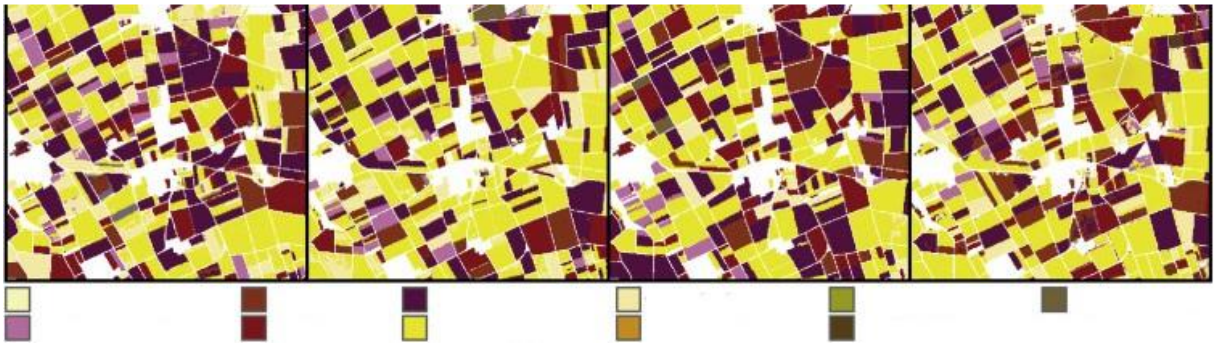

Figure 10: Visualisation of output data

Table 14: Output data for the monitoring of organic farming practices

Short description	File type	Expected size	Frequency
NDVI	GeoTIFF	1-2MB/file	~4/month
Yield estimation	GeoTIFF	1-2MB/file	~4/month
Parcel based Decision (Binary)	GeoTIFF	-	~2/growing season

2.7 Soil Organic Carbon monitoring

2.7.1 Name of the service – provider

The Soil Organic Carbon service (from now on SOC micro-service) is part of ENVISION and will deliver several EO-based services able to support the implementation of Flemish SOC business case (Soil Condition) and the provider is EV ILVO.

As presented in the following figure, the SOC service (see use cases highlighted with blue color) will deliver:

- Image tiles present the SOC spatial distribution (WMS)
- Estimated SOC mean values per agricultural parcel as those exist in LPIS systems. The agricultural parcels will behave as existing geometry objects and will be selected by the end-users.

- Estimated SOC mean values per area of interest (AOI). AOI will be created by the end-users and will act as new geometry objects. Estimated SOC mean values per administration boundaries, for example, the Flemish Region, and per crop type. Existing administrative boundaries will be used as input to the micro-service.

Additionally, the SOC micro-service will provide two extra abilities:

- To calculate and return SOC changes for a specific period. Condition for this is the existence of SOC maps that represent the SOC values in the range of the requested period. This function will support the PAs to effective monitoring of the SOC in time, classify and encode the changes using CAP implementation policy criteria. The Flemish SOC business use case will be the pilot for the definition of service functions and the first finalization of the implementation criteria. For example, it will be possible for the end-user to request the presentation of the SOC changes for a particular agricultural parcel for the period between 2018 and 2023. SOC microservice return data can be combined with other data, in a way to identify correlations between SOC degradation and other parameters like farm management practices.
- To provide SOC measurement data, coming from in-situ measurements and metadata that describe the applied collection methodology, etc. The incoming SOC data together with the new satellite data streams can be used by the SOC image providers, in this case, EV ILVO and LV, to further evaluate and tuning the SOC model and deliver new SOC images with higher accuracy/quality.

2.7.2 Description of the service

As presented in the following figure with blue color, the SOC service aims to:

- Deliver Verified top-soil (0-10 cm) qualitative Soil Organic Carbon estimations,
- Visualize SOC spatial variability at parcel, area, and regional level, and to
- Support the further collection of SOC measurement data, as a way to improve the SOC model and to validate its results.

The service will be used to monitoring CAP's soil requirements (in terms of soil organic carbon) and support the maintenance of soil organic matter level relative to the current and future CAP requirements.

End-users may use service results to get insights and information on tillage, drainage, and overall agricultural management practices.

During the project and specific at the Flemish SOC Business case, end-users will have access to the results of the SOC services throughout the web application. The web application will provide the needed functionality to ENVISION customers or better say end-users (like Flemish farmers, or LV or other policymakers) to monitor the SOC conditions (parent or general use case) under the following possible interaction ways (see orange children use cases):

- Children or specialized use case A: Per agricultural parcel
- Children or specialized use case B: Per agricultural area or AOI (polygon)
- Children or specialized use case C: Per administration boundary (for example the Flemish region) and per crop type.

Depending on the specialized use case needs, the end-user will have the ability to:

- Select the SOC monitoring period, for example, the period 2018-2021[36] and the period 2022.
- Define on a map with digitization the AOI who wants to monitor.
- Select on a map and/or select from a list the administration boundary who what to monitor.
- Select on a map and/or select from a list which agricultural parcels what to monitor.
- Select for which crop types what to aggregated the SOC results.

After this and by using the UI (see green use cases in Figure 1), the end-user will be able to:

- See the displayed in graph SOC values for each specialized use case.
- See the displayed in graph SOC changes for each specialized use case.
- See the SOC values at pixel size in a map for each specialized use case.
- To upload SOC measurements.

To perform the above functions the UI needs to consume data coming from the SOC service and more specific needs to:

- Get SOC values
- Get Map tiles
- Post SOC values

The above use cases will have a dependency form the SOC services. Below there is a more extended description, trying to provide details for their functionality.

3.7.2.1 Service: SOC WMS

The specific service supports the Request for map tiles specialize use case. The TMS will provide SOC tiles presenting the SOC distribution in agricultural areas[37] at high[38] spatial resolution. The service will provide the consumers with the ability to select which SOC map will be presented (the period of the SOC product map). The service can be used at the different use cases (see above specialize use cases) as a background map, to visualize the SOC distribution at a pixel level.

3.7.2.2 Service: SOC values per agricultural parcel

The specific service supports the Calculate SOC values use case. The service will provide SOC mean values, for selected agricultural parcels as those exist in spatial databases like the LPIS. By using the specific service, it will be possible for the end-user to extract the SOC mean or median value, as that exists in a spatial area that is covered by an agricultural parcel. The service will allow the selection of multiple functions like median, min, max, SD, as a way to provide values that present the SOC variability inside the parcel area. The service will provide the ability to the consumer to select from which SOC maps or period the results will be calculated. By default, the service will return all the values as those exist in the available SOC maps for all periods. The service will be integrated as backend services in the SOC Web application of the ENVISION platform and it will be tested by the end-users during the Flemish

[36] SOC maps can be created taking into account one or several satellite images. If the map is a result of one satellite image then the map is a snapshot and we can reference to it by using the satellite image take time, for example if the satellite image was taken on the 28-8-2020, the the SOC map represent the SOC conditions as it was in 28-8-2020. If the map takes into account a set of satellite images, the the map represent the SOC conditions for the period that the satellite images cover, for example if the satellite set cover the year 2018, then the SOC map represents the SOC conditions for the year 2018 or for the period 2018. Using satellite image collections deliver more accurate results and better coverage (for more please see SOC estimation methodology)

[37] Agricultural areas can be filter by using LPIS data or CORINE data

[38] A high spatial resolution satellite image, can be between 10m to 30m. S2 bands have a resolution of 10m, 20m

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

SOC business case. Service results will be visualized in graphs and together with the SOC TMS will provide info regarding the soil conditions at parcel level in terms of SOC. For better visualization results, the service may return an image that represents SOC values within the area of a parcel. GDAL libraries may be used for the generation of the clip georeferenced images. The service will provide SOC values per parcel only when the SOC Maps contain pixels with values that cover the total parcel area (or a threshold of 90%).

3.7.2.3 Service: SOC per area

The specific service supports the Calculate SOC values use case. The service will act similar to the previous one (per parcel) but in this case, the geometry input object will be provided by the end-user either by giving a set of ordered coordinates or/and by digitizing the area of interest on the map (UI). For performance reasons, limitations will exist in the area of the provided polygon (max allowed area). Coordinates need to be validated in terms of order and projection system.
This service will give the ability to the end-users to estimate the SOC values for areas, in a more dynamic way. That functionality serves environmental monitoring needs and is useful for policymakers when they perform medium size SOC monitoring.

3.7.2.4 Service: SOC per region and per crop

The specific service supports the Calculate SOC values use case. The service acts similarly to the SOC for parcels calculating the SOC median values for the area but also per crop type (aggregate summary results per crop type). More specific:

- Input parameters to the service are geometries that represent administration boundaries (for example the Flemish region).
- Service results will provide SOC values using statistical functions like median for the total region area and per crop types as those exists in the specific region.

The per crop functionality is useful for the evaluation of the conditions, at specific sectors for a period. This functionality can be used by policymakers, agricultural institutes, consultants, etc.

3.7.2.5 Service: SOC changes

The specific service supports the Calculate SOC values use case. The micro-service, as explained at intro needs to calculate and return SOC changes for a specific period. Even if this function can be performed at the UI level by combining SOC results from multiple periods, the micro-service design will take into account the specific need and will apply, whenever is possible, the CAP business logic at the micro-service level.

3.7.2.6 Service: Store SOC values

The specific service supports the Store SOC values use case and needs to be able to:

- Receive and validate SOC measurements and their metadata, due to the dependency with the Post SOC value use case.
- Deliver the validated measurements to SOC map providers, like ILVO, supporting this way the Update SOC model use case.

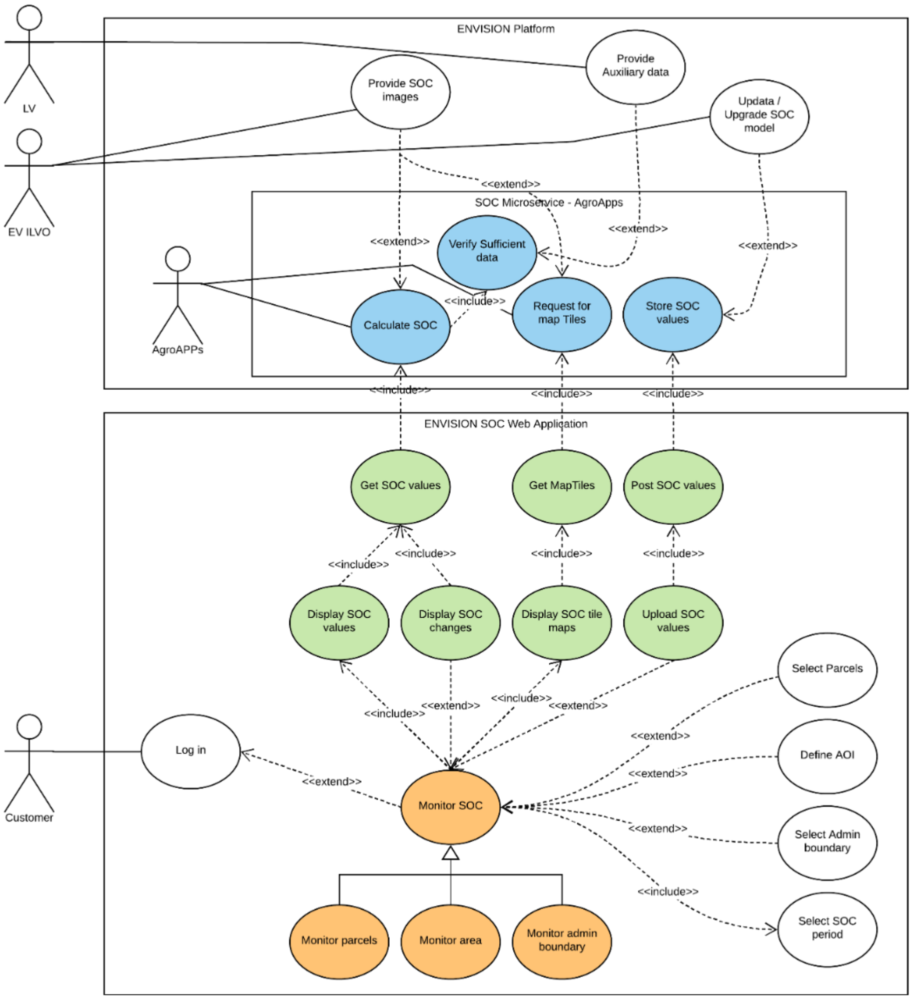

Figure 11: A UML use case diagram that represents a) the Flemish SOC Business case main actions (end-user or customer actions) and b) how the SOC service(s) collaborate with the SOC Web Application and other ENVISION actors.

2.7.3 Dependencies of the service from other services

The SOC services need as an input the following data:

- SOC Carbon images (Maps) presenting topsoil soc values at high spatial resolution (10-30 meters).
- LPIS data.
- Administration boundaries.

3.7.3.1 External dependencies/ Data Acquisition plan

The following dependencies exist:

1. Data coming from SOC map provider. The SOC map providers need to deliver georeferenced images (with EPGS of 3857) that represent the SOC qualitative values at pixel size. For the needs of the Flemish SOC business use case, the development of the SOC images for the Flemish region will be done by ILVO. The methodology that EV ILVO is going to follow for the development of the SOC images for the Flemish region.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

2. Vector LPIS data represent agricultural parcels for subsidy periods. The data need at least to contain the geometry and the crop type description, at a higher level. Crop types can be used to aggregate SOC results. For most EU countries those data are open.
3. Administration boundaries. Those are vector data that present the boundaries of administration regions. Merging of low-level boundaries may result in higher-level boundaries.

All vector data needs to follow specific geometry quality standards.

2.7.4 Input data

The input data to the microservice are:

- SOC Carbon images (Maps) presenting topsoil SOC values at high spatial resolution (10-30 meters).
- LPIS data.
- Administration boundaries.

The bare soil synthetic layer is the basis for the production of a SOC map, together with SOC measurements that can be used for the modeling. The resulting SOC map will contain one main band with the SOC values.

The LPIS data are vector data, with crop values and codes. For the identification of the parcels by the end-user, it will be possible to use the parcel code and the period. PAs, and more specific LV for the needs of the Flemish SOC Business use case, can use the parcel code to link the parcel with the Application Aid Declaration and the Farmers. That is mainly related to the provided functionality at the UI level.

2.7.5 Output data

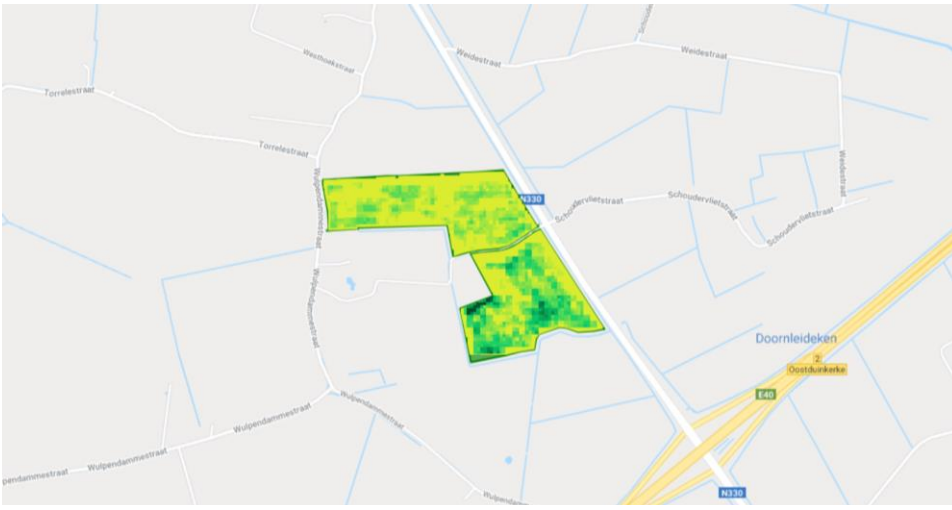

Figure 12: An example of a visualization of the SOC values at pixel level inside agricultural parcels, as those, exist in the Flemish LPIS system.

Table 15: Output data for the SOC monitoring

Short description	File type	Expected size	Frequency
Synthetic bare soil layer obtained by	Multilayer Raster (GeoTIFF)	35GB	1/year (Flanders region

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

Sentinel-2 time series (10 m resolution)			
SOC content for croplands (10 m resolution)	Raster GeoTIFF	1.5GB	1/year (Flanders region
Average SOC content for each agricultural parcel (including other statistics)	Shapefile	300MB	1/year (Flanders region

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

3 ENVISION's Development Pipeline

3.1 Software Development methodology

Software development projects are resource intensive, requiring money and time in order for a final product to be delivered on a timely manner. A way of better managing such project is through the implementation of a software development methodology that would promote a disciplined project management to structure, plan and control the process of developing a system without schedule deviations as well as extra effort and costs.

The development of ENVISION system will follow an Agile approach. Agile methodologies help teams respond quickly to changing requirements and better adapt the planning approach. The BEACON development team will follow the Scrum[39] methodology which is the most popular agile development framework.

Agile methodology is an umbrella term that includes more agile methodologies (e.g. Scrum, Xp, Crystal, FDD and DSDM)[40]. It is based on the combination of iterative and incremental approaches and it is open to changing requirements over time and encourages continuous feedback from the end-users. Agile methodology aims to shorten the time period between the decision-making process and the feedback gained from the beta users. Therefore, Agile methods focus on flexibility, constant improvement and speed. However, it can be hard to establish a solid delivery date or the final product can be quite different from what was initially described.

The most popular agile framework is Scrum, which is followed for incremental product development using one or more cross-functional, self-organising teams. It provides a structure of roles, meetings, rules and artifacts. The development teams are responsible for creating and adapting their processes within this framework. Scrum uses fixed-length iterations, called Sprints, which are typically two weeks to 30 days long. Scrum teams attempt to build a potentially shippable (properly tested) product increment on every iteration.

Based on this approach, the ENVISION team (WP2, WP3 and WP4) created User Stories (US) that were part of the user requirements identification phase (recorded on the D2.2). A US is describing in a simple language a requirement at a very high level and contains the right amount of information for the development team to estimate the effort needed to implement it. Wireframes will be produced and circulated to the end-users in order to receive feedback and finalise the first version of the ENVISION platform. This feedback will be collected and will form the product backlog (aka the list of the issues/ suggestions that should be completed as part of the project). This backlog will be prioritised based on the importance of each feature and the dependencies between the US. However, the prioritisation of the issues is a constant task and it depends on several criteria (severity, type, necessity). An issue can be characterised as completed only when development is complete, testing has taken place, all bugs are fixed, review has also performed and the product owner closes the issue.

After prioritization is completed, the development team along with the product owner estimate the effort required for each registered issue. A list of issues is part of a sprint, which approximately lasts for two weeks and after its completion, a product is released (Minimum Value Product - MVP). Just before the beginning of the sprint, the development team commits to the registered issues that will

[39] http://scrumreferencecard.com/scrum-reference-card/

[40] https://en.wikibooks.org/wiki/Introduction_to_Software_Engineering/Process/Agile_Model

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

be completed within the next two-week period and discuss further details of them in order to form the acceptance criteria and break the issue down in smaller tasks. Before every release, a more detailed test run, called regression testing, is performed in order to ensure that the released version is bug-free.

In order to assist this process, the development team uses Jira, a software development tool that is used for planning, tracking and managing agile and software development projects. It is easy to be customized based on the project and team's workflow in order to release great software. Furthermore, slack is used from the technical team to constant communication.

3.2 Testing methodology

A testing methodology will be followed throughout the duration of the development phase of the ENVISION platform. By having a thorough testing methodology, it will be guaranteed that the development of the ENVISION platform will be secured from the very low-level technical details all the way to user interface. This approach will be constantly updated based on the project phase and/ or on any specific testing requirements.

After the completion of a sprint, the development team will allocate a specific amount of time testing the issue and the US that have been finalized. The US will be tested based on the defined acceptance criteria and the requirements list. Whether bugs are found during the testing procedure, then they will be registered into the tracking system (Jira) and they will be fixed, if it is feasible, before the end of the sprint. In some cases and after a common decision among the development team and the product owner some of the bugs might be transferred to another sprint based on the severity and the impact of such an action. Once all the bugs are fixed, a final review of the released version will take place to assure that this version is ready to be delivered. As a final step, a regression testing will be performed to ensure that the version is functionable and bug-free.

3.2.1 API & Integration Testing

The REST API handles the communication between the various modules of the ENVISION platform. For this reason, a round of tests will be performed to ensure its successful operation. Therefore, the API's endpoints will be tested by sending requests with specific parameters and making sure that proper data and status codes will be returned in the response.

After each component is ready, it will be integrated into the ENVISION platform and an integration testing will be performed by passing the necessary inputs to the component, validating that the expected outputs are produced. One of the main concerns of the development team is to ensure that the communication between the new components and the already integrated system will work according to the expectations and if the specific component is not working, then the system will not fail as a whole.

3.2.2 User Interface Testing

The ENVISION platform will support all the latest versions of all common browsers (Chrome, Firefox, Safari, Internet Explorer, mobile browsers) in order to ensure that navigation and the user experience will not change based on the used browser.

The ENVISION project has received funding from the European Union's Horizon 2020 research
and innovation programme under grant agreement No 869366

3.2.3 User Acceptance Testing (UAT)

Business cases users will be responsible for the UAT who will verify that the toolbox was developed based on their needs and requirements. Feedback taken from the MVP testing and the pilot implementation will be taken into consideration and the platform will be modified and adjusted accordingly.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

4 Conclusions

This deliverable aims to provide the ENVISION system's architecture and the services specifications in detail, the technologies stack and development processes that will be followed for the development of the ENVISION system.

This document's target audience is mainly the technical partners of the consortium so that we all have a common understanding of ENVISION's architecture. It will act as a reference and a source of information for all partners.

End of Document

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 869366.