Big Data to Enable Global Disruption of the Grapevine-powered Industries

D8.3 - Integration and Operation with real-life Software Systems

DELIVERABLE NUMBER	D8.3
DELIVERABLE TITLE	Integration and Operation with real-life Software Systems
RESPONSIBLE AUTHOR	Simone Parisi

Co-funded by the Horizon 2020
Framework Programme of the European Union

GRANT AGREEMENT N.	780751
PROJECT ACRONYM	BigDataGrapes
PROJECT FULL NAME	Big Data to Enable Global Disruption of the Grapevine-powered industries
STARTING DATE (DUR.)	01/01/2018 (36 months)
ENDING DATE	31/12/2020
PROJECT WEBSITE	http://www.bigdatagrapes.eu/
COORDINATOR	Nikos Manouselis
ADDRESS	110 Pentelis Str., Marousi, GR15126, Greece
REPLY TO	nikosm@agroknow.com
PHONE	+30 210 6897 905
EU PROJECT OFFICER	Ms. Annamária Nagy
WORKPACKAGE N. | TITLE	WP8 | Grapevine-powered Industry Application Pilots
WORKPACKAGE LEADER	ABACO
DELIVERABLE N. | TITLE	D8.3 | Integration and Operation with real-life Software Systems
RESPONSIBLE AUTHOR	Simone Parisi
REPLY TO	s.parisi@ABACOgroup.eu
DOCUMENT URL	http://www.bigdatagrapes.eu/
DATE OF DELIVERY (CONTRACTUAL)	30 June 2019 (M18), 30 June 2020 (M30 2nd Version), 31 December 2020 (M36 – Final Version)
DATE OF DELIVERY (SUBMITTED)	28 June 2019 (M18), 30 July 2020 (M31 2nd Version), 31 December 2020 (M36 – Final Version)
VERSION | STATUS	3.0 | Final
NATURE	R (Report)
DISSEMINATION LEVEL	PU (Public)
AUTHORS (PARTNER)	Simone Parisi (ABACO), Florian Schlenz (Geocledian)
CONTRIBUTORS	Evanthia Bozou, Giannelou Iliana, Petri Alina, Kouriantaki Eirini, Mpokogianni Ifigeneia, Kaloudis Stathis, Kontogiannis Theodoros, Giannis Stoitsis, Mihalis Papakonstadinou, Francesca Tsaropoulou (Agroknow)
REVIEWER	Maritina Stavrakaki (AUA)

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.1	Initial ToC and document structure	29/5/2019	Simone Parisi (ABACO)
0.2	Chapters 1, 2, 3	15/6/2019	Simone Parisi (ABACO)
0.3	Chapter 4	19/6/2019	Florian Schlenz (GEOCLEDIAN)
0.4	Chapters 5,6	22/6/2019	Simone Parisi (ABACO)
0.5	Input from pilot partners	24/6/2019	Florian Schlenz (GEOCLEDIAN), Aikaterini Kassimati (AUA), Panagiotis Zervas (Agroknow)
0.7	Peer-review comments	26/6/2019	Maritina Stavrakaki (AUA)
0.8	Pre-final draft for 1st version	27/6/2019	Simone Parisi (ABACO)
0.9	Internal Review	27/6/2019	Maritina Stavrakaki (AUA)
1.0	Final review of v1	28/6/2019	Simone Parisi (ABACO)
1.1	Draft version of v2	25/5/2020	Florian Schlenz (GEOCLEDIAN)
1.2	Update Chapters 1, 2, 3	2/6/2020	Simone Parisi
1.3	Update Chapter 4	4/6/2020	Florian Schlenz (GEOCLEDIAN)
1.4	Update Chapters 5,6	5/6/2020	Simone Parisi (ABACO), Florian Schlenz (GEOCLEDIAN)
1.5	Internal Review	11/6/2020	Maritina Stavrakaki (AUA)
1.6	Updates based on internal review	11/6/2020	Simone Parisi (ABACO)
2.0	Final version of v2	30/6/2020	Simone Parisi (ABACO)
2.4	Update Chapters 1, 2, 3	29/11/2020	Simone Parisi (ABACO)
2.5	Update Chapter 4	30/11/2020	Florian Schlenz (GEOCLEDIAN)
2.9	Update Chapter 5	16/12/2020	Evanthia Bozou, Iliana Giannelou, Alina Petri, Eirini Kouriantaki, Ifigeneia Mpokogianni, Stathis Kaloudis, Theodoros Kontogiannis, Giannis Stoitsis, Mihalis Papakonstadinou, Francesca Tsaropoulou (Agroknow)
2.9	Internal review	31/12/2020	Maritina Stavrakaki (AUA)
3.0	Final version of v3	31/12/2020	Simone Parisi (ABACO)

PARTICIPANTS		CONTACT
Agroknow IKE (Agroknow, Greece)		Nikos Manouselis Email: nikosm@agroknow.com
SIRMA AI (SAI, Bulgaria)		Todor Primov Email: todor.primov@ontotext.com
Consiglio Nazionale DelleRicherche (CNR, Italy)		Raffaele Perego Email: raffaele.perego@isti.cnr.it
Katholieke Universiteit Leuven (KULeuven, Belgium)		Katrien Verbert Email: katrien.verbert@cs.kuleuven.be
Geocledian GmbH (GEOCLEDIAN, Germany)		Stefan Scherer Email: stefan.scherer@geocledian.com
Institut National de la Recherché Agronomique (INRA, France)		Pascal Neveu Email: pascal.neveu@inra.fr
Agricultural University of Athens (AUA, Greece)		Katerina Biniari Email: kbiniari@aua.gr
Abaco SpA (ABACO, Italy)		Simone Parisi Email: s.parisi@abacogroup.eu
SYMBEEOSIS EY ZHN S.A. (Symbeeosis, Greece)		Konstantinos Rodopoulos Email: rodopoulos-k@symbeeosis.com

ACRONYMS LIST

BDG	BigDataGrapes
WP	Work Package
D	Deliverable
DSS	Decision Support System
REST API	Representational State Transfer Application Program Interface
SVIs	Spectral Vegetation Indices
Vit	Vitality
NDVI	Normalized Difference Vegetation Index
NDRE1	Normalized Difference Red Edge Index (v1)
NDRE2	Normalized Difference Red Edge Index (v2)
NDWI	Normalized Difference Water Index
SAVI	Soil Adjusted Vegetation Index
EVI2	Enhanced Vegetation Index 2
CI-RE	Chlorophyll Index - Red Edge
RGB	Red, Green, Blue
BDG	BigDataGrapes

EXECUTIVE SUMMARY

The deliverable (D8.3) focuses on the task **T8.3 Integration with existing real-life Software Systems:** this task focuses on the integration of the BigDataGrapes software stack and data model into a market-ready Farm Management System produced by ABACO, SITI4Farmer, augmenting its functionalities to support the implementation of the envisaged Use Cases.

This deliverable aims to incorporate by prototyping onto SITI4Farmer, the relevant functionalities of the BigDataGrapes software stack, further used in the piloting sessions.

Also, in this framework, an integration with FOODAKAI, an existing software on food security that is complementary to the capabilities of ABACO SITI4Farmer, has been performed. In fact, food security depends also on meteorological and agronomical issues that are the focus of SITI4Farmer.

Within the BigDataGrapes scope, the farm management system integration refers to the integration of weather-soil-plant monitoring data via data exchange over Rest-APIs in JSON format. Another important integration domain is represented by the GEOCLEDIAN Satellite data and the related vegetation indices. The real-life software system allows to define (draw and edit) farming plans at field level on top of which time series of indices are calculated.

This document is structured as follows: after a brief introduction, Chapter 2 focuses on the farm management pilot and the related data recorded. Chapter 3 is about datasets and their relationships with use case scenarios. Chapter 4 describes SITI4Farmer and how the BigDataGrapes technologies are integrated into the system. The exploited BigDataGrapes technologies are related to satellite data from GEOCLEDIAN and data visualization improvements with the support of KU LEUVEN. Chapter 5 describes FOODAKAI and how the BigDataGrapes technologies are integrated into the system.

Finally, Chapter 6 describes the next steps and conclusions

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The deliverable D8.3 – Integration and Operations with real-life Software Systems, focuses on the integration of ABACO software SITI4farmer into BigDataGrapes task.

The task focuses on the incorporation of the BigDataGrapes software stack in an existing Farm Management System produced by ABACO, SITI4Farmer, augmenting the functionalities required for the implementation of the envisaged Use Cases tanks in cooperation with the technological partners of the project.

The exploited BigDataGrapes technologies are related to satellite data from GEOCLEDIAN and data visualization components with the support of KU LEUVEN. CNR developed a water stress prediction module based on Machine Learning.

Apart from the Farm Management system, in the second review version of the deliverable an integration with an existing software on food security, that is dependant from what happens at Farm level has been introduced. In this framework, an integration with FOODAKAI existing software which is complementary to the capabilities of ABACO SITI4Farmer has been performed. FOODAKAI has been discussed in section 5, while the other sections are focused on SITI4Farmer and GEOCLEDIAN software stacks.

Within the Big Data Grapes context, the new version of the software SITI4Farmer named ABACO Farmer is currently being updated based on the findings and developments of the project, more specifically with functionalities regarding satellite indices data storage and display, field sensors data storage and display, and with new data visualization widgets, the DSS, calculated from both satellite and field sensors, that can be easily used by piloting farmers. The improvement in data visualization includes providing practical decision support in farming activities, such as suggesting water quantity to be supplied and so on.

The capabilities of the real-life software span from Precision Farming techniques and good agricultural practice capturing, in a corporate database, the information and data of the company and its daily activities. The system integrates this data with other information from heterogeneous sources.

The data feed is thoroughly geo-referenced into the corporate knowledgebase onto which a set of analysis algorithms are executed. These algorithms process information and indicators in order to support a smart management of the company.

SITI4farmer follows European Union's guidelines for Precision Agriculture, defined as a way to "apply the right treatment in the right place at the right time", a farming management concept based upon

- observing
- measuring
- responding

to inter and intra-field variability in crops, or to aspects of animal rearing.
The goal is to define a decision support system (DSS) for whole farm management with the goal of optimizing returns on inputs whilst reducing environmental impacts.

2. FARM MANAGEMENT PILOT

2.1 INTRODUCTION & SPECIFIC GOALS

The ABACO and Geocledian Farm Management Pilot is focused on developing a unique system that satisfies these needs:

- Farm Management System with all functionalities to support the farmer in his day-by-day activities and to help gathering field data
- Host data from various sources and provide tools and functionalities for comparisons and easy data management
- Data exchange: (a) a "day by day" data producer, to feed the generated data into the other BDG components and (b) a "data consumer" to use the information from the other BDG components.
- Data visualization: the data relevant for the farmer should be displayed in a way that provides an added value and new insights for farming activities.

Two wine makers were identified as actors in this pilot. They are involved in the pilot in two ways:

- They are supported in their work by the developed products and, apart from the farm management system itself, such products include sensors and other measurement techniques that provide decision support data.
- On the other hand, these actors help design the new system by providing feedback and know-how about their needs and activities. They can also give insights on how to disseminate results, approaches and ideas delivered by the BigDataGrapes Project.

2.2 SITE DESCRIPTION

The 2 wineries are located in Tuscany, Italy:

Figure 1 Tuscany region map and the position of the two piloting farms

Company Name: CASATO PRIME DONNE CIRCA
Address: Località Casato – Montalcino, Tuscany, IT
GPS Coordinates: 43.088196° N 11.464319° E
Internet Site: www.cinellicolombini.it

12 HA of wineyards of Brunello of Montalcino

Figure 2 Casato Prime Donne vineyards aerial image

Company Name: CANTINA IL PALAZZO
Address: Loc. Antria, Arezzo, Tuscany, IT
GPS Coordinates: 43.502773, 11.904402
Internet Site: www.tenutailpalazzo.it

35 HA of Wineyards of CHIANTI D.O.C.

Figure 3 IL Palazzo vineyards aerial image

2.3 EQUIPMENT USED

ABACO supplied a version of SITI4farmer readily accessible by the 2 winemakers and the project partners.

SITI4farmer is a web-based system that handles:

- Preparing of the graphical crop plan or farming plan
- Managing farming practices and phenology phases
- Analyzing indices
- Providing dashboards to support decisions (agro-meteorology and vegetation)
- Keeping farm data organized and accessible
- Recording field data with the SITI4land app
- Printing and exporting data

Furthermore, it is able to integrate weather data and services from different sources, and it can use open databases and local land registries made available by everyone that expose data through web services.

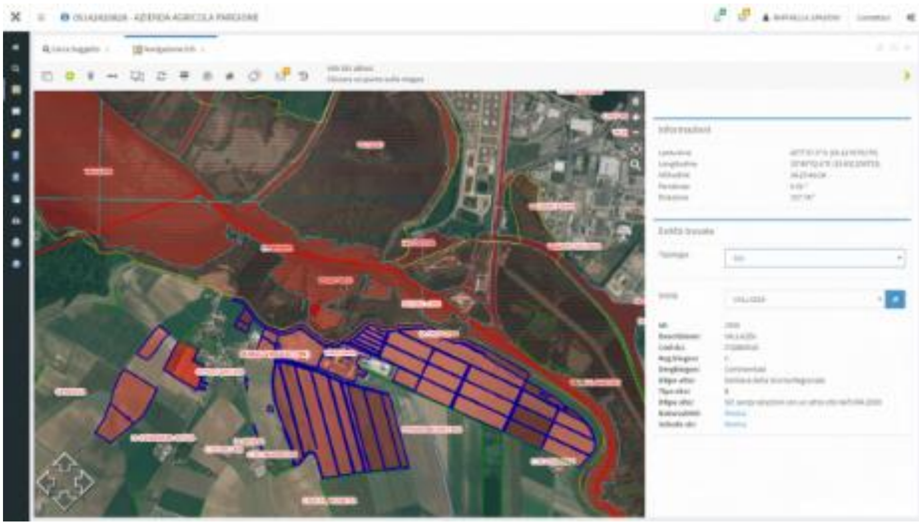


Figure 4 SITI4farmer screen view

In order to make full and comprehensive measurements in the fields, automatically as much as possible, ABACO acquired 2 Weather and sensor stations and integrated the data in their system. The Sensor Stations have been purchased directly by ABACO, and installed at the two farms, after a dedicated study with ABACO's experts, to find the right place and position in accordance with the expected measurement quality.

The weather stations have been set up with a data transmission via radio to a central server which transmits the data directly to SITI4farmer. They are equipped with:

- Modem, aerial, battery, solar panel
- Rain Gauge Module
- Temperature and humidity sensors
- Wind direction system
- Wind speed measurement sensors
- Solar Radiation sensor
- Single Leaf Temperature Sensor
- Infrared Temperature Module
- IR Temperature sensor
- Leaf Wetness Sensor Module with a 5-meter cable
- Drill & Drop Sensors (Temperature and soil moisture sensors)

Figure 5 Sensor & Weather Station (left) and Rain Gauge Module (right)

Geocledian is acquiring, processing and delivering Copernicus Sentinel-2 and USGS Landsat-8 images for all sites during the pilot run time directly to SITI4Farmer (more details in section 4.1).

2.4 TIMELINE

ABACO's main Tasks and Operations, that have been performed to achieve the goals of the pilot are:

- Formal Engagement of the winery companies
- Collecting information of fields, terrain, product quality
- Analysis for the sensors set up on the right spot and configuration
- Deployment of the system SITI4farmer for the 2 companies
- Development of the interfaces and interoperability with the central system of the sensors stations
- Measurements and monitoring of field activities
- Integration with Geocledian services
- Improvement in data visualization with KU-Leuven

Throughout the pilot duration, Geocledian has acquired and processed the described satellite data of all sites. Visible images and Vegetation Index Maps together with the newly developed EO data products are produced in their Processing platform Ag|knowledge and the data is provided to all project partners in near real-time. The following figure provides a rough overview of ABACO's and Geocledian's main tasks:

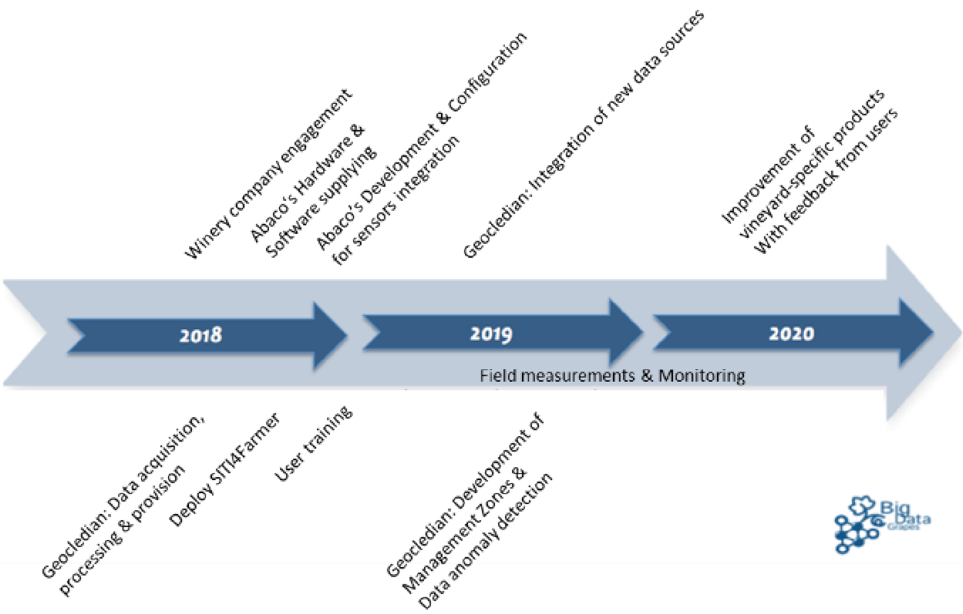


Figure 6 Timeline of Farm Management piloting plan

2.5 MEASUREMENTS

ABACO, with its SITI4farmer system, is engaged to provide the big picture of farm management data, bonding information from external environments, space and ground activities. This means to measure all environmental variables as presented in the previous paragraph, and in the following sections, together with all information coming from the farm activities, for example: treatments/fertilizing (when, what, how much), ground handling, or tasks related to the culture management; and with data from the position, shape, terrain, geographically localized. Geocledian is delivering a variety of satellite data products via REST API to SITI4Farmer (see section 4.1 for details).

2.6 ENVISAGED OUTCOMES

The Farm Management Pilot is focused on developing SITI4Farmer further into a unique system that satisfies these needs:

- Farm Management with all the functionalities to support the farmer in his day by day activities and gather data from the field
- Hosting data from different sources with proper tools and functionalities for comparisons and easy data management
- Data exchange. A “day by day” data producer, to feed the generated data into the other BDG components and make use of the incoming information from the other BDG components.
- Data visualization. The data relevant for the farmer should be displayed in a way that provides an added value and new insights to the farmer for his activities.

In the frame of the pilot, Geocledian is further developing their initial data processing platform Ag|knowledge into a Big Data Processing Platform that allows the scalable production, provision and analysis of large scale data sets of new vineyard-specific products for all test sites of the project so that they can be integrated into farm management systems, like ABACO’s SITI4Farmer. The combination of remote sensing with in-situ field and weather data enables the development of new Farm Management products. ABACO makes use of the output from Geocledian, from sensors, and from the users of the system, in order to create knowledge maps and data systems to relate the crop quality with all the other variables.

3. SITI4FARMER & USE CASES

The SITI4FARMER Portal, at the beginning of the project, consisted of few essential modules that mainly allowed data overview at different scales:

LAND DATA OVERVIEW

- Historical meteo
- Meteo Forecast
- Climatology
- Pedology maps
- Pedological survey
- Ortophoto base map

FARM DATA OVERVIEW

- Asset management
- Warehouse and Stock

FIELD DATA OVERVIEW

- Crop plan
- Farming activities
- Product management
- Mandatory registries
- Record activities on the field, with SITI4Info APP
- Agrometeorological indices
- Earth Observation

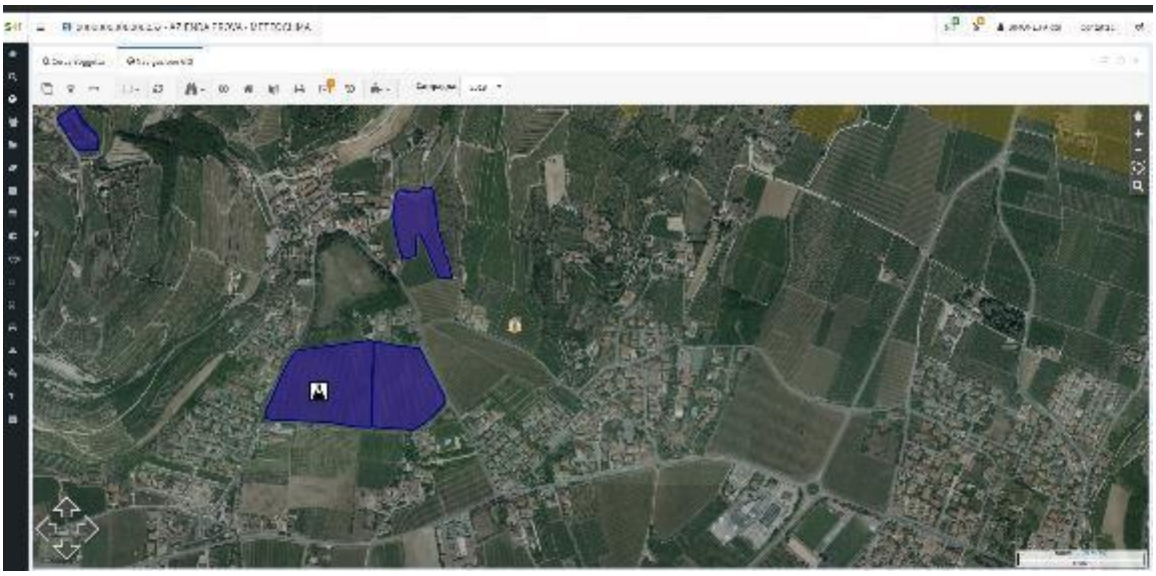


Figure 7 The GIS Navigation system in SITI4FARMER

Starting from the software prototype described above, the ABACO and Geocledian activity in the Farm Management Pilot is focused on developing a unique system that satisfies these needs:

- Farm Management with all the functionalities to support the farmer in his day by day activities and gather data from the field
- Hosting data from different sources with proper tools and functionalities for comparisons and easy data management
- Data exchange. A "day by day" data producer, to feed the generated data into the other BDG components and make use of the incoming information from the other BDG components.
- Data visualization. The data related to the farmer should be displayed in a way that provides an added value and new insights to the farmer for his activities giving precise suggestion for best practice activities.

In particular, precision farming features like the production of variable rate maps represent a powerful instrument for improving quality and productivity in the vineyards.

The pilot's Use Case scenarios have been developed to describe the new capabilities that have been developed for SITI4FARMER. The work on two use case scenarios is as follows:

The first scenario is on the Optimization of Farm Practices in the Vineyard. Management Practices such as irrigation, fertilization and phytochemicals are regularly over- or underestimated with respect to the real plant needs. Especially in the case of overestimation in quantities, there is a negative countereffect towards the plants. The fertilization or phytochemical spraying actions can be supported by satellite data and new precision agriculture practices are based on the evidence that there are interaction effects between weather, management practices and vegetation/fruit qualities. By means of the Precision Agriculture module in SITI4FARMER and in particular new data modelling features, variable rate best practices have been implemented and tested.

The second scenario is on Earth Observation Data Anomaly detection & classification. EO data can serve as a valuable source of information for precision farming. But in order to make efficient use of Earth Observation (EO) data for Farm Management applications, it is crucial to be able to differentiate between data issues and anomalies. Anomalies detection is possible through the detection of deviations between Expectation and Observation. Inputs that can support this are the Static Heterogeneity of the field (Management Zones) and typical patterns of expected crop development for the observed environmental conditions. Classification of anomalies should be able to differentiate between Data errors (clouds, shadows, atmospheric disturbances) and Farm Management related issues (Pests, diseases, vegetation stress through missing water or fertilizer or weather-related damage).

4. SITI4FARMER SYSTEM EVOLUTION

The SITI4Farmer prototype that existed at the beginning of the project is briefly described in section 3. The following sections provide an overview on the system evolution during the project.

4.1 BIG DATA GRAPES DATA

SITI4Farmer acts as an input into the BigDataGrapes data model. Moreover, SITI4Farmer also acts as an output from semantic data modelling by Ontotext.
As regards to Big Data, an essential input is represented by Satellite data provided by Geocledian with its analytics capabilities.

4.2 SATELLITE DATA

ABACO provides Landsat 8 and Sentinel 2 data from Geocledian for each lot registered in the system SITI4Farmer. Every lot has an ID that is registered when a new lot geometry is saved. As the new ID / lot is generated, the downloading of satellite index time series referred to the campaign year in which the lot is registered starts.
In a few minutes, satellite indices based on Landsat 8 and Sentinel 2 data time series are calculated and cut on the lot geometry. All the image raster data and lot statistics on the vegetative indices are provided by means of Rest-API standard in JSON Format as well as geotiff/pngs. The JSON Format is then easily translated in tables and plots displayed in the SITI4Farmer dashboard. In the next section, the work that has been done on the satellite image processing side is described.

Geocledian has acquired, processed and delivered all available USGS Landsat 8 and Copernicus Sentinel 2 satellite images for 2013 – 2020 above a certain cloud cover threshold for all available parcels in this pilot. Geocledian's Cloud Processing Platform provides the field monitoring service ag|knowledge that allows the automatic crop monitoring for fields with multispectral products. Ag|knowledge provides a REST API allowing easy access and integration of satellite remote sensing data and analytics into agricultural applications. The API provides access to field monitoring products for registered parcels (i.e. fields or parts of land). The data for each parcel are immediately updated as soon as new measurements are available. For all of these products, time series and a full history of the last 5 years are available. Currently, these data products are available:

Table 1 The data products currently available in Ag|knowledge

Name	Description
Visible	A visible true colour image of the parcel (RGB).
Vitality	A simplified NDVI product optimized for vegetation vitality visualization.
Variations	A simplified relative NDVI product optimized for vegetation vitality variations visualization.
Normalized Difference Vegetation Index (**NDVI**)	NDVI is a quantitative vegetation monitoring tool used as indicator for the vitality of a crop in particular for the live green vegetation. NDVI correlates with the amount of leaf area (LAI) of active, healthy, green vegetation. Quantifies vegetation by measuring the difference between near-infrared (vegetation strongly reflects) and red light (vegetation absorbs). It is the most widely used vegetation index. Overall, NDVI is a standardized way to measure the amount of healthy vegetation, although it has the disadvantage to saturate at high leaf area levels and therefore shows limited variation in dense fields with high biomass. It minimizes topographic effects. Reference: Rouse et al. 1974

Normalized Difference Red Edge Index (v1) (**NDRE1**)	Substitution of NDVI's red band with NDRE's red edge band (730nm). The Red Edge Indices are designed to estimate chlorophyll content in the canopy. More sensitivity in vegetation with high LAI. Less sensitivity to open water. There are several different formulas for the NDRE index. Reference: Clevers & Gitelson 2013
Normalized Difference Red Edge Index (v2) (**NDRE2**)	Substitution of NDVI's red band with NDRE's red edge band (700nm). The Red Edge Indices are designed to estimate chlorophyll content in the canopy. More sensitivity in vegetation with high LAI. Less sensitivity to open water. There are several different formulas for the NDRE index. Reference: Clevers & Gitelson 2013
Normalized Difference Water Index (**NDWI**)	NDWI is less susceptible to atmospheric scattering than NDVI but does not remove completely the background soil reflectance effects, similar to NDVI. Because the information about vegetation canopies contained in the SWIR channel is very different from that contained in the VIS channel, NDWI is considered as an independent vegetation index. It presents enhanced sensitivity to vegetation water content & water stress. There are also other "NDWI" indices with different meaning. Reference: Gao 1996
Soil Adjusted Vegetation Index (**SAVI**)	The index minimizes soil brightness influences from spectral vegetation indices involving red and near-infrared (NIR) wavelengths. It is interesting in sparse vegetation canopies or early growing stages. Reference: Huete 1988
Enhanced Vegetation Index 2 (**EVI2**)	The enhanced vegetation index (EVI)2 is a vegetation index designed to enhance the vegetation signal with improved sensitivity in high biomass regions and improved vegetation monitoring through a de-coupling of the canopy background signal and a reduction in atmosphere influences. Reference: Jiang et al. 2008
Chlorophyll Index - Red Edge (**CI-RE**)	The index is used to calculate the total chlorophyll content of the leaves. The CIgreen and CIred-edge values are sensitive to small variations in the chlorophyll content and consistent across most species. Apart from the very high Chlorophyll content, it also presents Nitrogen sensitivity and thus canopy Chlorophyll and Nitrogen contents can be derived from this index. Reference: Clevers & Gitelson 2013

Throughout the pilot duration, Geocledian acquires and processes the described satellite data of all sites. Visible images and Vegetation Index Maps are produced, and the data provided to all project partners in near real-time via REST API as JSON, png and Geotiff files. For every parcel geometry registered in our system, we deliver the time series of satellite images together with timeseries statistics on all the vegetation indices. The data can be visualized with visualization components, e.g. a web widget that is already integrated in S4F and used by KUL for dashboard development.
After being downloaded to the Ag|knowledge platform, the data is preprocessed, atmospherically corrected, cloud masked, and ready-to-use images are produced for every registered parcel. The full product overview is available here: https://sites.google.com/site/geocledian/home/product-overview

The following figures show vegetation index time series for one parcel of the Tenuta il Palazzo vineyard:

Figure 8 A Sentinel-2 RGB image for a parcel of the Tenuta il Palazzo vineyard, Italy. All available Landsat 8 and Sentinel-2 image acquisitions for 2017 – 2020 are also shown.

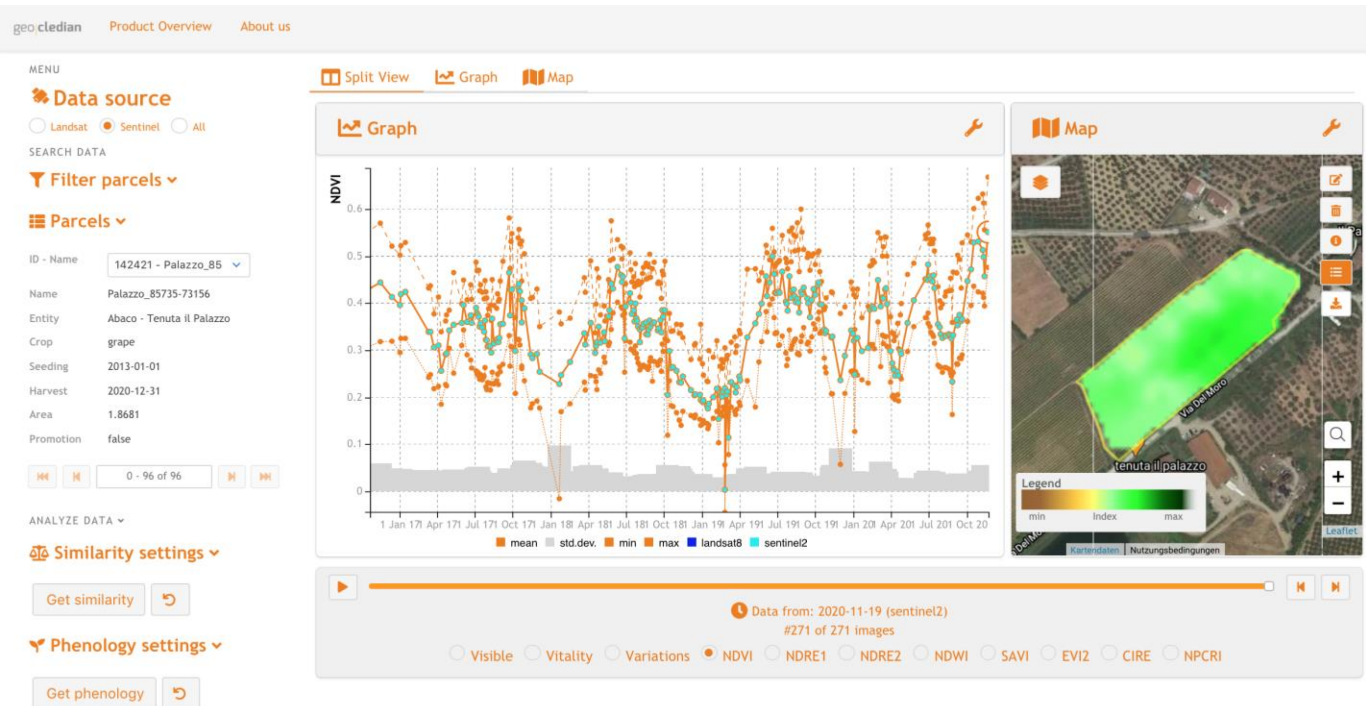


Figure 9 A Sentinel-2 NDVI time series for 2017 – 2020 for a parcel of the Tenuta il Palazzo vineyard, Italy.

Figure 10 A Sentinel-2 NDWI time series for 2017 – 2020 for a parcel of the Tenuta il Palazzo vineyard, Italy. This NDWI is known to be related to leaf water content.

Figure 11 A Sentinel-2 CI-RE time series for 2017 – 2020 for a parcel of the Tenuta il Palazzo vineyard, Italy. The Chlorophyll Index-Red Edge (CI-RE) is reported to be highly correlated with canopy Chlorophyll and Nitrogen contents.

At the start of the project, the first version of Ag|knowledge, a basic image delivery service based only on the first 3 products (Visible, Vitality, Variations), was integrated in SITI4Farmer allowing the user to see these products for his fields without any user interaction.

Extensive tests in the first project year have shown that Ag|knowledge at project initiation had bottlenecks concerning the data processing capacity and scalability, especially when large amounts of new parcels have been registered in the system. On top of that, certain features had to be improved in order to enable new data product generation and increase the product quality.
Therefore, a series of developments have been implemented and deployed successfully to improve data download and processing, performance monitoring, scalability and data visualization and to enable the delivery of the new data products and vegetation indices that were developed for ABACO. In the frame of this project, the field monitoring service Ag|knowledge is further developed from a basic image delivery service into an Agricultural Big Data Processing Platform that allows the scalable production, provision and analysis of large-scale data.

More specifically these tasks have been carried out in the frame of the pilot:

Scalability

- Development and implementation of processing performance monitoring tools for Ag|knowledge
- Performance analysis of certain Ag|knowledge components, e.g. the satellite scene or parcel processing
- Architectural refactoring of the initial system into modular subsystems to enable scalability and processing optimizations
- Development of scalable modules with CNR (technical details reported in WP4 deliverables)
- Increase the scalability of the user management of Ag|knowledge

Data acquisition, processing & provision

- Development and implementation of extended data download and processing routines (e.g. Delivery of all 12 S2 spectral bands to enable the new vegetation indices; Data Reprocessing campaign for all pilot sites to complete the datasets with these spectral bands)
- Acquisition and processing of all available Sentinel-2 datasets for all sites
- Data quality improvements (e.g. atmospheric corrections and cloud detection)
- Provision of these datasets to the project partners via API and visualization tools

Analytics

- Development and implementation of 7 new vegetation indices and data products as requested by ABACO
- Development and implementation of new API endpoints for data delivery
- Development of a Data anomaly detection component to detect data issues (e.g. related to cloud and cloud shadows)
- Work on EO-based water stress detection
- Development & Testing of a prototype for the automatic generation of Farm Management Zones from EO data

Data Visualization and analysis tools

- Development and implementation of data visualization tools for data review and analysis that are used by KUL in the visualization components (see Github for Open Source visualization components: https://github.com/geocledian)
- Development, implementation and release of Grapevine specific colour scales for all vegetation indices.

Geocledian has developed a new customizable color scale specifically for vineyards for all vegetation indices which reflects the different dynamic range of LAI of the grapevine plants. It has been integrated in the irrigation dashboard (section 4.6). The next figure shows a comparison between the old vegetation index colour scale for NDVI (middle) and the newly developed **grapevine specific vegetation index color scale** (right).

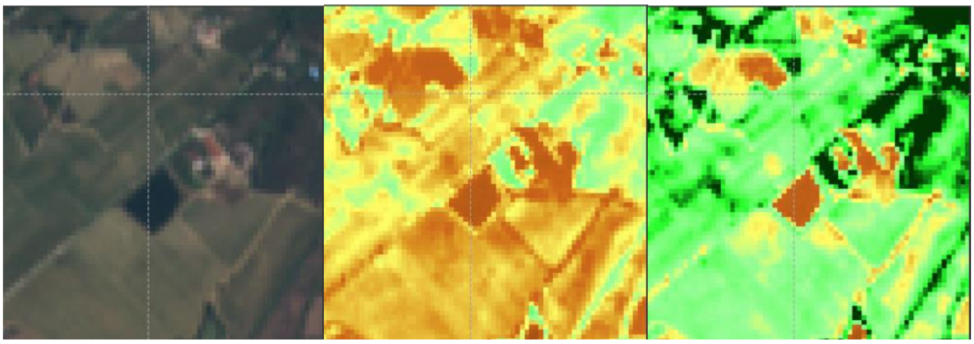


Figure 12 An S2 RGB image from Il Palazzo vineyard on 2020-06-27 (left). It can clearly be seen that on this date (peak of season) the vineyard is mostly green. The old NDVI colour scale (middle) did not reflect this intuitively because most fields are represented by brown colours (due to the low LAI levels of grapevine compared to other crops). The newly developed colour scale (right) corresponds better to a farmer's view on his plants in the field. It has been tuned to fit to the lower LAI levels of grapevine for the vineyards represented in the BDG project and can be user-defined. Water and buildings are still shown in brown as expected.

Geocledian has also developed a prototype for the automatic generation of **Farm Management Zones** from EO data. The farm management zones map provides the Field Management Zones that are needed to apply precision farming actions, more specifically Variable rate application (VRT). By combining all satellite images available during a user defined period of crop development, the zones with similar characteristics are derived. Zones with higher fertility are shown in green, followed by the medium zone colored in yellow, zones with lower fertility are displayed in red. The Management Zones Map is important for field management. It can be used as input to more advanced Zoning maps or used directly for fertilization or irrigation planning (e.g. in Siti4Farmer):

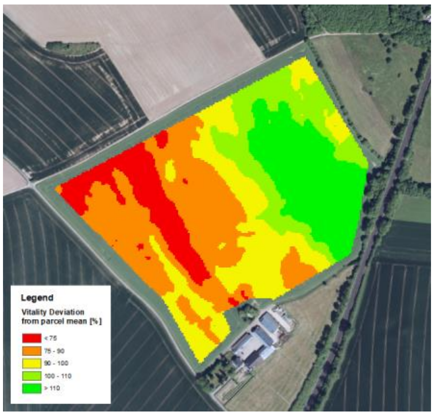


Figure 13 A Farm Management Zones map derived from Sentinel-2 NDVI images for 2019.

The management zones map can derive from images from a user defined time span or the automatically derived season duration. For this feature, phenology markers automatically derive to determine the start and end of a season (green markers in the next figure) based on the vegetation development and used to define the timespan incorporated in the zones map.

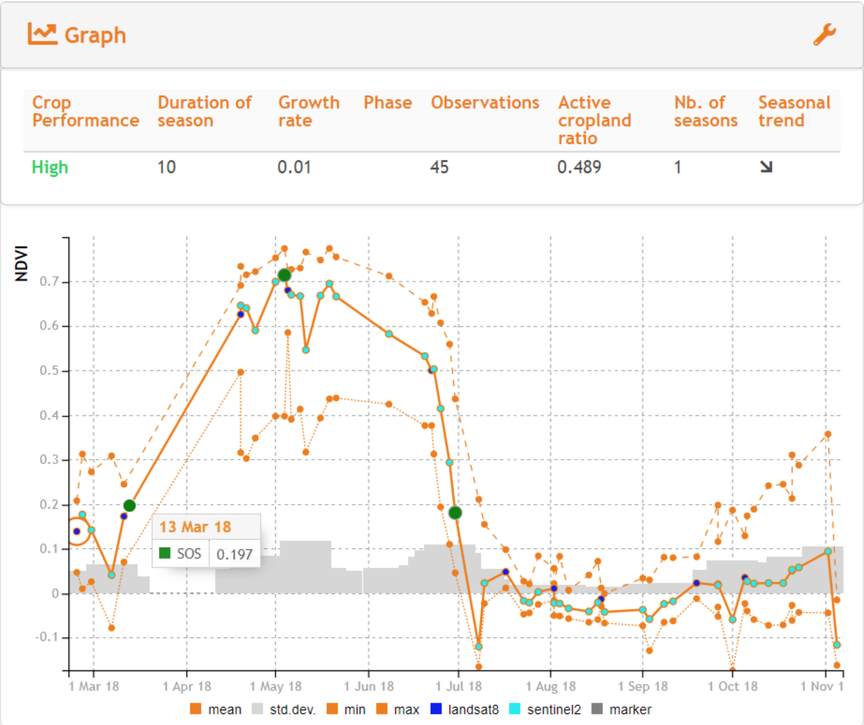


Figure 14 Phenology markers (green dots) that can be used to automatically define the start and end of a season for which management zones can be derived.

As reported in literature, NDWI is correlated with vegetation water content (Serrano et al 2019, Ihouma 2017, Moreno et al 2013, Gao 1996). To better understand the information that NDWI can deliver for grapevine fields, we investigated the relationship between **vegetation water stress** and NDWI. The following plot shows the timeseries of leaf temperature measurements (green) and air temperature (red) done at one plot of Il Palazzo vineyard in 2019 by ABACO. The purple line shows the NDWI derived from S2 imagery for this plot. It is known that the leaf temperature increases when the plant experiences water stress because the cooling effect of evapotranspiration decreases. The highest temperature days in the period June/July were reported to show significant water stress. NDWI shows a negative correlation to the leaf temperatures which is as expected and reported by other authors (e.g. Serrano et al 2019). Unfortunately, the data set was too small to go into deeper analysis at this point. But it was decided to include NDWI in the Irrigation dashboard reported in section 4.6 to be able to study this in more depth.

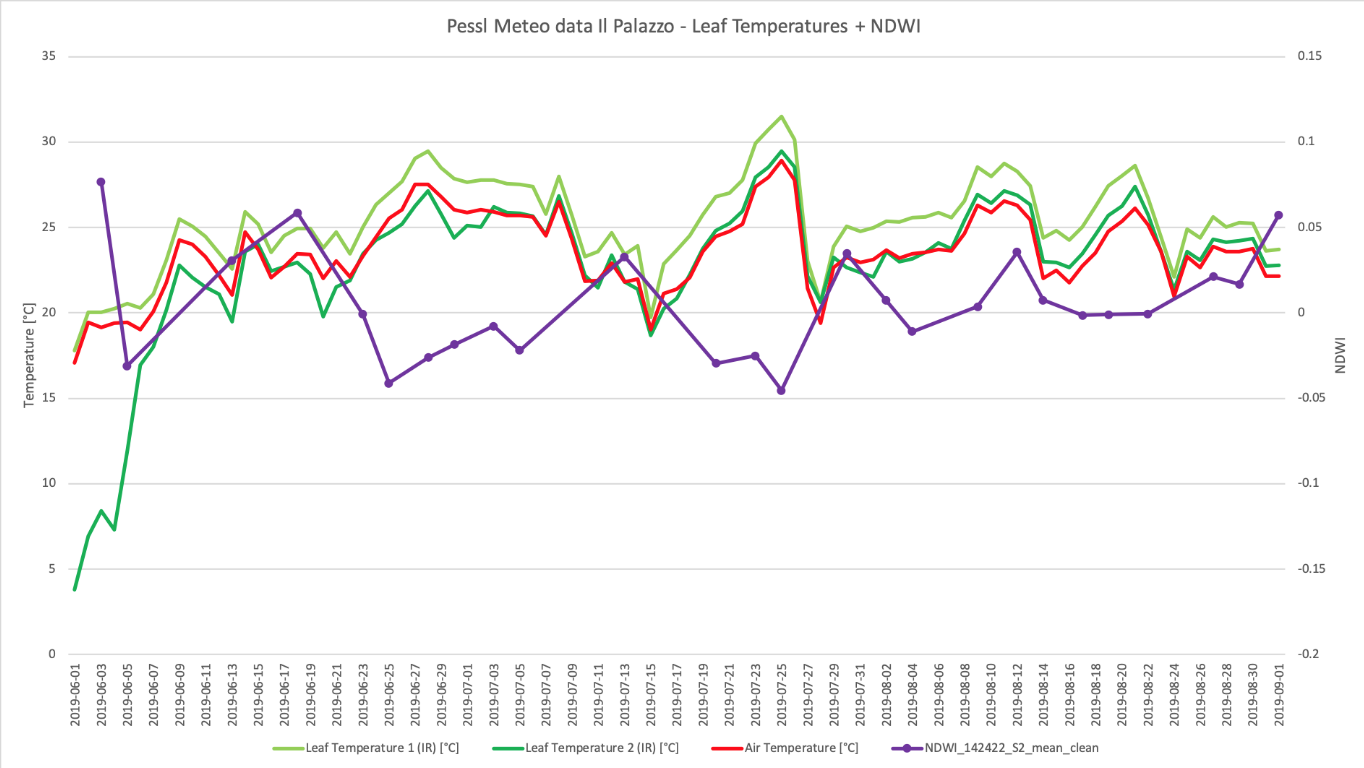


Figure 15 Timeseries of leaf temperature measurements (green) and air temperature (red) done at one plot of Il Palazzo vineyard in 2019 by ABACO. The purple line shows the NDWI derived from S2 imagery for this plot.

The following components have been deployed successfully:

- Extended data download and processing components
- New processing performance monitoring tools
- New vegetation indices and data products component (now including new Grapevine specific colour scales for all vegetation indices and a Farm Management Zones product generated from EO data)
- New API endpoints for data delivery
- New data visualization tools for data review and analysis

To tackle the data processing optimization and scaling issues, Geocledian started an intensive collaboration with CNR in the frame of the project. This activity comprised these tasks:

- performance monitoring and analysis of the system;
- identification and analysis of performance bottlenecks;
- architectural refactoring of the initial system into modular subsystems to enable scalability and processing optimizations;
- development of scalable platform components.

This activity has been reported in more detail in WP4 deliverables.

4.3 SATELLITE DATA MODULE EVOLUTION IN SITI4FARMER

In the meantime, SITI4Farmer, using services provided by Geocledian which offer Earth Observation (Open BIG DATA) capabilities, has included a dashboard for satellite indices display.

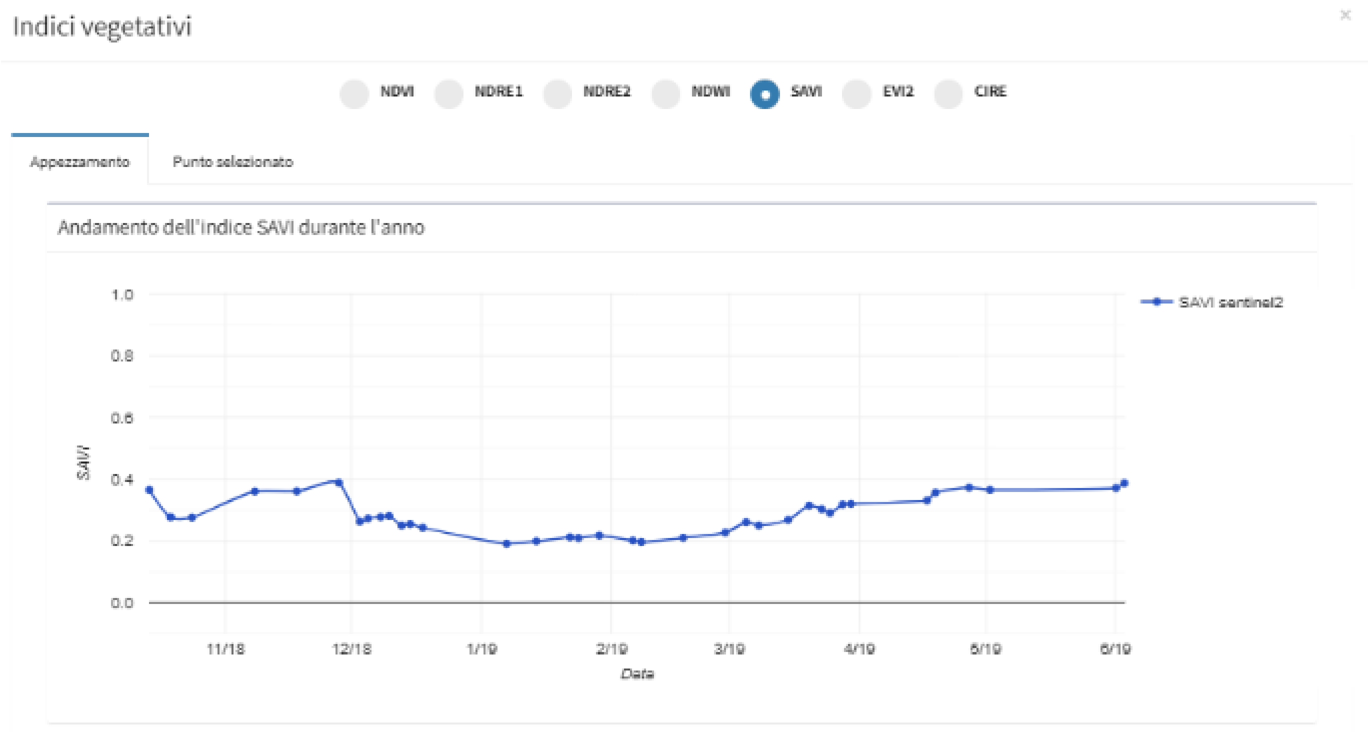
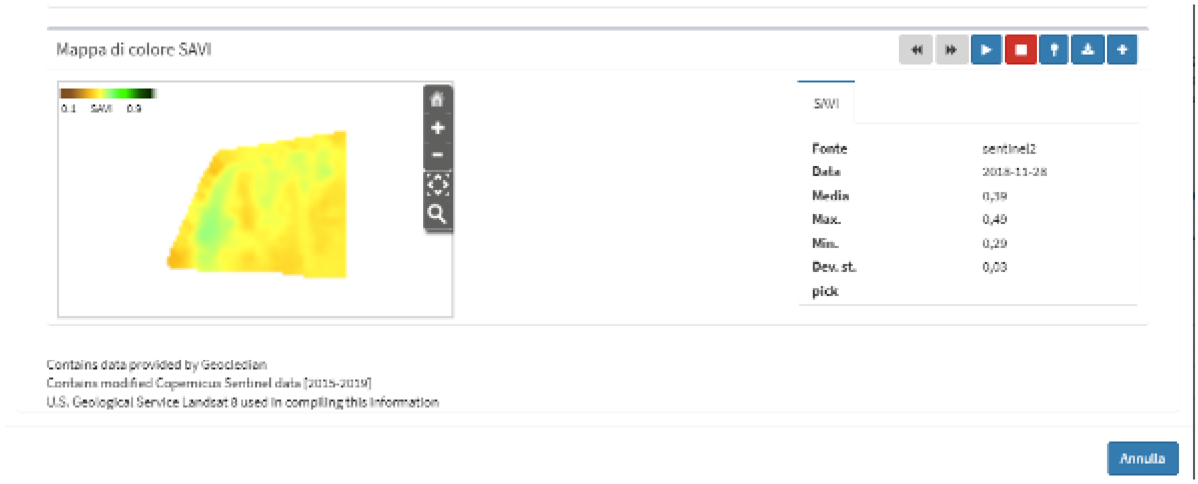


Figure 16 For a selected date in the Satellite indices time series the raster image of the parcel is displayed with statistics data



In particular the following remote sensing indices are calculated from multi-spectrum images Landsat 8 and Sentinel 2 (Copernicus project).

- Normalized Difference Vegetation Index (NDVI)
- Normalized Difference Red Edge Index (v1) (NDRE1)
- Normalized Difference Red Edge Index (v2) (NDRE2)
- Normalized Difference Water Index (NDWI)
- Soil Adjusted Vegetation Index (SAVI)
- Enhanced Vegetation Index 2 (EVI2)
- Chlorophyll Index - Red Edge (CI-RE)

Each index focuses on different kind of support on plant stress detection. In detail, the NDWI index represents a very useful support for plant water stress alert, then NDRE indices are very well correlated to Nitrogen plant content and they are essential inputs for Prescription Maps for the fertilization (see Precision Farming Module).

Moreover, the above-mentioned indices, combined with atmosphere-plant-soil data (from PESSL stations), can provide Standard Alerts on:

- Soil temperature suitable for sowing
- Insufficient quantity of ground water
- Not uniform vegetative growth
- Vegetative stage not in line with forecasts
- Approaching of critical temperatures
- Accumulation of thermal and radiative insufficient resources to technological maturity / harvest

For each type of ALERT, the user can define:

- the monitoring data sources
- the alarm thresholds (defined by default, but modifiable)
- the number of standard ALERT

4.4 PRECISION FARMING MODULE

Variable rate technology (VRT) of inputs is a key component of precision agriculture, providing economic benefits to growers in the form of reduced use of fertilizers, agrochemicals, and irrigation water, while having a positive environmental impact.

Thanks to improved satellite data, ABACO released a Precision Farming module in their GIS Navigator tool of the SITI4Farmer agriculture solution. The module enables users to manage variable rate input (of chemicals) and crop management practices taking into consideration various inputs.

The SITI4Farmer Precision Farming has the capability to rasterize, i.e. to provide a continuous data layer from scattered data, such as soil survey data all over plots and farm.

This rasterized data is one of the several inputs used to build a prescription map together with:

- Geocledian Satellite vegetative indices
- Drone raster maps
- Yield maps
- Soil surveys raster data

The preparation and adjustment of prescription maps, also known as variable rate maps, is also possible thanks to another component of SITI4Farmer called “Raster Calculator”.

Users can do the re-classification of a map, that means subdividing the output into different customizable classes, that can be managed by the variable rate machinery, such as a fertilizing trail. The export format to the variable rate machinery is currently done through a shapefile.

As regards to new implementations, the module will be improved so as to provide an even smarter and more intuitive navigation by the farmer. The command layout will allow precise and fast variable rate map editing and the visualized output will be also available in PDF format, to enable also a manual use of variable rate maps.

Big Data enables ABACO to further study and execute Data Modelling tasks. The data coming from Satellite datasets are valuable inputs for new data models that can run in the R or Python languages on the Raster Calculator platform inside SITI4Farmer.

Data modelling, in the framework of Decision Support System (DSS), aims to calculate more precise, accurate, and differentiated doses of fertilizer, pesticides, water and other inputs by means of variable rate maps.
The variable rate maps can now be exported directly to the most popular machinery cloud systems as My JohnDeere, in order to be easily imported into the machinery for variable rate activities. This functionality can be addressed into the Agriculture 4.0 rules.

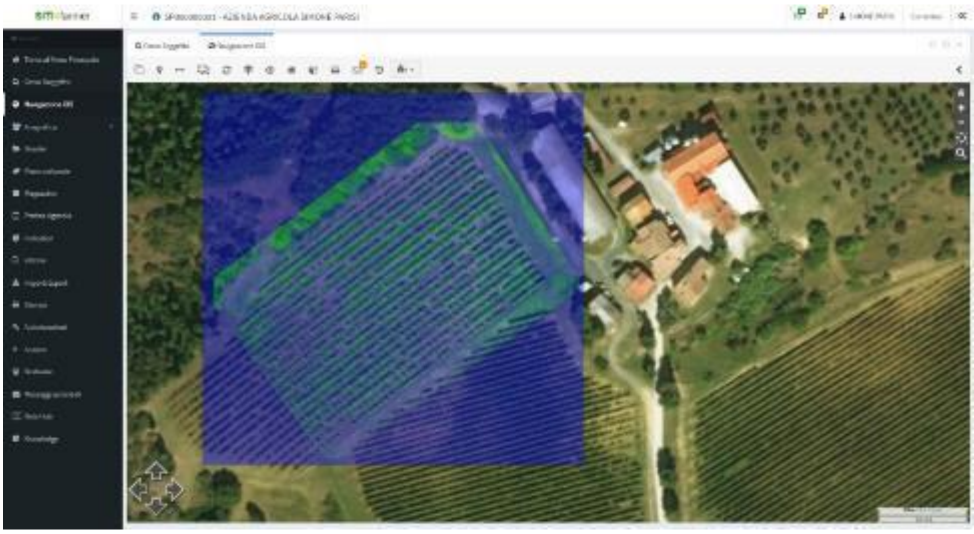


Figure 17 Import of Drone (UAV) raster data

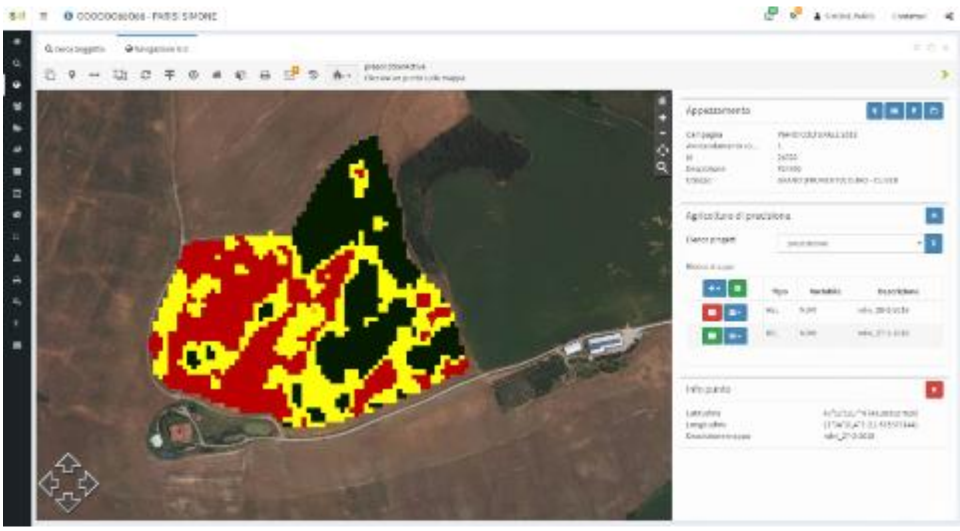


Figure 18 Reclassified Map from NDRE2 satellite data

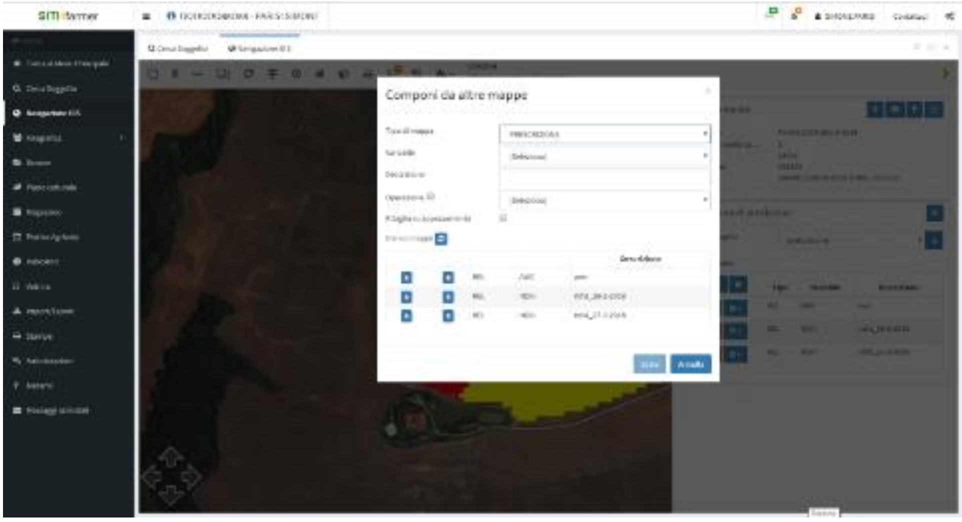


Figure 19 Raster calculator interface. Allows model run on different raster input data

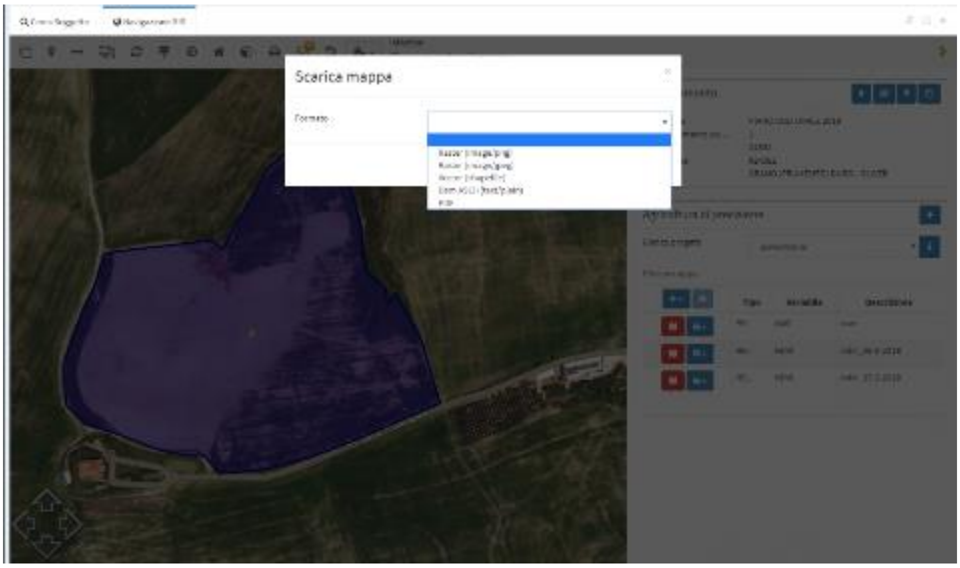


Figure 20 Map export on vectorial or raster formats

Figure 21 Import of a prescription map on a Trimble VR computer

4.5 SENSOR DATA CONNECTIVITY – THE SENSOR PROXY

Weather station data recorded in piloting farms are connected to the SITI4Farmer portal thanks to the development of The Sensor Proxy procession, enabling connectivity by API-REST protocol, where the data is disseminated by means of SITI4Farmer dashboards.
PESSL Instruments' sensors are connected to the SITI4Farmer portal by API-Rest in order to display data in a dedicated dashboard widget and to use such data for the DSS models.

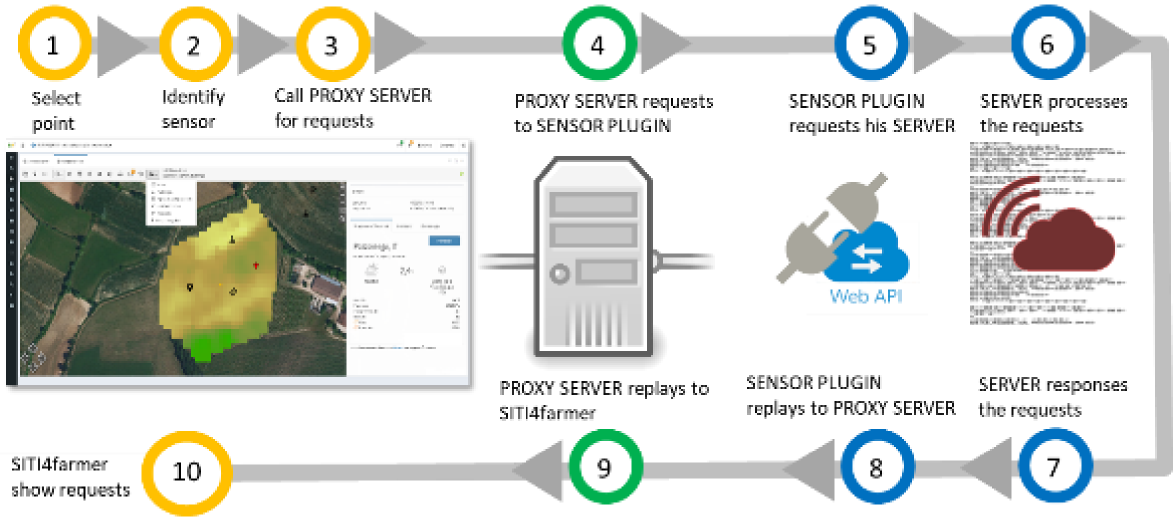


Figure 22 Sensor Proxy data fluxes scheme

4.6 THE IRRIGATION DASHBOARD

As for any kind of agricultural production establishment, a wine farm is subject to critical business decisions. Among the ones agronomists and managers face, the most important is related to increase of productivity versus fruit (grape) quality. From this standpoint, any reduction in resources used to reach such business outcome is paramount.
Firstly, water requirements must be addressed in the direction of emergency irrigation instance reduction. To reach this objective, it is important to detect the moment in which stress level is critical for the plant and which part of the vineyard really needs irrigation.
In order to help the agronomist and in general the vineyard management operative, a dashboard has been defined through which for any selected parcel (vineyard) a user has complete view on the historical and forecast data usable for the irrigation DSS designed by KU Leuven.
As for any model and Information system, it is fundamental to verify the following conditions, which were successfully confirmed in our two pilot farms:
How much of the data available are representative of the real condition of the environment in which the vineyard is?
What is the relation between water stress level alerts and the real status of the vineyard?
Which part of the vineyard is mostly homogeneously consistent with water stress level?
In detail, the historical data are inclusive of satellite indices and ground sensor data. For both sources a specific subclass may be selected. In our example a specific type of index (NDVI, NDWI, NDRE, CI-RE, EVI) is selectable among the satellite ones available.
The same has been done for sensor data. Selection is available among the various parameters gauged by the weather/soil stations. Historical data time range is customizable by users.
As regards to satellite indices specifically, a given image of dates can be displayed so as to highlight any area affected by water stress.

Real-time water stress level alerts are computed through a reservoir-based water balance model – provided by ABACO - and outcomes derive from soil moisture sensor data and satellite indices analysis – provided by Geocledian.
Depending on specific properties, soil has a capacity of storing water from precipitation and irrigation. There are also reducing factors such as evapotranspiration and percolation. Specifically, evapotranspiration can depend on vine variety. These variations are stored in specific Db tables.
A main finding of the user interviews on the dashboard has been that the EO-based vegetation index images should be improved because they are mostly brown throughout large parts of the season and therefore cannot be interpreted by the farmer intuitively. This is because grapevine has a lower LAI than other crops so that standard colour scales can lead to misinterpretation. To overcome this, Geocledian has developed, tested and implemented a new customizable color scale specifically for vineyards for all vegetation indices which reflects the different dynamic range of LAI of grapevine. It has been integrated in the dashboard.
CNR developed a machine-learning component that performs water balance prediction using meteorological data from weather stations and soil data. The analysis has been performed on two vineyards under study by the BigDataGrapes project. The two vineyards are the "Casato Prime Donne" and "Il Palazzo" estates. On each vineyard, two fields have been equipped with two weather and soil stations that record several parameters at different levels of granularity. Data from the weather stations are collected by ABACO.
CNR applied a set of machine-learning techniques in order to solve the following regression task: given an observation, how to predict its associated water level. The problem has been modelled by employing:

- linear models
- regularized linear models, i.e., ridge, bayesian ridge and lasso
- gradient-boosted regression trees
- random forests
- voting methods
- SVM regressor
- neural networks

For all the methods, CNR employed a standard k-fold cross validation methodology (with k = 5). The final performance reported is the average of the five performances on the test sets. The validation set is used for performing early stopping of the training of the models in order to avoid overfitting the data. The best performance is achieved using a Random Forest regressor.

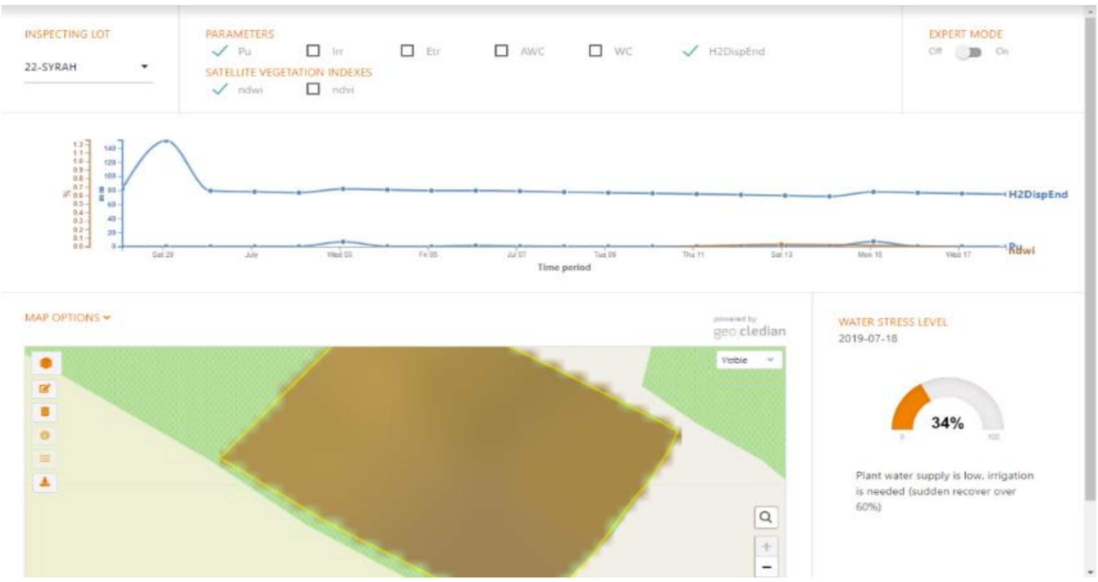


Figure 23 Irrigation dashboard screenshot. Water stress level percentage and irrigation suggestions on the right side are shown, together with satellite index data, water balance and evapotranspiration data. Expert mode enable further data display

4.7 THE NEW IRRIGATION DASHBOARD IN ABACO FARMER

In 2020, SITI4Farmer has undergone a deep revolution on the user experience and in particular on the end-to-end logics. First of all, the name of the agriculture platform changed into ABACO Farmer and all the modules have a new architecture.

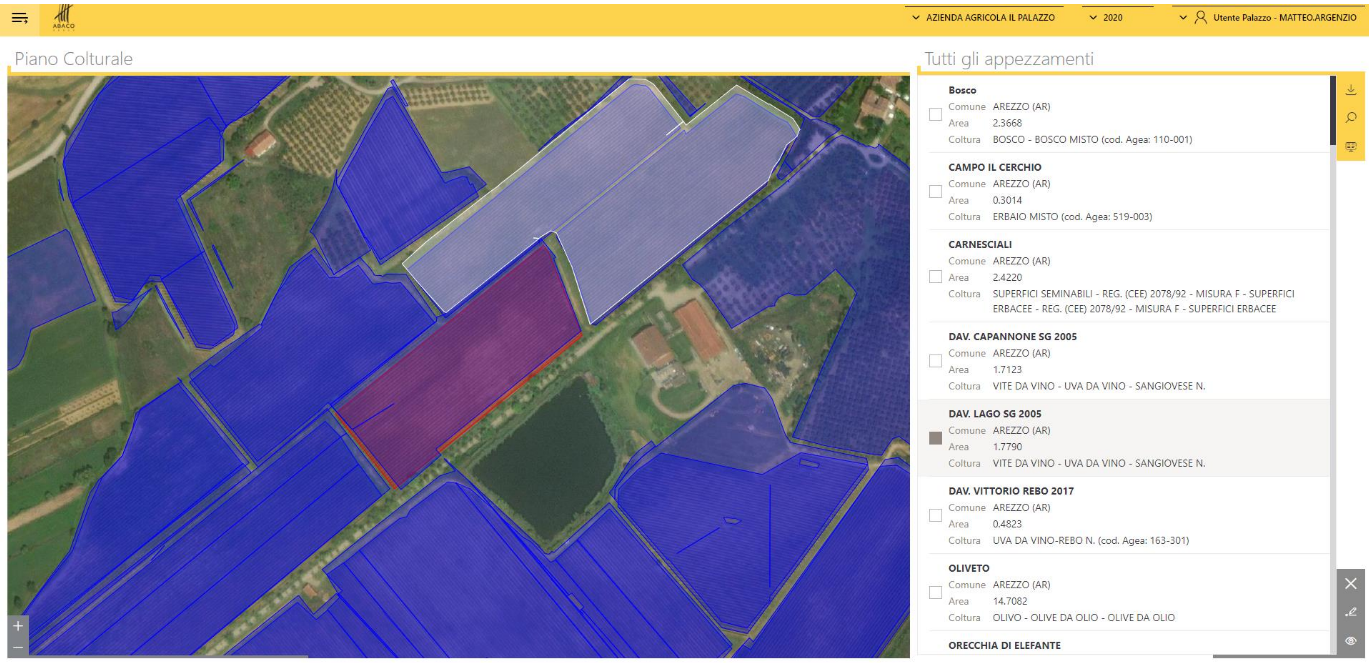


Figure 24 Crop plan module displaying in example "IL PALAZZO" farm

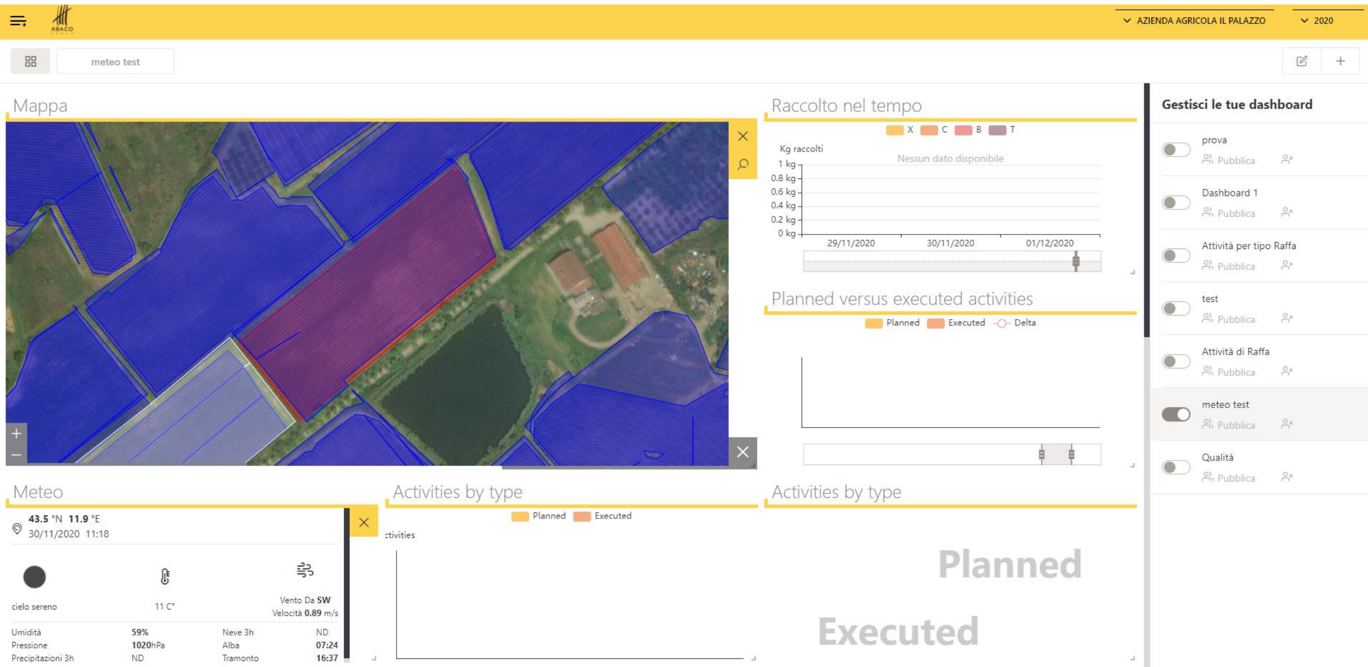


Figure 25 The Dashboard module example, with dashboard editing option

Thanks to the new "dress" of the platform, the integration of the irrigation dashboard in ABACO Farmer must pass through new tiles and a new user experience too for this section. It can be called, a complete integration on a farm management system.

In the Dashboard module, the user can specify on the base of a complete set of pre-built plots, his own control panel, by which the user can decide the feature that he/she wants to have under control. For example: the costs of activities, their durations, weather station real time data and of course the Water stress (Irrigation) DSS control panel.

Specifically, the new Irrigation Dashboard allows to have in a glance the information on the timeseries that determine the balance of the soil water reservoir, the loss term (Evapotranspiration) and the gain terms (Rainfall and Irrigation). Moreover, a GIS-based interface and an interactive plot help to choose the satellite index image in order to check spatial variability of water stress in the field.

4.7.1 TILE 1: EVAPOTRANSPIRATION, PRECIPITATION, IRRIGATION

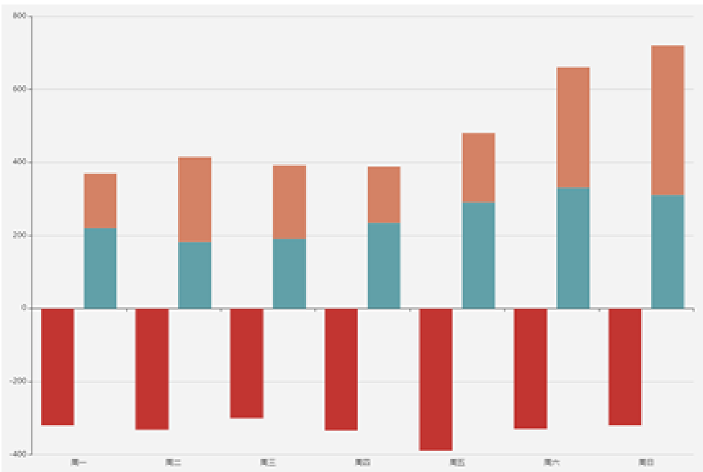


Figure 26 The irrigation, rainfall and evapotranspiration timeseries based on daily, monthly and yearly time span.

4.7.2 TILE 2: WATER BALANCE

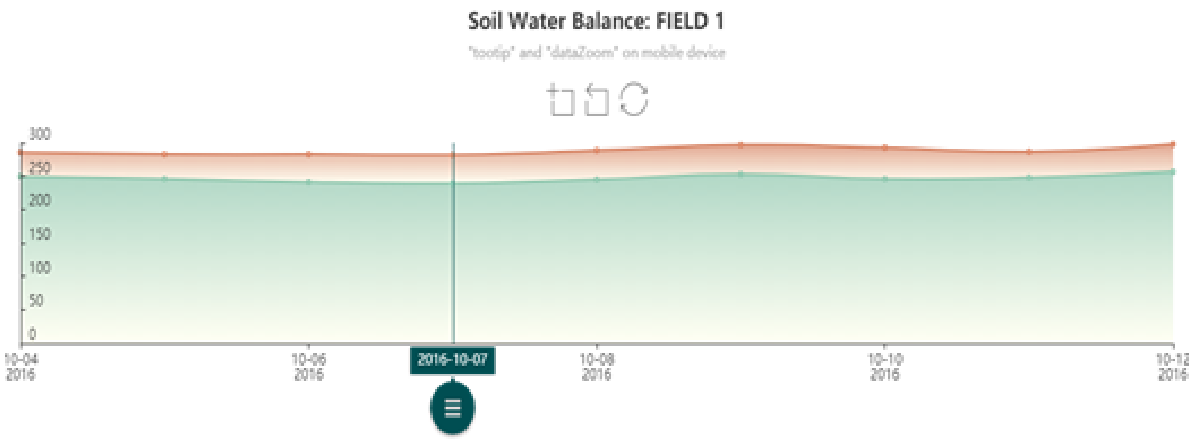


Figure 27 Trend of the soil water reserve status in a selected period

4.7.3 TILE 3: SATELLITE INDICES IMAGERY

Figure 28 Satelite indices imagery

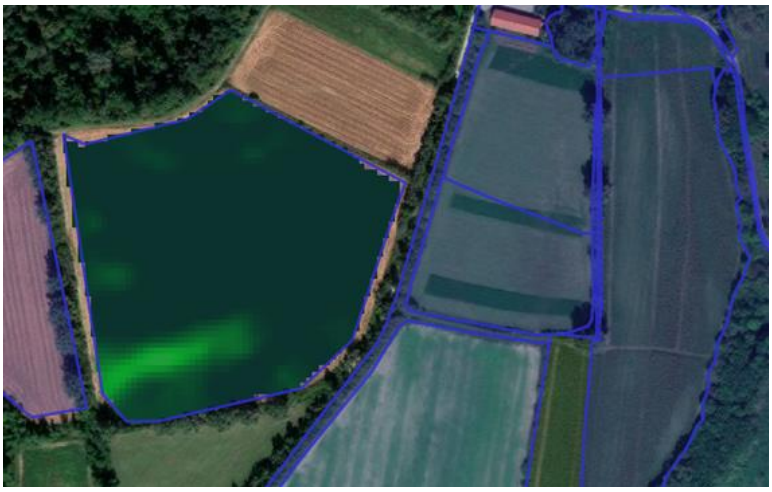


Figure 29 The tile allows the user to select the satellite shot from all available ones shown on thumbnails, and the related plot showing the average values and its standard deviations. The selected shot is then shown on the gis module projected on the base map.

4.7.4 TILE 4: SYNTHETIC SOIL WATER AVAILABILITY PERCENTAGE

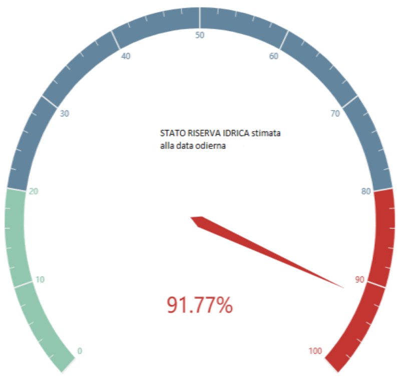


Figure 30 The gauge meter allows the user to know the updated status of the soil water reservoir in percentage and suggestion on irrigation activities. the update refers to the last data received when the dashboard is opened.

4.8 PRESCRIPTION MAPS DASHBOARD

The dashboard allows the visualization of the satellite indices images available on the selected plot, and after selecting the requested satellite data, the system calculates automatically the reclassed map.

The reclassification is based on percentiles computation. The percentiles threshold depends on the number of classes the user decides to obtain. The calculated map describes the class level of NDVI vigour, if the map has been obtained using NDVI data, or describes water stress levels, in case the map has been retrieved using NDWI data.

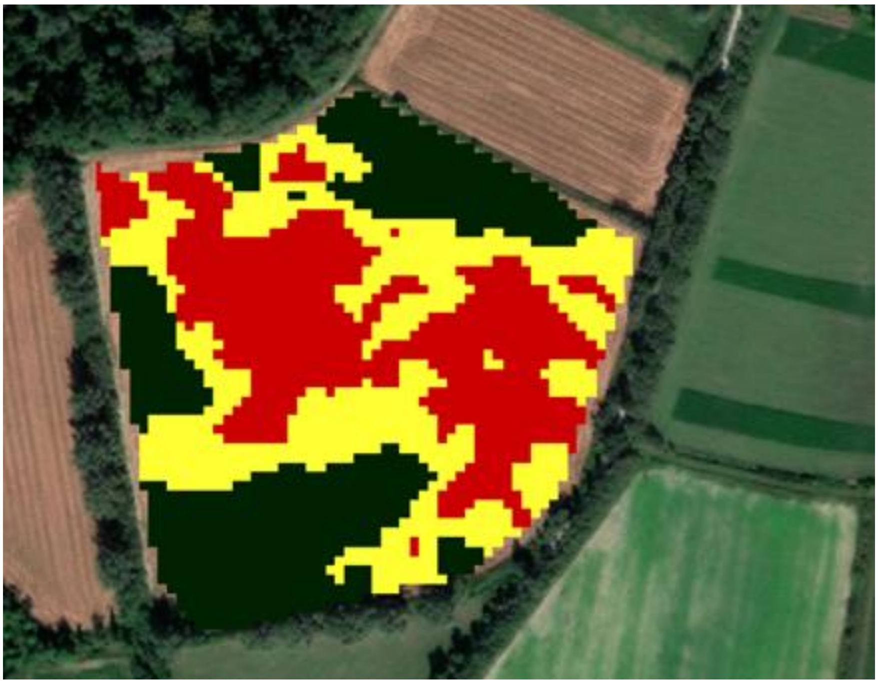


Figure 31 Water stress levels, in case the map has been retrieved using NDWI data

5. FOODAKAI EVOLUTION

This section focuses on the incorporation of the BigDataGrapes software stack in FOODAKAI, a Risk Scanning and Monitoring system, augmenting it with the functionalities required for the implementation of the Food Protection Use Case.

FOODAKAI is an online SaaS data analytics platform that serves food companies globally in the context of risk monitoring, assessment and prediction. Under the BDG project, efforts were focused on deploying and testing FOODAKAI in the context of integrating datasets and services from pilots and extending FOODAKAI capabilities through the BDG platform. Moreover, the Agroknow team has performed a series of focused group and consultation meetings with several companies of the food industry, such as Gallo Winery, Conagra, Campbell, Pepsico, Coca Cola, Hershey and Lamb Weston. The meetings were held during large food safety events like the GMA Science Forum and the GFSI conference. During these meetings, the FOODAKAI team validated the need for new FOODAKAI extensions powered by the Big Data Platform that will enable risk predictions in the supply chain. The main objective of the extended pilots is to enhance the current digital solution with new modules that will address further needs of the grape and wine supply chain. The enhancement will mainly focus on the further development of Agroknow's Big Data platform with new software modules that will enable advanced data analysis and risk prediction using machine learning and deep learning methods.
FOODAKAI is the food safety intelligence platform and provides access to more than 100.000 food safety incidents that are announced by official sources. The food incidents go back to 1980 and cover more than 170 countries covering all the hazard types and all the product categories.

The user of FOODAKAI can have a personalized dashboard that will focus only on the hazards, products and origins of interest, while still be able to access all the data. Additionally, FOODAKAI is able to send alarms for the food recalls and border rejections that are relevant to the products and ingredients of the user.

Furthermore, FOODAKAI allows the generation of statistics that the user needs to identify the risks in the supply chain to compare products, ingredients, and origins, generate the statistics and save the results. All diagrams are interactive, dynamic with access to the original announcement. One of the features of FOODAKAI is to provide a live risk assessment for all products of interest and to identify which are the products of high risk based on the recent trends.

In the following sections the integration of the Big Data Grapes software stack in FOODAKAI is analysed and the new operations enabled are presented.

INCORPORATION OF BIGDATAGRAPES PLATFORM IN THE FOOD PROTECTION PILOT

In the following table we provide an overview of operations of the FOODAKAI software system which were enhanced to support the Food Protection Pilot.

Table 2 Overview of operations of the FOODAKAI software system

Operation (Module)	How FOODAKAI system was enhanced	Which part of the Big Data Platform was used
Dashboard and risk monitoring	Enhanced from 27 data sources in 2018 to more than 50 data sources in 2020 for monitoring food safety incidents announced globally.	Scanning every few minutes more than 50 data sources. The Web Scraping component of the Big Data Platform was used to collect the data.

Integration of third party data	FOODAKAI was augmented by the integration of third party systems that the food industry is already using.	The Big Data Platform data layer was used to integrate food safety data from other systems.
Hazards analysis	A new module was developed to allow the analysis of chemicals in ingredients and raw materials like grapes and raisins 102 Million of records are now available in FOODAKAI, making it the only system on the market that provides access to such a large dataset.	Collection of lab data from 34 countries for years 20. Processing of 102M of data points by the Data processing and Knowledge classification component of the Big Data Platform
Reports	Development of the Reporting module that can be used to support content marketing activities	The processing and indexing component of the Big Data Grapes components were used
Food Fraud & Adulteration	Integration of data sources for Food Fraud and adulteration identified in raw materials and ingredients like grapes and raisins	The Web scraping component of the data layer was used to collect food fraud and adulteration data from sources. The Knowledge Classification component was used to enrich the data with fraud and adulteration terms. A special taxonomy for fraud was developed in the context of the project and was integrated in the Knowledge classification component of the Big Data Platform.
Risk assessment	Enhancement of risk assessment module by the integration of risk estimation for countries and incidents.	The trend analytics component of the Intelligence layer of the Big Data Platform was developed and used
Risk prediction	Price and Recall prediction models were developed and integrated in prediction dashboards of the FOODAKAI.	The prediction component of the Big Data Platform was used to deploy prediction models The experimental panel of the Intelligence layer was used to train and parameterize the price and recalls prediction models.

In the next section we are presenting more details about the integration and how it upgraded the services that the FOODAKAI system provides.

5.1 DATA COVERAGE UPGRADE

In the context of the BDG project, the FOODAKAI platform was significantly enhanced in terms of data coverage. More specifically, the volume of the data that is currently available through the platform was increased from 200.000 records in 2017 to 246.063.102 in 2020 with a 5% data growth rate every month. The Geographical coverage of the data for the food protection pilot is presented in the following figure.

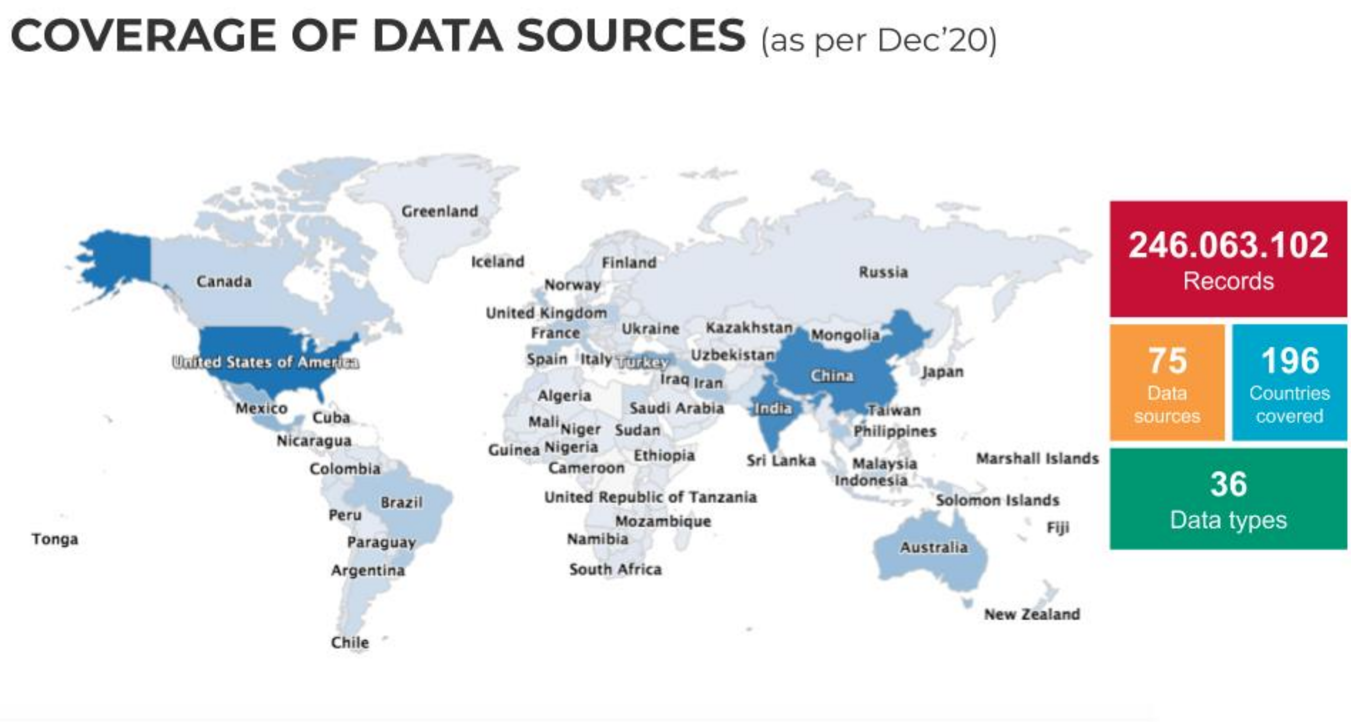


Figure 32 Data coverage expansion

The Big Data technologies enabled the processing and interconnection of several different data types that are valuable for the food protection use case (figure 33). In addition to the food safety incidents, in the context of the BDG project, Agroknow made available to its customers new data types such as fraud cases, news, lab tests, prices, and inspections. Sensor readings from raisins storage facilities were also processed.

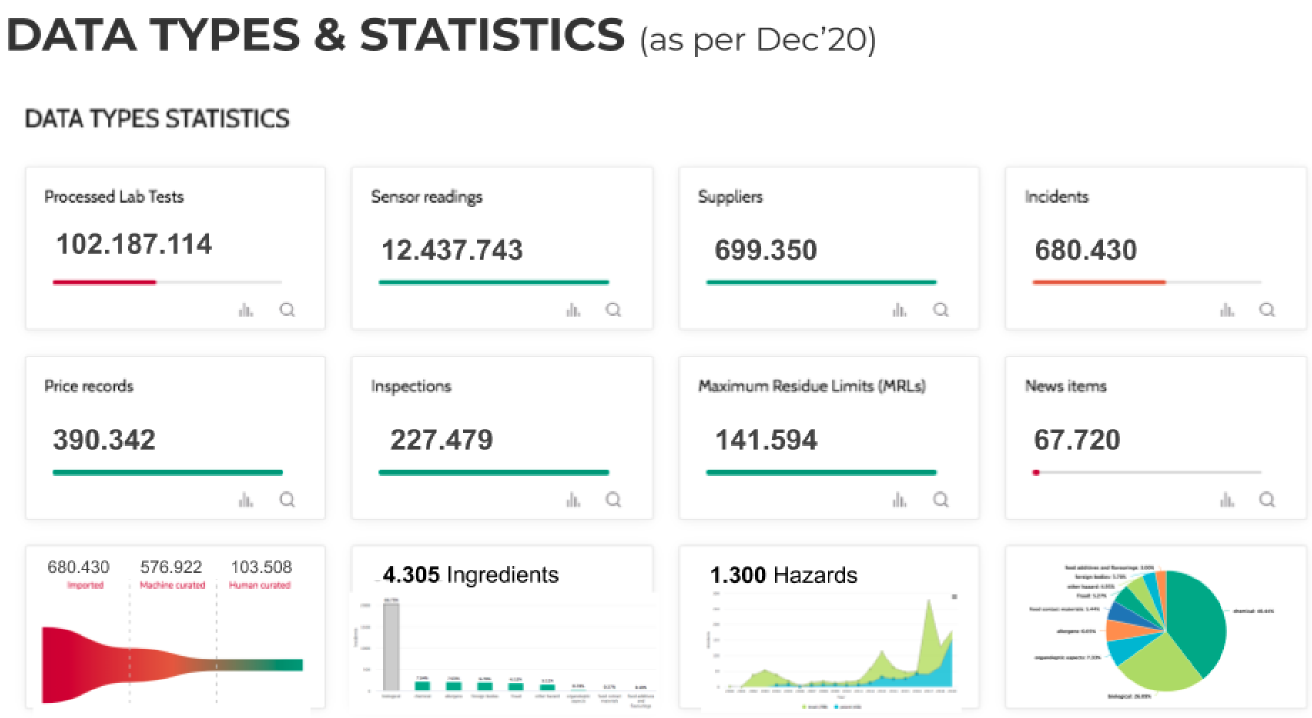


Figure 33 Data types expansion

5.2 TAILOR-MADE DASHBOARD

Using the Big Data Platform capabilities for the integration of new data types and more data, Agroknow developed a tailor-made dashboard that can be customized to the supply chain of a food company. More specifically, the end use can provide information about the ingredients, products and suppliers that he wants to monitor and the dashboard will be adjusted to his preferences showing only the data about recalls, border rejections, fraud cases, regulations and news that are relevant to the supply chain of his company. An early warning module was developed to send alerts every time that there are incidents linked to the ingredients, products and suppliers of the food company.

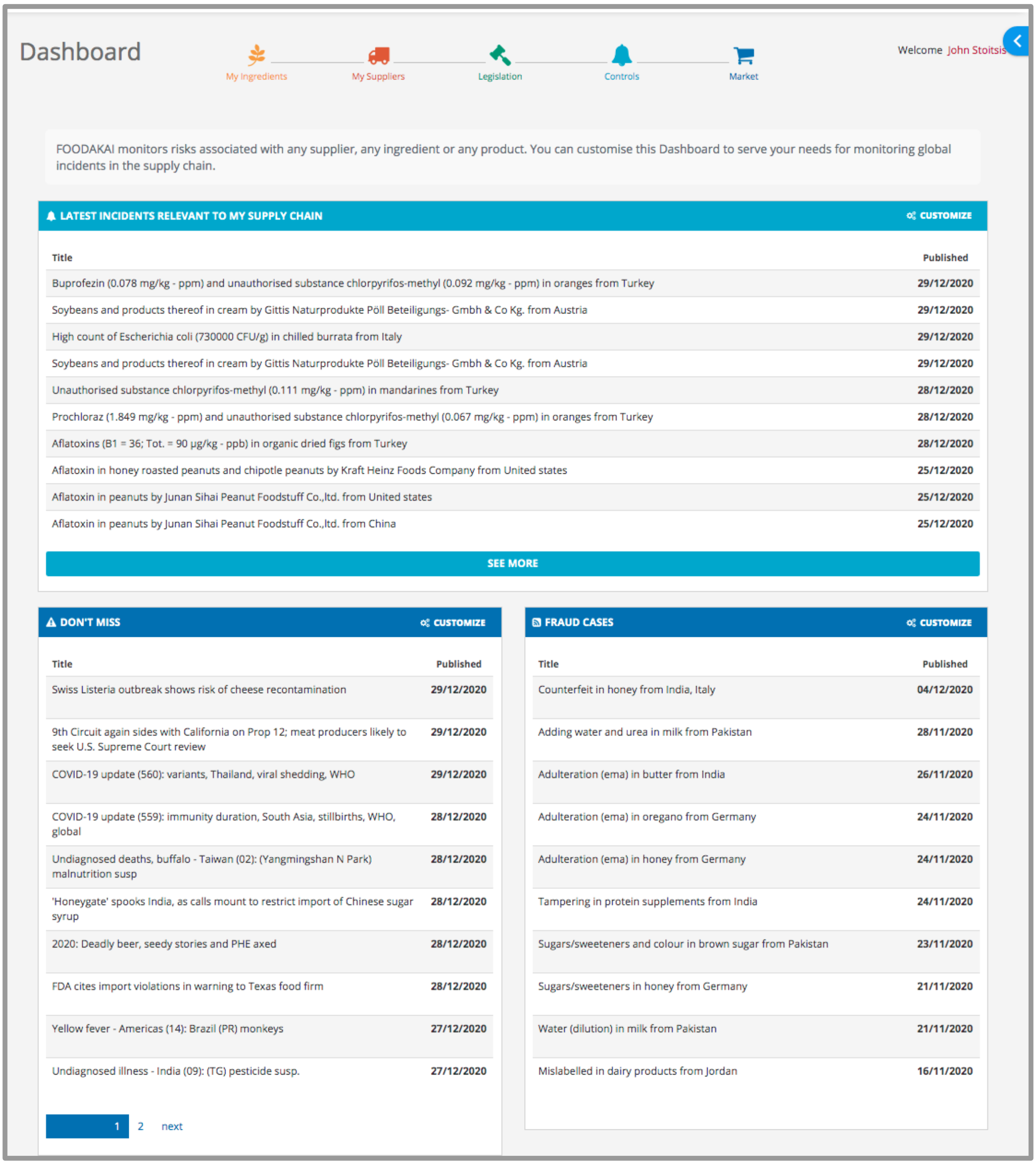


Figure 34 FOODAKAI tailor-made dashboard

5.3 KNOWLEDGE CLASSIFICATION POWERED HAZARD ANALYSIS

All the interconnected food safety and fraud data can be used in the platform to perform a deep and fast hazards analysis for thousands of ingredients, including grapes, raisins and wines. The hazards analysis is based on a hazards hierarchical taxonomy that includes over 1.400 terms and a products hierarchical taxonomy that includes more than 4.300 terms. All the collected incidents are enriched with hazards and products terms using text classification algorithms.

In the following figure we are presenting the hazards taxonomy that was developed to classify the hazards that are found in the grape products.

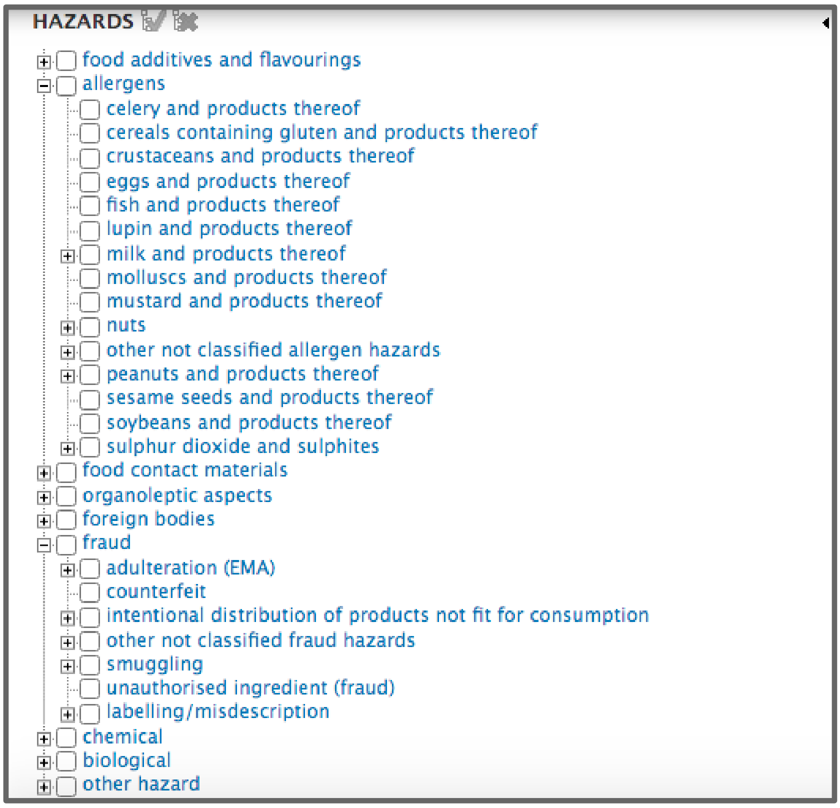


Figure 35 Part of the hazards hierarchical taxonomy that was developed by the food safety experts for the classification of food safety and fraud incidents.

In the following diagram the part of the product taxonomy that was developed for the grape and wine products is presented.

Figure 36 Parts of the hierarchical taxonomy that was developed by food safety experts for the classification of the food safety and fraud incidents.

An example of hazards analysis for a product like wine is presented in the following figure.

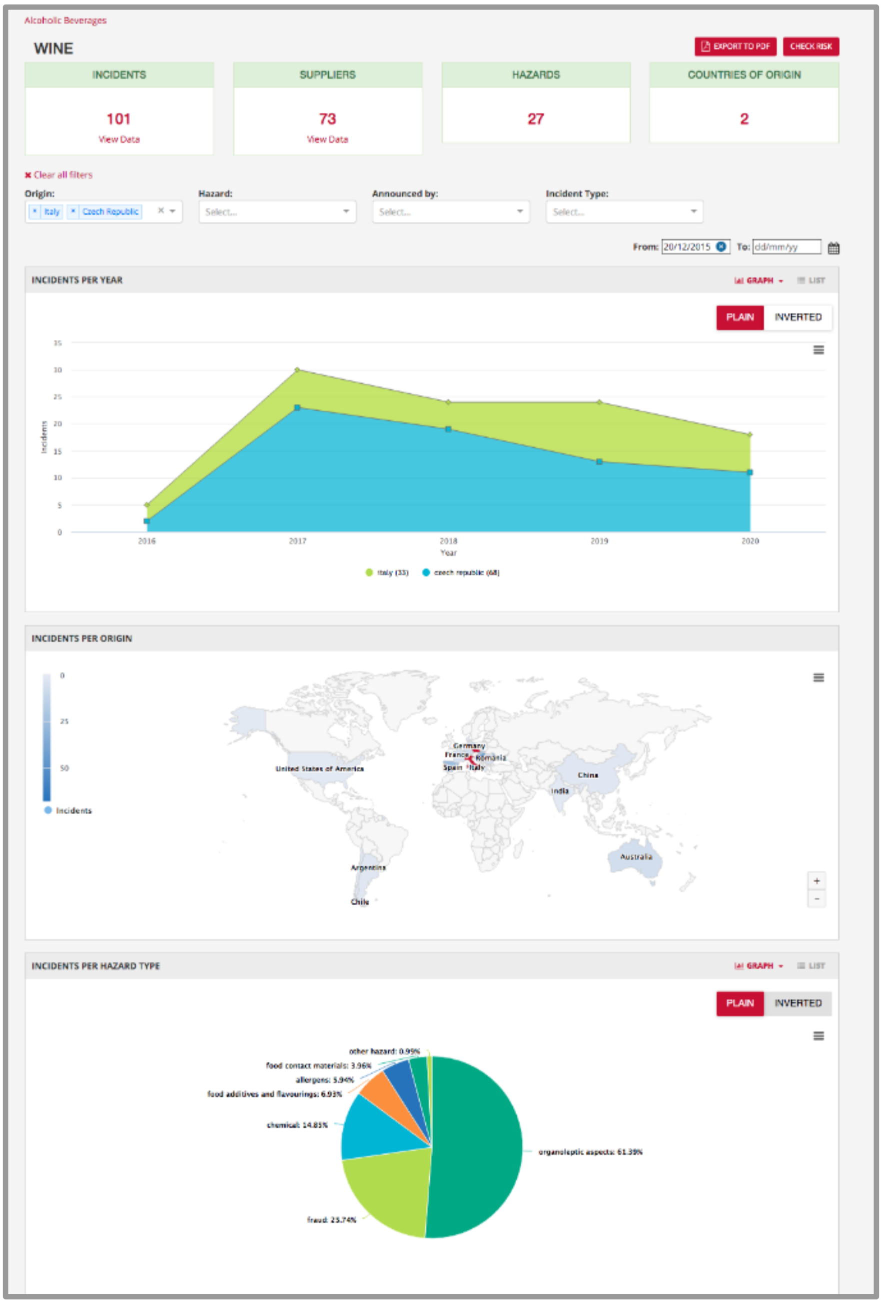


Figure 37 Example of hazards analysis

5.4 DATA-POWERED HAZARD REPORTS

The data collection and processing operations of the Big Data Platform were used by Agroknow to develop a module that allows the creation of data-powered reports for hazards in ingredients, raw materials and suppliers. The end user can create groups of ingredients and map his ingredients to terms of products taxonomy. Such hazards reports are used by the food companies to share regularly important insights with their management and their teams. The reports are also used by Agroknow's marketing team to create content that attracts the interest of the food safety experts. An example of such hazards report for wines from Italy and France is presented in the following figure.

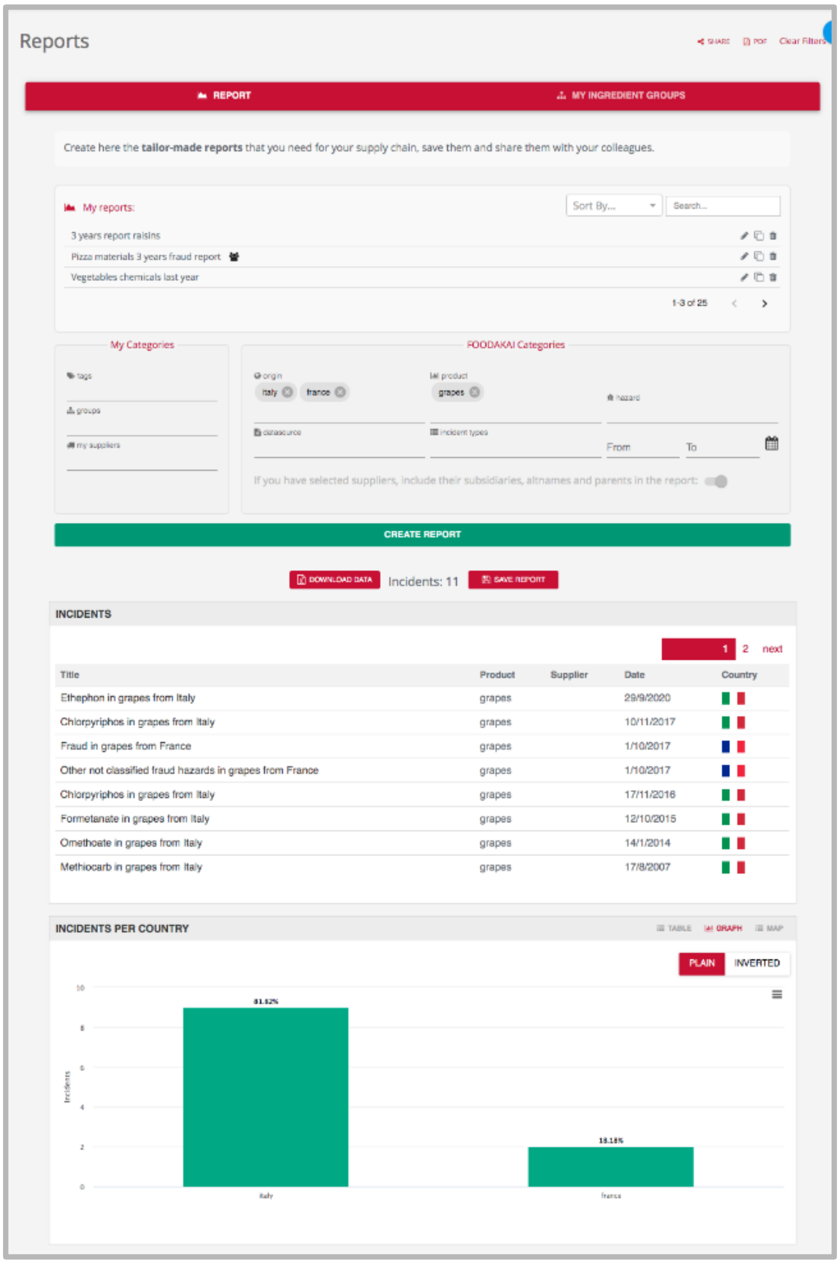


Figure 38 Example of a data powered hazards report for wine products

5.5 LAB TESTING DATA ANALYTICS

In the context of the food protection pilot, we collected and processed more than 100 million laboratory testing results that are published by 33 countries in Europe and the United States. Using the knowledge classification component, we classified these results to corresponding hazards and products in order to deliver useful trends and analytics to the food safety experts that are using FOODAKAI. The laboratory testing analytics are used to identify chemical residues found in thousands of ingredients and raw materials such as grapes. A laboratory testing analytics dashboard was developed by the Agroknow team using the Big Data Grapes software stack for data collection, processing and visualization. The laboratory testing dashboard is presented in the following diagram.

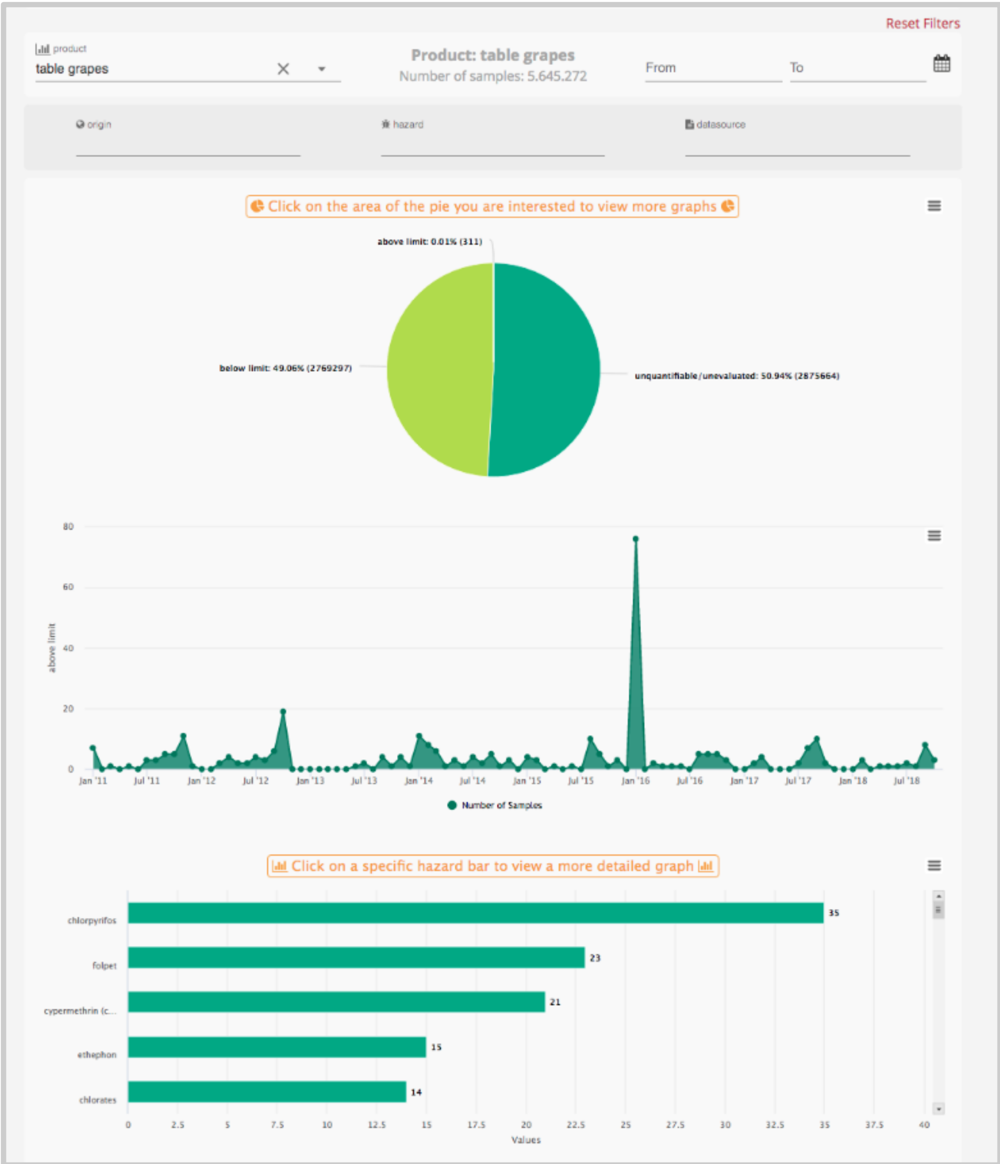


Figure 39 Lab testing results dashboard.

5.6 AUTOMATED FOOD FRAUD REPORTS

One of the use cases in the Food protection pilot focused on integrating a content-based marketing mechanism to FOODAKAI, in order to attract more food safety experts. To that direction Agroknow developed a mechanism that gets insights from the system and sends it automatically to end users in the form of a monthly food fraud report (figure 40).

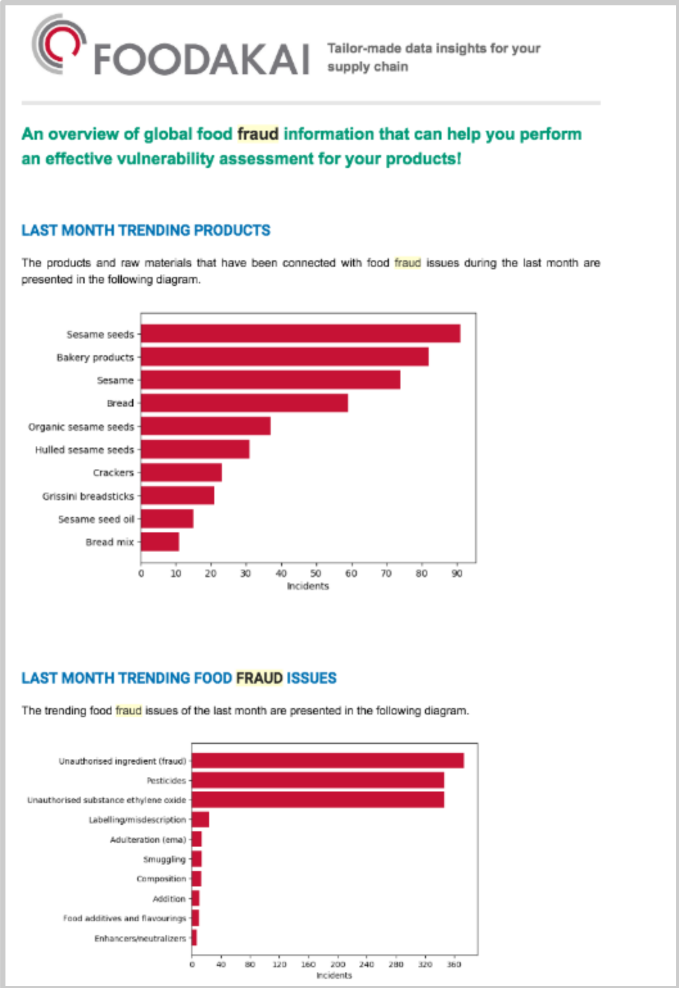


Figure 40 Example of a food fraud report sent to the end-users of FOODAKAI

The automated food fraud reports are generated monthly using the food fraud news items collected from the Knowledge Centre for Food Fraud and Quality (https://knowledge4policy.ec.europa.eu/food-fraud-quality_en) and several professional websites.

5.7 RISK ESTIMATION MODULE

Using the risk estimation models of Big Data Platform's intelligence layer, Agroknow developed a risk estimation module that is based on the frequency and the severity of the incidents. Agroknow developed a Risk API in the intelligence layer of the Big Data Platform that was used by the FOODAKAI front end. The risk estimation module can be used to identify new and increasing risks for any ingredient, finished product and supplier as presented in the following diagram. The end-user can create his finished products and suppliers by adding the ingredients used for the production of the product and ingredients managed by the suppliers, respectively.

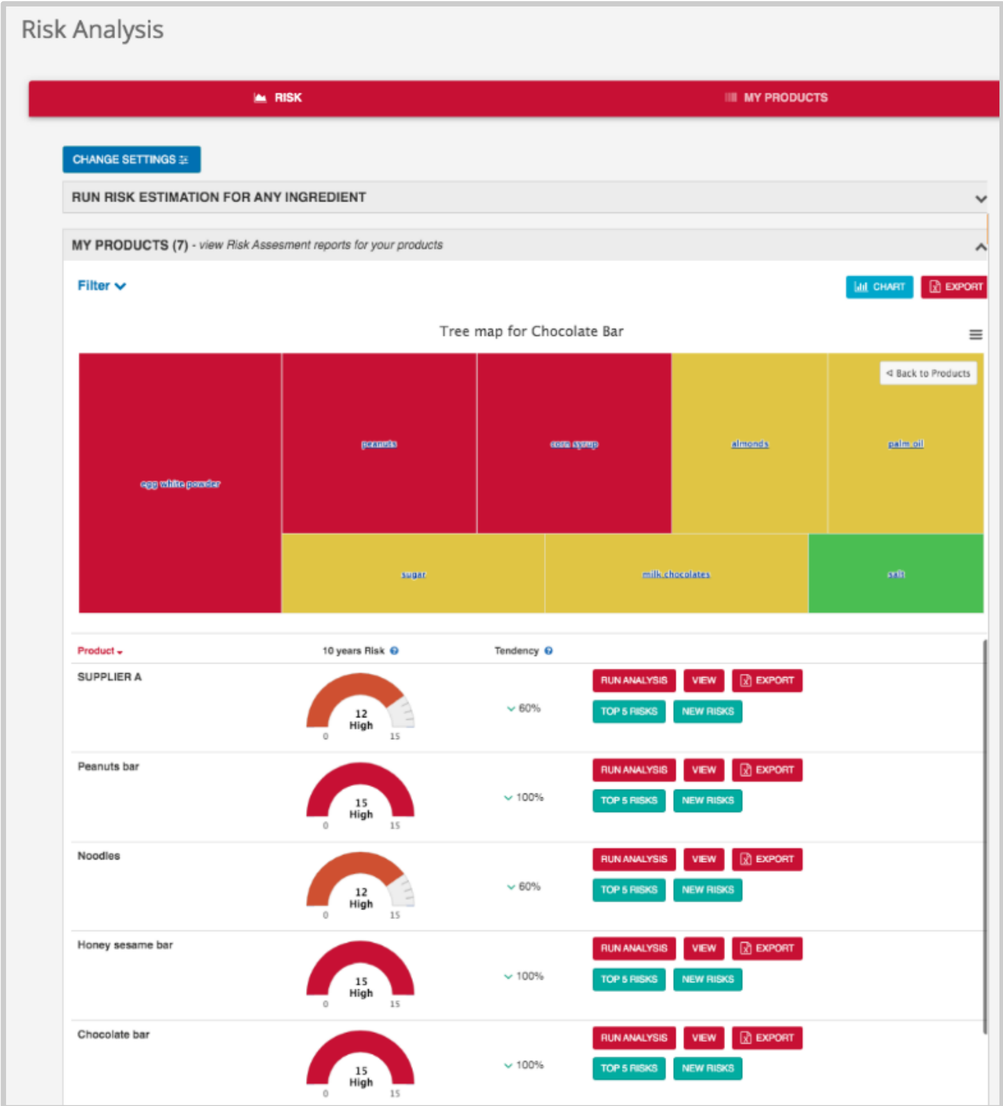


Figure 41 Risk estimation dashboard for ingredients, finished products and suppliers

5.8 GLOBAL RISK PREDICTION DASHBOARD

Agroknow used the prediction models of the Big Data Platforms's Intelligence layer to develop a global predictions dashboard.

The main goal of the risk prediction is to enable the prevention of food safety incidents for an industry like the grapevine. The main steps of how you can enable the prevention of incidents using predictive analytics are presented in the following diagram.

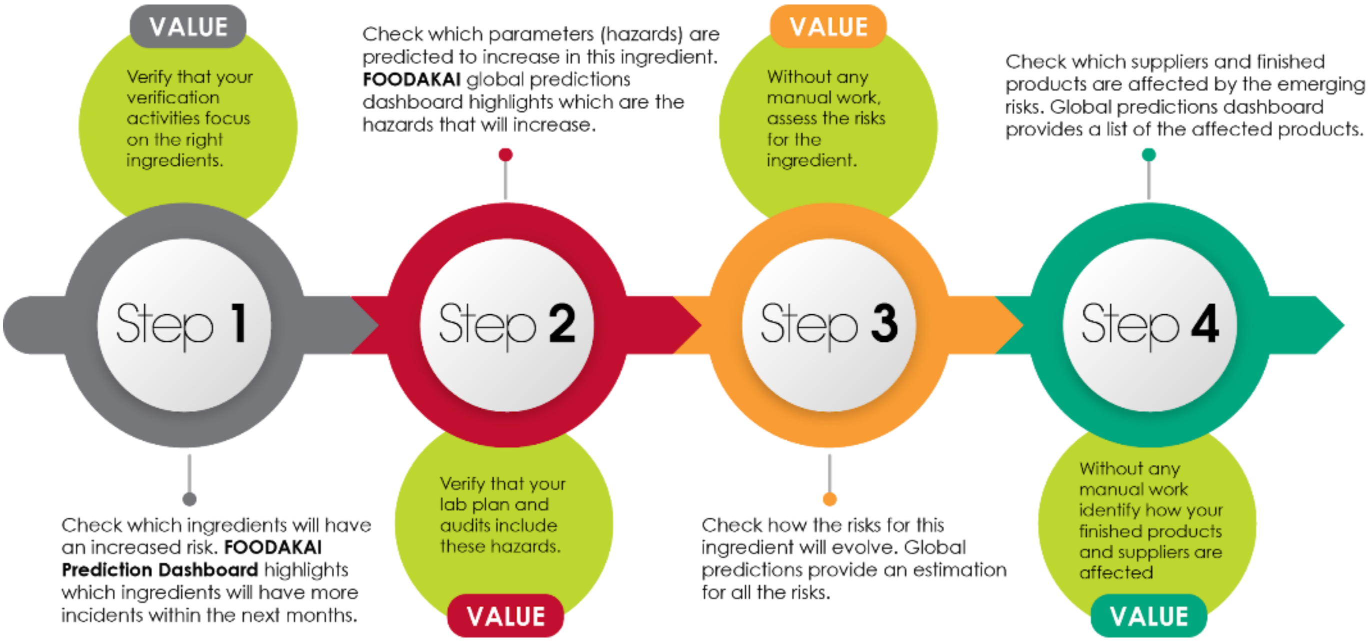


Figure 42 Main steps to enable prevention of incidents

The FOODAKAI prediction dashboard helps the end user to learn which of the ingredients used by the company will be affected by the increase of incidents and which are the most important risks that he needs to monitor.

We are using our high-quality data for recalls and rejections to train our models and generate tailor-made predictions for your supply chain. The prediction models are updated periodically using the latest update of the data about recalls and border rejections.

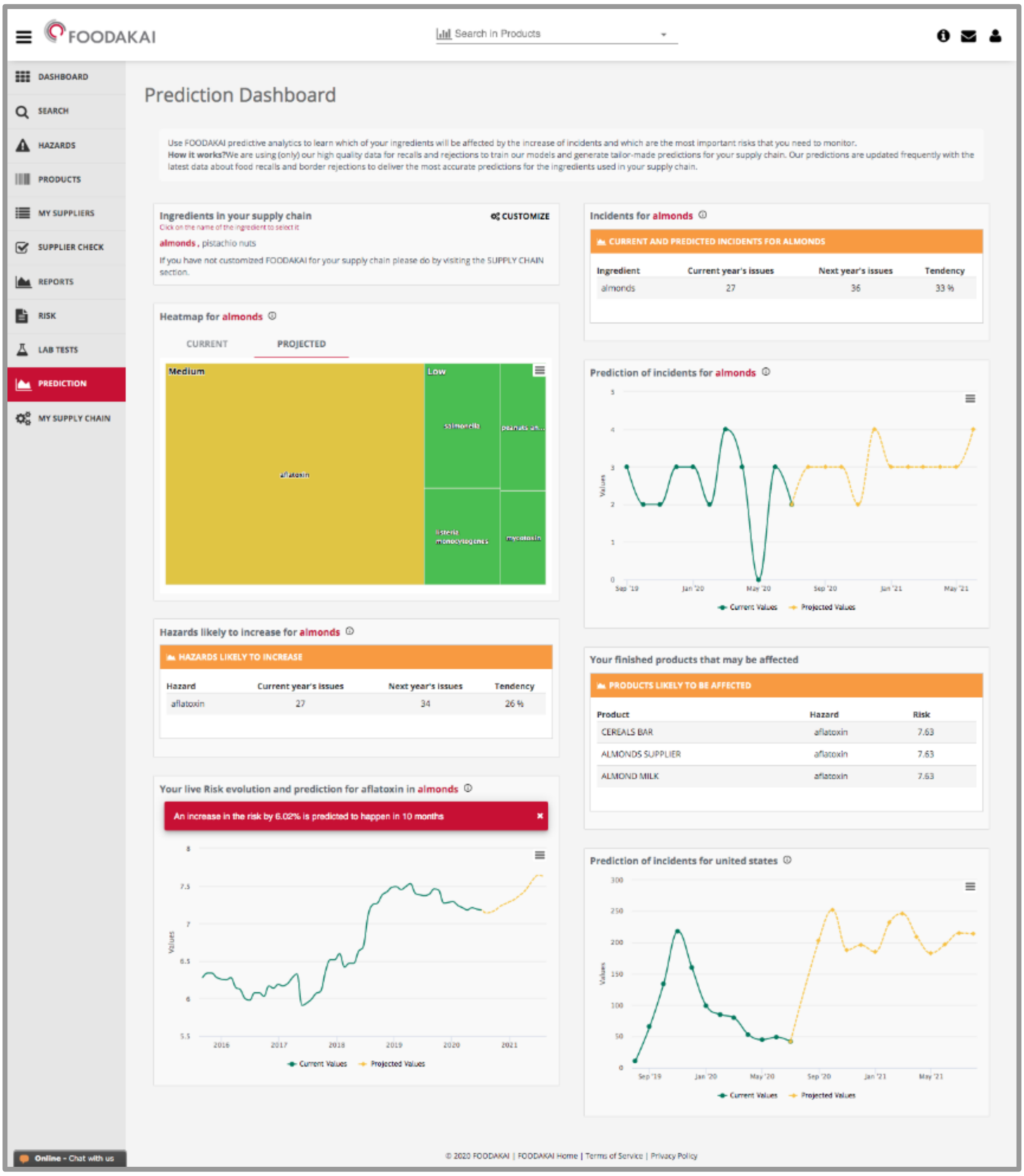


Figure 43 Global Predictions dashboard for ingredients and products

More specifically the dashboard includes blocks that provide

- **Prediction of the incidents trend** for the key ingredients used in company's supply chain. The predictions were computed using prediction models that are trained with tailor-made datasets for the specific ingredients. The actual and predicted values are displayed in a line chart.
- **Prediction of the ingredients risk** that will be increased due to the increase of the number of incidents for the ingredients. The actual and predicted values are displayed in a line chart.
- **A table with ingredients** for which the incidents are likely to be increased within the next months. The table presents both the current year incidents and the anticipated incidents for the next year.

- **A table with hazards** that are likely to be increased in the ingredients that the company is using. The table presents both the current year incidents and the anticipated incidents for the next year.
- **The risk heatmap for the ingredients** highlighting the top risk hazards based on the current and predicted values.
- **Finished products** that will be affected due to the increase of the risk in ingredients used in the products
- **Prediction for the risk of the countries or region** from which the company is sourcing ingredients. The actual and predicted values are displayed in a line chart.

The predictions for the incidents and for the specific hazards are produced using the recalls prediction algorithms that were developed in the context of the Big Data Grapes project. The visualizations included in the dashboard will be validated through interviews with end users of FOODAKAI that are working in food companies.

Personalized predictions are sent to the end users of the FOODAKAI in the form of an email, as presented in the following figure.

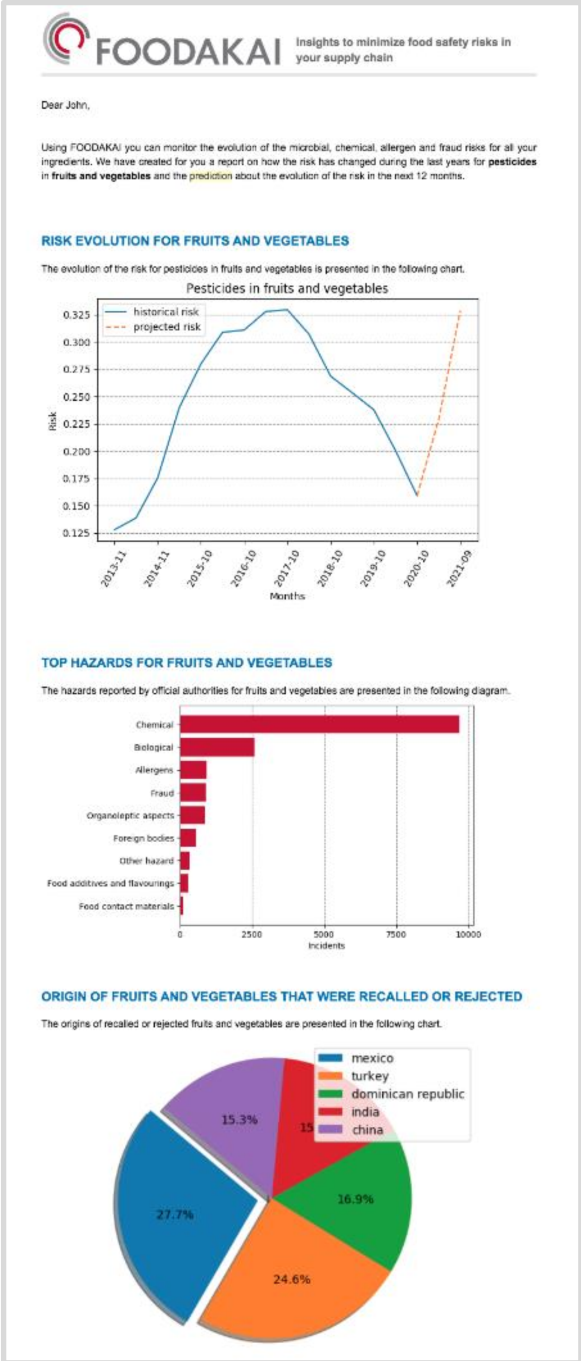


Figure 44 Predictions for user's ingredients sent through email

5.9 INTEGRATION OF THIRD PARTY DATA

Using the Big Data Grapes software stack, Agroknow integrated third party food safety data to develop a tailor-made risk dashboard for a specific food company of the beverages and snack industry. To integrate the tailor-made risk dashboard in FOODAKAI system, Agroknow used the data collection components to collect the third party data, the intelligence components to estimate the risk trends and the visualization components to deliver then estimated risk in visual format to the end users. The tailor-made risk dashboard using the third party data is presented in the following figure.

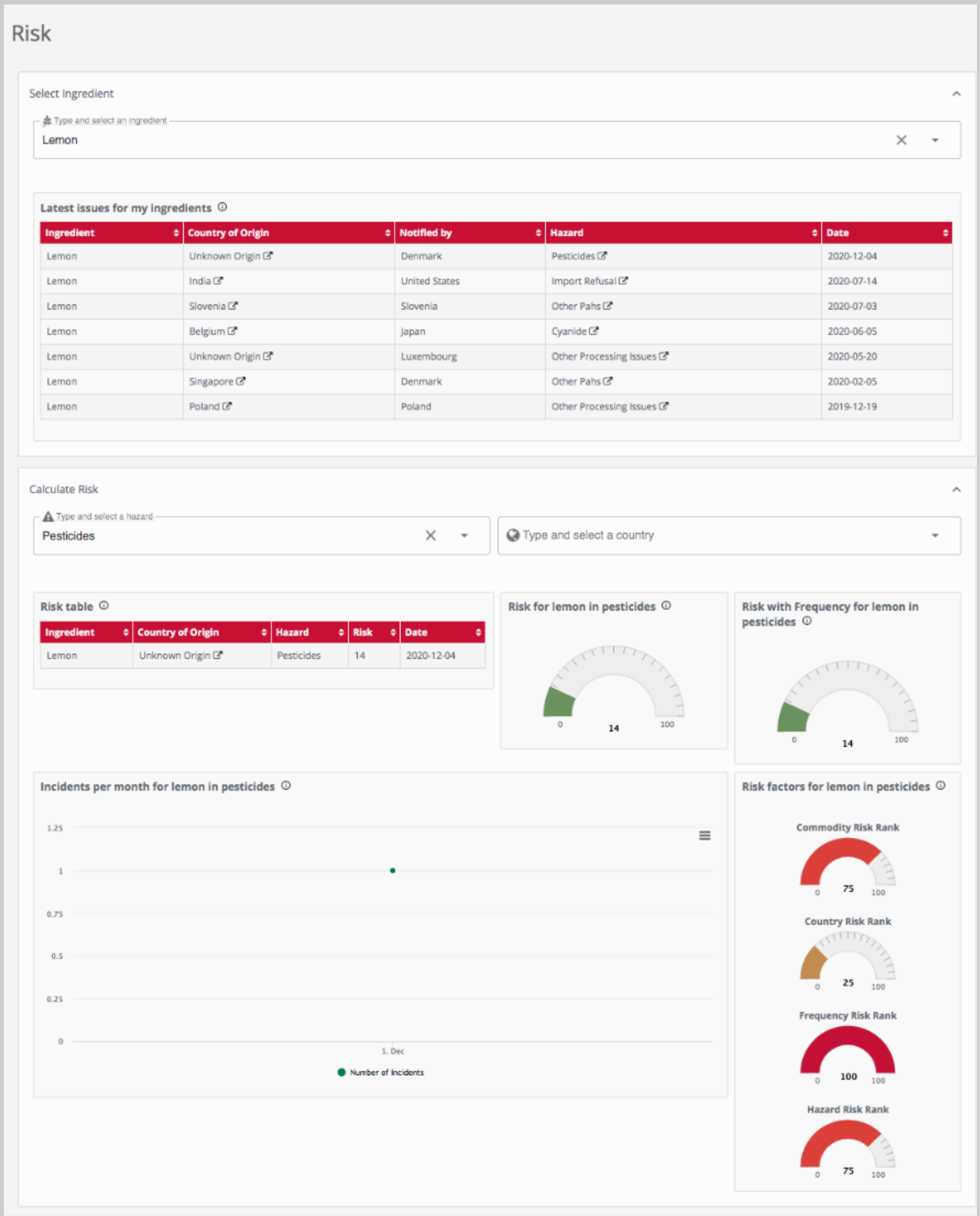


Figure 45 Tailor-made risk dashboard built using Big Data Grapes software stack and third-party data

5.10 PESTICIDES RESIDUES REPORT FOR THE AGROCHEMICAL INDUSTRY

By integrating the laboratory testing data from the pesticides monitoring programs of 34 countries, we managed to develop reports for the residues in foods and target specific industries such as the agrochemical. We combined the border rejections data and the lab testing data to test specific hypotheses about the factors that are affecting the increase of the residues in foods such as grapes. The report was presented to one of the main stakeholders of the agrochemical industry. The result from one such hypothesis is presented in the following diagram.

Figure 46 Example of hypothesis that was tested using the pesticides residues data

5.11 SOCIAL DATA INTEGRATION

In order to integrate the consumer's view about food safety and fraud issues, we expanded the Big Data Platform with components that can be used to collect data from social media. With this integration the Big Data Platform is collecting posts and comments in Greek from social media such as Facebook and Twitter. Sentiment analysis is performed for each text to identify positive, negative and neutral views of the consumers. This is a high velocity data source that was integrated in the Big Data Platform.

We run a specific use case about the recall of a honey product by a retailer in Greece. The historical food safety and fraud data were used to predict the hazards or fraud issues that can lead to a honey product recall. The social data were used to study the impact of the recall to the brand reputation. The case was presented to the retailer.

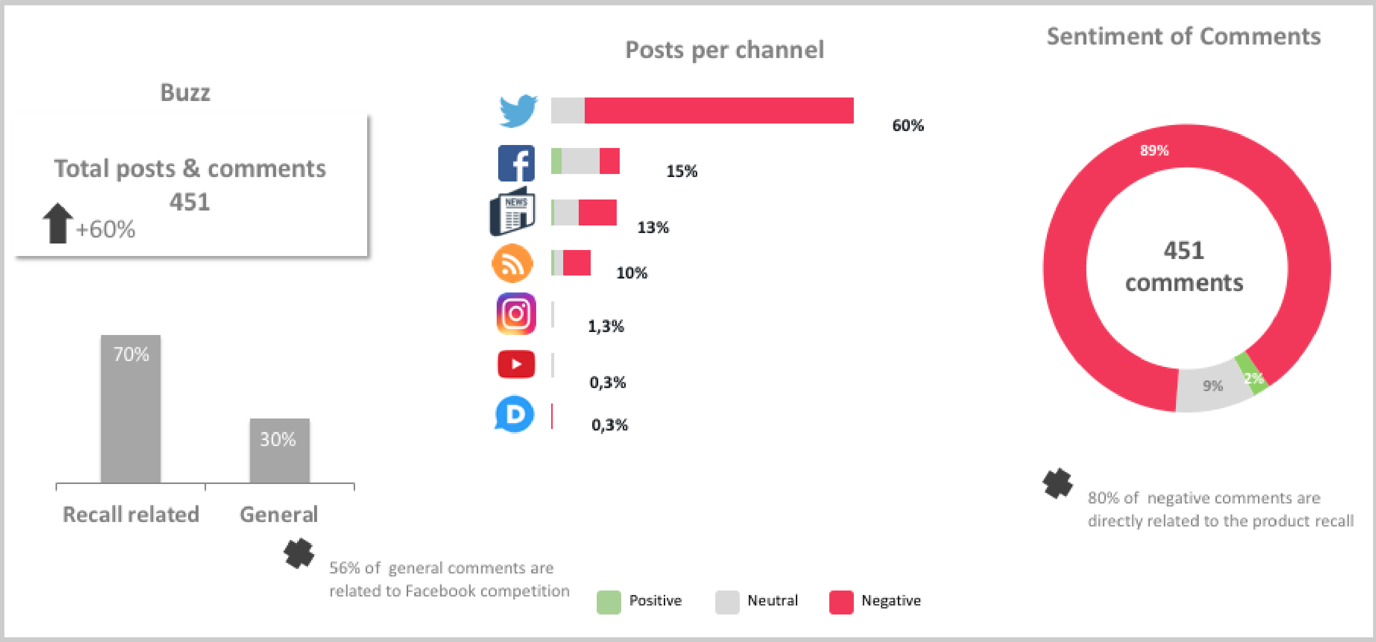


Figure 47 Example of social data report for an impact of a honey product recall

5.12 DATA API FOR EXPOSING GLOBAL FOOD SAFETY DATA

Using the Big Data Grapes software stack, Agroknow developed a Data API for sharing global food safety data with third party systems such as the SITI4FARMER. The data exposes the enriched version of the different types of the food safety data that is collected and processed by the Big Data Platform. The Data API was shared with startups and companies from the food industry to collect feedback about the coverage and the methods that are provided by the API. An example of a request that can be sent to the API using curl is presented in the following figure. This curl call searches for data:

- that are food safety incidents,
- after 2020-01-01,
- that involve grapes,
- originating from France

```
curl -X POST "https://api.foodakai.com/search-api-1.0/search/" -H "accept: application/json" -H
"Content-Type: application/json" -d "{\"apikey\": \"{apikey}\",\"from\": \"2020-01-
01\",\"entityType\": \"incident\",\"detail\": true,\"strictQuery\": {\"products.value\":
\"grapes\",\"origin.value\": \"france\"}}"
```

5.13 EXPLOITATION OF BIG DATA PLATFORM'S DATA API WITH STARTUP COMPANIES

This section focuses on the exploitation of the Big Data Platform' Data API that was developed in the context of the BigDataGrapes project. The Data API was used by 5 start ups and the use cases for each company are analysed in this section.

5.13.1 THE CASE OF ISMOOD

Intro about the company

isMOOD (https://www.ismood.com/) is providing an online platform for social media mining focusing on specific geographical regions. It provides the services to companies from several sectors such as financial, food, health, retailers and insurance. It monitors the brand of the companies and provides analytics and sentiment analysis based on the data that is collecting from social networks and media sites.

Contact point

Anna Kasimati, CEO.

The use case

The main objective of the use case was to study the impact in the brand of a food company due to food recalls.

isMOOD is working with large food brands such as Nestle and Coca Cola. To provide an integrated service to food companies for brand monitoring, isMOOD needed access to data about food recalls. More specifically, the goal was to study the impact of food product recalls to a company's brand. isMOOD used the Big Data Platform Data API, to get the information of product recalls and border rejections for all the food companies that it serves. BigDataGrapes technical partners helped isMOOD Data engineers to create a bridge that uses the Data API to pull information from the BigDataPlatform and link this information with the social data for a specific company.

5.13.2 THE CASE OF CREMEGLOBAL

Intro about the company

Creme Global (https://www.cremeglobal.com/) is a data modelling, visual analytics and computing company working to enable better decision making in a complex world. Creme Global is working with many of the biggest companies in the world. They help organisations to understand the context of their data and provide unequalled services in scientific modelling, data analytics and new model development utilising advanced cloud computing technology and curated data sets. They are working with large brands from the food sector such as Cargill, Unilever and Kerry. They have developed a data platform, namely Data Foundry, that was used by the Food Industry Intelligence Network to collect food fraud and authenticity data from the laboratory tests that food companies are conducting.

Contact point

Brendan Ring, Commercial Director

The use case

The collaboration with Creme Global focused on two aspects:

a) On how data from the BigDataPlatform about food safety and fraud incidents can be integrated into their Data Platform

b) How the risk assessment models that they are building for assessing the exposure to chemicals can utilize the data from the BigDataPlatform that was setup in the context of the BigDataGrapes project.

The **objective** of the use case was to manage concerns related to perceived increase in risk to consumers due to a particular pesticide residue alert.

Example scenario: One of the many global regulatory authorities detect a pesticide residue on a produce (e.g. a fruit like grapes, vegetable, dairy etc.) at above the Maximum allowable Residue Level (MRL). The regulatory authority issues an alert on their publicly available website. This alert causes anxiety with some consumers / lobby groups. Even though it was the agri producer or processor that used incorrect quantity of pesticide, it triggers a negative reaction to the pesticide manufacturer. The pesticide manufacturing company(s) must provide clear communication based on solid science to mitigate the concerns and de-escalate the issue.

The means by which to provide this communication can be achieved by:

1. Calculating the typical exposure a population would have to a particular pesticide residue.
2. Determining what the change in exposure would be due to the increased concentration as referenced in the alert.
3. Compare the two charts and show the impact related to the alert

5.13.3 THE CASE OF PROVISIONANALYTICS

Intro about the company

ProvisionAnalytics (https://www.provisionanalytics.io/) has developed a data platform that captures and interprets all the food data in one place. Using their platform they have built a food safety compliance and traceability solution, to help food businesses to go 100% paperless.

Contact point

Michael Gibbons, VP Product

The use case

The main objective of the use case was to empower the ProvisionAnalytics' data platform with global food safety data of the Big Data Platform.

More specifically, ProvisionAnalytics' Engineers used the Data API of the BIg Data Platform to pull information about food safety incidents linked to the products and suppliers of their customers. The use case focused on developing the integration of the global food safety data for specific supply chains in North America such as the ones for seafood and protein. The dashboards the ProvisionAnalytics provides to their customers were enhanced with the information of food recalls and border rejections providing improved evaluation reports for ingredients and suppliers. As the platform of ProvisionAnalytics focus very much on traceability, we used the information of UPC and LOT number to interconnect the data from the two platforms.

5.13.4 THE CASE OF MESH INTELLIGENCE

Intro about the company

Mesh Intelligence (https://www.meshintel.com/) provides services for predicting food safety and food supply chain risks. They work with the world's leading organizations and help them to predict, plan for, manage and avoid costly food safety and supply chain disruption.

Contact point
Shub Debgupta, CEO

The use case
The main objective of this use case was to use the global food safety incidents to predict risks in the supply chain. More specifically, Mesh Intelligence is focusing on developing risk prediction models that are based both on the company's internal data and external data for the global supply chain. It's engineers were seeking for more datasets in order to build more accurate prediction models. This is how they approached Agroknow and asked if we could provide the global food safety data in the form of Data API. The use case focused on the dairy sector and specifically on milk and milk products. They used the food safety incidents that the Big Data Platform is collecting and exposing through the Data API in order to develop models that may predict biological hazards. The outcomes of developing prediction models for the case of dairy sector was presented to one of the largest dairy companies on the world.

5.13.5 THE CASE OF BRITESCAN

Intro about the company.
BriteScan (https://britescan.com/) team started in 2010 as one of the most successful natural product DNA testing labs in the United States. That lab conducted testing and method development for some of the world's largest retailers, dietary supplement manufacturers, herb & spice and food companies in the world including federal government agencies like the US Food and Drug Administration (FDA), US Department of Agriculture (USDA), National Institute of Standards and Technology (NIST), US Fish and Wildlife Service (USFWS), and National Center for Complementary and Integrative Health (NCCIH) and the National Institutes of Health (NIH). In 2015 they sold that lab to an international standard setting organization. In 2018, Britescan tean discovered the power, speed and affordability of Artificial Intelligence and they launched the BriteScan portable testing device.

Today BriteScan's platform provides fast, affordable testing to government bodies, third party testing labs, manufacturers, food & beverage companies, cannabis growers and oil producers on a global basis.

Contact point
Danica Reynaud, CEO

The use case
BriteScan is currently looking for ways to enhance the detection capability of their device. To that direction they need access to more data about fraud and adulteration cases identified in the ingredients that are tested for authenticity using their device. The data of fraud and adulteration cases can be used to train the artificial intelligence models that are used to classify the samples to authentic or not. More specifically, they wanted to focus on the herbs and spices as well as olive oil. To get the critical mass of fraud and adulteration data that are needed to train the algorithms, they have used the Data API of the Big Data Platform that was developed in the context of the BigDataGrapes. The results of integrating more data in the BriteScan for the case of herbs and spices was evaluated and presented to a very large manufacturer.

5.14 DEVELOPMENT OF OPEN DATA COUNTRY RISK INSIGHTS AND INTEGRATION OF THE COVID RELATED DATA SOURCES SAFETY DATA

The COVID-19 situation has disrupted the global supply chain in several ways (https://agroknow.com/fsqa-digital-transformation/fears-of-global-food-shortages-on-the-major-food-production-countries-around-the-world/). As a response to the COVID-19 situation, Agroknow developed a live and open country risk dashboard (https://open.agroknow.com/) combining the data from the global supply chain with the data about the pandemic status in each country. To develop the risk dashboard, we integrated into the Big Data Platform that was developed using BigDataGrapes software stack, daily open data about COVID-19 cases that are published by European Centre for Disease Prevention and Control. This COVID-19 data was linked with the food safety profile data of each country. The main goal was to help the food safety experts working in food companies, national authorities, research centers and universities to have the most updated view of each country and to be able to explore the correlation of the pandemic with the food safety profile of the country. The risk dashboard provides open access to the data collected and processed in the context of the BigDataGrapes projects and analytics about COVID-19 status in the country and the status of the food safety incidents for the products that are produced in the country (figure).

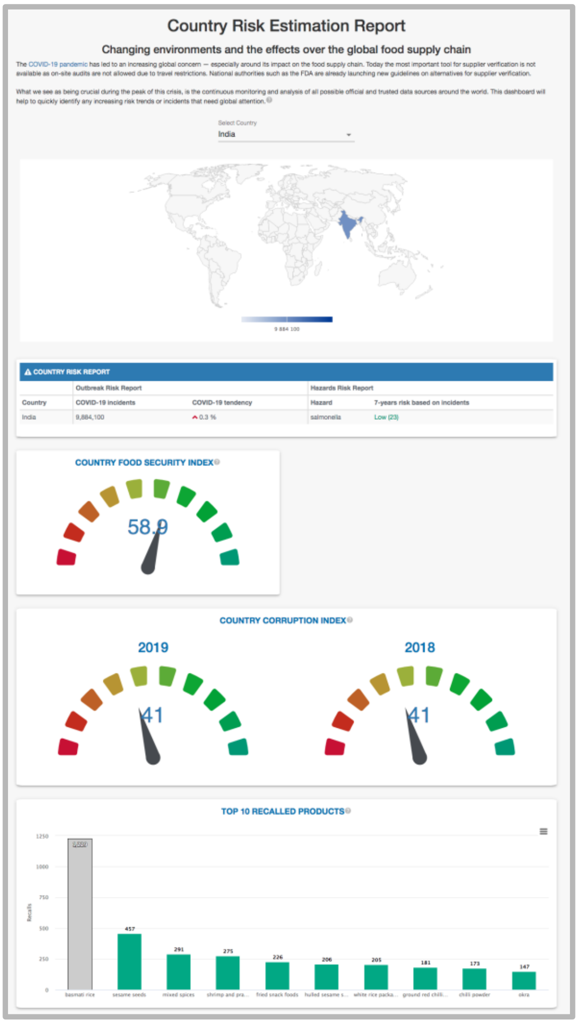


Figure 48 The COVID-19 live and open risk dashboard for the food industry

6. NEXT STEPS AND CONCLUSIONS

As seen in the previous chapters, the software SITI4Farmer is going to be updated with a new platform named ABACO Farmer with new interfaces for functionalities regarding satellite indices data storage and displaying, field sensors data storage and displaying, and finally the DSS improved by those satellite and field sensors data with a new data visualization, that can be easily readable by the farmers.

In the context of Precision Farming, it is essential to note that the BigDataGrapes project aims to also demonstrate that this new technology suits small-medium farms and not only big ones like in the ones in USA-Canada: this will be possible through the software solution from ABACO. In fact, the most important obstacle for Precision Farming is the cost of machinery. Within the BigDataGrapes project, we have demonstrated that is not necessary, at first time, to procure new hi-tech equipment, since the farmer can simply take effective and precise decisions starting from the knowledge of plants-soil conditions.

Furthermore, in the context of BigDataGrapes project, it should be taken into consideration that vineyards require a specific agronomic management to satisfy the real needs of this crop, in relation to the spatial variability within the vineyard. The introduction of new technologies in order to support vineyard management allows the efficiency and quality of production to be improved and, at the same time, reduces the environmental impact.

Figure 49 Variable Rate map for fertilization in Viticulture. 3D parcels representation simulation

7. REFERENCES

D. Kent Shannon, David E. Clay and Kenneth A. Sudduth. Precision Agriculture Basics, 2018 Published by: American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America, Inc.

Rouse, J.W., Haas Jr., R.H., Schell, J.A., Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. In: Third ERTS-1 Symposium, NASA, Washington, DC, pp. 309–317.

Gao, B., 1996. NDWI – A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water From Space. Remote Sens. Environ. 58, 257-266.

Huete, A.R., 1988. A soil-adjusted vegetation index. Remote Sens. Environ. 25, 295–309.

Jiang, Z., Huete, A.R., Didan, K., Miura, T., 2008. Development of a two- band enhanced vegetation index without a blue band. Remote Sensing of Environment 112, 3833–3845.

Clevers, J.G.P.W., Gitelson, A.A., 2013. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. International Journal of Applied Earth Observation and Geoinformation 23, 344–351 .

Ihuoma, S.O., Madramootoo, C.A., 2017. Recent advances in crop water stress detection. Computers and Electronics in Agriculture, 141, 267-275.

Morenoa, A., Masellib, F., Chiesib, M., Genesiob, L., Vaccarib, F., Seufertc, G., Gilaberta, M.A. , 2014. Monitoring water stress in Mediterranean semi-natural vegetation with satellite and meteorological data. International Journal of Applied Earth Observation and Geoinformation, 26, 246–255.

Serrano, J., Shahidian, S. and Marques da Silva, J., 2019. Evaluation of Normalized Difference Water Index as a Tool for Monitoring Pasture Seasonal and Inter-Annual Variability in a Mediterranean Agro-Silvo-Pastoral System. Water, 11, 62.