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

D8.5 - Evaluation Report and KPI Assessment

DELIVERABLE NUMBER	D8.5
DELIVERABLE TITLE	Evaluation Report and KPI Assessment
RESPONSIBLE AUTHOR	Aikaterini Kasimati (AUA)

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	Agricultural University of Athens
DELIVERABLE N. | TITLE	D8.5 | Evaluation Report and KPI Assessment
RESPONSIBLE AUTHOR	Aikaterini Kasimati (AUA)
REPLY TO	akasimati@aua.gr
DOCUMENT URL	http://www.bigdatagrapes.eu/
DATE OF DELIVERY (CONTRACTUAL)	30 June 2019 (M18), 31 Dec 2020 (M36, Updated Version)
DATE OF DELIVERY (SUBMITTED)	28 June 2019 (M18), 31 Dec 2020 (M36, Updated Version)
VERSION | STATUS	2.0 | Final
NATURE	Report (R)
DISSEMINATION LEVEL	Public (PU)
AUTHORS (PARTNER)	Aikaterini Kasimati (AUA)
CONTRIBUTORS	Spyros Fountas (AUA), Vasileios Psiroukis (AUA), Maritina Stavrakaki (AUA), Florian Schlenz (Geocledian), Simone Parisi (ABACO), Coraline Damasio (INRAE), Pantelis Natskoulis (Symbeeosis), Theodoros Kontogiannis (Agroknow), Giannis Stoitsis (Agroknow), Mihalis Papakonstadinou (Agroknow), Francesca Tsaropoulou (Agroknow), Iliana Giannelou (Agroknow), Alina Petri (Agroknow), Eirini Kouriantaki (Agroknow), Ifigeneia Mpokogianni (Agroknow), Nyi-Nyi Htun (KULeuven)
REVIEWER	Coraline Damasio (INRAE)

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.1	Initial ToC and document structure	10/03/2019	Aikaterini Kasimati (AUA), Vasileios Psiroukis (AUA)
0.4	Evaluation Report and KPIs - Template	07/06/2019	Aikaterini Kasimati (AUA), Vasileios Psiroukis (AUA)
0.8	Individual Piloting Evaluation Reports and KPIs lists	17/06/2019	Aikaterini Kasimati (AUA), Vasileios Psiroukis (AUA), Spyros Fountas (AUA), Maritina Stavrakaki (AUA), Florian Schlenz (Geocledian), Simone Parisi (ABACO), Coraline Damasio (INRA), Pantelis Natskoulis (Symbeeosis)
0.9	Internal Review	26/6/2019	Coraline Damasio (INRA)
1.0	Partners review, final comments and edits	28/06/2019	Aikaterini Kasimati (AUA), Vasileios Psiroukis (AUA)
1.2	Internal Technological Evaluation (I)	29/04/2020	All BDG Partners
1.5	End-users Surveys	30/10/2020	Aikaterini Kasimati (AUA), Maritina Stavrakaki (AUA), Simone Parisi (ABACO), Coraline Damasio (INRAE), Pantelis Natskoulis (Symbeeosis), Theodoros Kontogiannis (Agroknow), Giannis Stoitsis (Agroknow), Mihalis Papakonstadinou (Agroknow), Francesca Tsaropoulou (Agroknow), Iliana Giannelou (Agroknow), Alina Petri (Agroknow), Eirini Kouriantaki (Agroknow), Ifigeneia Mpokogianni (Agroknow), Nyi-Nyi Htun (KULeuven)
1.7	Internal Technological Evaluation (II)	04/11/2020	All BDG Partners
2.0	Partners review, final comments and edits	20/12/2020	Aikaterini Kasimati (AUA), Nyi-Nyi Htun (KULeuven), Giannis Stoitsis (Agroknow)

PARTICIPANTS		CONTACT
Agroknow IKE (Agroknow, Greece)		Nikos Manouselis Email: mailto: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 (INRAE, 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

AA1 (DPPH)	Antioxidant Activity 1 (2,2-DiPhenyl-1-PicrylHydrazyl)
AA2 (ABTS)	Antioxidant Activity 2 (2, 2'-Azino-Bis-3-ethylbenzoThiazoline-6-Sulfonic acid)
BA	Biological Activity
BDG	BigDataGrapes
D	Deliverable
DSS	Decision Support System
ECa	Electrical Conductivity
FAIR	Findable, Accessible, Interoperable, Reusable, as set of principles acting as an international guideline for high quality data stewardship
FRAP	Ferric Reducing Antioxidant Power
L8	Landsat-8
LAI	Leaf Area Index
MAC	Maceration
NDRE	Normalized Difference Red Edge Index
NDVI	Normalized Difference Vegetation Index
pH	Potential of Hydrogen
RI	Refractive Index
S2	Sentinel-2
SVIs	Spectral Vegetation Indices
TFC	Total Flavonoids Content
TMC	Total Microbial Count
TPC	Total Phenolic Content
UAE	Ultrasound Assisted Extraction
WP	Work Package
Y&M	Yeast and Moulds count

EXECUTIVE SUMMARY

The deliverable D8.5, “Evaluation Report and KPI Assessment”, aims to provide a report on the results of the application piloting sessions, in line with the defined experimental protocols and in accordance with the evaluation methodology. Based on the experimental and evaluation protocols defined in the context of T8.2, pilot partners monitor the execution of the application pilots and evaluate their results qualitatively and quantitatively against the Key Performance Indicators (KPIs) defined by the relevant protocols. The objective of this deliverable is to provide, in the first part, an overview and a first evaluation regarding each of the five pilots’ progress, stating and explaining the current status of development, the successfulness of implementation and the achieved performance of the BigDataGrapes Pilots. This is the updated and final version of this deliverable, including the evaluation of the pilots.

Deliverable D8.5, “Evaluation Report and KPI Assessment”, is based on the individual Qualitative and Quantitative Evaluations of the following pilots: Table and Wine Grapes Pilot (AUA), Wine Making Pilot (INRAE), Farm Management Pilot (ABACO & Geocledian), Natural Cosmetics Pilot (Symbeeosis) and Food Protection Pilot (Agroknow). This document reports an overview of the application piloting sessions’ results, both qualitatively and quantitatively against the defined KPIs, for each of the five pilots. Information was directly provided by the pilot leaders to ensure the specificity of the guidelines and the smooth progress and evaluation of all piloting sessions.

The document is structured as follows: Chapter 1 serves as an introduction to the deliverable whereas Chapter 2 provides an overview of the five pilots’ Qualitative and Quantitative evaluation. Each of the pilots’ Qualitative Evaluation Summary is divided in five sections: the pilot’s evaluation summary, plan progress, status of implementation, impact and potential pilot modification or required plans for improvement. For the Quantitative evaluation a KPIs list along with baseline values have been defined, including the Domain Specific KPIs only for those pilots that this applies and the Technological KPIs for all pilots. This part of the deliverable provides tangible results. Chapter 3 presents the structure of the BigDataGrapes Survey, the planning of the questionnaire, which was distributed to the end-users between the Intermediate and Summative phases (M30 - M34). Finally, the last chapter of the deliverable consists of the overall conclusions regarding the evaluation of the pilots, including discussion and suggestions that could improve their impact.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

Data-driven approaches have the potential to improve decision making in different industries and settings. This is also the case for the grapevine-powered industries, where it is evident that a rich, large-scale and diverse data pool is needed for carrying out the foreseen research and industry-centred activities. Therefore, throughout the lifecycle of the project, BigDataGrapes will continuously collect and monitor sensor data derived from all experimental sites accessible by the pilot partners, generating an expansive and diverse collection of datasets. These streams of data and datasets serve as the basis for carrying out research and technical work and are being used as the testbed for enabling the implemented technical components to efficiently handle the volume and intricacies of these data, acquired from realistic in-field conditions. A data marketplace demonstrator using these data assets harvested, serves as the project's experimentation environment where testing and adjustments of the proposed technical solutions can be carried out in a realistic setting. As the project progresses, the data pool is continuously enriched in volume and range, in accordance with the needs and requirements of the covered use cases.

Moving from testing in laboratory conditions to testing in real-world settings, BigDataGrapes has designed and is executing application pilots pertaining to the defined Use Cases, under WP8, "Grapevine-powered Industry Application Pilots". This work package is responsible for the pilots' planning and preparation, the definition of the experimental and evaluation protocols to be followed, the execution of the pilots and ultimately, the collection and evaluation of the pilot results and their assessment over indicators defined by the end users. In this context, this document provides a pilots' evaluation, a report on the results of the executing application piloting sessions during the BigDataGrapes project lifetime, in line with the defined experimental protocols and in accordance with the evaluation methodology.

The development of fully defined demonstrators for each of the grapevine-powered industry use cases allow to showcase and evaluate the BigDataGrapes platform and components in the context of specific end-user requirements from the different pilots. This evaluation is centered on realistic, strictly defined experiments that reflect real-life operations of the related industries. A detailed evaluation plan has been produced and is being followed during pilot execution, following an iterative approach of assessment that will be performed according to the proposed three-phase human-centred activities, organised in the following phases:

Formative phase, leading up to the "Use Case Definition & Assessment Planning" (M9): Industry-centred requirements and the concrete use cases where the BigDataGrapes solution will be applied and tested against these requirements will be defined through WP2 and WP8. During this phase, suitable data and processes for fulfilling the requirements of the specific use cases will be identified and relevant piloting activities will be defined.

Intermediate phase, leading up to "Functional Assessment Sessions" (M18): The first round of controlled pilot trials, will implement a first version of the pilots, using the first versions of newly developed BigDataGrapes components. These will be restricted piloting trials in terms of scale and complexity. The objectives of these trials are (a) to provide data for the assessment of early BigDataGrapes components and (b) to refine the pilots themselves into their subsequent iterations.

Summative phase, leading up to "Operational Assessment" (M36): The final phase entails the validation of the BigDataGrapes components in real-life conditions and with realistic complexity. The components will be used throughout the timespan, with developments in the technologies incorporated opaquely in the operational platform. A summary of the operation of the system and the respective pilot observations will be delivered, followed by a final Evaluation report where the performance of the system will be assessed against the established evaluation criteria and the appropriate Key Performance Indicators.

2 QUALITATIVE AND QUANTITATIVE EVALUATION

Evaluation within Task 8.5 is to be both formative and summative. The former is essentially self-assessment and will be carried out by all partners through filling the "Qualitative and Quantitative Evaluation", a part of the evaluation that provides tangible results via this deliverable. The summative evaluation will involve external as well as internal evaluation in the form of "BigDataGrapes Pilots' Survey".

This part of the deliverable is dedicated to the qualitative and quantitative evaluation of the BDG pilots. All five pilot partners (Table and Wine Grapes Pilot- AUA, Wine Making Pilot- INRAE, Farm Management Pilot- ABACO-Geocledian, Natural Cosmetics Pilot- SYMBEEOSIS, Food Protection Pilot- Agroknow) provided a report on the results of the application piloting sessions, in line with the defined experimental protocols and in accordance with the evaluation methodology in the context of T8.2, in order to evaluate their progress, stating and explaining the current status of development, the successfulness of implementation and their achieved performance.

The Evaluation consists of two parts the "Qualitative Evaluation Summary" and the "Quantitative Evaluation against the KPIs". Each of pilots' Qualitative Evaluation Summary is divided in five sections:

- ✓ Pilot's Evaluation Summary, including specific objectives, achievements/results and problems/challenges
- ✓ Plan Progress, providing details on the conducted tasks in the reporting period
- ✓ Status of Implementation, including the actors involved, methodology, deployed components and gathered data and format
- ✓ Impact of each piloting session
- ✓ Potential Pilot Modifications or required plans for improvement.

For the Quantitative evaluation a KPIs list along with baseline values have been defined, including the Domain Specific KPIs and the Technological KPIs for all pilots. The overall goal of the Technological KPIs is to assess the data generation process within the pilots and specific Use Case Scenarios. These KPIs include big data metrics such as the 3Vs (Volume, Variety, Velocity). The baseline values for the KPIs have been defined by the pilot partners, using various sources, such as collected data (both from the first year of the application pilot's lifetime and/or historical data owned by the piloting partner), available literature, statistical data, sector average, and expert knowledge. The baseline values are used to calculate the effect of the pilots.

The aim of the evaluation process is to clearly distinguish methodological and functional testing from real-world evaluation by actual users, to ensure that technical progress is directed towards improving realistic applications of the technologies. A detailed evaluation for the BDG technical outcomes under the specified use cases, are recorded in the form of an evaluation report for each pilot execution, briefly describing the involvement of the pilot and presenting the obtained results against the evaluation criteria and the appropriate Key Performance Indicators (KPIs).

2.1 TABLE AND WINE GRAPES PILOT INDIVIDUAL EVALUATION (AUA)

2.1.1 Qualitative Evaluation Summary

In order to report the Table and Wine Grapes Pilot's progress, AUA has completed the following table with the necessary information, regarding the current status of development, the successfulness of implementation, its impact and potential modifications.

AUA	Table and Wine Grapes Pilot Qualitative Evaluation Summary

Pilot Evaluation Summary

Specific Objectives

Deriving meaningful knowledge from many relevant, yet heterogeneous data sources is very important and will act as the basis for future decision-making processes. Throughout the lifecycle of the project, AUA will continuously collect and monitor sensor, farming and phenological data derived from all test sites located in Greece.

Soil properties, climate conditions and cultivation techniques constitute significant variables, which affect the quality of the final product. In particular, soil data (soil texture, soil electrical conductivity etc.) and weather data (average temperature, humidity etc.) affect both crop quality data (sugar content, anthocyanins content, phenolic compounds concentrations etc.) and crop quantity data (crop yield, berry weight and size etc.).

Some of the goals to be achieved through this sensor and farming data collection, is to denote associations and correlations between precision agriculture information and phenological data and grape and wine chemical analysis. Location-specific data will be used as auxiliary sources and will lead to the supply of vegetation indexes corrected for vineyard cultivation practices, more accurately determined vegetation stages and input to plant performance and grape quality indicators among others. Finally, the ultimate goal is to correlate the aforementioned data with earth observation data to examine the effectiveness of applying machine learning techniques and eventually train the relevant machine learning components.

The Table and Wine Grapes pilot is linked to the Prediction and Farm Management Use Cases and more specifically to four Scenario Hypothesis, namely B1. Yield Prediction, B3. Crop Quality Prediction for Optimizing Post Harvest Treatments of Table Grapes, C1. Optimisation of Farm Practices in the Vineyard, C2. Management Zones Delineation for Vineyards, as shown in Figure 1.

Table 1: Scenario Hypothesis served by the Table and Wine Grapes Pilot

Use cases	Use Case Scenarios	Partner
A. Data Anomaly Detection & Classification	A. Earth Observation Data Anomaly Detection & Classification	A. ABACO-GEOCLEDIAN
B. Prediction	B1. Yield Prediction B2. Predicting Biological Efficacy B3. Crop Quality Prediction • for Optimizing Post Harvest Treatments of Table Grapes (B3-1) • for Optimizing Winemaking (B3-2)	B1. AUA B2. Symbeeosis (APIGEA) B3-1. AUA B3-2 INRA
C. Farm Management	C1. Optimization of Farm Practices in the Vineyard C2. Management Zones Delineation for Vineyards	C1. ABACO-GEOCLEDIAN AUA C2. AUA

Achievements/Results

The first step to the successful implementation of the piloting activities for the Table and Wine Grapes pilot was to define the appropriate "Data and

<table>
<tr>
<td></td>
<td>
Datasets" to be collected throughout the project's lifetime. The second step was to update and refine the Scenario Hypothesis and create a link to the data. Thus, during the first crop season/experimental period, which belongs to the Formative phase, suitable data and processes for fulfilling the requirements of the specific Use Cases were identified and relevant piloting activities took place.
This includes the collection of the following data:

High quality topographic data have been collected from all three (3) test sites and topographic and elevation maps have effectively been generated.
Georeferenced apparent soil Electrical Conductivity (ECa) data has been collected from two (2) test sites (Kontogiannis Estate and Palivou Estate) and could effectively be divided into management zones.
Vegetation Indices-based maps have been generated from all six measurements and the different phenological growth stages of grapevines have been successfully monitored for the first season.
Weather and soil data have been recorded effectively since the installation of the IoT stations on the fields.
Yield data have been collected for the first harvesting period.
Qualitative and quantitative data have been gathered from all three sites, which will be further statistically analysed and correlated with the data coming from the field for the first season.
Access to Landsat-8 and Sentinel-2 satellite data for the test sites under the Table and Wine Grapes Pilot has been provided by Geocledian, in order to enable temporal and spatial observations and analysis through their field monitoring service ag|knowledge.

After the completion of the data acquisition procedures by the end of the first crop season/experimental period data management activities took place, including Data Normalisation, Data Modelling, Semantics Annotation and transformation to RDF among others.

Problems/Challenges
Soil electrical conductivity mapping and in general ECa non-touch sensing systems are extremely sensitive to interferences from nearby metallic objects. Out of the three (3) experimental sites of the Table and Wine Grapes pilot, Palivou estate and Kontogiannis estate are open-field vineyards, allowing for successful ECa data collection. In Fasoulis estate, however, vines are covered with a light nest-like fabric approximately 0.5 meters from the top of the canopy, placed on metallic pillars across the entire vineyard. For this reason, no accurate ECa data could be collected from this site.

Another hardship we encountered was a hardware malfunction of the EM-38 MK2 sensor on the first data collection visit to the experimental sites. The handheld logger, used to wirelessly store data and control the sensor's functions, lost Bluetooth contact with the sensor during data collection on Palivou estate, demanding a direct cable connection between logger and sensor. This way, the sensor was no longer mount-able to the tractor, forcing the AUA team to perform the data collection manually.
</td>
</tr>
<tr>
<td>Pilot Plan Progress</td>
<td>Measurements related to the Table and Wine Grapes pilot successfully took place throughout the first crop season of the pilot on the three (3)</td>
</tr>
</table>

experimental sites. As it is natural, emphasis was given during the summer months (May through September) when grapevines grow and produce grapes and therefore allowing for crop monitoring data collection, which is the procedure that generates the greatest volume of data of the pilot. More specifically, an overall plan of the experimental methodology of the pilot is presented below:

At first, the boundaries of the vineyards were geo-referenced using GPS technology at the very beginning of the project, as soon as the experimental fields were chosen. Time-stable zones have been formed using soil ECa mapping, assisted by elevation mapping using the RTK-GPS. These data related to the boundaries, management zones and elevation will be used throughout the course of the project. These measurements mentioned above have been performed by the AUA Precision Agriculture lab.

- *Dataset Name:* Topographic data and elevation maps
- *Data Description:* Spatial data (boundaries and elevation data)
- *Data collection frequency: a single measurement per crop season*

Soil and weather data have been continuously collected since Day 1 of the project. The setup of the IoT stations and the data acquisition have been performed by the AUA Precision Agriculture lab.

- *Dataset Name: IoT Stationary data*
- *Data Description: Soil moisture and meteorological parameters*
- *Data collection frequency: Constant data stream throughout the entire year*

Canopy reflectance data and vegetation indices have been recorded with the use of Crop Circle ACS-470, SpectroSense2 and Crop Circle RapidSCAN CS-45 sensors six times per season/summer starting at the beginning/middle of May, so that the phenological development of the grapevine, which is divided into nine (9) principal growth stages, will be followed in the best way. The measurements mentioned above have been performed by the AUA Precision Agriculture lab.

- *Dataset Name: Canopy sensing and vegetation indices*
- *Data Description: Canopy reflectance and vegetation indices data*
- *Data collection frequency: six (6) per crop season, with a minimum of two (2) sensors in each data collection*

Landsat-8 and Sentinel-2 satellite data have been collected for the same periods with the canopy reflectance measurements, again six times per season/summer. The satellite images have been provided by Geocledian.

- *Dataset Name: Satellite data*
- *Data Description: Landsat-8 and Sentine-2 imagery of the test sites*
- *Data collection frequency: six (6) per crop season*

Some of the qualitative and quantitative characters of the grapevines, such as pH, soluble solids, total titratable acidity, antioxidant capacity by DPPH, FRAP assay, and aminoacids, have been tested three times over a season. Finally, the rest of the qualitative and quantitative characters have been assessed at the end of each season, when harvesting. The qualitative and quantitative measurements have been performed by the AUA Vitis lab.

	- *Dataset Name:* Classical analytical techniques (HPLC) - *Data Description:* Phenolic composition data - *Data Frequency: Laboratory analyses take place once every year, after harvest* The grape and berry mechanical properties (weight, length, width, density etc.), berry deformation, berry detachment, density, grape volume, berries diameter, berries weight for table grapes will be measured at post-harvest. - *Dataset Name: Grape and Berry Mechanical Properties* - *Data Description: Lab Measurements of berries' properties* - *Data Frequency: Laboratory analysis' take place once every year, after harvest* Finally, yield data has been collected and yield mapping has been performed during the harvest period, by both AUA labs. - *Dataset Name: Yield Mapping* - *Data Description: Yield data* - *Data collection frequency: 1/crop season*
Status of Implementation	*Actors Involved* AUA PA lab and Vitis Lab have successfully collected data throughout the first crop season from the three (3) experimental sites (Palivou estate, Kontogiannis estate and Fasoulis estate), while Geocledian provided the satellite imaging along with their data analysis service for the same sites. Agroknow, KULeuven and Ontotext participated as Tech providers, with Agroknow and Ontotext acting as model providers, while KULeuven was responsible for data visualisation. Finally, all pilot partners and tech partners provided insights in order to formulate and define the scenario hypothesis. *Methodology* The first step to the successful implementation of the piloting activities for the Table and Wine Grapes pilot was to define the appropriate "Data and Datasets" to be collected throughout the project's lifetime. The second step was to update and refine the Scenario Hypothesis and create a link to the data. Thus, during the first crop season/experimental period, which belongs to the Formative phase, suitable data and processes for fulfilling the requirements of the specific Use Cases were identified and relevant piloting activities took place. For the data collection, the methodology used is as described in D8.2 - Experimental Protocols and Evaluation Methodology. After the completion of the data acquisition procedures, by the end of the first crop season/experimental period, data management activities took place, including Data Normalisation, Data Modelling, Semantics Annotation and transformation to RDF among others. *Deployed Components* <u>Precision Agriculture Lab</u> • EM38-MK2 probe (Geonics LTD, Mississauga, ON, Canada) (Figure 1). Data collection is supported by the DAS70-AR Data Acquisition System (Archer Data logger). The EM38 measures apparent soil ECa in millisiemens per metre (mS/m) in the root zone at 0.5 and 1.0 m depth and the in-phase

ratio of the secondary to primary magnetic field in parts per thousand (ppt).

Figure1: EM38-MK2 (left) and Archer Data logger (right)

- HiPer V RTK GPS (Topcon Positioning Systems Inc., Livermore, CA, United States) (Figure 2). Records topographical data, such as field boundary points, and elevation data. The final output can be a KML, KMZ file. This measurement has been performed a single time throughout the course of the project, at the beginning of the table and wine grapes pilot, prior to all other measurements.

Figure 2: Topcon HiPer V RTK GPS

- Crop Circle ACS-470 (Holland Scientific Inc., Lincoln, NE, United States) (Figure 3). This remote sensing tool is measuring the radioactive transfer and the biophysical characteristics of plant canopies. It is an active crop canopy sensor that provides basic reflectance information from plant canopies and soil as well as classic spectral vegetative index data (NDVI, NDRE etc.). Vegetative indices measurements have been done in two different canopy parts, by the side and at upper canopy of the vines, after mounting the equipment to a winegrowing tractor.

Figure 3: Crop Circle ACS-470

- SpectroSense2+ GPS (Skye Instruments Ltd, Landrindod Wells, UK) (Figure 4). Used to estimate the Leaf Area Index (LAI) and NDVI vegetation indices.

Figure 4. SpectroSense2+

- Crop Circle RapidSCAN CS-45 (Holland Scientific Inc., Lincoln, NE, United States). Used to estimate vegetation indices such as NDVI and NDRE indices (Figure 5).

Figure 5: Crop Circle RapidSCAN AC-45 handheld reflectance sensor

- Software such as Surfer 11 (Golden Software), ArcGIS (ESRI, Redlands, CA, USA), Global Mapper for the generation of thematic maps.
- Two Vantage Pro 2 weather stations (Davis Instruments Corp., Hayward, CA, United States) (Figure 6) with rain sensor, to detect rainfall, anemometer to measure wind speed and direction, air temperature sensor, air humidity sensor, barometer to monitor atmospheric pressure. The basic equipment can be supplemented with sensors for UV and solar radiation. The automatic weather station will be installed inside the vineyard. Weather information is being recorded throughout the growing season.

Figure 6: Vantage Pro2 Weather Station (left) and Decagon EC-5 soil moisture sensor (right)

- Four Decagon EC-5 soil moisture sensors (METER Group, Inc., Pullman, WA, USA) (Figure 6) recording throughout the growing season the humidity and temperature of the soil.

<u>Laboratory of Viticulture</u>

- Soluble solids will be determined using an ATAGO N1-a refractometer with a 0-32 Brix measurement range at 0.28 Brix increments.
- Total titratable acidity will be measured by titration with a 0.1 N NaOH solution and will be expressed as tartaric acid.
- The quantitative and qualitative analysis of the substances, which exist in berries, must and wines such as, organic acids, sugars, phenolic compounds, amino acids, anthocyanins, volatile compounds, etc. will be performed using an HPLC Shimadzu Nexera comprising a gradient pump Shimadzu Nexera X2, a ProStar model 410 AutoSampler, and a ProStar model 330 Photodiode Array Detector on a reversed-phase Waters C18 x select (250 mm x 4.6 mm, 5 mm) column.
- Antioxidant activity (2,2-diphenyl-1-picrylhydrazyl, DPPH) will be evaluated by the free radical scavenging activity of DPPH using a modified colorimetric method, while the reduction of the DPPH radical will be determined by measuring the absorption at 517 nm in a UV/Vis spectrophotometer (Perkin Elmer, Lambda 25, Beaconsfield, Bucks, U.K.). The absorption of the antioxidant activity (Ferric Reducing Antioxidant Power, FRAP) will be measured at 593 nm.
- Assimilation rate (photosynthesis) and stomatal conductance will be obtained by measurement of inlet and outlet CO_2 and H_2O relative concentration using a portable photosynthesis system (Li-6400XT, Li-Cor, Lincoln Nebraska, USA).
- Chlorophyll concentration will be measured on the leaves using a SPAD 502 (Konica Minolta, Europe).

Site Description

Three test sites have been chosen for data collection for BigDataGrapes in Greece. These are situated in the regional unit of Corinthia, in the north-eastern part of Peloponnese. The following have been selected: for winemaking Palivou Estate and Kontogiannis Estate and for table grapes Fasoulis Estate.

Palivou Estate: is located in Nemea, planted with *Vitis vinifera* L. cv. 'Agiorgitiko' and 'Merlot' for winemaking. The row orientation is northeast-southwest, and the training/trellis system is VSP (vertical shoot positioned)- cane pruning, double Guyot.

Figure 7: Palivou Estate test site (Google Earth Pro)

Kontogiannis Estate: in Ancient Corinth having the same VSP -double Guyot or double Royat- training/trellis system planted with 'Roditis', 'Savatiano', 'Mavroudi' and 'Soultanina' for winemaking. Its row orientation is north to south.

Figure 8: Kontogiannis Estate test site (Google Earth Pro)

Fasoulis Estate: situated in Nemea, cultivated with 22 different table grape varieties, where each line has a different variety. The orientation is southeast to northwest.

Figure 9: Fasoulis Estate test site (Google Earth Pro)

Gathered Data and Formats

- Spatial data (boundaries and elevation data)
    - Dataset using for the pilot: Topographical and elevation maps
    - Data collection frequency: Once at the start of the pilot
    - Associated data model/format: csv, xls, xml
    - Data size: MB

- Geo-referenced apparent soil electrical conductivity (ECa)
    - Dataset using for the pilot: ECa maps
    - Data collection frequency: Once at the start of each year
    - Associated data model/format: csv, xls, xml
    - Data size: MB

- Canopy reflectance and vegetation indices using proximal sensors
    - Dataset using for the pilot: Vegetation Indices maps
    - Data collection frequency: Six (6) times during each crop season
    - Associated data model/format: csv, xls, xml
    - Data size: MB

- Canopy reflectance and vegetation indices using satellite imagery
    - Dataset using for the pilot: Vegetation Indices maps
    - Data collection frequency: Six (6) times during each crop season
    - Associated data model/format: GEOTIFF, png
    - Data size: GB

- Qualitative and quantitative characters; Grape and berry mechanical properties (weight, length, width, density etc.), berry deformation, berry detachment, density, grape volume, berries diameter, berries weight.
  For wine and table grapes: soluble solids, pH, total titratable acidity, total phenols and anthocyanins, total flavonoid content, total flavanol, flavonol, flavone content, tannins, antioxidant capacity (trans-resveratrol, piceid, ε-viniferin) by DPPH, FRAP assay, aminoacids
    - Dataset using for the pilot: Grape and Berry Mechanical Properties

  - Data collection frequency: Once at the end of each crop season
  - Associated data model/format: csv, xls, xml
  - Data size: MB

- Full phenolic profile of grapevine varieties in correlation with the phenological stages to improve the quality of viticultural products
  For table grapes: leaf analysis, foliar chlorophyll contents photosynthetic pigment content of the leaves, water potential correlated to the proline content
  - Dataset using for the pilot: Phenolic composition data
  - Data collection frequency: Once at the end of each crop season
  - Associated data model/format: csv, xls, xml
  - Data size: MB

- Yield data
  - Dataset using for the pilot: Yield variation maps
  - Data collection frequency: Once every year, during harvesting
  - Associated data model/format: csv, xls, xml
  - Data size: MB

- IoT stationary data
  - Dataset using for the pilot: Soil moisture data, meteorological parameters
  - Data collection frequency: Constant data stream throughout the year
  - Associated data model/format: csv, xls, xml
  - Data size: MB

Table 2: Table and Wine Grapes Pilot Data and Datasets

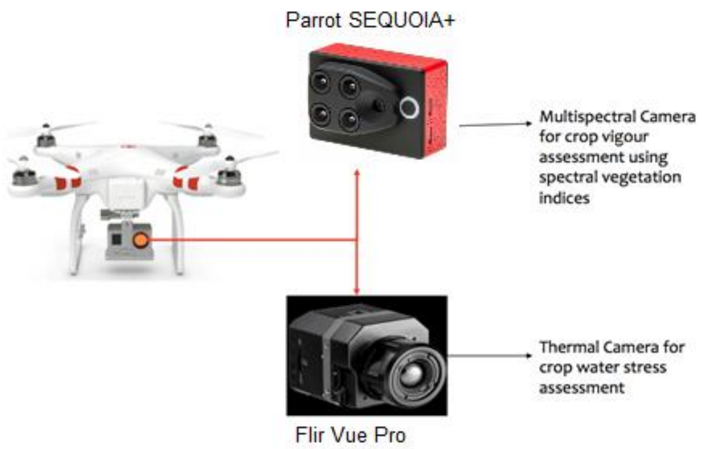
Name	DataSet Description	Priority	Provenance	Data Type Format	Data size
Yield Mapping	Yield data	Essential	Laboratory equipment	csv, xls	MB
Grape and berry mechanical properties	Measurements	Essential	Laboratory equipment	csv, xls	MB
Classical analytical techniques (HPLC)	Phenolic composition data	Essential	Laboratory equipment	csv, xls	MB
Topographic da ta and elevation maps	Spatial data (boundaries and elevation data)	Essential	Remote sensing	csv, xls, xml	MB
Canopy sensing and vegetation indices	Canopy sensing data	Essential	Proximal sensors	csv, xls	MB
IoT stationary data	Soil moisture data, meteorological parameters	Essential	IoT data	csv, xls	MB
Drone imagery	Drone images	Essential	Multispectral and thermal cameras	GEOTIFF	GB
Crop Calendar	Records of crop groth stages and agricultiral operations	Essential	Log files	doc, xls	MB
Eca sensing	Geo-referenced soil electrical conductivity data	Essential	Proximal sensors	csv, xls	MB

Impact

Some of the goals to be achieved through this sensor and farming data collection, is to denote associations and correlations between precision agriculture information and phenological data and grape and wine chemical analysis. Location-specific data will be used as auxiliary sources and will lead to the supply of vegetation indexes corrected for vineyard cultivation practices, more accurately determined vegetation stages and input to plant performance and grape quality indicators among others. Finally, the ultimate goal is to correlate the aforementioned data with earth observation data to examine the effectiveness of applying machine learning techniques and eventually train the relevant machine learning components.

The Table Grape and Wine pilot aims to achieve the following:

- Allow for data-based predictions to be made with accuracy levels higher than ever before. Combination of improved resolutions (spectral, spatial and temporal) of remotely sensed images, coupled with more precise on-the-ground multiple data sources, such as soil, weather and vegetation indices derived from sensors can create powerful models able to deliver predictions taking into consideration a plethora of factors and variables, drastically increasing the overall accuracy of the estimations.
- Optimize the timing of table grapes harvest and storage based on data from multiple sources. They will also receive production data and assure them whether the production covers the specific standards that are set by the super markets. A powerful system that allows growers to efficiently plan the

	harvest and storage of their table grapes will greatly benefit them and increase the overall production quality of the sector. • Assess soil, weather and vegetation data as recorded from the pilots by unlocking the value of big data for irrigation and fertilization in agriculture and promote potential real-time decision support tools. • Develop tools that automatically generate a number of options for delineating management zones, taking into consideration two or more variables and optimizing the process.
Pilot Modifications	✓ During the first year of piloting activities no measurements with the Laser Scanner LMS100 LiDAR sensor were performed. ✓ Addition of drone imagery during the second crop season: Two (2) Phantom 4 Pro drones (Dà-Jiāng Innovations, Shenzhen, Guangdong, China) (Figure 10) equipped with a multispectral Parrot Sequoia+ camera (Parrot SA, Paris, France) and a Flir Vue Pro thermal camera (FLIR Systems Inc., Wilsonville, Oregon, United States) to collect aerial imagery data and generate Vegetation Indices and Irrigation/Water Activity maps respectively.  Figure 10: Phantom 4 Pro drone w/ multispectral and thermal camera

2.1.2 Quantitative Evaluation Against KPIs

Domain Specific KPIs

AUA has created the list of domain specific KPIs for the Table and Wine Grapes Pilot and has defined their baseline values, which are presented in the following table.

Table 3: Table and Wine Grapes Pilot Domain Specific KPIs Catalogue

Variable	Definition	Units	2018 Baseline	2019	2020
Total soluble solids	The minimum sugar content of the must at harvest for the	Brix	20	20	20

	production of red dry wine				
Total titratable acidity	The minimum total titratable concentration of the must at harvest for the production of red dry wine	g tartaric acid/L of must	3.5	3.5	3.5
Total anthocyanin content	Minimum total anthocyanin content for the production of red dry wine	mg malvidin/g of fresh skin	3.00	3.00	3.00
Selective harvesting	The purpose is to achieve different harvest dates depending on the grape quality characters per plot/cell instead of harvesting the entire vineyard on the same date	Number of harvesting dates per plots per vineyard	1	1	1

Technological KPIs

Additionally, in order to perform a complete quantitative evaluation for the Table and Wine Grapes Pilot, a Technological KPIs list along with baseline values have been defined by AUA.

Table 4: Table and Wine Grapes Pilot Technological KPIs Catalogue

Variable	Definition	Units	2018 Baseline	2019	2020
Focusing Big Data					
Volume	Variation in raw data volume – Proximal sensor data	MB	26	19.5	13.5
Volume	Variation in raw data volume – Weather data	MB	5	8.5	8.4
Volume	Variation in raw data volume – Earth observation data	GB	300	350.75	346.0**
Volume	Variation in raw data volume – Drone Imagery	GB	*	37.8	35
Volume	Variation in raw data volume – Yield and Quality	MB	2	2	2
Variety in Data Source Types	Number of different data source types	Data sources	15	17	17
Variety in Data	Number of different types of data (in	Datasets	10	12	12

	different resolutions)				
Velocity	Speed of data generated – Proximal sensor data	MB/crop season	26	19.5	13.5
Velocity	Speed of data generated – Weather data	MB/year	5	8.5	8.4
Velocity	Speed of data generated – Earth observation data	GB/crop season	125	146.15	157.25
Velocity	Speed of data generated – Drone Imagery	GB/crop season	-	37.8	35
Velocity	Speed of data generated – Yield and Quality	MB/crop season	2	2	2

*Drone imagery is expected to add up to another 200GB from 2019, when the data collection starts.

2.2 WINE MAKING PILOT INDIVIDUAL EVALUATION (INRAE)

2.2.1 Qualitative Evaluation Summary

In order to report the Wine Making Pilot's progress, INRAE has completed the following table with the necessary information, regarding the current status of development, the successfulness of implementation, its impact and potential modifications.

INRAE	Wine Making Pilot Qualitative Evaluation Summary
Pilot Evaluation Summary	*Specific Objectives* Our specific objectives are related to the following topics: - to have a device to improve data quality (correction) and make FAIR data - to have a better understanding of 'How data from the field can affect the wine quality?' and 'How vine water status can affect the wine quality?' - to discover knowledge in order to design new viticulture / winemaking systems The Wine Making pilot is linked to the Prediction Use Cases and more specifically to the Scenario Hypothesis B3-2. Crop Quality Prediction for Optimizing Wine Making. *Achievements/Results* Climatic database • Results: precipitation, water height, evapotranspiration, humidity, insolation, wind. French Network of Grapevine Repositories

	• Results: morphological description, genetic profile, accessions, location, aptitudes, details. PHIS • Results: leaf area, plant height, biomass, plant width and images/ plant trait extractions. SilexVitioeno • Results: soil characteristics, vineyards, plots, sub-plots, grapevines, different treatments of experiment, cropping management, grape/ berry properties, yield, observations regarding the grape characteristics during the growing season. Monitoring of winemaking operations • Different stages of the winemaking process, duration of winemaking steps, dates etc. ALFIS • Alcoholic fermentation kinetics. Laboratory analysis • Chemical and physical analysis of must and wine: pH, alcohol, total acidity, volatile acidity, residual sugars, etc. Sensory analysis • Wine flavour profiles, scores regarding aromas presence. After the completion of the data acquisition procedures, data management activities took place, including Data Normalisation, Data Modelling and Linking, Semantics Annotation and transformation to RDF among others. *Problems/Challenges* Data "linkability": The main challenge is related to traceability from field to wine. Indeed, diverse persons are involved in data collection at different scales and it is not always easy to make links between all operations (especially between winemaking and field data). Moreover, data are not always in our information system. Our challenge is to make connections between all data to be able to have a transversal approach from the vineyard to the final product.
Pilot Plan Progress	Data are coming from diverse sources. The main data sources identified are: Climatic database - *Climatic data available from 1989 to 2020.* - *Partners involved: INRAE.* - *The goal is to connect environmental or climatic data to information related to the winemaking process or vines.* French Network of Grapevine Repositories - *The French Network of Grapevine Repositories (RFCV) includes 36 regional partners involved in the preservation of grapevine genetic resources and selection. More than 180 repositories are distributed to the grape producing regions in France.* - *The network partners are the regional stakeholders involved in grapevine conservation and selection in France. They play a pivotal role in the conservation and valorisation of our viticultural heritage.* PHIS - *Three experiments have been achieved on the phenotypic platform in 2012, 2013 and 2014 (the PhenoArch phenotyping platform).*

	- *Partners involved: the researches at the LEPSE unit (INRAE).* - *Measurement goal: context of water scarcity and global climate changes, the researches at the LEPSE aim at analysing and modelling the responses of plants to drought and high temperatures as well as their genetic variability at the intra- and inter-specific levels.* SilexVitioeno - *SilexVitioeno Pech Rouge is an information system filled by people from Pech Rouge and about plots, vine stock, and berries all along grapevine lifecycle.* - *Partners involved: INRAE, experimental unit of PechRouge.* - *The measurement goal is to characterize vines and their environment to optimize crop management.* Monitoring of winemaking operations - *The monitoring of winemaking operations is obtained with a form filled after harvest to record all information linked with the must studied.* - *Partners involved: Workers from Pech Rouge.* - *The goal of this monitoring is the traceability to compare and follow every steps of the winemaking process.* ALFIS - *Alfis is a SQL database started in 2004; this is an automatic monitoring connected to an information system.* - *Partners involved: INRAE unit called Science for Oenology ("SPO").* - *The goal is to make fermentations under controlled conditions with an online acquisition of fermentation kinetics to understand must and yeast behaviour and optimise the process of alcoholic fermentation.* Laboratory analysis - *Laboratory analysis is done on each sample of must and wine.* - *Partners involved: experimental unit of Pech Rouge and another INRAE unit "Science for oenology" SPO.* - *The purpose is to have chemical and physical analysis of must and wine and look at their evolution* Sensory analysis - *Sensory analysis is done by wine experts on some wine samples before and after bottling.* - *Partners involved: Science for Oenology unit.* - *Measurement goal: to obtain wine flavour profiles.* All this data gathered will help us to build hypothesis and create semantic models. All data are recorded but they still need to put in order and adapted to the Big Data Grapes project.
Status of Implementation	*Actors Involved* INRAE has successfully collected data throughout the first crop season. More specifically, in Viticulture from Pech Rouge Nicolas Saurin (Team leader) and the Team Viticulture / Quality Grapes. In Winemaking from Pech Rouge Jean-Michel, Alain Samson (Team leader) and the Team Innovative Technology/ Oenology. Agroknow, CNR, KULeuven and Ontotext participated as Tech providers, with Agroknow and Ontotext acting as model providers, CNR on the data analytics, while KULeuven was responsible for data visualisation. Finally, all pilot partners and tech partners provided insights in order to formulate and define the scenario hypothesis.

Methodology

Viticulture:

Research topics are mainly related to the agro-climatic and social context of the South of France. It can be summarised as follows: hot climate, dryness, climate change, pH and K+ elevation, grapes and wine acidity diminution, irrigation, varieties selection, vineyard cultural practices, diversification, sensors. The following is an example of the main steps for vineyard experiments:

- Phenology: budding, flowering, veraison (observation, visual counting).
- Hydric status: water potential (pressure chamber): 1 time a week from June to harvest.
- Apex: apex growth: 1 or 2 times a week (from mai/ June until harvest).
- Exposed leaf area (calculation with width, height and foliage porosity): once after growth stop (in July, August).
- Maturity: sample of 200 berries with berry weight analysis, refractometric indexes (sugar content), total acidity, pH, assimilable nitrogen, and for red wines anthocyanins and total polyphenol index. 3 sampling dates (at 2 or 3 dates and at harvest).
- Aroma precursors, depending on the project.
- Yield components: bunch number, weight per grapevine at harvest.
- Cut wood weight: number of shoots, shoots weight per grapevine in winter (during pruning, after harvest).

Winemaking:

First, information related to harvest is recorded. Then, different operations are monitored during the winemaking process. For example, some observational data are done concerning the different product features and also observational results of some attributes for a particular product stage such as grape, initial must, must after alcoholic fermentation, and finished wine (such as sensorial analysis achieved by judges).

Here, Figure 11, is an example of the different steps of winemaking done at Pech Rouge for red wines.

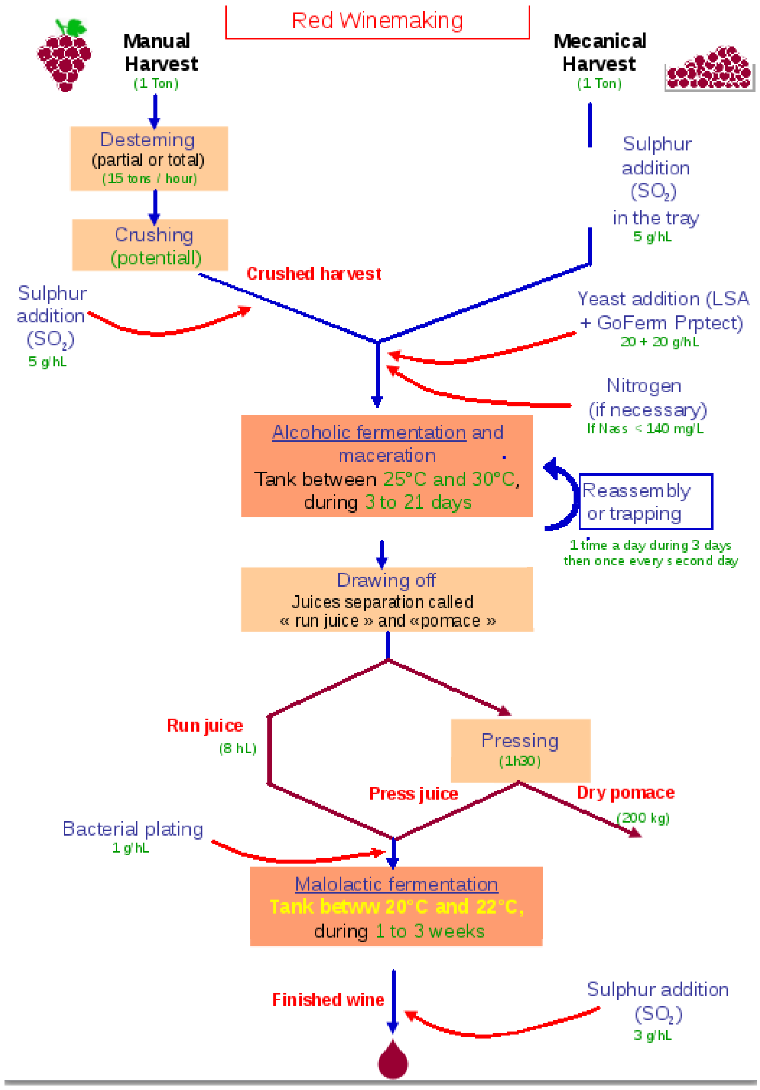


Figure 11: Steps for winemaking taking place in Pech Rouge

Deployed Components

Viticulture:

The land field of Pech Rouge includes a total area of 170 ha of land planted with 38 hectares of vines, distributed in three areas. The INRAE Pech Rouge Experimental Unit also contains analytical laboratories, technological tools.

Winemaking:

The land field of Pech Rouge includes a total area of 170 ha of land planted with 38 hectares of vines, distributed in three areas. The INRAE Pech Rouge Experimental Unit also contains analytical laboratories, technological tools and finally a Sensory Analysis Laboratory, which enables the tasting of different wines.

- ✓ Laboratory equipment
- ✓ Expert analysis
- ✓ Experimental technological facilities: A technological facility dedicated to grape extraction, grape processing and winemaking. A technological facility for delayed fermentations allowing alcoholic fermentations under controlled conditions with on-line acquisition of fermentation kinetics.

- ✓ Winery
- ✓ Packaging facility
- ✓ Barrel cellar
- ✓ Wine bar
- ✓ Laboratory analysis
- ✓ Sensorial analysis

Gathered Data and Formats

Viticulture:

- ✓ The gathered data using this measurement technique: climatic data, genetic data, and information about vineyards, plots, sub-plots, grapevines, and different treatments of experiment.
- ✓ Defined frequency of data collection: every years but variables assessed depend on projects goals.
- ✓ Database: SilexVitiOeno, CLIMATIK, PHIS, French Network of Grapevine Repositories, format: xls/ csv, size: MB.

Winemaking:

- ✓ The gathered data using this measurement techniques: different operations during the winemaking process: recorded observational data concerning the different product features. Observational results of some attributes for a particular product stage such as grape, initial must, must after alcoholic fermentation, and finished wine.
- ✓ Defined frequency of data collection: every years but variables assessed depend on projects goals.
- ✓ Database: files coming from Pech Rouge experimenters/ laboratories science for oenology unit, Alfis.
- ✓ Format: xls, hand-written, size: MB.

Table 5: Wine Making Pilot Data and Datasets

Name	DataSet Description	Priority	Provenance	Data Type Format	Data size
Genetic Data	Genetic profile, Morphological description, origin, etc.	Essential	French Network of Grapevine Repositories (Database of the collections)	csv or api	MB
Soil characteristics	Texture, pH etc.	Essential	Field measurement	xls	MB
Plot management	Treatments/fertilizing (when, what, how much), ground handling, or tasks related to the culture management, pest control, water status, yield etc	Essential	Field measurement	Pdf,doc, xls	MB
Climatic data	Rainfall, temperature, radiation etc.	Essential	Field measurement	xls	MB
Grape and berry mechanical and chemical properties	Anthocyanin content, weight, length, width, density etc.	Essential	Field measurement	xls	MB
Qualitative and quantitative characteristics of must	Sugar content, alcohol, pH etc.	Essential	Laboratory equipment		MB
Winemaking activities	Bioconversion of sugar into ethanol and CO_2, Monitoring of alcoholic fermentation and sugar content, yeast characteristics etc.	Essential	Laboratory equipment	xls, pdf	MB
Sensory analysis	Expert panel of tasters' sensory analysis (wine bitterness, astringency, phenol content, aroma etc.)	Essential	Expert analysis	xls	MB
Wine commercial information	Number of bottles produced, number of bottles sold	Additional Data	Selling point	xls	MB

Examples:

- Data collection Phenological stages (Budding, flowering and veraison).
  - Sampling: 15 grapevine per repetition.
  - Frequency: 2 dates per stages.
- Maturity (weight of 200 berries, sugar, total acidity, pH, assimilable nitrogen, anthocyanins, polyphenols).
  - Sampling: 200 berries per treatment + 200 berries for anthocyanins and polyphenols.
  - Frequency: 14 days before harvest, 7 days before harvest and at harvest.

	• Aroma precursors o Sampling: 20 bunch of grape/treatment. • Yield o Sampling: 15 grapevines / repetition at harvest. • Leaf Area Index (LAI) o Sampling: 4 transepts per treatment.
Impact	The main impact of the Wine Making piloting sessions is related to data gathering at its experimental site. Links have been identified but also lack of data connection. A direct impact will be noticeable on our way of working, storing and managing data.
Pilot Modifications	No need for modifications has appeared yet for the Winemaking Pilot.

2.2.2 Quantitative Evaluation Against KPIs

Domain Specific KPIs

INRAE has generated the list of domain specific KPIs for the Wine Making Pilot and has defined their baseline values, which are presented in the following table.

Table 6: Wine Making Pilot Domain Specific KPIs Catalogue

Variable	Definition	Units	2018 Baseline	2019	2020
Product yield per plot	Kg per plot of grapes harvested	Kg/plot	1425	1347	1198
Product yield	Kg per ha of grapes harvested	Kg/ha	4107	4721	4115
Wine volume per kilogram harvested Red wines	Red wines It corresponds to the volume of wine produced per kilogram of harvested grape	L/kg	0.65	0.67	0.66
Wine volume per kilogram harvested White and rosé wines	White and rosé wines. It corresponds to the volume of wine produced per kilogram of harvested grape	L/kg	0.29	0.39	0.26
Residual sugar content in wine – red wines	Sugar content in wine. We have to check that the value of residual sugar is below 2 g/L. Red wines Calculation: 100 * nber of conformed wine/total wine number	%	100	100	100
Residual sugar content in wine –	Sugar content in wine. We have to check that the value	%	100	100	100

white and rosé wines	of residual sugar is below 2 g/L. White and rosé wines Calculation: 100 * nber of conformed wine / total wine number				
Volatile acidity after alcoholic fermentation for red wines	It must be 0,10 <x < 0,98. Red wines Calculation: 100 * nber of conformed wine/total wine number	%	100	100	100
Volatile acidity after alcoholic fermentation for white and rosé wines	It must be 0,10 < x < 0,88 White and rosé wines Calculation: 100 * nber of conformed wine/total wine number	%	100	100	100
Color intensity (darkness) for red wines – visual analysis Before bottling	The purpose is to have a dark color for red wines. The ratio calculated corresponds to the number of judges who found the wine dark / total number of judges who are able to detect the characteristic	0 < Ratio < 1	0.66	0.57	0.33
Color intensity (clearness) for white and rosé wines – visual analysis Before bottling	The purpose is to have a clear wine for white and rosé wines. The ratio corresponds to the number of judges who found the wine clear / total number of judges who are able to detect the characteristic	0 < Ratio < 1	0.52	0.60	0.56
Fruity flavor Before bottling	The fruity flavor is well desired for all wine types. The ratio corresponds to the number of judges who detected this aroma / total number of judges	Ratio	0.48	0.55	0.49

Technological KPIs

Additionally, in order to perform a complete quantitative evaluation for the Wine Making Pilot, a Technological KPIs list along with baseline values have been defined by INRAE.

Table 7: Wine Making Pilot Technological KPIs Catalogue

Variable	Definition	Units	2018 Baseline	2019	2020

			First year of the project		
Focusing Big Data					
Volume	Variation in raw data volume – Plot Management	MB	2.4	2.4	2.4
Volume	Variation in raw data volume – Climatic data	KB	265	265	265
Volume	Variation in raw data volume – Grape and berry mechanical and chemical properties	KB	139	144	61
Volume	Variation in raw data volume – Qualitative and quantitative characteristics of must and wine	KB	408	336	492
Volume	Variation in raw data volume – Winemaking activities	MB	8.0	7.6	5.6
Volume	Variation in raw data volume – Sensory Analysis	KB	544	388	1550
Volume	Variation in raw data volume – Satellite Data	GB	47	55	44.5**
Velocity	Speed of data generated during harvesting period	GB/ harvesting period, 4 months	14.9	18.3	16.2**
Velocity	Speed of data generated – Satellite data	GB/month	2.6 S2 1.3 L8	3.3 S2 1.3 L8	2.9 S2** 1.2 L8**
Variety in Data Source Types	Number of different data source types	Data sources	18	19	20
Variety in Data	Number of different types of data (in different resolutions)	Datasets	9	10	11
Data transformation	Number of rdf triplets, from raw data	Number	0	62157	207190

Data linked	% of data linked, data connection – dataset linked divided by the total number of datasets	%	11%	67	100
Level of FAIR-ness	Fair data assessment tool especially for winemaking activities	RDA SHARK evaluation (David et al., 2019)	12/18 Never 6/18 If Mandatory 0/18 Sometimes 0/18 Always	3/18 Never 7/18 If Mandatory 7/18 Sometimes 1/18 Always	1/18 Never 0/18 If Mandatory 10/18 Sometimes 7/18 Always
Big Data Process Metrics					
Data Normalization (Homogenization)	Steps number needed for data to be available for analysis and processing Winemaking activities	Number	7	5	3

It is important to underline that these variables make sense if they are well described in ontologies using semantic web to be able to do machine learning on them.

** measured until November 30, 2020

2.3 FARM MANAGEMENT PILOT INDIVIDUAL EVALUATION (ABACO-GEOCLEDIAN)

2.3.1 Qualitative Evaluation Summary

In order to report the Farm Management Pilot's progress, ABACO, with the help of Geocledian, has completed the following table with the necessary information, regarding the current status of development, the successfulness of implementation, its impact and potential modifications.

ABACO	Farm Management Pilot Qualitative Evaluation Summary
Pilot Evaluation Summary	*Specific Objectives* The ABACO and Geocledian 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 relevant for the farmer should be displayed in a way that provides an added value and new insights to the farmer for his activities.

The ultimate goal for this pilot is to provide precise monitoring of the vegetation – atmosphere – soil system and manage best practice activities, through the improvements of SITI4farmer, with the support of specific DSS tools and precision farming features.

The Farm Management pilot is linked to the Data Anomaly Detection & Classification and the Farm Management Use Cases and more specifically to three Scenario Hypothesis, namely A. Earth Observation Data Anomaly Detection & Classification, C1. Optimisation of Farm Practices in the Vineyard, C2. Management Zones Delineation for Vineyards.

Achievements/Results
The pilot has compiled best practice activities and has started to read properly information provided by sensors installed in fields.
Geocledian has acquired and processed Copernicus Sentinel-2 and USGS Landsat-8 images for all sites during the first 18 months of the pilot run time and made them available to all partners. Geocledian's Processing Platform provides the service ag|knowledge that allows the automatic crop monitoring for fields with advanced products based on all spectral bands of these satellites. The processing platform has been substantially improved and the development into a Big Data Processing platform has started. New products are available.

Problems/Challenges
Challenges have been encountered on training the DSS and precision farming tools.

Pilot Plan Progress

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
- Setup of SITI4farmer for the 2 companies
- Development of the system sensors station interfacing
- Measurements and monitoring of field activities

Piloting activating (in particular the last one) will be part of the entire project and is considered as an activity in continuous improvement, where data and results will be presented to the other partners regularly.

Macro Activity	Time
Geocledian: Data acquisition, processing & provision	Q3-Q4.2018
Winery Company formal engagement	Q3.2018
Abaco's Hardware & Software supplying	Q3.2018
Deploying of SITI4farmer	Q4.2018
Abaco's Development & Configuration for sensors integration	Q4.2018
Training of user on the system	Q4.2018
Geocledian: Integration of new data sources (new vegetation indexes)	Q1-Q2.2019

Field Measurements & monitoring	Q1.2019 to Q4.2020
Geocledian: Development of Management Zones & data anomaly detection	Q1-Q4.2019
Geocledian: Improvement of vineyard specific products with feedback from users	Q1-Q4.2020

Geocledian has acquired and processed the described satellite data of all sites up to M18. Visible images and Vegetation Index Maps can be produced in our Processing platform and the data is available to all project partners in near real-time. The satellite data processing platform has been substantially improved.

The pilot aims at providing a precise monitoring of the vegetation – atmosphere – soil system and manage best practice activities with the support of specific DSS tools and precision farming features. It has compiled best practice activities and started to read properly information provided by sensors installed in fields.

Status of Implementation

Actors Involved

Farm Management Pilot (ABACO-Geocledian) and farm management data collected from the test sites in Italy, with Geocledian providing the satellite and SVIs datasets and KU Leuven for the visualisation of the dashboard.

Methodology

Abaco is going to release a version of its product; SITI4farmer is ready to be used in the field by 2 different winemakers, owners of the test cites in Italy and all the project partners.

SITI4farmer is able to (with just an internet connection trough browser):

- Prepare the graphical crop plan
- Manage farming practices and phenology phases
- Analyse indices and dashboards to support decisions (agro-meteorology and vegetation)
- Keeping farm data organized and accessible
- Recording field data with the SITI4l and app
- Printing and export data

Furthermore, it’s able to integrate weather data and services from different sources, also, it can use open databases and local land registries made available by everyone that has an exposed service.

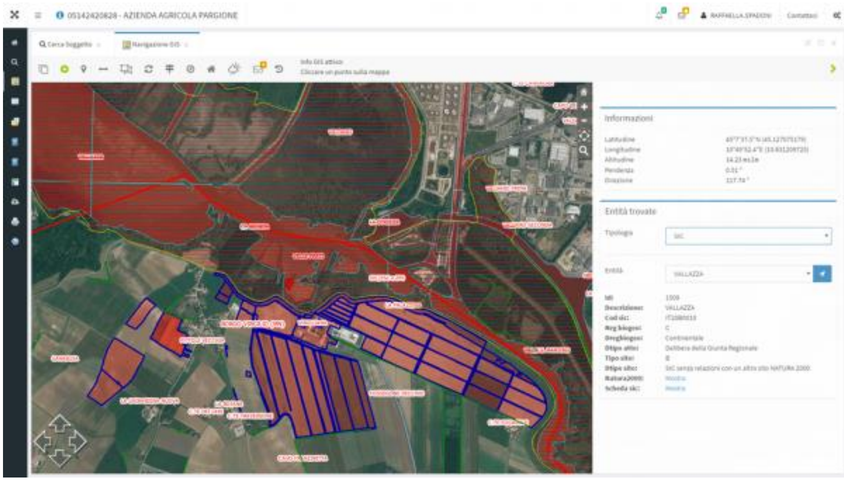


Figure 12: SITI4farmer screen view

Geocledian has acquired and processed the described satellite data of all sites up to M18. Visible images and Vegetation Index Maps can be produced in our Processing platform and the data is available to all project partners in near real-time. The satellite data processing platform has been substantially improved. A series of developments have been implemented and deployed successfully to improve data download & processing, performance monitoring, scalability and data visualization and to enable the delivery of the new data products and vegetation indexes 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 & analysis of large-scale data. Developments on Management Zones & data anomaly detection have been started.

Deployed Components

In order to make full and comprehensive measurements in the fields, automatizing as much as possible, Abaco has acquired and integrated within the system 2 sensors stations dedicated to this purpose. Sensor Stations were purchased directly from Abaco and installed within two farms.

Sensors and weather station are set to working via radio with a central server, and transmit 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
- Instruments Leaf Wetness Sensor Module with 5 meters of cable
- Drill & Drop Sensor (Temperature and soil moisture sensors)

Figure 13: Sensor & Weather Station

Figure 14: Rain Gauge Module

Concerning Geocledian, these components have been deployed successfully in the frame of the improved satellite data processing platform Ag|knowledge:

- Extended data download and processing components
- New processing performance monitoring tools
- New vegetation indexes and data products component
- New API endpoints for data delivery
- New data visualization tools for data review and analysis

Site Description

The approach consists on the involvement of 2 wineries, making them an active part of the project, collecting data from the field, in automatic and manual manners, and therefore contribute to the results.

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 Vineyards of Brunello of Montalcino

Figure 15: 12 HA of vineyards of Brunello of Montalcino

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 Vineyards of CHIANTI D.O.C.

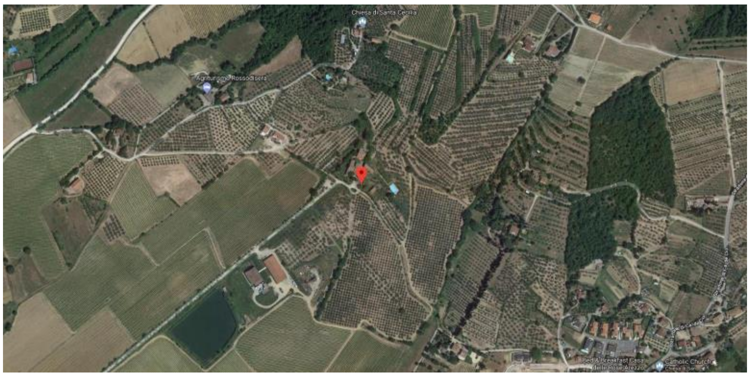


Figure 16: 35 HA of Vineyards of CHIANTI D.O.C.

Gathered Data and Formats

- Several Satellite vegetation indices based on Sentinel 2 and Landsat 8 data (NDVI, NDRE1, NDRE2, NDWI, EVI2, SAVI, CIRE)
- Pessl Instruments Weather and Soil sensors: Air temperature and humidity, wind speed and direction, precipitation, soil temperature and water content at different deepness (from 0 to 50 cm), leaf and bunches IR Temperature, Leaf moisture.
- Weather and soil data are monitored every hour, Satellite data depends on data availability in function of coverage with an average of 6 days frequency,
- Rest-API JSON format for both Satellite and Weather-Soil station

Table 8: Farm Management Pilot Data and Datasets

Name	DataSet Description	Priority	Provenance	Data Type Format	Data size
Sentinel-2	Sentinel-2A/B MSI visible & NIR bands, NDVI time series & advanced products	Essential	Copernicus EO Programme, ESA	JSON, GEOTIFF, PNG	150 GB/year*site
Landsat-8	Landsat-8 OLI visible & NIR bands & advanced products	Essential	USGS, NASA	JSON, GEOTIFF, PNG	6 GB/year*site
VHR data	VHR satellite data, e.g. TripleSat VHR optical bands	Additional	TBD, e.g. 21AT	TBD, e.g. GEOTIFF	TBD
Chemical and physical info on grapes	Antocyanins, Ph, Brix values during maturation	Additional	Excel table file	XLS	TBD
Day by Day Activities in term of treatments, fertilization, field operation	Diary where farmer or operators can record and /or plan all the activities on their fields	Essential	SITI4farmer	Text file	TBD
Plot and Fields information georeferenced	Information of Plots position, shaping, cultures, type of seed, dates, and everything related on the culture and the farm itself (form official and not official point of view)	Essential	Form the field through SITI4farmer	Test files	TBD
Relative Humidity	Relative humidity (RH) is the ratio of the partial pressure of water vapor to the equilibrium vapor pressure of water at a given temperature	Essential	Field Sensors	Decimal Data	TBD
Air Temperature		Essential	Field Sensors	Decimal Data	TBD
Global Solar Radiation	It's the power per unit area received from the Sun in the form of electromagnetic radiation in the wavelength range of the measuring instrument	Additional	Field Sensors	Decimal Data	TBD
Wind Speed and Direction		Essential	Field Sensors	Decimal Data	TBD
Soil Temperature		Additional	Field Sensors	Decimal Data	TBD
Soil Moisture	Measurement of the water in the large and intermediate size pores that can move about in the soil and be easily used by plants	Essential	Field Sensors	Decimal Data	TBD
Precipitation	Rainfall measurements	Essential	Field Sensors	Decimal Data	TBD
Infrared Surface Temperature	Temperature Surface calculated with infrared measurements	Essential	Field Sensors	Decimal Data	TBD

Impact	High impact on DSS provided to the pilots

Pilot Modifications	We decided to not purchase VHR data for the time being as the open data used currently yields more information about vineyards than expected at project start. We first want to concentrate on the exploitation of these information like water stress and phenological parameter derivation from dense open data time series.

2.3.2 Quantitative Evaluation Against KPIs

Domain Specific KPIs

ABACO has acquired the list of domain specific KPIs and their baseline values for the Farm Management Pilot from the IL Palazzo test site in Italy, which are presented in the following table.

Table 9: Farm Management Pilot Domain Specific KPIs

Variable	Definition	Units	2018 Baseline	2019	2020
Harvested Area	ha of harvested area	ha	29	34	34
Product Yield	Kg per ha of grapes harvested, wine produced, raisins produced	Kg/ha	100	90	80
Grape Product Quality			High	High	High
Production Costs	Costs in Euros per year	Euros/year	5500	5500	5600
Organic Fertilizer Use	Kg fertilizer used per kg grapes harvested per year	Kg/kg y	500	500	500
Organic Pesticides Use	Kg pesticides used per kg grapes harvested per year	Kg/kg y	18	23	30
Irrigation Cost	Euros per ha	Euros/ha	0	0	50
Fertilization Cost	Euros per kg	Euros/kg	400	500	650
Pesticides Cost	Euros per kg	Euros/kg	600	600	700
Labour Cost	Euros per hour	Euros/hr	2500	2600	2600

Technological KPIs

Additionally, in order to perform a complete quantitative evaluation for the Farm Management Pilot, a Technological KPIs list along with baseline values have been defined by ABACO and Geocledian.

Table 10: Farm Management Pilot Technological KPIs Catalogue

Variable	Definition	Units	2018 Baseline	2019 (Up to M18)	2020
Focusing Big Data					
Volume	Variation in raw data volume – Sentinel2	GB	130.5	139.5	127.5**
Volume	Variation in raw data volume – Landsat8	GB	54	57	60**
Volume	Variation in raw data volume – Pessl Instrumens	MB	2.5	3.3	-
Variety in data	Sentinel 2	Number of scenes	174	186	170**
Variety in data	Landsat 8	Number of scenes	54	57	60**
Variety in data	Pessl Instrumens	Hours	24h * 90 days	24h * 365 days	
Variety in Data Source Types	Data sources (Sentinel 2, Landsat 8, Pessl Instruments)	-	3	3	-
Variety inter Data	All variables measured (Satellite vegetation Indices, Soil data, Weather data, Canopy data)	Datasets	47	47+7 (new satellite indices)	-
Velocity	Speed of data generated – Sentinel 2	GB / month	10.88	11.63	11.59**
Velocity	Speed of data generated – Landsat 8	GB / month	4.50	4.75	5.45**
Velocity	Speed of data generated – Pessl Instrument	GB / month	0.25	0.3	-

**** measured until November 30, 2020**

2.4 NATURAL COSMETICS PILOT INDIVIDUAL EVALUATION (SYMBEEOSIS)

2.4.1 Qualitative Evaluation Summary

In order to report the Natural Cosmetics Pilot's progress, SYMBEEOSIS has completed the following table with the necessary information, regarding the current status of development, the successfulness of implementation, its impact and potential modifications.

<table>
<tr><th>SYMBEEOSIS</th><th>Natural Cosmetics Pilot Qualitative Evaluation Summary</th></tr>
<tr><td>Pilot Evaluation Summary</td><td>Specific Objectives
There is a need in extracting the most out of pharmaceutical plants for both economic and environmental reasons. A real challenge is to add high value to by-products. Wine making produces a lot of by-products that may have a significant biological value if there are adequate data concerning farm management. These data can lead to decisions concerning the processing of by-products in order to produce high added value active ingredients for cosmetics and food supplements.
The scenario hypothesis presumes that precision farming and control of parameters linked to the quality of wine may provide by-products of superior quality. In particular, the pilot intends to gather samples of vineyard by-products across the Greek territory and more specifically vine leaves of two different grape varieties (Agiorgitiko and Mandilaria) and test their phytochemical profile and biological value after extraction.
The Natural Cosmetics pilot is linked to the Prediction Use Cases and more specifically to B2. Predicting Biological Efficacy.

Achievements/Results
The preparation (i.e., maceration and ultrasound assisted extraction) of vine leaf extracts and testing of their biological efficacy for each sample took place at the laboratory of the collaborative to Symbeeosis Company APIVITA S.A. – Natural Cosmetics, located in Industrial Park of Markopoulo Mesogaias in Greece. At the laboratory conducted extractions under the two different methods and the following measurements of biological activity (BA): pH, RI, TPC, TFC, Total Microbial Count, Yeasts & Moulds, DPPH & ABTS assay. The collected data from the natural cosmetics pilot provided the necessary information for the evaluation of the quality of each sample, linked with the special characteristics of the vineyard of origin (SVIs data). Finally, the correlation analyses of BA parameters with the vineyards characteristics pointed out which information should be taken into account for building the models that will support the Decision Support System (DSS).

Problems/Challenges
The BA parameters data intended to be correlated with weather data but due to limited input it was preferred to do the correlation with satellite data for vegetation indices (SVIs).</td></tr>
<tr><td>Pilot Plan Progress</td><td>Start month: M1, duration: 36M
Partners involved: SYMBEEOSIS (BA parameters Data), Geocledian (SVIs datasets), CNR (data correlation analysis), Ontotext (data modelling), Agroknow (data management, transformation, uploading to the software stack), AUA (WP leader)
Third parties involved: APIVITA S.A. (laboratory analyses of samples)
Pilot’s objective: The collected data from the natural cosmetics pilot will provide the necessary information for the evaluation of the quality of each sample, linked with the special characteristics of the vineyard of origin. The goal is to face the challenge: “how data from the field can be linked to the biological efficacy of final products - an application on wine making by-products”.</td></tr>
</table>

- Progress of activities: The collected data from the natural cosmetics pilot provided the necessary information for the evaluation of the quality of each sample, linked with the special characteristics of the vineyard of origin (SVIs data). The correlation analyses of BA parameters with the vineyard's characteristics pointed out which information should be taken into account for building the models that will support the DSS.

Status of Implementation

Actors Involved

Natural Cosmetics Pilot (SYMBEEOSIS) and the BA data collected from samples all around Greece, with Geocledian providing the SVIs datasets, CNR for the data correlation analysis, Ontotext for data modeling, Agroknow for data management and their appropriate transformation for uploading to the software stack, and KU Leuven for the visualisation of the dashboard.

Methodology

A. Sample Collection

For the first year of the project, sixteen regions of the Greek territory have been chosen for sample collection, i.e., dried vine leaves of two different grape varieties (Agiorgitiko and Mandilaria). Also, samples of both grape varieties from the vineyard of Hellenic Agricultural Organization "DIMITRA" located in Attica will be tested. The dispersion and origin of the samples is shown in the following map, where the samples of Agiorgitiko are pictured in green and the samples of Mandilaria in red.

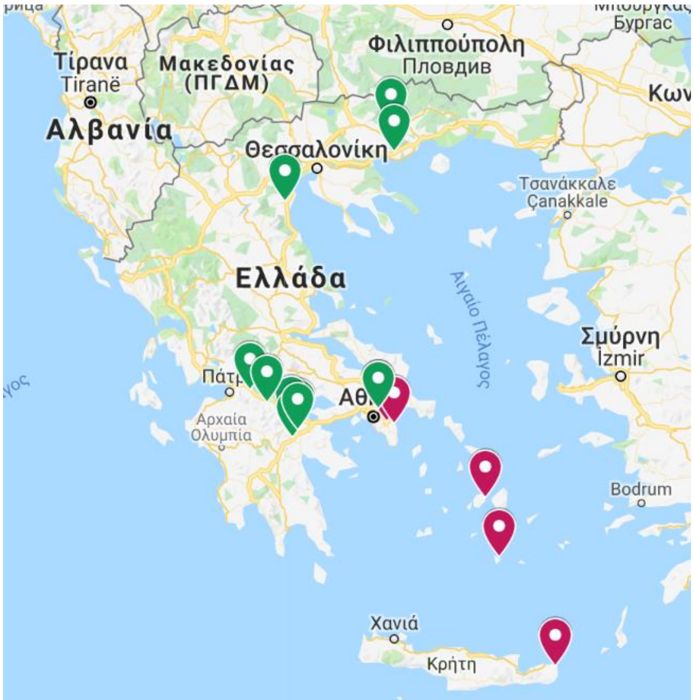


Figure 17: Dispersion of samples across the Greek territory

In the following table there is a list of the vineyards chosen for sample collection and their location.

Table 11: Vineyards chosen for sample collection

Vineyard	Grape Variety	Region	City

1	Semeli Wines	Agiorgitiko	Peloponnese	Nemea
2	Pavlidis Estate	Agiorgitiko	Northern Greece	Drama
3	RIRA Vineyards	Agiorgitiko	Peloponnese	Aigio
4	Vassaltis Vineyards	Mandilaria	Aegean	Santorini
5	Strofilia Estate Winery	Agiorgitiko	Peloponnese	Stimfalia
6	Papagiannoulis Winery	Agiorgitiko	Northern Greece	Katerini
7	Tetramythos Wines	Agiorgitiko	Peloponnese	Ano Diakopto
8	Skouras Domaine	Agiorgitiko	Peloponnese	Argos
9	Moraitis Winery	Mandilaria	Aegean	Paros
10	Toplou Winery	Mandilaria	Crete	Sitia
11	Aoton Winery	Mandilaria	Attica	Peania
12	Biblia Chora Estate	Agiorgitiko	Northern Greece	Kavala
13	Papagiannakos Domaine	Mandilaria	Attica	Markopoulo
14	Hellenic Agricultural Organization "DIMITRA"	Mandilaria	Attica	Lykovrisi
15	Hellenic Agricultural Organization "DIMITRA"	Agiorgitiko	Attica	Lykovrisi
16	Palivou Estate	Agiorgitiko	Peloponnese	Nemea

B. Laboratory testing

The preparation of vine leaf extracts and testing of biological efficacy of each sample will take place at the laboratory of collaborating Company APIVITA S.A. – Natural Cosmetics, located in Industrial Park of Markopoulo Mesogaias in Greece.

Figure 18: Collaborating Company's (APIVITA) laboratory

Deployed Components

- Extractions will be conducted using Elma S60H Elmasonic Ultrasonic Bath.

Figure 19: Elma S60H Elmasonic

- The measurement of pH will be conducted with a seven compact pH meter, METTLER-TOLEDO.

Figure 20: pHmeter, METTLER-TOLEDO

- The measurement of refractive index will be conducted with a Digital Refractometer RX-a- series ATAGO

Figure 21: Digital Refractometer ATAGO

- A NUVE Incubator and a Laminar Telstar BO-II-A will be used for the measurement of total microbial count with classic development of micro-organism in petri-dishes.

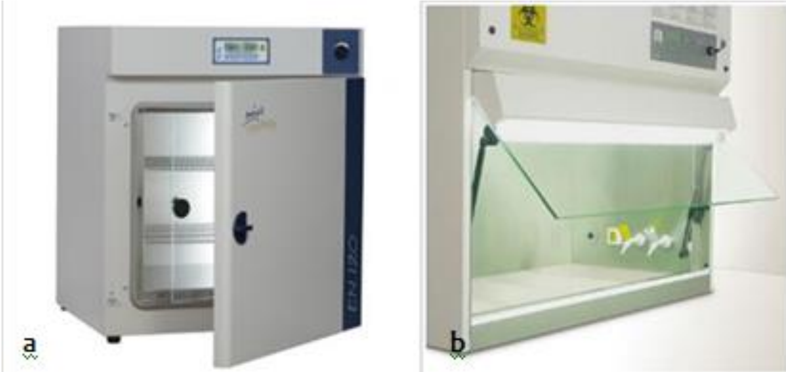


Figure 22: (a) NUVE Incubator, (b) Laminar Telstar BO-II-A

- A Memmert will be used for the measurement of yeasts and moulds with classic development of micro-organism in petri dishes.

Figure 23: Memmert Universal Oven 055 UN/UNm

- A UV 1800 SPECTROPHOTOMETER, SHIMADZU EUROPA will be used for the measurement of antioxidant activity (DPPH & ABTS assay), total phenolic content and total flavonoid content.

Figure 24: UV Spectrophotometer

A Nanoquant, infinite M200 Pro, TECAN will be used for additional measurements antioxidant activity and total phenolic and flavonoid content.

Figure 25: Nanoquant, infinite M_{200} Pro

Table 12: Laboratory components; Testing frequency one sampling per year; Analysis duration 3 months

Name	Experimentation Controlled Variables Description
Elma S 60 H Elmasonic Ultrasonic Bath	Ultrasonic bath for extraction
SevenCompact pH meter METTLER-TOLEDO	Bench pH-meter for pH measurement
Digital Refractometer RX-a- series ATAGO	Digital Refractometer for measurement of Refractive Index
NÜVE EN 055 Incubator	Incubation chamber for classic development of bacteria
Memmert Universal Oven 055 UN/UNm	Incubation chamber for classic development of yeast & fungi
Laminar Telstar BO-II-Advance	Class II Biological Safety laminar air flow cabinet for microbiological analyses
Elma S 60 H Elmasonic Ultrasonic Bath	Ultrasonic bath for extraction

Gathered Data and Formats

Table 13: Natural Cosmetics Pilot Data and Datasets

Name	DataSet Description	Priority	Provenance	Data Type Format	Data size
SVIs Data	Sentinel-2A/B MSI spectral bands, vegetation indexes (NDVI, NDRE1, NDRE2, NDRE3 NDWI, SAVI, EVI2, CIRE, NCPRI)	Essential	Copernicus EO Programme, ESA	json, geotiff, png	TB
Agiorgitiko Samples UAE (11 samples)	Data on biological efficacy of samples of Agiorgitiko dried vine leaves, developed with Ultrasound Assisted Extraction	Essential	Laboratory testing	csv, xls	MB
Agiorgitiko Samples MAC (11 samples)	Data on biological efficacy of samples of Agiorgitiko dried vine leaves, developed with Maceration	Essential	Laboratory testing	csv, xls	MB
Mandilaria Samples UAE (5 samples)	Data on biological efficacy of samples of Mandilaria dried vine leaves, developed with Ultrasound Assisted Extraction	Essential	Laboratory testing	csv, xls	MB
Mandilaria Samples MAC (5 samples)	Data on biological efficacy of samples of Mandilaria dried vine leaves, developed with Maceration	Essential	Laboratory testing	csv, xls	MB
Weather Data	Weather data on the regions selected for sample gathering	Essential	Open source data	csv, xls	GB

Impact	The collected data from the natural cosmetics pilot provided the necessary information for the evaluation of the quality of each sample, linked with the special characteristics of the vineyard of origin (SVIs data). The correlation analyses of BA parameters with the vineyards' characteristics and weather data pointed out which information should be taken into account for building the models that will support the DSS.
Pilot Modifications	The BA parameters data intended to be correlated with weather data but due to limited input it was preferred to do the correlation with satellite data for vegetation indices (SVIs). Another shortcoming was the relatively low by-products samples collected for BA analysis, which is scheduled to be overcame by the next seasons sampling.

2.4.2 Quantitative Evaluation Against KPIs

Domain Specific KPIs

SYMBEEOSIS has generated the list of domain specific KPIs for the Natural Cosmetics Pilot and has defined their baseline values, which are presented in the following table.

Table 14: Natural Cosmetics Pilot Domain Specific KPIs Catalogue

Variable	Definition	Units	2018 Baseline	2019	2020
Agiorgitiko Samples/ parcel	Number of samples per vineyard (parcel)	Number	1	1	1
Mandilaria Samples/ parcel	Number of samples per vineyard (parcel)	Number	1	1	1

Agiorgitiko Samples	Samples of vine leaves to be analysed	Number	16	16	16
Mandilaria Samples	Samples of vine leaves to be analysed	Number	16	16	16
UAE and MAC efficiency	Percentage of extract from incoming raw material	%	>60	>60	>60
Extract pH	Ranges for acceptable pH	pH	> 3.5	> 3.5	> 3.5
Extract RI	Ranges for acceptable % for RI	%	22±4	22±4	22±4
Extract TMC	Ranges for acceptable Total Microbial Count	CFU	< 10	< 10	< 10
Extract Y&M	Ranges for acceptable Yeats and Moulds counts	CFU	< 10	< 10	< 10
Processing Time	Overall time for extraction, required analysis, and assessment of new product	Months	3	3	3

Technological KPIs

Additionally, in order to perform a complete quantitative evaluation for the Natural Cosmetics Pilot, a Technological KPIs list along with baseline values have been defined by SYMBEEOSIS.

Table 15: Natural Cosmetics Pilot Technological KPIs Catalogue

Variable	Definition	Units	2018 Baseline	2019	2020
		Focusing Big Data			
SVIs Volume Data	Sentinel-2 A/B MSI visible & NIR bands, NDVI time series	GB	301.5	294.0	316.5**
SVIs Volume Data	Landsat 8 A/B MSI visible & NIR bands, NDVI time series	GB	160.0	163.0	143.0**
Agiorgitiko and Mandilaria Samples UAE BA parameters Volume	Data on biological efficacy of samples of Agiorgitiko and Mandilaria dried vine leaves, developed with Ultrasound Assisted Extraction	KB	58	81	120
Agiorgitiko and Mandilaria Samples MAC parameters Volume	Data on biological efficacy of samples of Agiorgitiko and Mandilaria dried vine leaves,	KB	58	81	120

	developed with Maceration				
Variety in Data Source Types	BA parameters, SVIs	Data sources	8	8	8
Variety in Data	Number of different types of data (in different resolutions)	Datasets	6	6	6
SVIs Velocity Data	Sentinel-2 A/B MSI visible & NIR bands, NDVI time series	GB/month	25.13	24.50	28.77**
SVIs Velocity Data	Landsat 8 A/B MSI visible & NIR bands, NDVI time series	GB/month	13.33	13.48	13.00**
BA Parameters Velocity	Speed of data generated – BA Parameters	KB/season	58	23	39
Weather Data Velocity	Speed of data generated - WD	MB/seasons	17.6	17.5	17.4
Big Data Process Metrics					
Data Normalization (Homogenization)	Time needed for data to be available for analysis and processing	Months	3	3	3

** measured until November 30, 2020

2.5 FOOD PROTECTION PILOT INDIVIDUAL EVALUATION (AGROKNOW)

2.5.1 Qualitative Evaluation Summary

In order to report the Food Protection's progress, Agroknow has completed the following table with the necessary information, regarding the current status of development, the successfulness of implementation, its impact and potential modifications.

AGROKNOW	Food Protection Pilot Qualitative Evaluation Summary
Pilot Evaluation Summary	*Specific Objectives* Food protection, including safety and fraud, is one of the most critical parameters in food production highly affecting the food companies from the financial and brand point of view. Agroknow is providing a digital solution for the food industry that delivers trends and risk estimation for raw materials, ingredients and finished products. This solution, FOODAKAI, is helping the Quality Assurance (QA) and Food Safety (FS) experts working in the food industry to identify risk in their supply chain. The current solution is limited to alarms, statistics, simple trends and search mechanisms. Agroknow has performed a series of focused group and consultation meetings with several companies of the food industry, such as Gallo Winery, Conagra,

Campbell, Pepsico, Hershey and Lamb Weston. The meetings were held during large food safety events like the GMA Science Forum. During these meetings Agroknow team validated the need for new FOODAKAI extensions that will enable risk predictions in the supply chain.

Thus, the main objective of this pilot is to enhance the current digital solution with new modules that will address the need for risk prediction in 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.

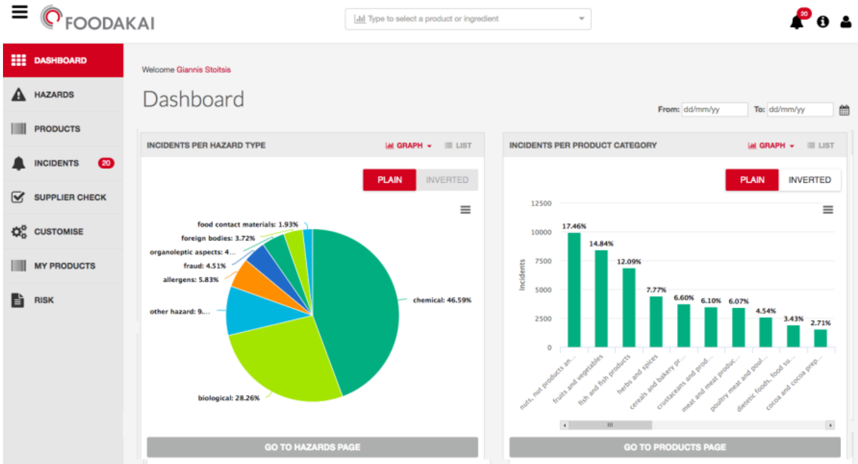


Figure 26: Food Protection dashboard of FOODAKAI system

The specific goals of the food protection pilot are:

- To develop a software module that will be able to predict emerging and increasing risks for chemical hazards in the grapes and wines supply chain.
- To develop a price prediction dashboard that will include algorithms able to predict the prices of agricultural products, including grapes.
- To develop a food fraud dashboard that will help experts working in the food industry to perform an effective vulnerability assessment for products, focusing also on recall predictions.
- To develop a marketing automation module that will facilitate the exploitation of the food safety and fraud solutions that will be developed by Agroknow in the context of the project.

Achievements/Results

The pilot has achieved the aforementioned objectives, achieving integration of the aforementioned components into FOODAKAI. These are available through the risk assessment and recall prediction dashboards.

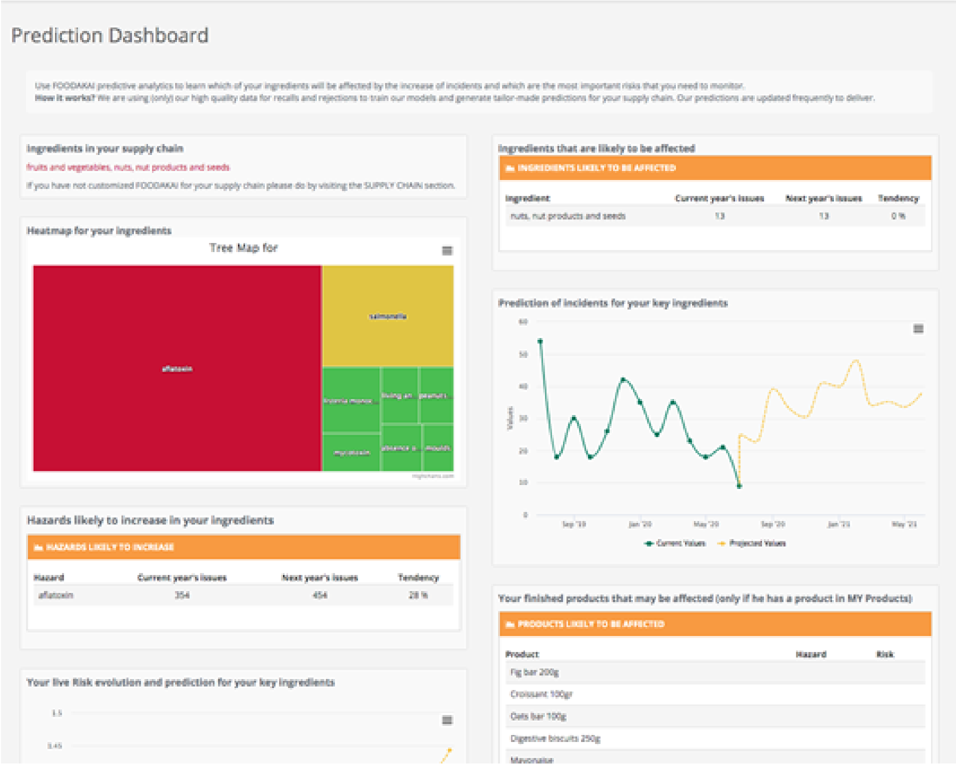


Figure 27: The FOODAKAI prediction dashboard

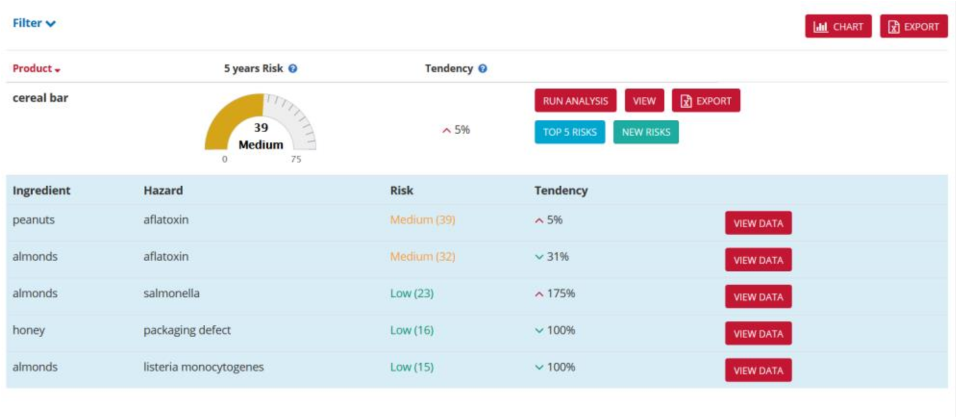


Figure 28: The FOODAKAI risk assessment dashboard

Pilot Plan Progress

- Start month: M17, duration: 19M
- Partners involved: Agroknow (Data, data correlation analysis, data modelling, data management, transformation, uploading to the software stack).
- Pilot's objective: The objective of this pilot was to evaluate the integrated risk prediction and assessment dashboards in the FOODAKAI platform as well as assess the effectiveness of the price prediction visualizations.
- Progress of activities: The pilot has successfully completed with the evaluation activities, involving its target users.

Status of Implementation

Methodology

A. Big Data Collection & Processing

During this step all the datasets that are needed to develop the prediction services have been identified. The following data types have been integrated into the Big Data Platform

- **Food recalls and border rejections** that include incidents for hazards and fraud identified in raw materials, ingredients and food products
- **Pricing data** for the agricultural products
- **Surveillance studies data** about hazards and fraud in food products on the market
- **Laboratory testing data** of the food companies and producers
- **Country risk and corruption index** data

After collecting the data, we applied big data processing techniques available at the BDG software stack to filter, classify and enrich the data. Enrichment included the automatic annotation with terms from multilingual semantic vocabularies for products and hazards.

B. Big Data Analysis

Based on the real industry scenarios we identified the main parameters that need to be predicted. Based on these findings we tested several machine learning and deep learning algorithms to predict parameters like the chemical risk, pricing and fraud. The goal was to combine different datasets to achieve the optimum performance for the prediction of the risk in raw materials (e.g. grapes) and finished products (wine).

C. Prediction Dashboards

Interview and focus group sessions with the end users from industry have been organized to validate the digital services that were to developed and integrated in the FOODAKAI system. Mockups of the final services will be developed and will be validated with the end users. Based on the final version of the mockups, Agroknow developed the functionalities and the user interface. The risk and recall prediction dashboards were then integrated in the FOODAKAI system and evaluated by the end users.

D. Marketing automation

By applying the BDG technologies we aimed to produce food safety and fraud insights that can be used to develop powerful marketing content. We focused on the development of two processes

a. A process to enable the semi-automatic creation of data reports for specific product categories e.g., grapes-based products.
b. A process for personalized marketing messages that will be based on the data reports and on the target prospect profile.

The two processes are part of a generic marketing workflow that take into account the profile of the end user. Algorithms that classify the targeting users in specific categories were defined and developed. Automated mechanisms for sending the data-powered emails have been designed and developed.

Gathered Data and Formats

A large number of different data sources and data types have beenwill be used to enable the predictions. We will used textual information and numerical data. Textual information includes mainly announcements about food recalls and border rejections whereas numerical data includes lab test results and pricing data. The main source of our datasets is the open data published by the governments and the private data for lab testing that will be provided by the companies.

Table 16: Food Protection Pilot Data and Datasets

Name	DataSet Description	Priority	Provenance	Data Type Format	Data size
Lab testing data	Results of the laboratory testing that the food companies perform to identify the presence of hazards	Essential	Laboratory testing	xsl, csv	GB
Pricing data	Pricing data published by Statistical Authorities both at local	Essential	Open Data published by Authorities	xsl, csv, API	GB
	and global level				
UK Food Standards Agency	Food recalls published by the UK Food Standards Agency	Essential	Open Data published by Authorities	Html	MB
RASFF – Rapid Alert System for Food and Feed	Food recalls and border rejections information published by EU	Essential	Open Data published by Authorities	XML, Html	MB
Food Standards Australia New Zealand	Food recalls and border rejections information published by the Asutralian Gov	Essential	Open Data published by Authorities	Html	MB
FDA Recalls, Market Withdrawals, & Safety Alerts, warning letters, import refusals and inspection citations	Food recalls and border rejections information published by the United States Gov	Essential	Open Data published by Authorities	OpenFDA Food API, html, xsl, json, xml	GB
EFET - Hellenic Food Safety Authority	Food recalls published by the National Food Safety Authority of UK	Essential	Open Data published by Authorities	Pdf, Html	MB
Japanese Imported Foods Inspection Services	Food recalls and border rejections information published by the Japanese Gov	Essential	Open Data published by Authorities	Html, xsl	MB

	Czech Agriculture and Food Inspection Authority	Food recalls published by the National Food Safety Authority of Czech Republic	Essential	Open Data published by Authorities	Html	MB
	Healthy Canadians food alert information website	Food recalls published by the Canadian National Food Safety Authority	Essential	Open Data published by Authorities	Html	MB
	Food Safety Authority of Ireland	Food recalls published by the Irish National Food Safety Authority	Essential	Open Data published by Authorities	Html	MB
	German Food Safety: warnings and information to the public	Food recalls published by the German National Food Safety Authority	Essential	Open Data published by Authorities	Html	MB
	Hong-Kong-Center for Food Safety	Food recalls published by the Irish National Food Safety Authority	Essential	Open Data published by Authorities	Html	MB
	Open Food Facts	Crowd sourced open data about food products and their composition	Essential	Open Data published by Companies	Html	MB
	ProMED-mail	Dataset about foodborne outbreaks, animal and plant diseases with global coverage	Essential	Open Data published by ProMED-email	Html	MB
Impact	Risk assessment is a very critical part of a food safety system in order to prevent food safety incidents in the supply chain, in general, and the wine supply chain in particular. Today, Quality Assurance and Safety Experts that are working in food companies are using risk estimation approaches that are based on static data such as literature and guidelines published by National Authorities. Such risk estimation approaches are not taking into account the emerging and increasing risks of the global supply chain and cannot predict the risk. This results in several serious food safety incidents that may impact public health, can cause large financial loss for farmers and industries and can damage their “brand” and lose customers. This Demonstrator is providing a digital solution for the food industry that delivers trends and risk estimation for raw materials, ingredients and finished products. This solution is helping the Quality					

	Assurance (QA) and Food Safety (FS) experts working in the food industry to identify risk in their supply chain to work more effectively
Pilot Modifications	The Food Protection pilot was an integral addition to the BigDataGrapes piloting sessions.

2.5.2 Quantitative Evaluation Against KPIs

Domain Specific KPIs

For the cost reduction we developed a Return-on-Investment calculator and we used it to estimate the cost reduction in collaboration with the food companies that are using the FOODAKAI risk estimation and prediction. Agroknow has generated the list of domain specific KPIs for the Food Protection Pilot and has defined their baseline values, which are presented in the following table.

Table 17: Food Protection Pilot Domain Specific KPIs Catalogue

Variable	Definition	Units	2018	2019 Baseline	2020
Cost reduction	Reduce cost of running risk estimation, including travelling costs	Euros/year	N/A	0.5M	0.25M
Productivity increase	Reduce the time that is needed to perform risk estimation and prediction	Hours	N/A	2 months every year	1 month per year (50% reduction)

Technological KPIs

Additionally, in order to perform a complete quantitative evaluation for the Food Protection Pilot, a Technological KPIs list along with baseline values have been defined by Agroknow.

Table 18: Food Protection Pilot Technological KPIs Catalogue – Lab Data

Variable	Definition	Units	2018 Baseline	2019	2020
Focusing Big Data					
Volume Data	The size of all the data that are stored in the Big Data Platform	GB	3	100	117.6
Variety in Data Source Types	The number of the data sources from which we are collecting information about food safety incidents	Data sources	10*	29	32

Variety in Data	The different types of data that we are processing in Big Data Platform	Datasets	1	1	1
Velocity Data	The growth of all the data types in the Big Data Platform	MB/month	9.9	8,000	1,460
Velocity Data	The growth of all the data types in the Big Data Platform	GB/month	0.099	8.08	1.46
Big Data Process Metrics					
Data Normalization (Homogenization)	Time needed for data to be available for analysis and processing	Months	0.1	2.5	0.4

Table 19: Food Protection Pilot Technological KPIs Catalogue –Food recalls and border rejections

Variable	Definition	Units	2018 Baseline	2019	2020
Focusing Big Data					
Volume Data	The size of all the data that are stored in the Big Data Platform	GB	2.32	2.55	2.8
Variety in Data Source Types	The number of the data sources from which we are collecting information about food safety incidents	Data sources	45*	45*	45*
Variety in Data	The different types of data that we are processing in Big Data Platform	Data Types	1	1	1
Velocity Data	The growth of all the data types in the Big Data Platform	MB/month	19.1	20.1	14.4
Big Data Process Metrics					
Data Normalization (Homogenization)	Time needed for data to be available for analysis and processing	Months	0.5	0.7	0.85

Table 20: Food Protection Pilot Technological KPIs Catalogue – Price data

Variable	Definition	Units	2018 Baseline	2019	2020
Focusing Big Data					
Volume Data	The size of all the data that are stored in the Big Data Platform	GB	2.27	2.5	2.68
Variety in Data Source Types	The number of the data sources from which we are collecting information about food safety incidents	Data sources	3	3	5
Variety in Data	The different types of data that we are processing in Big Data Platform	Data Types	1	1	1
Velocity Data	The growth of all the data types in the Big Data Platform	MB/month	19.3	14.75	15
Data Normalization (Homogenization)	Time needed for data to be available for analysis and processing	Months	0.45	0.65	0.75

3 BIGDATAGRAPES PILOT'S SURVEY

Each BigDataGrapes partner had to complete the Pilots' Survey as a part of the Human-centred Evaluation report. Based on the Experimental and Evaluation Protocols and the Use Case Scenarios, as they have been defined and identified, partners provided feedback on the execution of the application pilots in order to qualitatively evaluate their results (user satisfaction and involvement).

This part describes a user-centred assessment for the BigDataGrapes pilots. It includes surveys, divided in two levels of assessment, the "**Pilot Basis Evaluation**" and the "**Internal Technological Evaluation**", of heterogeneous sets of quantitative and qualitative indicators, metrics used to measure the effectiveness of the pilots. In order for this to be completed, an iterative approach of assessment was performed according to the proposed three-phase human-centred assessment activities. More specifically, the distribution of the survey took place a few months after the completion of the Intermediate phase (M30), which involves the first round of controlled pilot trials and implementations of the first versions of the newly developed BigDataGrapes components, and before the end of the Summative phase (M34), which entails the validation of the BigDataGrapes components in real-life conditions and with realistic complexity.

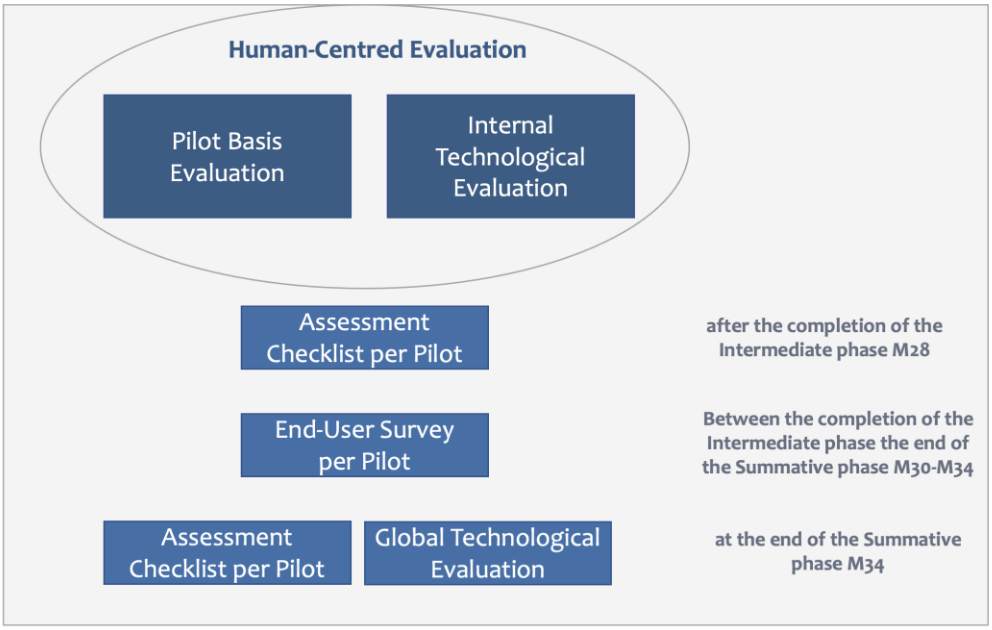


Figure 29: Three-phase iterative evaluation approach

Following the concept of gradual extension of functionality, intended audience and assessment of this scheme, the pilots interacted with the community and the pilot evaluators accordingly. Thus, this survey was distributed to all relevant stakeholders involved in the BigDataGrapes piloting activities. Feedback was asked from the end-user, with a focus on the "**Industry End-User**", including farmers, producers, owners of the reference vineyards, winemakers and other representatives from the food and cosmetic industry, as well as from the agricultural technological innovation industry and the BigDataGrapes pilot and technological partners. Finally, a small number of agronomists and researchers were asked to perform the evaluation, for additional feedback.

Table 21: BigDataGrapes Pilots' Survey Structure

BigDataGrapes Pilots' Survey	Section (Question Numbers)	Assessment Group	Designated Phase
Pilot Basis Evaluation	**Vineyard Information and Demographics** (Q1 – Q10)	End-user	The Evaluation was completed: After the completion of the Intermediate phase (M30) and before the end of the Summative phase (M34)
	End-User Survey (Q11 – Q42)	End-user	
Internal Technological Evaluation	**Assessment Checklist** (Q1 – Q5)	BDG pilot and tech partner	
	Global Technological Evaluation (Q6 – Q16)	BDG pilot and tech partner	

3.1 PILOTS BASIS EVALUATION

The BDG consortium has selected and defined nine (9) main Software Demonstration scenarios, reflecting the work that has been done in the five pilots, to focus on, fine-tune and showcase. These scenarios were selected to show how different software tools and components produced by the BDG project, together with the critical business decisions to be supported, relevant data and data sources, intelligence and data competence questions to be answered, and algorithm implementations, may support industrial end-users and other key stakeholders in the grapevine-powered industry in new innovative ways. The selected demonstrators and the corresponding partners responsible for their development, are:

1. **Table and Wine Grapes Pilot Software Demonstrator (AUA):**
   Grapevine responses to terroir

Based on sensor, farming and phenological data derived from all test sites located in Greece the user will be able to visualise different correlations between them and explain what affects grape quality and yield. Intelligence and data competence questions on how some attributes, such as soil properties, affect grape quality and yield. Data integration have been achieved using web semantic and ontologies to have all data connected in a knowledge graph. This Software Demonstrator has the potential to be used as an extension of the Farm Management Software SITI4farmer. Initial feedback will be retrieved from our industry end-users in order to improve the proposed BDG tools and components. The next step is to correlate the aforementioned sensor data with earth observation data to examine the effectiveness of applying machine learning techniques.

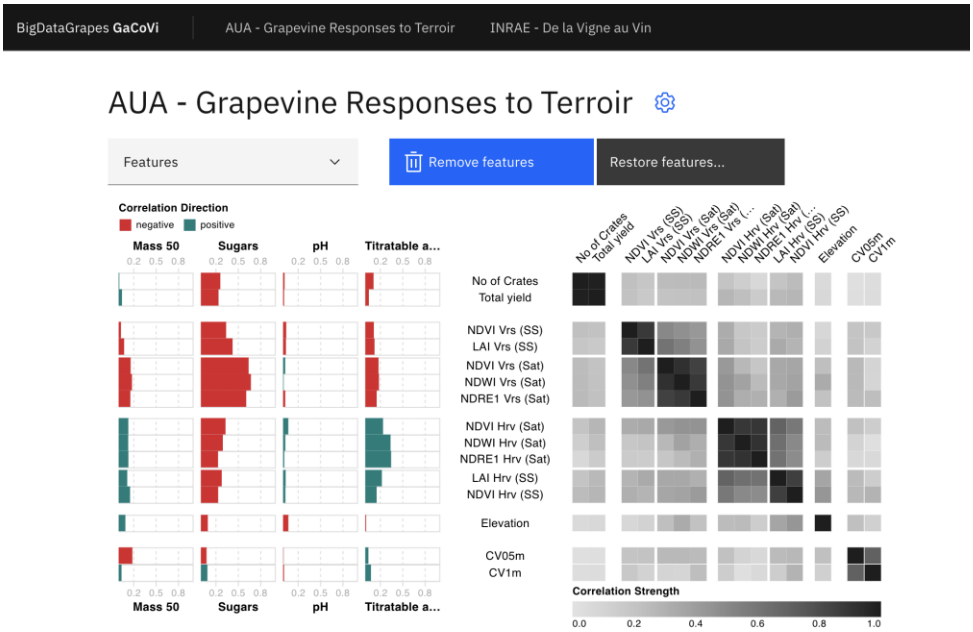


Figure 30: "Grapevine responses to terroir" demonstrator's visualisation

2. **Winemaking Pilot Software Demonstrators (INRAe)-prioritised**:

a. Counting Grapevine Leaves

Based on an existing large number of grapevine images derived from the PhenoArch platform, a machine-learning pipeline was developed, aiming at counting leaves from side-view grapevine images. The counting pipeline exploits deep learning techniques based on artificial neural networks to infer the number of leaves from each grapevine image. This scenario supports the counting and visualisation of the evolution of the existing grapevine leaves on a plant over time. The leaf count will enable to feed agronomic models used in decision support systems in agriculture. In addition, in precision agriculture image analysis can enable for site-specific application in the vineyard to reduce input and maximize profit. This Software Demonstrator has the potential to be used as an extension of the Farm Management Software SITI4farmer and also in OpenSilex software. Initial feedback will be retrieved from our research and industry end-users in order to improve the proposed BDG tools and components.

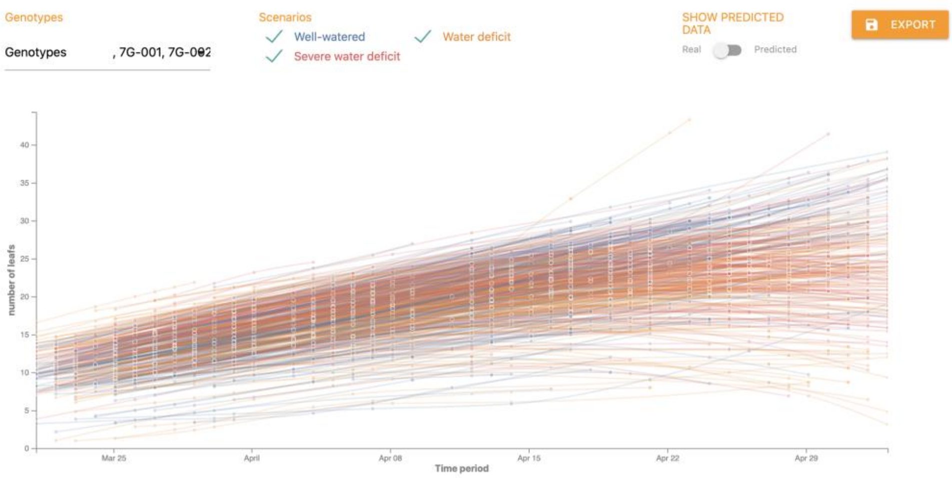


Figure 31: "Counting grapevine leaves" demonstrator's visualisation

b. From Vine to Wine: Parameters' Influencing Wine Quality

Based on existing large & heterogeneous data provided by different data sources gathered by diverse teams at each step of wine production, including climatic, soil, harvest and winemaking data, sensory and lab analysis, the user can visualise the influence of all of these parameters on the wine quality in response to a changing environment. In order to support discovery, access & visualisation of these linked data, data integration have been achieved using web semantic and ontologies to have all data connected in a knowledge graph. This Software Demonstrator has the potential to be used in OpenSilex software. Initial feedback will be retrieved from our industry end-users in order to improve the proposed BDG tools and components.

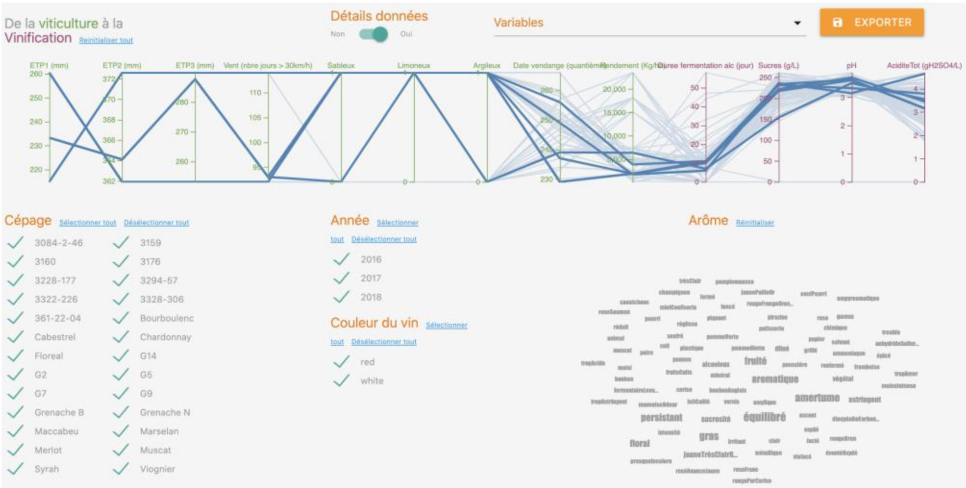


Figure 32: "From Vine to Wine" demonstrator's visualisation

c. Gacovi "From Vine to Wine"

Based on vine and wine data collected from the experimental unit of Pech Rouge, the user will be able to know the parameters of interest to obtain a target wine. The data available consist of field and climatic data related to plots, satellite data and lab analysis of must and wine which are quality data. The users can visualise existing associations and correlations between these data. They can know to what extent the field data is correlated with each other and influences the quality of the wine. Then, they can identify variables with weak or strong predictive power on a chosen quality variable. In order to support this tool, data integration and modelling have been achieved to have all data connected and linked. The next step was the calculation of correlations among parameters. This Software Demonstrator has the potential to be used as an extension of the Farm Management Software SITI4farmer and also in OpenSilex software. Initial feedback will be retrieved from our industry end-users in order to improve the proposed BDG tools and components.

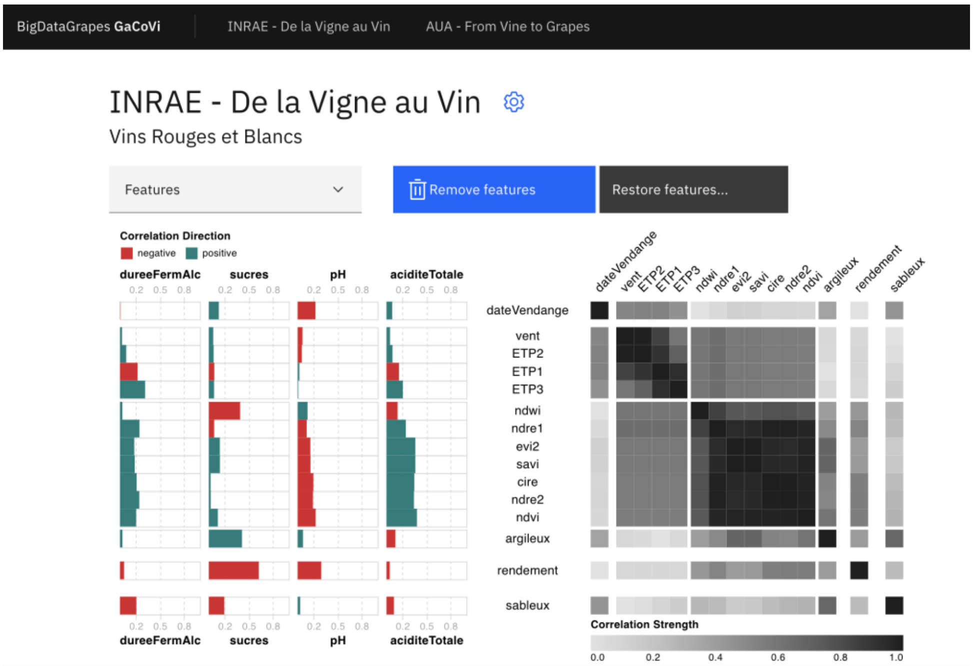


Figure 33: Gacovi "From Vine to Wine" demonstrator's visualisation

3. **<u>Farm Management Pilot Software Demonstrator (ABACO-GEOCLEDIAN)</u>: Water Availability and Irrigation Recommendations**

The software demonstrator aims to support optimization of the Irrigation, one of the best practices that has a high impact economically on food chain industry. As a matter of fact, the water consumption in irrigation is connected to the knowledge of the real needs of the crops also in relation to final production quality and quantity. For this reason, the Decision Support System outputs should be evaluated by expert agronomist and crop technicians. The DSS, where farmers are provided of complete weather stations, as in piloting farms are fed by data gauged exactly in the farm area. For most of other farmers and demonstrators the meteorological input data come from Regional Weather Networks that provides data for free in raster format. This data format allows to obtain meteorological data with a resolution of 1 km. Others important input of the DSS are the Irrigation input inserted in SITI4Farmer module by the users and the satellite data products from Geocledian, especially the water stress index. As the UI Dashboard mock-ups will be available, we expect to collect feedback for the demonstrator in terms of usability and reliability of the DSS outputs.

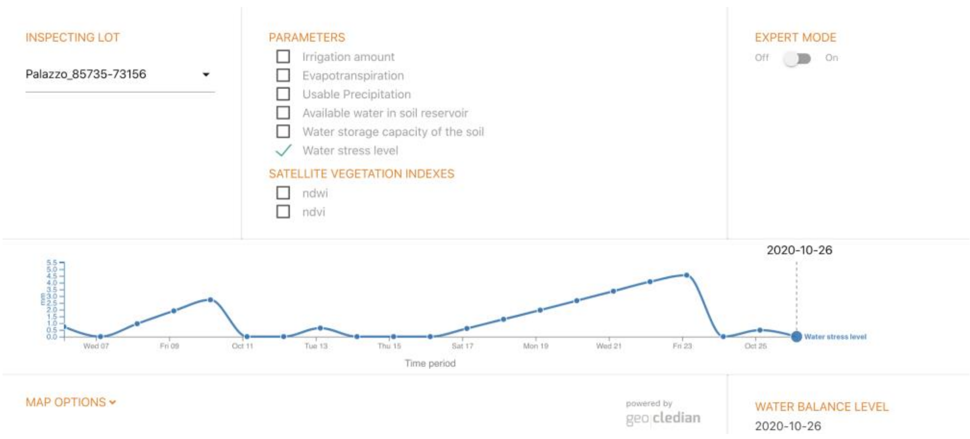


Figure 34: "Water Availability and Irrigation Recommendations" demonstrator's visualisation

4. **Natural Cosmetics Pilot Software Demonstrator (SYMBEEOSIS):**
   Grapevine By-Products Biological Efficacy Predictor

The Natural Cosmetics Pilot's Software Demonstrator will showcase the relative dashboard dedicated to grapevine practitioners and cosmetic industry end-users. The software will be visualised into a dashboard equipped with the appropriate tools that will support end-user on decision making and selection of the best grapevine by-products intended for natural cosmetic production. The present version of the software will be focused on grapevine leaves' biological efficacy from 16 Greek vineyards (private/industry data) which after correlation with weather and satellite vegetation indices data-sets (public/open data) will aim to predict the origin of vineyard that could supply the best quality leaves for next year's natural cosmetics production. The demonstrator will integrate intelligence data derived from satellites and meteorological stations for the targeted vineyards, laboratory analyses on the biological activity of two different type extracts from leaves, and finally will help end-user to answer competence questions related with incoming raw material quality, vineyards/samples' origin, correlation attributes between biological properties of sample and vineyard performance, and many other. In addition, efforts will be made to get the software linked with FOODAKAI platform, in order to incorporate food safety information on grapevine by-products as well. The performance of this first version of software demonstrator will be assessed and appraised from relative end-users selected mainly from cosmetic industry (8), but also from research organisations (2) and grapevine practitioners (1). The initial feedback from testing will be composed by usefulness of the selected modules, ease of handling, successfulness of implementation, and visualisation lay-out, while will provide suggestions on improvements that could be employed to next versions of demonstrator.

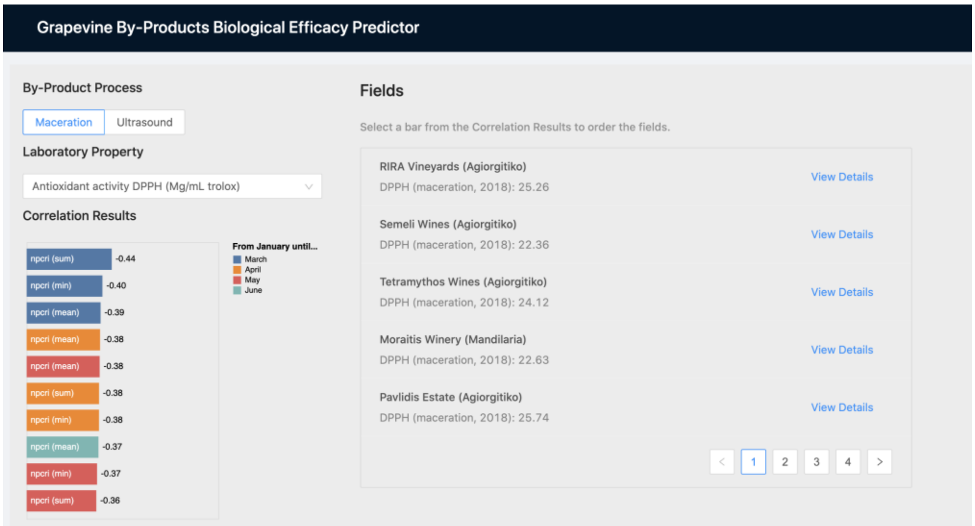


Figure 35: "Grapevine By-Products Biological Efficacy Predictor" demonstrator's visualisation

5. **Food Protection Pilot Software Demonstrator (AGROKNOW):**
Risk Assessment Module, Price Prediction Dashboard and Recall Prediction

Based on real industry cases, food protection pilot identify risk in food products in order to prevent issues (recall/border rejections) in a food supply chain. For this reason, it is critical to answer these question "which is the most important chemical risk for a food product", "Which are the ingredient (eg grape) with the higher risk", "Which are the countries with the higher risk for hazards and fraud", "Which are the increasing risk for my product (raisin)". A large number of different data sources and data types has been used. We used textual information that includes mainly announcements about food recalls and border rejections (FDA, RASAFF, FSSA, AUSTERLIAN FOOD SAFETY AUTHORITY). The main source of our datasets is the open data published by the governments. Moreover, private data that includes ingredients that a company uses for food production, list of suppliers and internal lab test. Risk assessment module, price prediction dashboard and recall prediction are developed to extend FOODAKAI. From industry stakeholders we needed feedback relevant to these questions "How critical is to have access to food recall and border rejections data", "How important is to access my supplier based on global food safety data ","How critical is to identify risk in my food products", "How useful was alerts module, supplier check and risk estimation"

A series of training sessions (29) were organised during the last three months of the pilot implementation, in order to introduce the Prediction Dashboard to food companies and validate it.

Table 22: Prediction Dashboard training sessions

Company Name	Number of training sessions
Unilever	2
PepsiCo	1
Conagra Brands, Inc.	2
Schreiber Foods	7
Haribo	2
Yili	1
Coca-Cola European Partners	2
Danone	1
Perfetti	1
Hershey	1
Mars	1
AB InBev	1
ABP	2
Buhler AG	1
WorldAware, Inc.	1
Zaragoza Logistics Center	1
Creme Global	1
André Simonazzi	1
Total training sessions on predictions	**29**

During the training session we presented to the participants how they can use the predictions dashboard to:
1) identify the hazards that they need to include in theirs testing plan by knowing the increasing risk trends for all their ingredients;

2) identify products that will be affected by the emerging and increasing risks;
3) identify the sourcing countries with the highest predicted risk for their ingredients.

Moreover, **a dedicated online workshop** was also carried out on the 8th of December,2020 with 13 members of the Conagra company. Conagra Brands, Inc. (founded in 1919) is an American packaged foods company headquartered in Chicago, Illinois. It is an approximately $8 billion company that combines a rich heritage of making great food with a sharpened focus and entrepreneurial spirit. Conagra makes and sells products under various brand names that are available in supermarkets, restaurants, and food service establishments.

Throughout the workshop we presented live and in real time the Prediction Dashboard we implement for the Food Protection Pilot and we had the change to ask specific questions to the participants about all the above functionalities. The next figures present the summary of the answers we received.

Figure 36: Prediction Dashboard suppliers coverage

Figure 37: Prediction Dashboard sources coverage

Figure 38: Prediction Dashboard timesaving

.Finally, to better understand our audience and have some critical insights about the pains that the prediction dashboard is going to relieve, we asked the participants about the value proposition they see from the adoption of a tool like this. It is clear that the majority of the participants have already identified the need to move to a more a proactive approach and use the right tools which will give them the ability to predict risks before they occur. The full set of the results are depicted in the following figure.

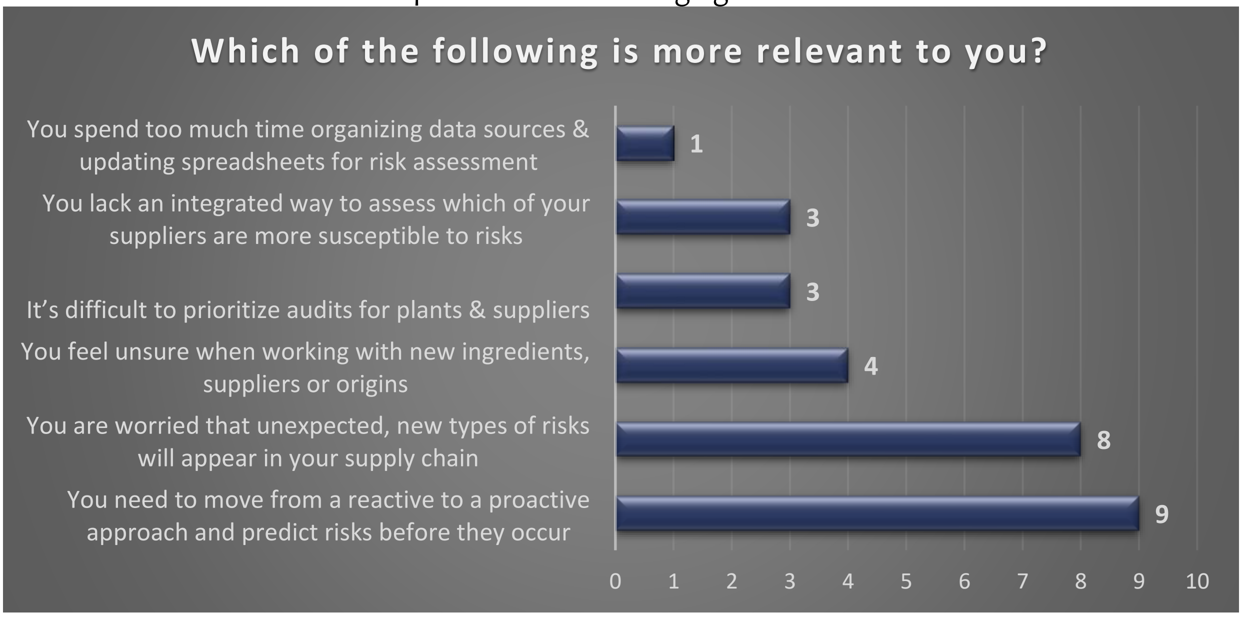


Figure 39: Prediction Dashboard relevancy

After this process the last version of the prediction tools was of the food protection pilot.

Price Prediction

Figure 40: "Price prediction" demonstrator's visualisation

Risk Assessment

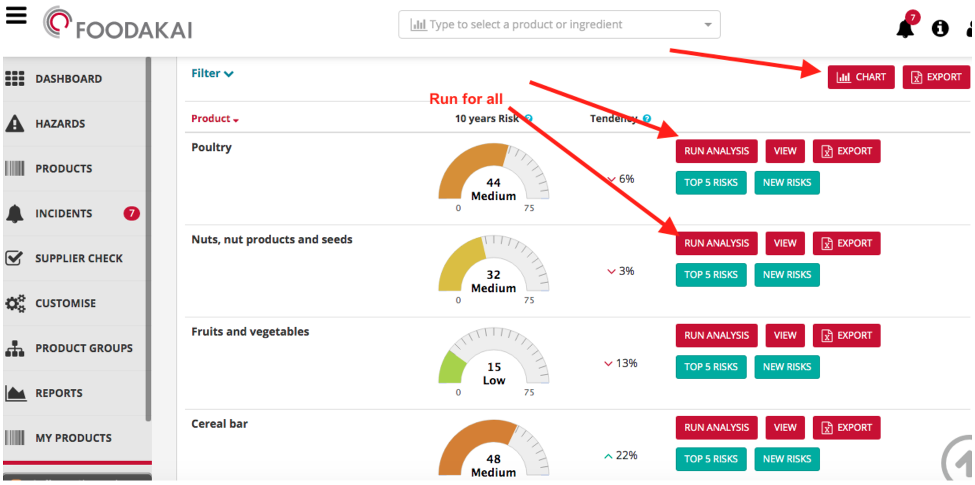


Figure 41: "Risk assessment" demonstrator's visualisation

Recall Prediction

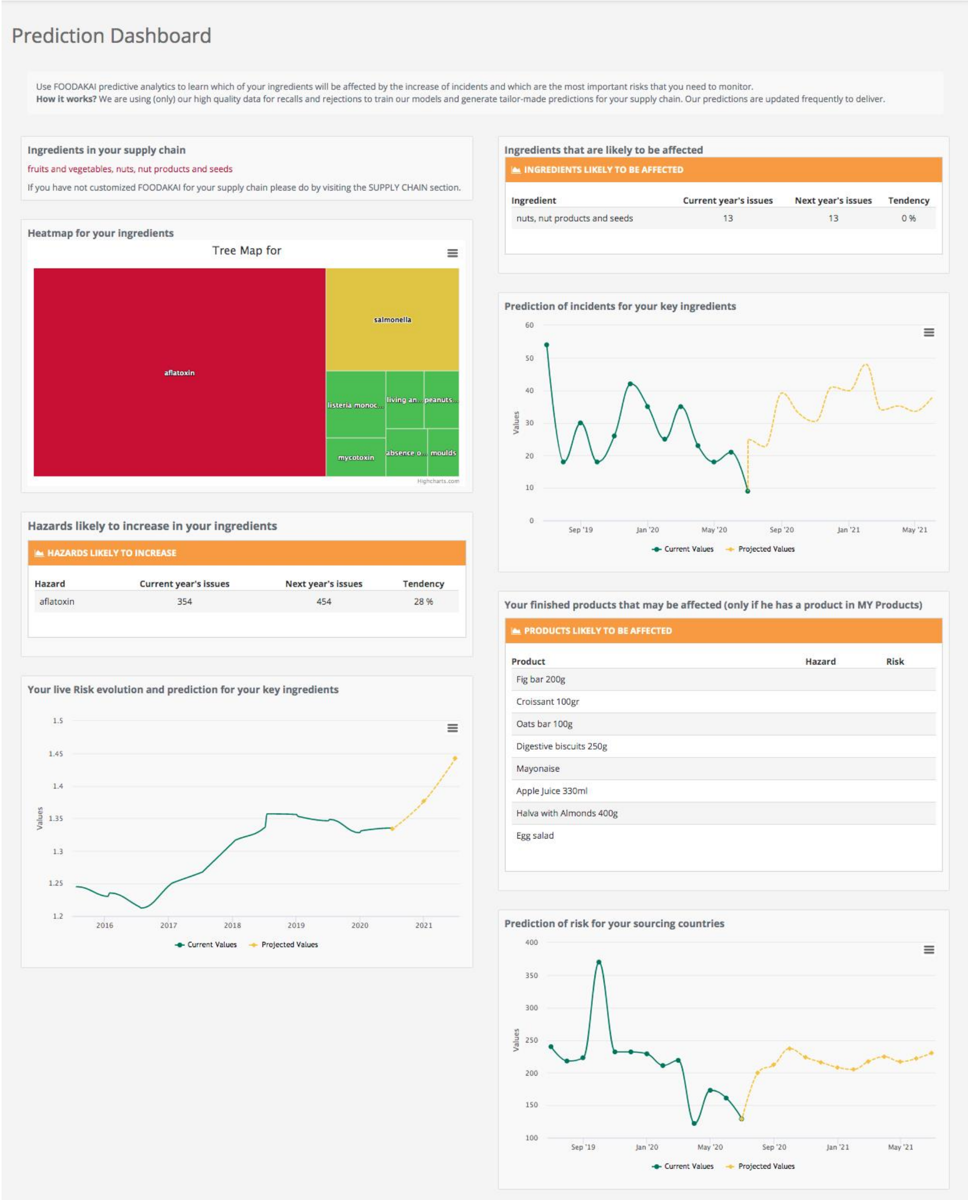


Figure 42: "Recall prediction" demonstrator's visualisation

Each pilot presented one or more of these Software Demonstrators to end-users coming from relevant to the grapevine-powered industry communities, in order to collect feedback using appropriate evaluation instruments that has a solid scientific basis on the trust, acceptance and adoption of information systems. Specifically, we used the UTAUT (unified theory of acceptance and use of technology) model (Venkatesh et al., 2003) to understand potential acceptance and adoption, the System Usability Scale (SUS) (Brooke, 1996) to measure usability and NASA-Task Load Index (TLX) (Hart, 1986) to measure workload of the proposed tools. Additionally, to measure end-user trust towards the decision support elements, we used custom questionnaires, based on Jian et al., 2000, in relevant pilots.

3.1.1 Vineyard Information and Demographics

The first part of the "Pilot Basis Evaluation" is an introductory section, the "Vineyard Information and Demographics" that incorporates questions related to the vineyard of interest, final product and end-user of each pilot. This part of the survey was intended to be completed by the end-users.

3.1.2 End-user Survey

In order to document how grapevine-powered business problems correspond to (big) data challenges, a set of detailed Use Case specifications, highlighting the data challenges that decision makers face, have been identified under WP2 - Grapevine-Powered Industries Big Data Challenges. The "Pilot Basis Evaluation" includes the "End-user survey" that is directly related to these Use Cases and their Demonstrator Scenarios, covering Data Anomaly Detection and Classification, Prediction and Farm Management. Metrics, such as, ease of use (usability), usefulness, efficiency, effort, satisfaction and intend to use in real-life, are assessed in this section. The first distribution of the survey to the end-users took place a few months after the completion of the Intermediate phase (M30-M34), which involves the first round of controlled pilot trials and implementations of the first versions of the newly developed BigDataGrapes components.

The "Pilot Basis Evaluation" included (but not be limited to) indicators from the following table:

Table 23: Indicators for Pilot Basis Evaluation for Assessment Group "End-user", after the completion of the Intermediate phase and at the end of the Summative phase.

Indicator	Examples
Usability (SUS)	
Ease of Use (Usability)	Measurement of an overall usability score using System Usability Scale (SUS) – 10 item Likert scale questionnaires with 5 option
Determinants of usage intention and behaviour (UTAUT)	
Performance expectancy	The degree to which an individual believes that using the system will help him or her to attain gains in job performance.
Effort expectancy	The degree of ease associated with the use of the system.
Social influence	The degree to which an individual perceives that important others believe he or she should use the new system.
Facilitating conditions	The degree to which an individual believes that an organisational and technical infrastructure exists to support use of the system.
Workload (TLX)	
NASA Task Load Index	Measurement of an overall workload score based on a weighted average of ratings on six subscales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration
Trust	
Trust in Systems	Custom 7-point Likert scale questions based on [4] which was constructed to measure trust in automated systems.

3.2 INTERNAL TECHNOLOGICAL EVALUATION

The "Internal Technological Evaluation" focuses on the evaluation of the technical solutions proposed within the pilots, regarding their effectiveness. Real-world settings allow showcasing and evaluating the BigDataGrapes platform and components in the context of specific end-user requirements from different areas. The developed BigDataGrapes methods and tools go beyond the state-of-the-art in Big Data management, processing and leveraging data value. Therefore, this section includes questions on Data Management, Data Analytics and Processing, Data Modelling and Semantics and Data Visualisation and User Interaction.

The "Internal Technological Evaluation" section includes the "**Assessment Checklist**" section, which is directed to both BDG pilot and technological partners and the "**Global Technological Evaluation**".

The first distribution of the "Assessment Checklist" survey to the BDG partners took place a few months after the completion of the Intermediate phase (M28, during the 5th virtual PMB), which involves the first round of controlled pilot trials and implementations of the first versions of the newly developed BigDataGrapes components. The second distribution happened two months before the end of the Summative phase (M34), which entails the validation of the BigDataGrapes components in real-life conditions and with realistic complexity. During this last evaluation session, the "Global Technological Evaluation" also took place in order to examine further some important technological indicators.

3.2.1 Assessment Checklist

The "Assessment Checklist", was to provide an overall evaluation of the technological progress of each of the five pilots' Software Demonstrators. The "Assessment Checklist" was conducted by the BDG pilot and technological partners and also contributed to the development and improvement of the following:

Table 24: Assessment Checklist after the completion of the Intermediate phase and at the end of the Summative phase.

Assessment Checklist	
Evaluation regarding	Example Questions
Data and Datasets	- Have the data sets and data sources required been identified, collected & shared? - Which data are private and which public? - Have those data been uploaded to the BDG platform? - Is there a way to update those data? - Are those data available on the website? - Are there projected data? - Is the data modelling completed?
Business Decisions and Competence Questions	- Which are the critical business decisions to be supported? - What are the Intelligence and data competence questions to be answered?
Algorithm Implementations	- Is the model algorithm completed? - Is the algorithm that fits the best identified? - Have numerical experiments been run?

	- Is the model algorithm integrated in the platform?
Software Tool Extension and UI	- Are the visualization components' mockups designed and developed? - Are the visualisations completed? - Is the visualization component integrated in the platform? - Is this Software Demonstrator an extension of SITI4farmer or FOODAKAI? Or stands alone?
Evaluation and Dissemination	- Have the end-users been identified? - Have initial feedback been collected from the end-users? - Has a content article describing the business case of the pilot been produced?

3.2.2 Global Technological Evaluation

The questions in the "Global Technological Evaluation" are formed in such a way in order to examine the FAIR-ness, scalability, resource optimization, flexibility, consistency, reliability, conformity and reusability within the pilots. Additionally, the size of the datasets that the pilots enable to deal with is another quantitative indicator.

- Resource management: management of data and tools.
- Data access: access and integration of data from various data sources with a focus on semantic issues.
- Machine learning: access to machine-learning approaches and also include support for modern machine-learning approaches like ensemble techniques (boosting, bagging and random forests) and deep learning.
- Flexibility, extensibility and openness: Integration of open-source data and libraries to the piloting activities.
- Data exploration and visualization: to provide interactive visualization.
- User interface: BigDataGrapes must provide coherent "look and feel" and support for the visual components for the application pilots.
- Reusability: the BigDataGrapes technology components should support reusability.

The "Global Technological Evaluation" will be conducted by the BDG technological partners.

Table 25: Indicators for the Internal Technological Evaluation for Assessment Group "BDG pilot and tech partner", at the end of the Summative phase.

Global Technological Evaluation	
Indicator	Example Questions
FAIR-ness	Do the BigDataGrapes pilots help in making research data and algorithms FAIR (Findable, Accessible, Interoperable, Reusable)? Do they support access and integrate data from various data

	sources and of different types (textual, SQL, RDF, images, location data, etc.)? Are there advantages compared to current research environments and data management practices?
Scalability	Do the BigDataGrapes pilots allow for improved scalability in data?
Resource Optimisation	Do the BigDataGrapes pilots enable management capabilities and resource optimisation (such as for security, compute resource management, governance and reuse)?
Flexibility	Do the BigDataGrapes pilots permit the addition of new data and enhanced functionality? Do they promote the transfer of knowledge by sharing components and data? Do they promote flexibility by the integration of open-source data and libraries?
Consistency	Are the BigDataGrapes pilots consistent and integrated in order to support an entire data analytics pipeline? Do they provide a seamless end-to-end experience, to make users more productive across the whole data and analytics pipeline?
Reliability	Are the BigDataGrapes components and tools developed for the pilots reliable enough for day-to-day use?
Conformity	Are results from the pilots conforming to user expectations and quality standards?
Reusability	Do the BigDataGrapes technology components support reusability?

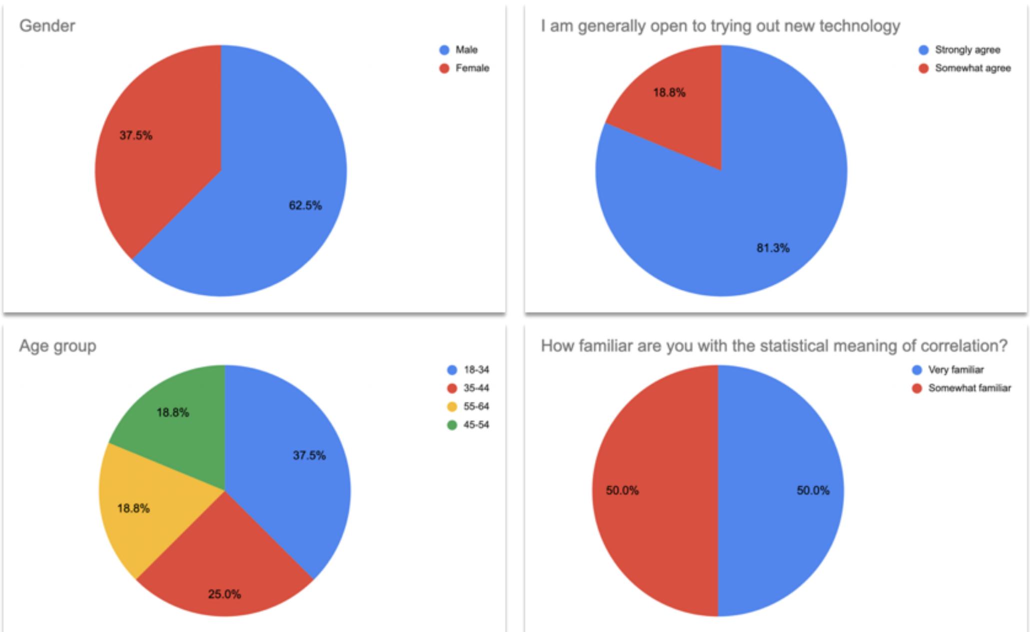
4 HUMAN-CENTRED EVALUATION RESULTS

4.1 RESULTS ON PILOTS BASIS EVALUATION

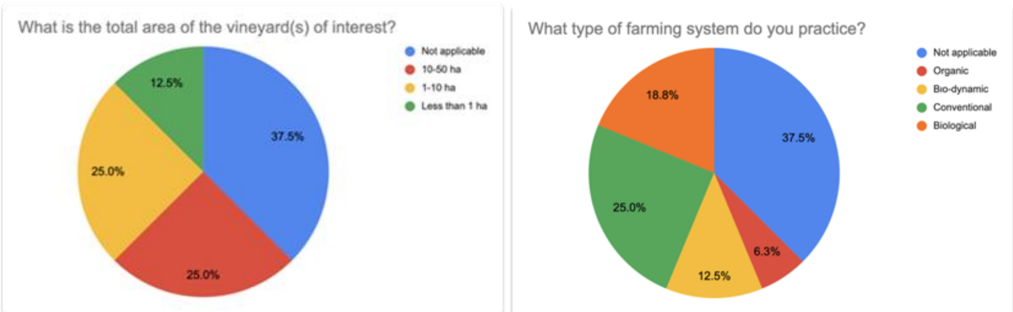
4.1.1 Table and Wine Grape Pilot Vineyard Information and Participant Demographics

Participant Demographics

Sixteen people participated in the Evaluation of the Table and Wine Grape Pilot, representing all aspects of the grapevine industry: vine growers, wine makers, agronomists, oenologists, as well as representatives from the research sector and the industry. The majority of the participants (62.5%) were male, of postgraduate education ageing from 18-44 years old. All participants proved to be open to try out new technology and quite familiar with the statistical meaning of correlation. Among the participants who owned a vineyard, they all had less than 50 ha of land, while all types of farming systems (conventional, organic, biodynamic) were among the answers provided.



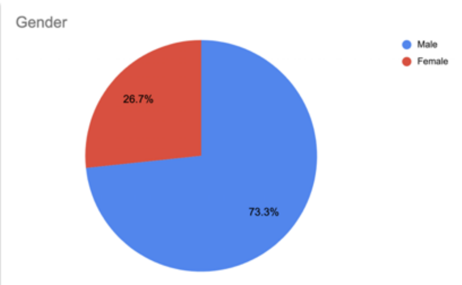
Vineyard Information

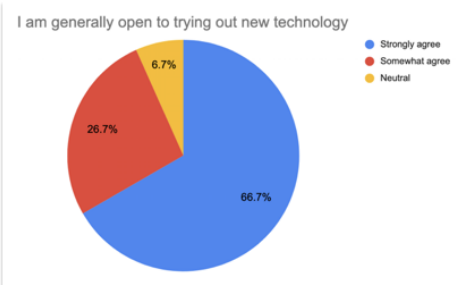


4.1.2 Wine Making Pilot Vineyard Information and Participant Demographics

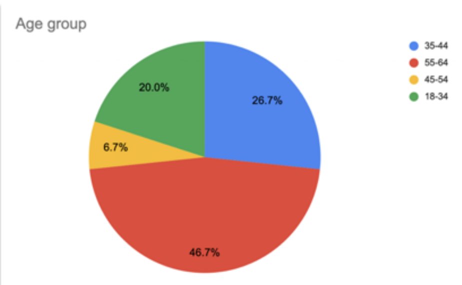
Participant Demographics

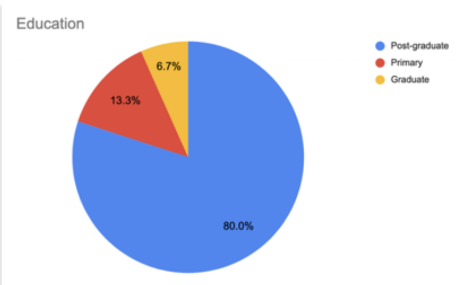
In terms of participants, a large panel of jobs from the wine industry were represented: vine growers, a head of a cooperative winery, engineers from R&D departments or consulting, etc. Moreover, researchers working on viticulture and winemaking in diverse research units were also doing the evaluation. First of all, most end users were male, postgraduate and almost half of them were between 55 and 64 years old. They were open to try new technology. This result is not surprising because they accepted our invite to this trial which was focusing on IT tools for professionals. Secondly, participants who owned their vineyard had a surface between 10 and 50ha conducted in a conventional farming system. For almost 50%, the question related to vineyard information is not applicable as people working in companies, in consulting or in the field of research do not have their own vineyard.



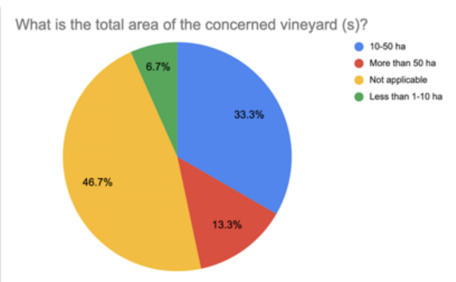


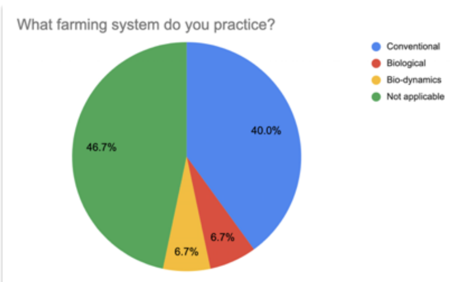
Name of organisation and role
INRAE
INRAE OEnologist
Wine-growing estate :« Domaine de la Baume », GCF Group / R&D manager
INRAE engineer
IFV – intern in data management and sensory analysis INRAE
...
R&D project manager
IFV, Engineer
INRAE
INRAE
INRAE
OPERATOR
SCA LES 3 GRAPPES, LE POUGET, PRESIDENT (head of the cooperative winery) Winegrower





Vineyard Information



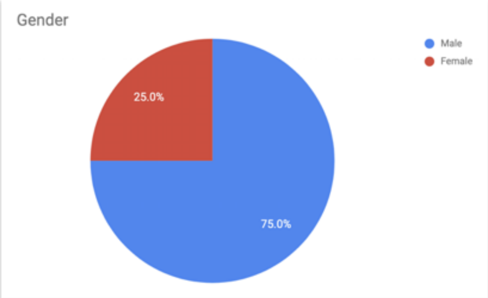


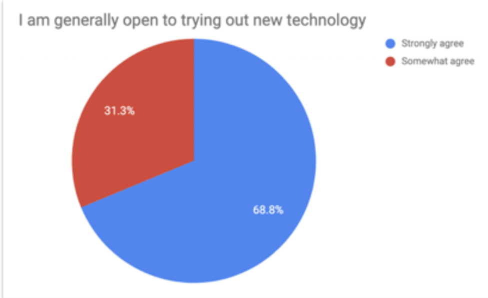
4.1.3 Farm Management Pilot Vineyard Information and Participant Demographics

Participant Demographics

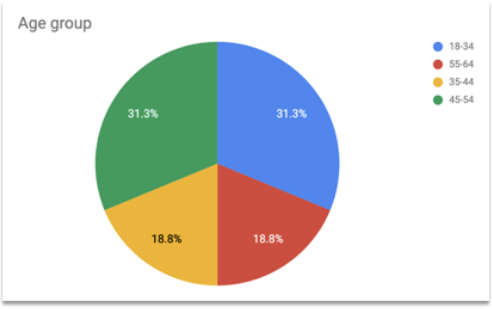
The participant demographics has been represented by wine makers agronomist and farmers, crop science university researchers and also software engineers. In particular the age of the participants in the majority of cases were male, below 45 years showing a particular predisposition to the new technologies. Participants were in the majority of cases graduated in crop science faculties and for over the 30 % not owners of vineyards.

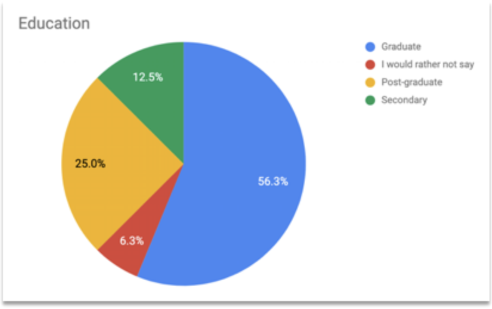
Among those owners, the vineyards surfaces were under 30 hectares, with conventional management. There were also agronomist working not only on vineyards but also with corn, tomato and wheat farms.



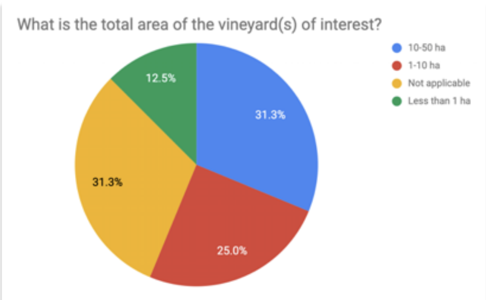


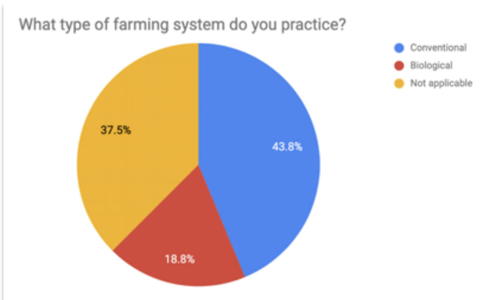
Name of organisation and role
agronomo
Agronomo consulente
Agronomo
Az.agr. Donatella Cinelli Colombini responsabile vigneti
Università
Non possiedo un'azienda
Cantine torrevilla
Cantine torrevilla
Cantine Torrevilla
Unimi
Università degli Studi di Milano. Researcher Crop Science Dept
Università degli Studi di Milano - DISAA
Torrevilla
Consorzio Agrario di Cremona
Abaco Agri-Food Department - Business Analyst
Abaco Factory-Tech Deparment - software architect





Vineyard Information



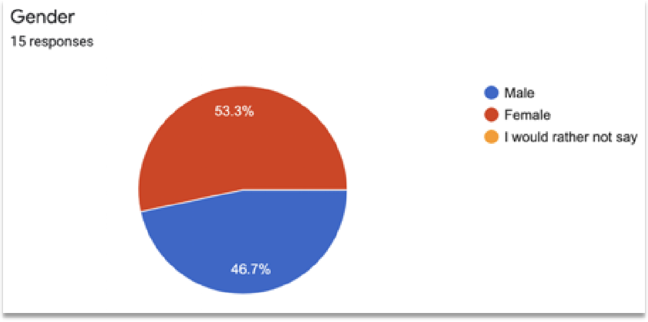


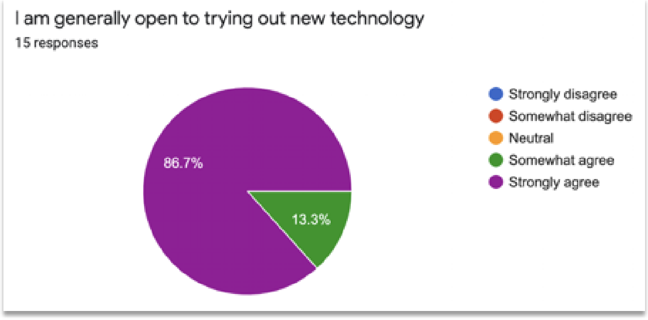
4.1.4 Natural Cosmetics Pilot Vineyard Information and Participant Demographics

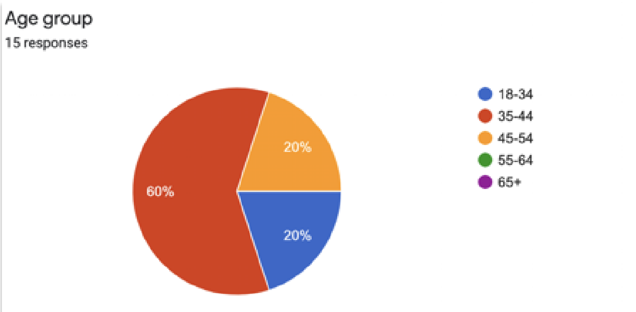
Participant Demographics

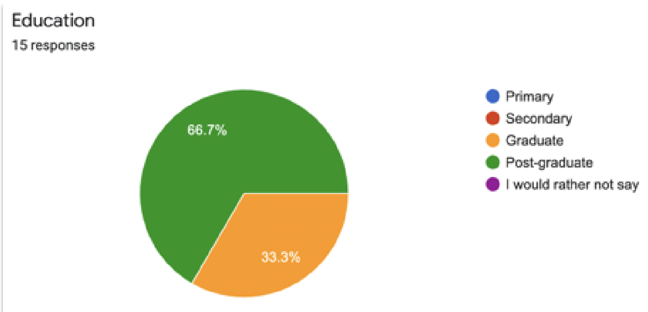
The participant individuals to the Evaluation Process run were composed by potential end-users of the Demonstrator of Natural Cosmetic Pilot, “Grapevine By-Products Biological Efficacy Predictor”. All end users were selected based on the fact that during their everyday practices are facing the competence questions on which the Pilot was developed, and thus can assess the advantages of such a DSS, dashboard’s visualisation and handling. A total of 15 end-users, with 10 of them being natural cosmetic industry’s end-users, 2 were researchers engaged with grapevine related disciplines, and 1 was a grapevine grower of a winery, have successfully completed the evaluation of the Demonstrator. Regarding the demographic of the sample, one third were graduated, while two thirds had postgraduate degrees, gender was equally distributed between male and female, all of them were younger than 54 years old with the majority between 35 to 44 years old, and all participants were open to try new technology. For the demonstrator assessment 93.3% stated that they would use the system frequently, 86.7% did not find the system complex, and 100% found it easy to handle. Although 26.7% of the sample would need the help of a technician and would need to learn a lot of things to use the platform, finally a 93.4% would learn to handle it very quickly and 86.6% felt very confident using it. All participants found the system easy to use and learn to operate it, with a clear and understandable interaction with it, and without feeling the need of additional knowledge or resources to successfully operate it.

Encouraging were also the findings that 80% found the system useful for their work, 74.4% could accomplish their tasks more easily, and 66.6% could increase their productivity, with it.



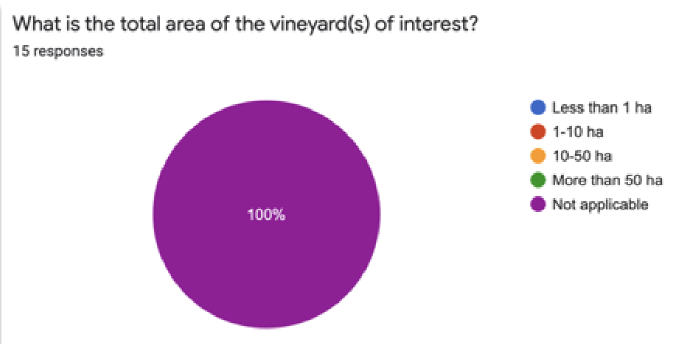


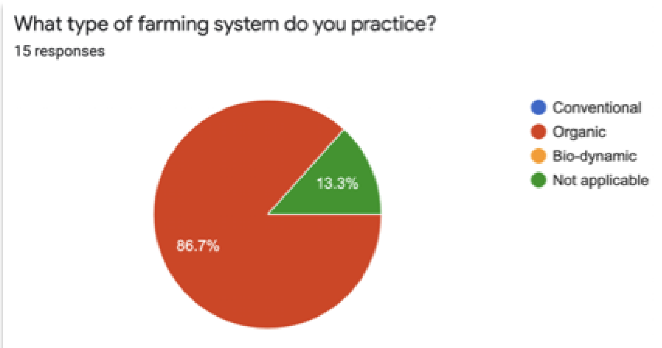




Name of organisation and role
APIVITA - RD DIRECTOR
SYMBEEOSIS - COO
QS Professional (Cosmetics) - Quality control
R&D MANAGER FREZYDERM S.A.
Frezyderm S.A Plant Manager
GR. Sarantis S.A. - Brand Manager
Hellenic Agricultural Organization/Researcher
The NuClab
The Nuclab, rnd cosmetologist
APIVITA, R&D & Extracts Development Scientist
APIVITA S.A. R&D Lab Scientist
Aoton Winery, Owner
Agricultural University of Athens; Lab of Food Microbiology & Biotechnology, Associate Researcher
The name of organisation is SYMBEEOSIS Eu Ζην SA and my role is "Raw Materials Procurement and Sustainable Sourcing Manager"
Sales associate/Cosmetic scientist at Beautylab the Store

Vineyard Information

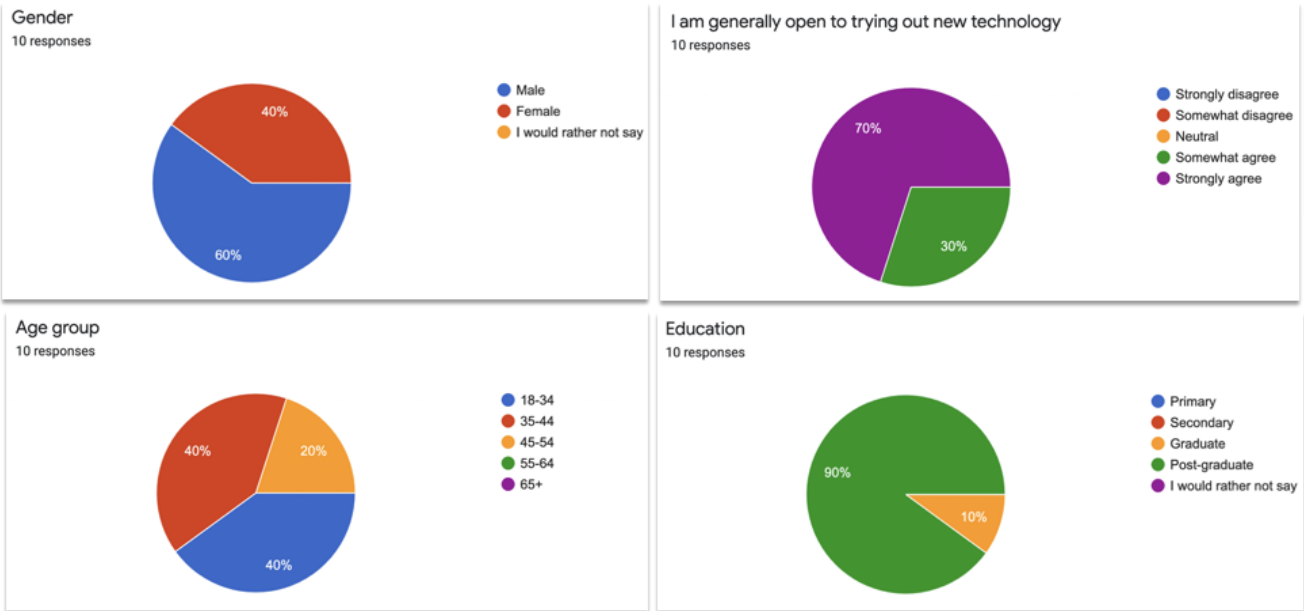




4.1.5 Food Protection Pilot Vineyard Information and Participant Demographics

Participant Demographics

The participants in the Food protection pilot, ten in total, were recruited through invitations to a variety of organizations so as to ensure the perspective of different types of experts, including food science and quality assurance experts (3), food scientists and researchers (5), as well as business and marketing professionals familiar to prediction processes as part of their work (2). Six men and four women participated in the evaluation, of diverse age groups, from 18 - 64 and all had reached a level of post-graduate education, except one. All participants reported interest and openness to try new technologies.



Name of organisation and role
Certification Body
Agricultural University of Athens, Assistant Professor
AUA - Research Associate
TUV AUSTRIA HELLAS - Food Safety Auditor
Cool Bananas Corporation, Marketing Manager
INRAE- pilot leader for BDG project
AUA
National Institute of Research and chemico-physicals analysis Tunisia, Ph.D student
Hellenic Agricultural Organisation - Institute of Technology of Agricultural Products
Teaching and Research Staff at Laboratory of Viticulture (AUA)

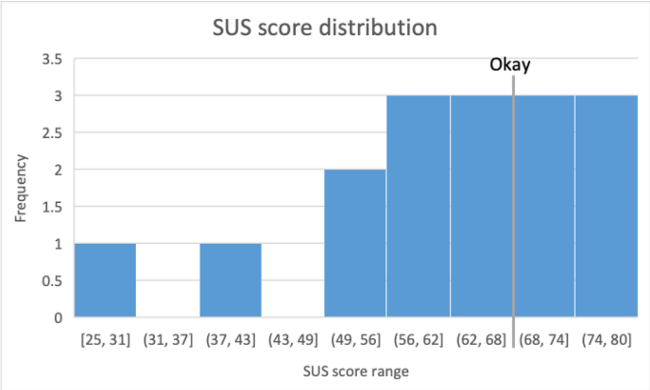
4.2 SYSTEM USABILITY SCALE (SUS)

In the following subsections, we present the results of the system usability scale questionnaire. Results of the remaining metrics measured as part of the evaluation will be disseminated in a scientific paper. At the time of writing, a detailed analysis of the results is being conducted, which will be submitted to the journal of Computers and Electronics in Agriculture for publication.

4.2.1 Table and Wine Grape Pilot - Correlation Task

The interface developed for the correlation task of the table and wine grape pilot can be found in Figure 30. The figure below shows results of the SUS questionnaire. A median of 66.3 puts the interface in a slightly lower percentile from the "Good" region of SUS scale, which is between 68 and 80.3. However, with the standard deviation of 14.7, we conclude that the SUS score does not significantly deviate from the acceptable region.

Regardless, further analyses using the participants' expertise and background as factors will help us gain a better understanding of this.



Median 66.3

SD 14.7

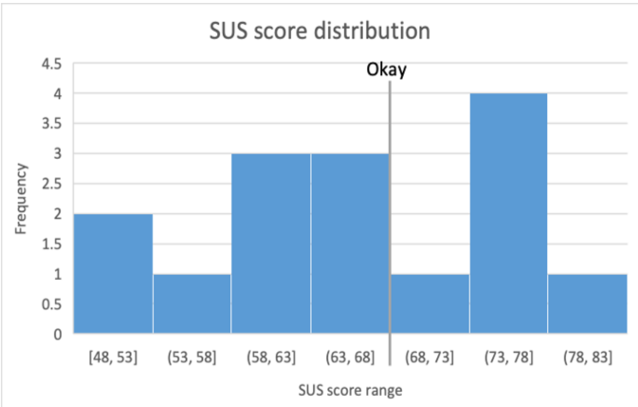
Reference

SUS Score	Adjective Rating
> 80.3	Excellent
68 – 80.3	Good
68	Okay
51 – 68	Poor
< 51	Awful

Figure 43: SUS score of the correlation interface developed for the table and wine grape pilot

4.2.2 Wine Making Pilot - Leaf Counting Task

The median of 67.5 puts the interface in an acceptable region of the SUS scale. Looking at the SUS score distribution, we can observe that the scores between 73 and 78 were given by a relatively large portion (4/15) of the participants but the majority (9/15 participants) yielded the scores below the 68th percentile.



Median 67.5

SD 10.6

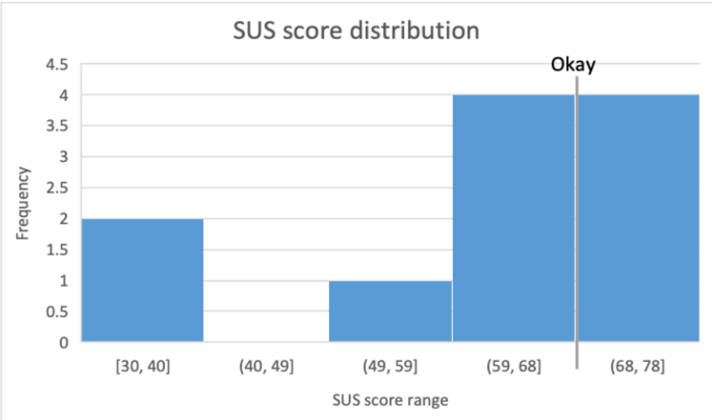
Reference

SUS Score	Adjective Rating
> 80.3	Excellent
68 – 80.3	Good
68	Okay
51 – 68	Poor
< 51	Awful

Figure 44: SUS score of the leaf counting interface developed for the wine making pilot

4.2.3 Wine Making Pilot - Correlation Task

Figure 45 shows results of the SUS questionnaire. This interface was exactly the same as the one developed for the table and wine grape pilot. Thus, it had a similar SUS score with the median of 65 and standard deviation of 15.1 across 11 participants. We therefore conclude that although the score does not significantly deviate from the acceptable region, further analyses will help us understand it better.



Median 65.0

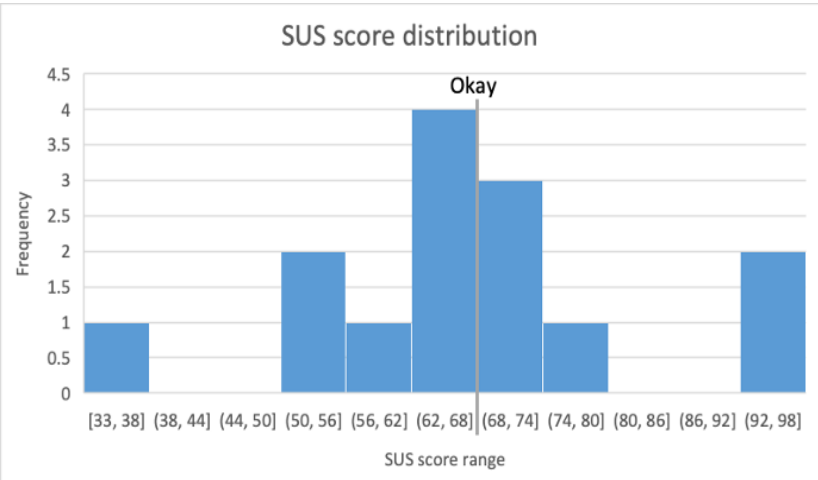
SD 15.1

Reference	
SUS Score	Adjective Rating
> 80.3	Excellent
68 – 80.3	Good
68	Okay
51 – 68	Poor
< 51	Awful

Figure 45: SUS score of the correlation interface developed for the wine making pilot

4.2.4 Wine Making Pilot - Vine to Wine Exploration Task

The interface developed for the vine to wine exploration task of the wine making pilot can be found in Figure 33. Figure 42 shows results of the SUS questionnaire. The median of 67.5 puts the interface in an acceptable region of the SUS scale. The SUS score distribution tells us that 8 out of the 14 responses put the interface below the 68th percentile.



Median 67.5

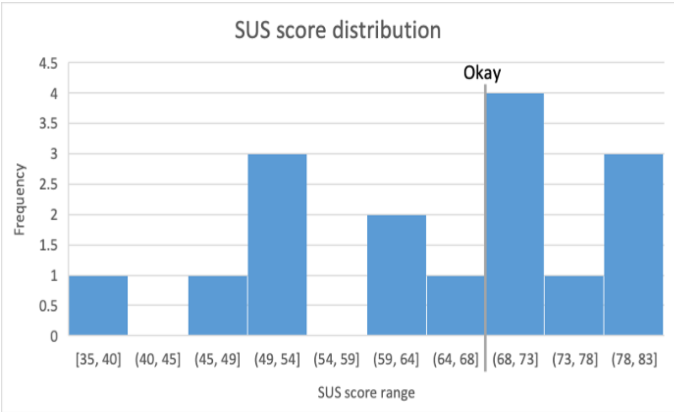
SD 15.9

Reference	
SUS Score	Adjective Rating
> 80.3	Excellent
68 – 80.3	Good
68	Okay
51 – 68	Poor
< 51	Awful

Figure 46: SUS score of the vine to wine exploration interface developed for the wine making pilot

4.2.5 Farm Management Pilot - Irrigation Task

The interface developed for the irrigation task of the farm management pilot can be found in Figure 34. Figure 43 shows results of the SUS questionnaire. The median of 68.8 puts the interface slightly above the acceptable region on the SUS scale. The SUS score distribution tells us that half (8/16) of the participants yielded the scores below 68.



Median 68.8

SD 14.1

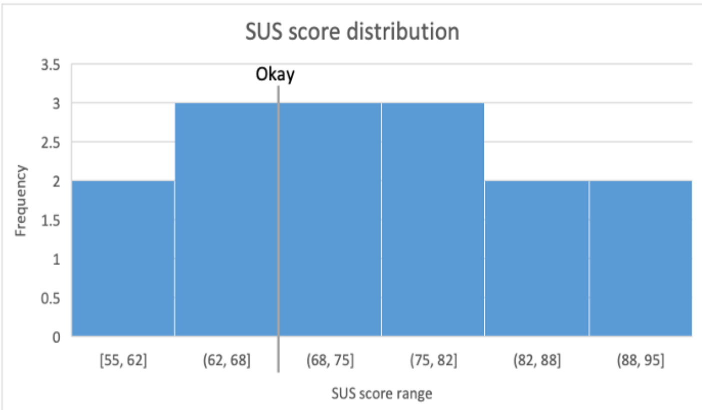
Reference

SUS Score	Adjective Rating
> 80.3	Excellent
68 – 80.3	Good
68	Okay
51 – 68	Poor
< 51	Awful

Figure 47: SUS score of the irrigation interface developed for the farm management pilot

4.2.6 Natural Cosmetics Pilot - Bio-efficacy Correlation Task

The interface developed for the bio-efficacy correlation task of the natural cosmetics pilot can be found in Figure 35. Figure 44 shows results of the SUS questionnaire. This interface, among the rest, received the highest SUS score with the median of 75 and the standard deviation of 11.3. Only 5 out of the 15 responses yielded the scores below 68, leaving the rest to put the interface in the "Good" and "Excellent" regions.



Median 75.0

SD 11.3

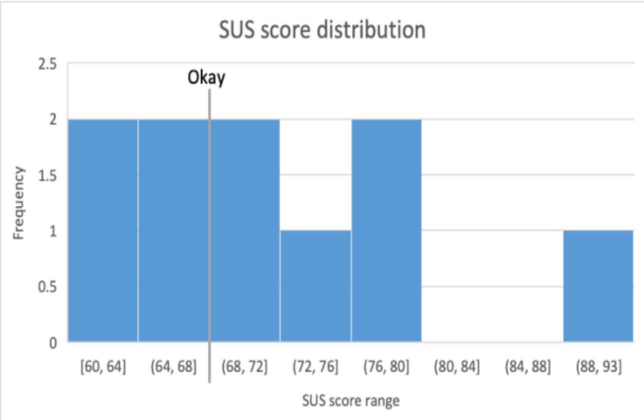
Reference

SUS Score	Adjective Rating
> 80.3	Excellent
68 – 80.3	Good
68	Okay
51 – 68	Poor
< 51	Awful

Figure 48: SUS score of the bio-efficacy correlation interface developed for the natural cosmetics pilot

4.2.7 Food Protection Pilot - Risk Assessment Task

The interface developed for the risk assessment task of the food protection pilot can be found in Figure 37. Figure 45 shows results of the SUS questionnaire. This interface received the second highest SUS score, following the highest scoring interface of the natural cosmetic pilot, with the median of 70 and the standard deviation of 9.3. The score distribution shows that 4 out of the 10 responses yielded the scores below 68, meanwhile the majority (5/10 participants) put the interface in the "Good" region and one put it in the "Excellent" region.



Median 70.0

SD 9.3

Reference

SUS Score	Adjective Rating
> 80.3	Excellent
68 – 80.3	Good
68	Okay
51 – 68	Poor
< 51	Awful

Figure 49: SUS score of the risk assessment interface developed for the food protection pilot

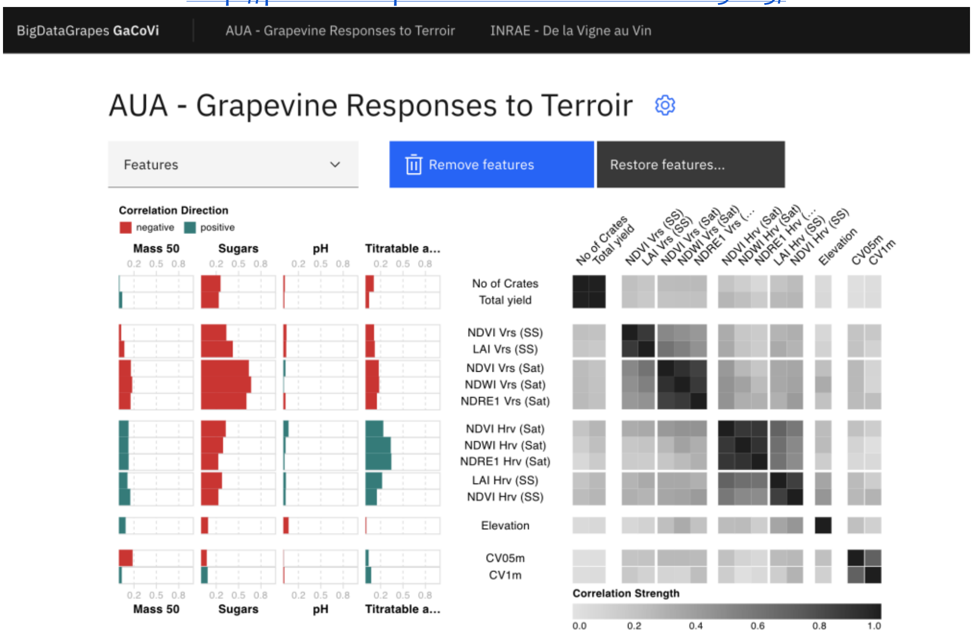
4.3 RESULTS ON INTERNAL TECHNOLOGICAL EVALUATION

4.3.1 Assessment Checklist Per Pilot

Table and Wine Grapes Pilot

Table and Wine Grapes Pilot Software Demonstrator Tasks	Notes
Grapevine responses to terroir	
Data and datasets	
Have you identified, collected & shared all data sets & sources required?	YES. These include: Yield Mapping Grape and berry mechanical properties Classical analytical techniques (HPLC) Topographic data and elevation maps Canopy sensing and vegetation indices IoT stationary data Drone imagery Eca sensing Sentinel-2 Landsat-8 All relevant steps have been completed
Which data are private and which public?	**Private:** Yield Mapping, Grape and berry mechanical properties, Classical analytical techniques (HPLC), Topographic data and elevation maps, Canopy sensing and vegetation indices, Drone imagery, Eca sensing **Public:** IoT Stationary data, Satellite data (Sentinel-2 and Landsat-8)
Have those data been uploaded to the BDG platform?	All our data are uploaded to Agroknow's BDG data platform

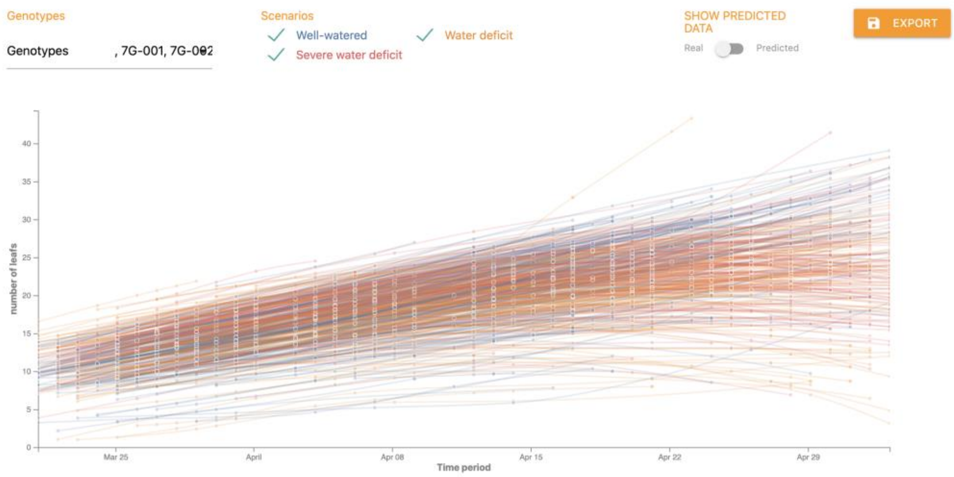
Is there a way to update those data?	YES, after being uploaded to the platform they can be also updated.
Are those data available on the website?	YES, after being uploaded to the platform they are also available on the website.
Do you have projected data?	No projected data. This is not relevant for the structured data.
Have you completed the data modelling?	The data modelling is completed and ingested.
Business decisions and Competence questions	
Which are the critical business decisions to be supported?	Based on sensor, farming and phenological data, the Table and Wine Grapes Pilot will answer questions concerning Farmers and Winemakers, on the existing associations and the correlations amongst these data. This Pilot will also answer questions to what extent the aforementioned data are correlated with earth observation data.
What are the Intelligence and data competence questions to be answered?	"Compare mechanical properties (and qualitative data) with soil conductivity data for a given plot" "Give me correlation between satellite images, NDVIs vegetation indices from proximal sensors and photosynthesis concentrations in plant level" "Which vineyard variables have strong predictive power on grape quality parameters?" "Which vineyard variables from different data layers capture similar information?"
Algorithm Implementations	
Have you completed the model algorithm?	YES, we have developed an algorithm that interacts over the data layers to find the coefficient correlations between pairs of features from different data layers. We have used Pearson product-moment correlation coefficients to perform the correlation analysis.
Did you find the best algorithm that fits best?	Aimed at finding the most meaningful setup of correlations between the data layers, we have performed a pairwise correlation analysis: (1) precision agriculture information (sensor data) and phenological data and grape chemical analysis (lab data); (2) Correlate the sensor data with earth observation data on vegetation indexes (NDVI) for similar dates. (3) Correlate sensor data and lab data with earth observation data.
Have you run numerical experiments?	YES, we have performed experiments showing in a heatmap the pairwise correlation coefficients between the features of the three data layer features. Specifically, we investigated the correlation between the following pair of data layers: (1) sensor data & lab data; (2) sensor data & satellite data; (3) sensor data & lab data. For the interpretation of the results, the coefficient values close 0 represent a low correlation between two layers of data. While, values close to 1 or -1 means a high correlation or high inverse correlation, respectively.
Is the model algorithm integrated in the platform?	The model algorithm is not yet integrated in the platform.
Software Tool Extension & UI	

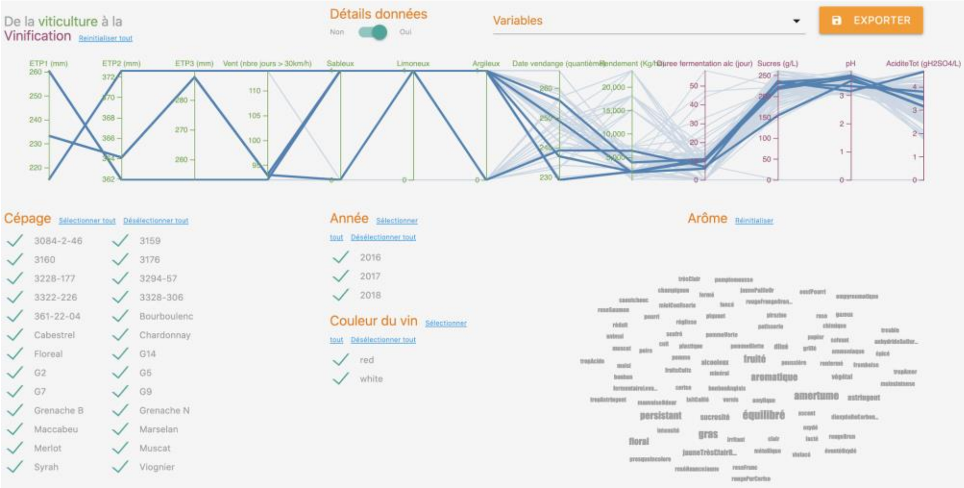
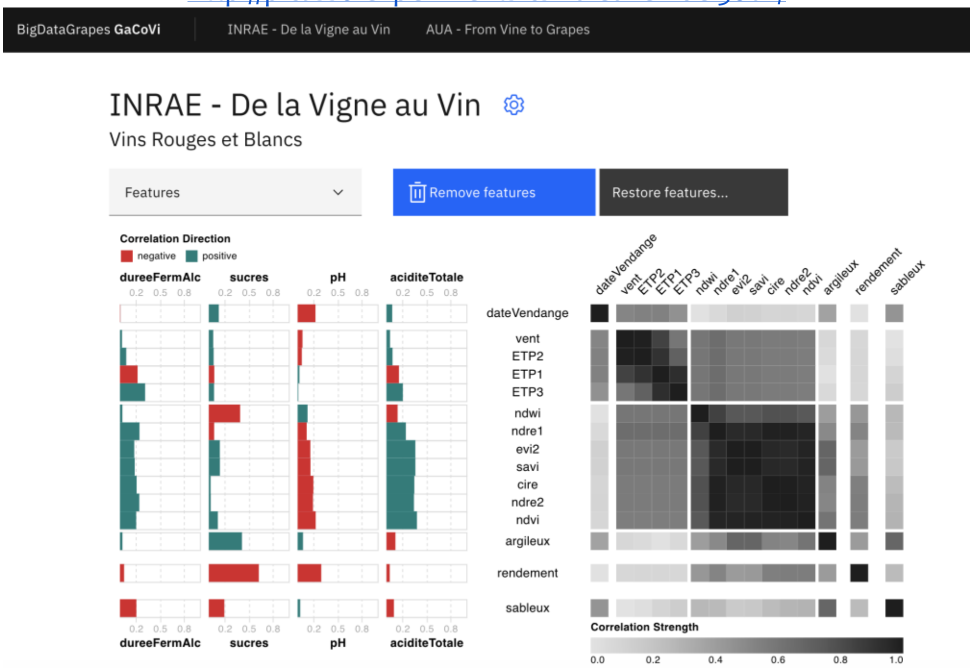
Have you designed and developed the visualization components' mock-ups?	YES
Have you completed the visualisations?	YES http://picasso.experiments.cs.kuleuven.be:3603/ 
Is the visualization component integrated in the platform?	The visualization component is not yet integrated in the platform.
This SD is an extension of SITI4farmer or FOODAKAI? Or stands alone?	First UI mock-up extension to SITI4farmer unless decided to go individually.
Evaluation & Dissemination activities	
Have you identified the end-users?	YES Farmers and Wine Industry End-Users: Palivos Estate Kontogiannis Estate Fasoulis Nursery Zacharias Winery Pirgakis Estate Tetramythos Winery Gaia Wines Agricultural and environmental solutions Industry End-Users: Agenso Research Organisation End-Users: Agricultural University of Athens
Did you have initial feedback from the end-users?	Completed (16 Surveys from 9 end-users)
Have you developed a content article describing the business case of the pilot?	YES Table and Wine Grapes Pilot: Grapevine Responses to Terroir Table and wine grape: Vineyard assessment to vineyard management

Wine Making Pilot

Wine Making Pilot Software Demonstrator Tasks	Notes
a. **Counting Grapevine Leaves** b. **From Vine to Wine: Parameters' Influencing Wine Quality** c. **Gacovi "From Vine to Wine"**	
Data and datasets	
Have you identified, collected & shared all data sets & sources required?	YES. They include: climatic data, geospatial information soil characteristics, phenotypic data, genetic data, plot management, grape and berry mechanical properties, harvesting data, qualitative and quantitative characteristics of must, winemaking activities, sensory analysis
Which data are private and which public?	**Public:** genetic data (needs citation), soil characteristics, Plot management, viticulture data, satellite data **Private:** Phenotypic data, Climatic data, grapevine images **Related to the wine:** public or private in function of the wine produced (list of wines) Grape and berry mechanical and chemical properties, Qualitative and quantitative characteristics of must, winemaking data, Sensory analysis
Have those data been uploaded to the BDG platform?	Metadata have been integrated to the Big Data Platform and Rdf data are into graphDB, and csv
Is there a way to update those data?	YES, manual updates in gsheets
Are those data available on the website?	YES
Do you have projected data?	YES for grapevine images (first demonstrator)
Have you completed the data modelling?	YES
Business decisions and Competence questions	
Which are the critical business decisions to be supported?	<u>Task 1: Counting Grapevine Leaves</u> The number of leaves is a key indicator of vegetative organogenesis in grapevine. - it enables to improve existing models of leaf emission rate - it enables to characterize genetic diversity of leaf emission rate - it enables to know the development stage (leaf number) of all plants on a given day <u>Task 2: From Vine to Wine: Parameters' Influencing Wine Quality</u> -**profits can be optimised** with a better understanding of parameters influencing wine quality, from the vine to the wine -it enables to know which variables impact aroma's production to **target specific wine profiles** -it enables to **adapt field management or winemaking process** to obtain the wine of interest <u>Task 3: Market penetration of a given wine in a new country- GaCoVi</u>

	From satellite data, field and lab analyses: to highlight existing associations and correlations between these data. Know to what extent the field data is correlated with each other and influences the quality of the must. Identify variables with weak / strong predictive power on a chosen quality parameter Knowing the variables of interests, leverages to obtain a target wine
What are the Intelligence and data competence questions to be answered?	There are many competence questions that need to be answered: Task 1: Can I retrieve the number of leaves given a grapevine? Can I have the location of the vine given a picture/ plant? (Link with environmental data) Can I know the variety of the grapevine given a picture? (Link with field management data (such as irrigation, etc.) Can I obtain the characteristics of the variety for a plant/ picture? (Link with genetic data) Task 2: Can I retrieve wines produced for a given plot? Can I retrieve plots or sublots of production given a wine? Can I retrieve sensory analysis for a given wine? Can I retrieve the winemaking process of a given wine? Can I retrieve must analysis (especially sugars, ph and total acidity) for a given wine? Can I retrieve fermentation length for a given wine? Can I retrieve genetic information of the variety of a given wine? Can I know if a plot was irrigated or not? Can I know if the soil of a given plot is composed of sand, silt or clay? Can I have the evapotranspiration of a given plot and a given date? Can I have the wind speed of a given period? Task 3: GaCoVi From which plot(s) does a wine come from? Which wine(s) comes from which plot(s)? Can we geolocate the parcel (on which a wine is produced)? Is it possible to get harvest data from a plot or a wine? Which vineyard variables have strong predictive power on must quality parameters? Which vineyard variables from different data layers capture similar information?
Algorithm Implementations	
Have you completed the model algorithm?	Task 1: In this task, we employ Deep Convolutional Neural Network (DCNN) to train a machine learning system for the estimation of the number of leaves of the vine. The data used for this task has been collected by the PhenoArch phenotyping platform which ran two experimentations, one in 2012 and another in 2013, collecting two datasets of plant images with the associated number of leaves of each plant image. Task 3: The visual analytics system that supports this task relies on the correlation coefficients between the different features. We have used

	Tetrachoric correlation to perform the correlation analysis when two dichotomous factors involved, Point-Biserial correlation when one dichotomous and one continuous factor involved, and Pearson correlation otherwise.
Did you find the best algorithm that fits best?	Task 1: Image Analysis is a well-known task that is usually addressed using neural networks. In particular, Deep Convolutional Neural Networks (DCNN) are a class of deep neural networks commonly applied to analysing visual imagery in order to extract useful information and interesting visual patterns. For the development of the DCNN in our component, we used the python Keras Library. We have tried many different configurations of the DCNN selecting the most effective one at the end.
Have you run numerical experiments?	Task 1: We modelled the problem as a regression task, i.e., given a vine image, predict its associated number of leaves. Each dataset is split into training, validation and test on a per-plant basis (i.e., all the images regarding a single plant are in a single fold). Training fold is used to fit the model, while validation fold is used for early stopping the training when overfit starts to occur and for hyper- parameters tuning. Finally, test fold is used to collect model performance, mainly in terms of Mean Absolute Error (MAE) and Mean Squared Error (MSE). Several neural network architectures have been tested (shape, size and number of layers), as well as different activation functions, optimization algorithms, and regularization techniques (e.g., batch regularises, dropout). As shown in Deliverable 4.3, the proposed solution is able to correctly count the number of leaves of each plant with an error lower than one leaf on average on the entire test fold.
Is the model algorithm integrated in the platform?	Task 1: YES, an API can be queried Task 3: The model algorithm is not yet integrated in the platform.
Software Tool Extension & UI	
Have you designed and developed the visualization components' mock-ups?	YES
Have you completed the visualisations?	YES Task 1: http://picasso.experiments.cs.kuleuven.be:3327/  **"Counting grapevine leaves" demonstrator's visualisation** Task 2:

	http://picasso.experiments.cs.kuleuven.be:3330/  **"From Vine to Wine" demonstrator's visualisation** Task 3: http://picasso.experiments.cs.kuleuven.be:3602/  **"Market penetration of a given wine in a new country" demonstrator's visualisation**
Is the visualization component integrated in the platform?	The visualization component is not yet integrated in the platform.
This SD is an extension of SITI4farmer or FOODAKAI? Or stands alone?	It will be integrated in OpenSilex (INRAE tool) unless decided to go individually.
Evaluation & Dissemination activities	
Have you identified the end-users?	Gerard Bertrand GCF group Cooperative winery "Les Trois Grappes" vinegrowers

	NYSEOS IFV INRAE SPO, science for oenology INRAE Pech Rouge, experimental unit
Did you have initial feedback from the end-users?	Completed (16 Surveys from 8 end-users)
Have you developed a content article describing the business case of the pilot?	yes Counting grapevine leaves in the BigDataGrapes project AI applications for wine and vine (full article on INFOWINE) A journey from vine to wine in pictures

Farm Management Pilot

Farm Management Pilot Software Demonstrator Tasks	Notes
Water Availability and Irrigation Recommendations	
Data and datasets	
Have you identified, collected & shared all data sets & sources required?	YES , the monitoring data are all available and shared. The best practices data are private data and cannot be shared. Sentinel-2 Landsat-8 Chemical and physical info on grapes are available but not shared Day by Day Activities in term of treatments, fertilization, field operation are available but not shared Plot and Fields information georeferenced are available and shared Relative Humidity available and shared Air Temperature available and shared Global Solar Radiation available and shared Wind Speed and Direction available and shared Soil Temperature available and shared Soil Moisture available and shared Precipitation available and shared Infrared Surface Temperature available and shared
Which data are private and which public?	**Public:** Satellite data (Sentinel-2 and Landsat-8); Crop plan is public. **Private** (but shared into the project)**:** Weather station data - Relative Humidity, Air Temperature, Global Solar Radiation, Wind Speed and Direction, Soil Temperature, Soil Moisture, Precipitation, Infrared Surface Temperature. Best Practice data are private and not shared.
Have those data been uploaded to the BDG platform?	YES, All our data are uploaded to Agroknow's BDG data platform

Is there a way to update those data?	YES using APIs. Data after being uploaded to the platform can also be updated.
Are those data available on the website?	YES, after being uploaded to the platform they are also available on the website.
Do you have projected data?	YES the Water Balance data have been projected
Have you completed the data modeling?	YES, data modelling is completed and ingested
Business decisions and Competence questions	
Which are the critical business decisions to be supported?	Increase productivity / grape quality Reduce the use of emergency irrigation Detect the moment in which the stress level is critical for the plant Which part of the vineyards really needs irrigation
What are the Intelligence and data competence questions to be answered?	"How much of the data available are representative of the real status of the environment in which the vineyard is?" "Which is the correspondence between the water stress alert level and the real status of the vineyard?" "Which part of the vineyard is mostly homogeneously correspondent with the water stress level?"
Algorithm Implementations	
Have you completed the model algorithm?	Yes. We 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 fields under study by the BigDataGrapes project. The two fields are on the "Casato Prime Donne" and "Il Palazzo" vineyards. The two fields have been equipped with two weather and soil stations that record several information at different levels of granularity. Data from the weather stations are collected by ABACO.
Did you find the best algorithm that fits best?	Yes, we applied a set of machine learning techniques for solving the following regression task: given an observation, predict its associated water level. The problem has been modeled 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
Have you run numerical experiments?	For all the methods, we employ 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 (RF). The performance of RF is very similar to another tree-based technique, i.e., Gradient Boosting regressor.
Is the model algorithm integrated in the platform?	YES, APIs have been shared with Agroknow
Software Tool Extension & UI	

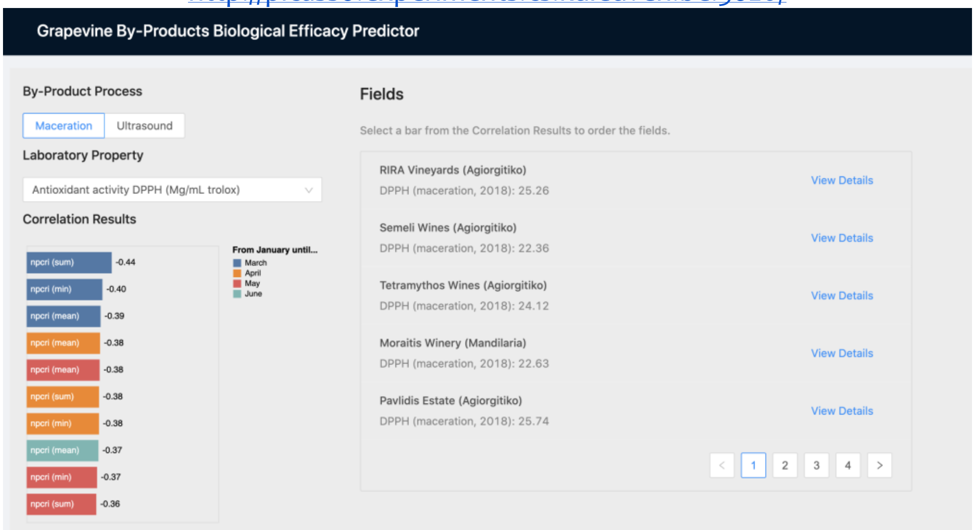
<table>
<tr><td>Have you designed and developed the visualization components’ mockups?</td><td>YES</td></tr>
<tr><td>Have you completed the visualisations?</td><td>Yes. But tweaking with feedback (KUL) still ongoing.
http://picasso.experiments.cs.kuleuven.be:3328/


“Water Availability and Irrigation Recommendations” demonstrator’s visualisation</td></tr>
<tr><td>Is the visualization component integrated in the platform?</td><td>We are streaming data through SITI4farmer and Geocledian APIs. For integration, source code is available on Github.</td></tr>
<tr><td>This SD is an extension of SITI4farmer or FOODAKAI? Or stands alone?</td><td>SITI4farmer</td></tr>
<tr><td colspan="2">Evaluation & Dissemination activities</td></tr>
<tr><td>Have you identified the end-users?</td><td>YES
Il Palazzo
Casato Prime Donne
Consorzio di Cremona
Cantina Torrevilla
Università di Milano
Consorzio di Siena / Università di Firenze
Abaco Agri-Food Department
Abaco Factory-Tech Department</td></tr>
<tr><td>Did you have initial feedback from the end-users?</td><td>Completed (16 Surveys from 7 end-users)</td></tr>
</table>

Have you developed a content article describing the business case of the pilot?	YES Finding answers to critical business questions through ABACO Farmer Can we see from space if a vineyard needs water?

Natural Cosmetics Pilot

Natural Cosmetics Pilot Software Demonstrator Tasks	Notes
Grapevine By-Products Biological Efficacy Predictor	
Data and datasets	
Have you identified, collected & shared all data sets & sources required?	YES Satellite data Weather data Laboratory analysis data
Which data are private and which public?	Public: Satellite data (Sentinel-2 and Landsat-8); Weather data, Laboratory data
Have those data been uploaded to the BDG platform?	YES
Is there a way to update those data?	YES
Are those data available on the website?	YES
Do you have projected data?	YES
Have you completed the data modelling?	YES, data modelling is completed and ingested
Business decisions and Competence questions	
Which are the critical business decisions to be supported?	Selection of the appropriate extraction method for by-products process Selection of the best quality by-products in terms of biological activity Compliance and assessment of incoming raw material quality (wine making by-products) by: Farm management; Weather conditions; Geo-climatic conditions - location) Best by-products exploitation Competent farm management
What are the Intelligence and data competence questions to be answered?	Which extraction process presents the highest BA parameters? Which grapevine variety presents the highest BA parameters? Which SVI has the highest correlation to a B.A. parameter? Which weather parameter has the highest correlation to a B.A. parameter? Which Parcel ID presents the highest correlation between SVIs and BA parameters? Which Parcel ID presents the highest correlation between weather data and BA parameters? Does an SVI value predict the B.A. parameter of a given parcel I.D. Which BA parameters are going to present the highest values for a certain Parcel ID?

<table>
<tr><th colspan="2">Algorithm Implementations</th></tr>
<tr><td>Have you completed the model algorithm?</td><td>YES. The algorithm uses the available features at the data layers to create vectors of size F x T, where F are the number of fields and T is the number of temporal points considered for the analysis. The objective of the algorithm is to compute how these properties correlate between them crossing data from the different vineyards. We have used Pearson product-moment correlation coefficients to perform the correlation analysis.</td></tr>
<tr><td>Did you find the best algorithm that fits best?</td><td>We have tried different setup of experiments using different timeframes to aggregate the satellite imagery indexes since the two data layers have different periodicity on the collection of their features. For example, new BA records are collected yearly, while the satellite images are taken almost weekly.
We use the following time frame to filter the satellite data for aggregation: from the beginning of the year of crop until end of <month> (of the same year) having as possible: March, April, May and June as possible values. Then, as an aggregation function, we use: min, max, mean, std, count and sum.</td></tr>
<tr><td>Have you run numerical experiments?</td><td>YES, we have performed experiments showing in a heatmap the pairwise correlation coefficients between the two data layer features. So that coefficient values close 0 represent a low correlation between two layers of data. While, values close to 1 or -1 means a high correlation or high inverse correlation, respectively.</td></tr>
<tr><td>Is the model algorithm integrated in the platform?</td><td>YES</td></tr>
<tr><th colspan="2">Software Tool Extension & UI</th></tr>
<tr><td>Have you designed and developed the visualization components' mock-ups?</td><td>YES</td></tr>
<tr><td>Have you completed the visualisations?</td><td>YES
http://picasso.experiments.cs.kuleuven.be:3620/


"Grapevine By-Products Biological Efficacy Predictor" demonstrator's visualisation</td></tr>
<tr><td>Is the visualization component integrated in the platform?</td><td>YES</td></tr>
</table>

This SD is an extension of SITI4farmer or FOODAKAI? Or stands alone?	FOODAKAI is scheduled to be integrated with the platform
Evaluation & Dissemination activities	
Have you identified the end-users?	YES (persons): Cosmetic Industry End-Users: APIVITA (2) The Nu Club (2) Symbeeosis (2) Frezyderm (2) Beauty Lab (1) QS Profesional (1) Sarantis (1) Farmers: Aoton Winery (1) Research Organisation End-Users: HAO-DEMETER (1) AUA (1)
Did you have initial feedback from the end-users?	Completed (15 Surveys from 10 end-users)
Have you developed a content article describing the business case of the pilot?	YES Can we predict which vineyard will supply the best quality leaves for natural cosmetics production?

Food Protection Pilot

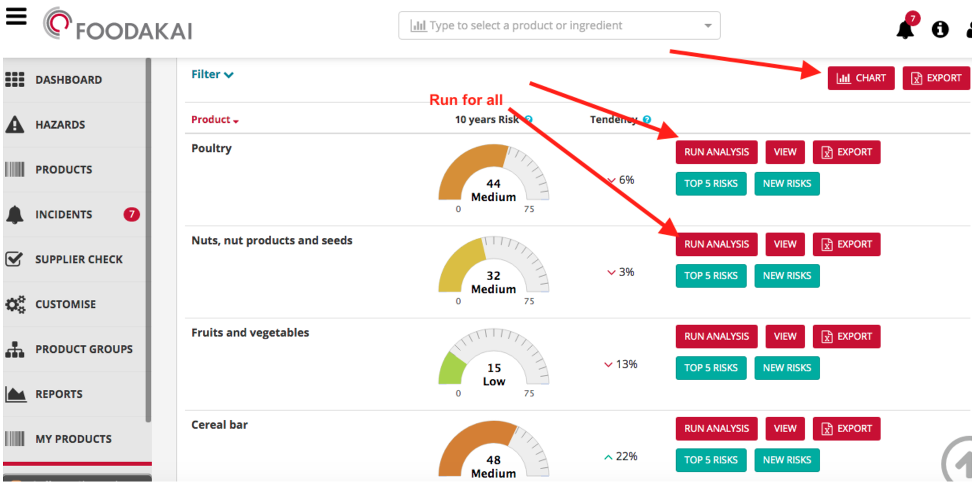
Food Protection Pilot Software Demonstrator Tasks	Notes
Risk Assessment Module, Price Prediction Dashboard and Recall Prediction	
Data and datasets	
Have you identified, collected & shared all data sets & sources required?	Yes Lab testing data Pricing data UK Food Standards Agency RASFF – Rapid Alert System for Food and Feed Food Standards Australia New Zealand FDA Recalls, Market Withdrawals, & Safety Alerts, warning letters, import refusals and inspection citations EFET - Hellenic Food Safety Authority Japanese Imported Foods Inspection Services Czech Agriculture and Food Inspection Authority Healthy Canadians food alert information website Food Safety Authority of Ireland German Food Safety: warnings and information to the public Hong-Kong-Center for Food Safety Open Food Facts ProMED-mail

Which data are private and which public?	Public: The original datasets published by the food safety authorities are publicly available and can be downloaded from their websites. Private: All the processed food recalls, border rejections, lab test, supplier's data are private and owned by Agroknow.
Have those data been uploaded to the BDG platform?	All the aforementioned datasets are integrated in AK's data platform (an instance of the BDG data platform)
Is there a way to update those data?	The data are continuously updated through are data integrators (crawlers). The data integration tool can be used to integrate new datasets.
Are those data available on the website?	All this data is available through the FOODAKAI platform, is available through a data API and will be available through the marketplace.
Do you have projected data?	No. We will create projected data in the context of T7.3.
Have you completed the data modelling?	YES, data modelling is completed and ingested
Business decisions and Competence questions	
Which are the critical business decisions to be supported?	Based on real industry cases, Food Protection pilot identify risk in food products in order to prevent issues (recall/border rejections) in a food supply chain.
What are the Intelligence and data competence questions to be answered?	"Which is the most important chemical risk for a food product?", "Which is the ingredient (eg grape) with the higher risk?", "Which are the countries with the higher risk for hazards and fraud?", "Which is the increasing risk for my product (raisins)?", "How did the price of my ingredient evolve over time?", "What are the projected changes in price for my ingredient?"
Algorithm Implementations	
Have you completed the model algorithm?	In the food protection pilot, we employ time series, i.e., sequences of per-product price observations, to train a machine learning system that allows us to predict the future price of the product, given an historical time window. We addressed it using neural networks.
Did you find the best algorithm that fits best?	Time series prediction is a well-known task that is commonly addressed using neural networks. In particular, Recurrent Neural Networks (RNN) are a specific kind of neural network that learns the properties of sequences of objects to predict their evolution. We employ Long Short-term memory (LSTM) networks to address this task. LSTM is a powerful RNN architecture with important application in time series prediction. For the development of the LSTM in our component, we used the python Keras library.
Have you run numerical experiments?	For the price prediction task, we model it as a regression problem, where, given a sequence of daily observations of prices, we want to predict the price that will be seen in the next day. We design a neural network architecture, which we find to be a LSTM network composed of one LSTM layer with 150 neurons. We then evaluate the performance of the LSTM on the test set, i.e., we provide a time window of data and the network outputs the price prediction that is then matched on the test time series to compute the error. We evaluate the performance of our LSTM by testing it against a strong baseline. The baseline we are proposing is always employed in time series analysis and prediction because it provides a way to compare the performance of a predictor against the performance of a price produced using the moving average of the previous

<table>
<tr><td></td><td>observed prices. As shown in Deliverable 4.3, LSTM is able to outperform the baseline for most of the products tested.
For recall prediction, we select four algorithms that have the best results and test them on different datasets. Then, the four algorithms are re-examined with differential data sizes to select the one with the best results in the majority of cases. We select K-nearest neighbours, Decision tree, long short-term memory networks, and the Prophet algorithm, i.e., an additive regression procedure. As shown in Deliverable 4.3, several experiments on different datasets report that the prophet algorithm presents the best results in the majority of cases. Important factors that affect the results are the size of the data, the homogeneity of the product categories and the seasonality.</td></tr>
<tr><td>Is the model algorithm integrated in the platform?</td><td>The recall prediction algorithm has been integrated in the platform. The price prediction integration is on-going.</td></tr>
<tr><td colspan="2">Software Tool Extension & UI</td></tr>
<tr><td>Have you designed and developed the visualization components’ mockups?</td><td>YES. The visualization mock-ups have been designed and validated and led to the implementation of the visualization components used in the pilot evaluation.</td></tr>
<tr><td>Have you completed the visualisations?</td><td>YES. The visualization components have been completed. Their integration is on-going.
Price Prediction
http://picasso.experiments.cs.kuleuven.be:3540/

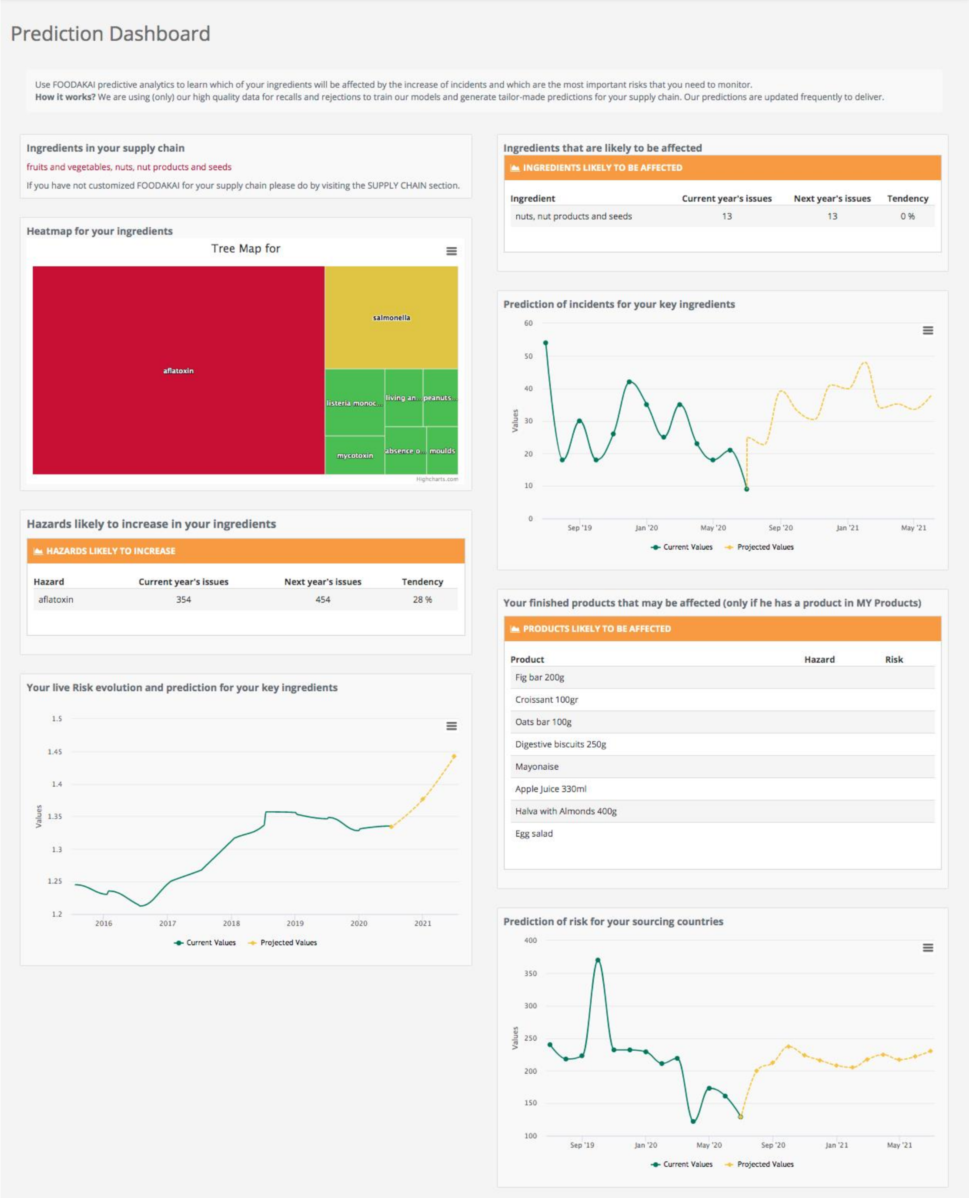
“Price prediction” demonstrator’s visualisation

Risk Assessment</td></tr>
</table>



“Risk assessment” demonstrator’s visualisation

Recall Prediction



	"Recall prediction" demonstrator's visualisation
Is the visualization component integrated in the platform?	The visualization component integration is on-going.
This SD is an extension of SITI4farmer or FOODAKAI? Or stands alone?	Foodakai
Evaluation & Dissemination activities	
Have you identified the end-users?	A list of companies and organizations has been identified as potential users of the pilot, based on their needs and activities. The following companies have been approached, among others, to evaluate the pilot, either at design and alpha stage or its final stage of development: Tuv Hellas Papadopoulos Biscuits Yili Coca Cola HBC Barry-Callebaut Conagra Ahold-Delhaize/ AB Vasilopoulos Ferrero Apivita Torres Gallo Winery Wines of Crete
Did you have initial feedback from the end-users?	Completed (20 Surveys from 12 end-users) Feedback from the target users have been collected with a variety of means, including demonstrations and presentations in the context of bigger events, training activities, interviews and focus groups discussions and tasks-based evaluation activities where they users had a chance of a hands-on experience with the pilot tool.
Have you developed a content article describing the business case of the pilot?	We have published several articles about risk assessment and prediction in our https://agroknow.com/blog/ which are linked to the Food protection pilo, for example https://agroknow.com/blog/frustrated-from-dealing-with-risk-assessment-as-a-manual-task/. We had also a presentation at the SmartCow project sharing the experience from BDG An additional post describing the pilot's business value was published in the BDG website: Can we really prevent food fraud and safety incidents from happening?

4.3.2 Global Technological Evaluation

The Global Technological Evaluation was completed by 14 respondents, the BDG pilot and tech partners. The majority of respondents (10+) agreed that BDG pilots have had the most impact on FAIR-ness, Scalability and Reusability. The overwhelming majority of respondents agreed that BDG allows for improved scalability in data, with 75% agreeing that data and algorithms are FAIR. They agreed that data from various data sources are supported with advantages compared to current research environments and data management practices. There was a mixed response in terms of reusability, with 57% agreeing that BDG technology components support reusability, 14% disagreed, and 29% remained neutral. The results for resource optimization showed a favourable outcome, with the majority of respondents (71%) agreed that BDG pilots enable management capabilities and resource optimization, with the remaining 29% taking a neutral stance. In terms of flexibility, 86% of respondents believed that BDG permits the addition of new data and enhanced functionality, with 14% remaining neutral. All respondents agreed that BDG promotes the transfer of knowledge by sharing components and data. More than 85% agreed that BDG promotes flexibility by the integration of open-source data and libraries, with the remaining 14% responding with neutral. The questions relating to the reliability indicator showed that 93% of respondents agreed (and 7% neutral) that BDG provides a seamless end-to-end experience, making users more productive across the whole data and analytics pipeline. When asked if the BDG components and tools developed were reliable for day-to-day use, 71% agreed, however, 21% remained neutral and 7% disagreed. When asked regarding whether the BDG pilots were consistent and integrated to support an entire data analytics pipeline, 78% of respondents agreed, 7% disagreed, and 14% were neutral. Finally, when asked whether respondents believed that the BDG pilots were conforming to user expectations and quality standards, 86% agreed, with the remaining 14% being neutral.

5 DISCUSSION AND CONCLUSIONS

This deliverable, the "Evaluation Report and KPI Assessment", belongs to WP8, "Grapevine-powered Industry Application Pilots". This work package is responsible for the planning and preparation of pilot, the definition of the experimental and evaluation protocols to be followed, the execution of the pilots and ultimately, the collection and evaluation of the pilot results and their assessment over indicators defined by the end users. The deliverable aims to provide a report on the results of the application piloting sessions, in line with the defined experimental protocols and in accordance with the evaluation methodology, providing an overview and a first evaluation regarding each of the five pilots' progress. It states and explains the current status of development, while the implementation and achieved performance of the BDG pilots are assessed.

Evaluation was to be both formative and summative. The former is essentially self-assessment and was carried out by all partners through filling the "Qualitative and Quantitative Evaluation", which consists of a total of five reports that displays the current status of the piloting activities and thus, it is providing tangible results. The summative evaluation involved external as well as internal evaluation in the form of "BigDataGrapes Pilots' Survey". This deliverable describes the structure of the BDG Survey, which was distributed to the end-user a few months after the completion of the Intermediate phase (M28) and before the end of the Summative phase (M34).

Throughout the piloting sessions, all five pilots successfully gathered data from their respective experimental sites. The individual reports have been analysed and the results have shown that the gross data volume resulted in a total of over ~ 3.5 TB throughout the projects lifetime. More specifically, all pilot partners used more than 90 different data sources to generate almost 70 unique datasets. One of the most popular characterization methods of Big Data is the "3V", representing Volume, Variety and Velocity of data generated respectively. From the pilots' data, it becomes obvious that out of the 3 "V"s, the Variety aspect has met the sufficient requirements, along with extended efforts to also cover sufficiently Volume and Velocity. The KPIs list was continuously updated during the project's lifetime.

The survey was divided in two levels of assessment, the "Pilot Basis Evaluation" and the "Internal Technological Evaluation", of heterogeneous sets of quantitative and qualitative indicators, metrics used to measure the effectiveness of the pilots. In order for this to be completed, an iterative approach of assessment was performed according to the proposed three-phase human-centred assessment activities.

The BDG consortium selected and defined eight (8) main Software Demonstration scenarios, reflecting the work that has been done in the five pilots, to focus on, fine-tune and showcase. These scenarios were selected to show how different software tools and components produced by the BDG project, together with the critical business decisions to be supported, relevant data and data sources, intelligence and data competence questions to be answered, and algorithm implementations, may support industrial end-users and other key stakeholders in the grapevine-powered industry in new innovative ways.

Following the concept of gradual extension of functionality, intended audience and assessment of this scheme, the pilots interacted with the community and the pilot evaluators accordingly. Thus, this survey was distributed to all relevant stakeholders involved in the BigDataGrapes piloting activities. Feedback was asked from the end-user, with a focus on the "Industry End-User". Representatives from 46 Industries/organisations participated with a total number of 83 Survey completed!

To conclude, the evaluation progress report submitted by the five pilots clearly prove that BDG project has a disruptive innovative potential to bring new and market driven ICT technologies into the grapevine-powered industries and that the majority of the pilots are performing their activities successfully and timely, showing professionalism and technical skills. The project's Evaluation Report and KPI Assessment presented in this report was a constantly updating roadmap to an optimal evaluation of the BDG pilots, tools and components, while being aligned with the project vision and objectives.

6 REFERENCES

Abed et al. (2015). In vitro assessment of cytotoxic, antioxidant and antimicrobial activities of leaves from two grape varieties collected from arid and temperate regions in Palestine. QScience Connect 4.pp. 1-9.

Anastasiou, E., Tsiropoulos, Z., Balafoutis, T., Fountas, S., Templalexis, C., Lentzou, D., & Xanthopoulos, G. (2017). Spatiotemporal stability of management zones in a table grapes vineyard in Greece. Advances in Animal Biosciences, 8(2), 510-514.

Balafoutis, A. T., Koundouras, S., Anastasiou, E., Fountas, S., &Arvanitis, K. (2017). Life Cycle Assessment of Two Vineyards after the Application of Precision Viticulture Techniques: A Case Study. Sustainability, 9(11), 1997.

Barreales et al. (2019). Effects of irrigation and collection period on grapevine leaf (Vitis vinifera L. var. Touriga Nacional): Evaluation of the phytochemical composition and antioxidant properties. Scientia Horticulturae 245, pp. 74–81.

Biniari, K., Gerogiannis, O., Daskalakis, I., Bouza, D., &Stavrakaki, M. (2018). Study of Some Qualitative and Quantitative Characters of the Grapes of Indigenous Greek Grapevine Varieties (Vitis vinifera L.) using HPLC and Spectrophotometric Analyses. NotulaeBotanicaeHortiAgrobotanici, 46(1), 97-106.

Bordiga M. (Ed.) (2015). “Valorization of wine making by-products” CRC Press.

Brooke J., SUS: a “quick and dirty’usability. CRC press, 1996.

Ferhi et al. (2019). Total Phenols from Grape Leaves Counteract Cell Proliferation and Modulate Apoptosis-Related Gene Expression in MCF-7 and HepG2 Human Cancer Cell Lines. Molecules, 24, pp. 612.

Fiume, M.M. (2012). Safety Assessment of Vitis Vinifera (Grape)-Derived Ingredients. Cosmetic Ingredient Review Expert Panel 2012. Pp. 1-35.

Hart S. G., “NASA Task Load Index (TLX) v. 1.0 Paper and Pencil Package,” pp. 1–19, 1986.

Jian J.-Y., A. M. Bisantz, and C. G. Drury, “Foundations for an empirically determined scale of trust in automated systems,” Int. J. Cogn. Ergon., vol. 4, no. 1, pp. 53–71, 2000.

Rasimah, C. M. Y., Halimah, B.Z., Azlina, A., 2011, Evaluation of user acceptance of mixed reality technology, Australasian Journal of Educational Technology, 27 (Special issue, 8), 1369-1387

Romain David, Laurence Mabile, Mohamed Yahia, Anne Cambon-Thomsen, Anne-Sophie Archambeau, et al.. How to assess FAIRness to improve crediting and rewarding processes for data sharing? A step forward towards an extensive assessment grid. RDA 13th (P13) Plenary Meeting, Apr 2019, Philadelphia, United States. ⟨https://www.rd-alliance.org/rda-13th-plenary-meeting-information⟩. ⟨10.5281/zenodo.2625721⟩. ⟨hal-02094678⟩

Stavrakaki, M., & Biniari, K. (2016). Genotyping and phenotyping of twenty old traditional Greek grapevine varieties (Vitis vinifera L.) from Eastern and Western Greece. Scientia Horticulturae, 209, 86-95.

Teixeira A, Baenas N, Dominguez-Perles R, et al. (2014) Natural Bioactive Compounds from Winery By-Products as Health Promoters: A Review. International Journal of Molecular Sciences, 15(9), 15638-15678.

Venkatesh V., M. G. Morris, G. B. Davis, and F. D. Davis, “User acceptance of information technology: Toward a unified view,” MIS Q. Manag. Inf. Syst., vol. 27, no. 3, pp. 425–478, 2003.

APPENDIX A- METHODS, GUIDELINES & RELATIVE MATERIAL

Selected Software Demonstrators and Targeted Audience

The following section is issued as a guide for the BigDataGrapes partners directly involved in the organisation of the Human-Centred Evaluation planned in Work Package 8, Task 8.5. The current Guidelines are not to be considered as a closed set of compulsory rules but rather as a general framework and guidelines for partners when preparing, organising, holding and reporting the Evaluation process. It is proposed though to all Pilots to keep similar methodology in the organisation so that global results can be extracted. The present Guidelines are the subject of an iterative review and discussion with partners up until its final agreement, and is accompanied by a set of templates, that partners will have to use.

The BDG consortium has selected and defined five (5) main Software Demonstration scenarios to focus on, fine-tune and showcase. These scenarios were selected to show how different software tools and components produced by the BDG project, together with the critical business decisions to be supported, relevant data and data sources, intelligence and data competence questions to be answered, and algorithm implementations, may support industrial end-users and other key stakeholders in the grapevine-powered industry in new innovative ways. The selected demonstrators and the corresponding partners responsible for their development, are:

1. **Table and Wine Grapes Pilot Software Demonstrator (AUA):**
   Grapevine responses to terroir
   Based on sensor, farming and phenological data derived from all test sites located in Greece the user will be able to visualise different correlations between them and explain what affects grape quality and yield. Intelligence and data competence questions on how some attributes, such as soil properties, affect grape quality and yield. Data integration have been achieved using web semantic and ontologies to have all data connected in a knowledge graph. This Software Demonstrator has the potential to be used as an extension of the Farm Management Software SITI4farmer. Initial feedback will be retrieved from our industry end-users in order to improve the proposed BDG tools and components. The next step is to correlate the aforementioned sensor data with earth observation data to examine the effectiveness of applying machine learning techniques.

2. **Winemaking Pilot Software Demonstrators (INRAe)-prioritised**:
   a. **Counting Grapevine Leaves**
   Based on an existing large number of grapevine images derived from the PhenoArch platform, a machine-learning pipeline was developed, aiming at counting leaves from side-view grapevine images. The counting pipeline exploits deep learning techniques based on artificial neural networks to infer the number of leaves from each grapevine image. This scenario supports the counting and visualisation of the evolution of the existing grapevine leaves on a plant over time. The leaf count will enable to feed agronomic models used in decision support systems in agriculture. In addition, in precision agriculture image analysis can enable for site-specific application in the vineyard to reduce input and maximize profit. This Software Demonstrator has the potential to be used as an extension of the Farm Management Software SITI4farmer and also in OpenSilex software. Initial feedback will be retrieved from our research and industry end-users in order to improve the proposed BDG tools and components.

   b. **From Vine to Wine: Parameters' Influencing Wine Quality**

Based on existing large & heterogeneous data provided by different data sources gathered by diverse teams at each step of wine production, including climatic, soil, harvest and winemaking data, sensory and lab analysis, the user can visualise the influence of all of these parameters on the wine quality in response to a changing environment. In order to support discovery, access & visualisation of these linked data, data integration have been achieved using web semantic and ontologies to have all data connected in a knowledge graph. This Software Demonstrator has the potential to be used in OpenSilex software and also as an extension of the Farm Management Software SITI4farmer. Initial feedback will be retrieved from our industry end-users in order to improve the proposed BDG tools and components.

c. **Gacovi "From Vine to Wine"**

Based on vine and wine data collected from the experimental unit of Pech Rouge, the user will be able to know the parameters of interest to obtain a target wine. The data available consist of field and climatic data related to plots, satellite data and lab analysis of must and wine which are quality data. The users can visualise existing associations and correlations between these data. They can know to what extent the field data is correlated with each other and influences the quality of the wine. Then, they can identify variables with weak or strong predictive power on a chosen quality variable. In order to support this tool, data integration and modelling have been achieved to have all data connected and linked. The next step was the calculation of correlations among parameters. This Software Demonstrator has the potential to be used as an extension of the Farm Management Software SITI4farmer and also in OpenSilex software. Initial feedback will be retrieved from our industry end-users in order to improve the proposed BDG tools and components.

3. **Farm Management Pilot Software Demonstrator (ABACO-GEOCLEDIAN)**:
   Water Availability and Irrigation Recommendations

The software demonstrator aims to support optimization of the Irrigation, one of the best practices that has a high impact economically on food chain industry. As a matter of fact, the water consumption in irrigation is connected to the knowledge of the real needs of the crops also in relation to final production quality and quantity. For this reason, the Decision Support System outputs should be evaluated by expert agronomist and crop technicians. The DSS, where farmers are provided of complete weather stations, as in piloting farms are fed by data gauged exactly in the farm area. For most of other farmers and demonstrators the meteorological input data come from Regional Weather Networks that provides data for free in raster format. This data format allows to obtain meteorological data with a resolution of 1 km. Others important input of the DSS are the Irrigation input inserted in SITI4Farmer module by the users and the satellite data products from Geocledian, especially the water stress index. As the UI Dashboard mock-ups will be available, we expect to collect feedback for the demonstrator in terms of usability and reliability of the DSS outputs.

4. **Natural Cosmetics Pilot Software Demonstrator (SYMBEEOSIS)**:
   Grapevine By-Products Biological Efficacy Predictor

   The Natural Cosmetics Pilot's Software Demonstrator showcases the relative dashboard dedicated to grapevine practitioners and cosmetic industry end-users. The software is visualised into a dashboard equipped with the appropriate tools that will support end-user on decision making and selection of the best grapevine by-products intended for natural cosmetic production. The software is focused on grapevine leaves' biological efficacy from 16 Greek vineyards (public/industry data), which after correlation with weather and satellite vegetation indices data-sets (public/open data), will aim to predict the origin of vineyard that could supply the best quality leaves for next year's natural cosmetics production. The demonstrator can integrate intelligence data derived from satellites and meteorological stations for the targeted vineyards, laboratory results on the biological activity of two different type extracts from leaves, and finally provides help to the end-user to answer competence questions related with incoming raw material quality, vineyards/samples' origin, correlation attributes between biological properties of sample and vineyard performance, and many other. In addition, efforts will be made to get the software linked with FOODAKAI platform, in order to incorporate food safety information on grapevine by-products as well. The performance of the present version of software demonstrator is furthermore assessed and appraised from relative end-users selected mainly from cosmetic industry (12), but also from research organisations (2) and grapevine practitioners (1). The initial feedback from demonstrator's testing has provided useful information on usefulness of the selected modules, ease of handling, successfulness of implementation, and visualisation lay-out, while has provided suggestions on improvements that could be employed to next versions of demonstrator.

5. **Food Protection Pilot Software Demonstrator (AGROKNOW):**

Risk Assessment Module, Price Prediction Dashboard and Recall Prediction

Based on real industry cases, food protection pilot identify risk in food products in order to prevent issues (recall/border rejections) in a food supply chain. For this reason, it is critical to answer these question "which is the most important chemical risk for a food product", "Which are the ingredient (eg grape) with the higher risk", "Which are the countries with the higher risk for hazards and fraud", "Which are the increasing risk for my product (raisin)". A large number of different data sources and data types has been used. We used textual information that includes mainly announcements about food recalls and border rejections (FDA, RASAFF, FSSA, AUSTERLIAN FOOD SAFETY AUTHORITY). The main source of our datasets is the open data published by the governments. Moreover, private data that includes ingredients that a company uses for food production, list of suppliers and internal lab test. All the components are developed are to extend FOODAKAI. Risk assessment module, price prediction dashboard and recall prediction have developed. From industry stakeholders we need feedback relevant to these questions "How critical is to have access to food recall and border rejections data", "How important is to access my supplier based on global food safety data ","How critical is to identify risk in my food products", "How useful was alerts module, supplier check and risk estimation"

Each pilot will present one or more of these Software Demonstrators to end-users coming from relevant to the grapevine-powered industry communities, in order to collect feedback using appropriate evaluation instruments that has a solid scientific basis on the trust, acceptance and adoption of information systems. Specifically, we will use the UTAUT (unified theory of acceptance and use of technology) model (Venkatesh et al., 2003) to understand potential acceptance and adoption, the System Usability Scale (SUS) (Brooke, 1996) to measure usability and NASA-Task Load Index (TLX) (Hart, 1986) to measure workload of the proposed tools. Additionally, to measure end-user trust towards the decision support elements, we will use custom questionnaires, based on Jian et al., 2000, in relevant pilots.

The approach selected is aiming at involving a wide range of stakeholders involved in the grapevine-powered industry. The pilots will interact with the industrial end-users' community with the survey being distributed to all relevant stakeholders involved in the BigDataGrapes piloting activities. For the purposes of this document, a stakeholder is a person that belongs to any 'stakeholder group' with an actual or potential interest in the economic, social or cultural use of the project outcomes, the BDG tools and components. The 'stakeholder approach', is used to measure organizational outcomes with a range of tools and methods appropriate for collecting feedback for an evaluation object. Therefore, the stakeholder approach incorporates an evaluation process that involves a representative group of end-users who have an interest in the key outcomes of the

BigDataGrapes. In the context of the BDG Evaluation, feedback will be asked from the end-users, with a focus to the industry end-users, including farmers, producers, owners of the reference vineyards, wine producers and other representatives from the food and cosmetic industry, as well as from the agricultural technological innovation industry and the BigDataGrapes pilot and technological partners. Finally, a small number of researchers and scientist will be asked to perform the evaluation, for supplementary feedback. The list of the end-users to be surveyed includes at least 8 industry end-users per pilot and in some of the pilots additional input from researchers and research institutions. BDG is aiming to receive two answered questionnaires per industry end-user.

Table A1: The list of the end-users surveyed per Pilot Software Demonstrator

Type of Organisation	End-User
Table and Wine Grapes Pilot Software Demonstrator	
Wine Producers	Palivos Estate Kontogiannis Estate Zacharias Winery Skouras Winery Lantides Winery Pirgakis Estate
Farmers	Nemea Wine Producers Cooperative Association
Industry	Fasoulis Nursery Agenso
Researchers	Agricultural University of Athens
Winemaking Pilot Software Demonstrators	
Industry	Nyseos
Wine Producers	Gerard Bertrand GCF Group Chateau Mercian Moët & Chandon
Researchers	IFV SPO Pech Rouge
Farm Management Pilot Software Demonstrator	
Wine Producers	Torrevilla Il Palazzo Casato Prime Donne Palivos Estate
Farmers	Consorzio Di Siena Consorzio Di Cremona
Industry	Abaco Agri-Food Department Abaco Factory-Tech Department
Natural Cosmetics Pilot Software Demonstrator	
Industry	Symbeeosis APIVITA The Nu Club Beauty Lab/The Store QS Professional

	Sarantis Frezyderm
Farmers	Aoton Winery
Researchers	Agricultural University of Athens Hellenic Agricultural Organisation
Food Protection Pilot Software Demonstrator	
Industry	Yili Coca Cola HBC Barry-Callebaut Conagra Abhold-Denze/ AB Basilopoulos Ferrero Apivita L' Uriage
Wine Producers	Gallo Winery Wines of Creet

Evaluation Methodology - Guidelines

There are several ways in which a survey may be delivered to its targeted audience. The table below gives an overview of these options. During the Evaluation of BigDataGrapes, partners will be given the freedom to use any delivery method that best helps to address their targeted audiences. The main requirement is to collect and report the results being collected to the Work Package Leader in a specific and structured electronic format (via Google Forms). All pilot partners will be provided with a document with pre-defined and quantified answers to be printed and distributed to the end-users as included in the Appendix. Nevertheless, the electronic version of the survey that will be available to all pilot partners, using Google Forms, and should be the primary choice for data collection. In the case pilot partners collect hand-written answers they are obliged to transfer them into digital format, using the Google Forms link that is provided. It is also of high importance to protect the end-users' identity and to follow the General Data Protection Regulation (GDPR). Each end-user will be assigned an ID number before the evaluation and all data will be anonymised. Each pilot leader will keep the information that is connecting each ID to an end-user in their local storage and/or a GDPR compliant cloud platform. This data should be kept securely and only accessible to the researchers involved in the evaluation. The following section presents the guidelines, templates and materials that have been developed to support the implementation of the evaluation trials.

Table A2: List of survey types

Survey Delivery Methods	
Oral feedback	Guided oral interview
Written physical feedback	Printed standardized questionnaire
Written online feedback	Standardized questionnaire via the internet

The distribution of the survey to the end-users, as part of the "Pilot Basis Evaluation" will take place over two separate occasions during the evaluation trials, on M29 and M34, and are expected to run as workshops or face-to-face interviews. However, prior to that, the "Internal Technological Evaluation" will be performed within the BDG consortium members, surveying pilot and technological partners. The first distribution of the survey to the end-users will take place a few months after the completion of the Intermediate phase, on M29, which involves the first round of controlled pilot trials and implementations of the first versions of the newly developed BigDataGrapes components. The second distribution will happen two months prior to the completion of the Summative phase, on M34, which entails the validation of the BigDataGrapes components in real-life conditions and with realistic complexity. On both occasions the evaluation trials will include the presentation of the Software Demonstrator(s) explaining the different phases and steps of the workflow. They are also expected to run as hands-on workshops during which representative end-users will have the opportunity to actually try one or more BigDataGrapes Software Demonstrators relevant to their work and interests.

Organisation:
Duration: 45-50 mins
Calling: It is strongly suggested to have direct contact with the attendees identified and selected, by emailing the end-users list or inviting them by telephone. It is recommended the Call to be made at least 2 weeks in advance to the holding of the Workshop and to provide information to potential attendees on the Software Demonstrators.
Venue: Venue allowing running the Demonstration activities.
Staff: Sufficient number of people from BigDataGrapes Partners to be involved in order to run the sessions. One or two staff people also to moderate the hands-on session.
Logistics:

- Laptop & a projector for the presentation.
- Laptops and tablets to perform the hands-on session.
- Materials: Agenda, BigDataGrapes brochure, list of attendees with contact details and evaluation form.

Program:
An outline of the different modules that need to be followed during the evaluation trials is provided to the pilot partners. This involves presentation sessions, including an introduction to the project, the piloting activities and a description of the Software Demonstrator logic, as well as workshop sessions with hands-on trial of software and completion of the evaluation questionnaire. Therefore, each evaluation trial is expected to include the following modules:

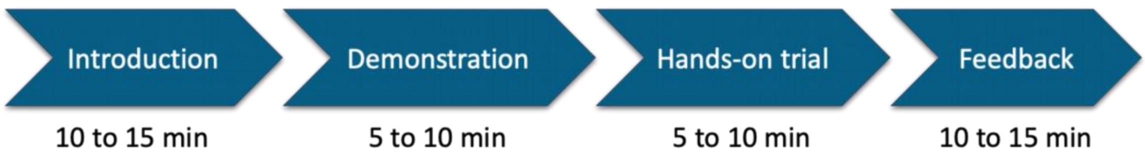


Figure A1: An overview of the procedure

a. **Introductory/opening session**
   - An introduction to the BigDataGrapes project and the piloting activities. Presentation of how the Software Demonstrators are illustrating the use of various BDG tools and components that support a complete lifecycle, from data collection and preparation, data analytics and processing, to the development of the BDG tools and components produced throughout the process.
   - Estimated duration: 10' to 15' maximum

- Supporting materials: A set of Powerpoint slides to be used as the basis for the introductory presentation (see Annex).

b. Presentation of Software Demonstrator(s)

- Explanation of the rationale behind one or more Software Demonstrators that will be presented and that are relevant to the work of the participants. Includes a presentation of the phases and steps of the workflow to be demonstrated, using PPT slides, that explains what happens at each stage, which data sets are being used, which critical decisions are supported and which software tools are being executed.
- Estimated duration: 5' to 10' maximum per Software Demonstrator
- Supporting materials: A set of Powerpoint slides or a pre-recorded video per Software Demonstrator explaining the rationale and the steps.

c. End-User hands-on trial

- Gives the participants the opportunity to execute a specific scenario on their own over the presented software tools. It should be based on a pre-defined script that helps participants perform specific steps and actions, similar to the demonstrated scenario. Gives them a real-life experience of the way that different software tools and services are being combined to execute the tasks within the presented scenario. Organisers are expected to closely monitor the process, both for supporting participants and for taking notes on verbal or non-verbal clues that are needed to be reported back to the project.
- Estimated duration: 5' to 10' maximum for the hands-on workshop
- Supporting materials: A Word document that includes the evaluation script with the specific steps that each participant should follow.

d. Evaluation feedback through the surveys

- This aims to give enough time to participants so that they can complete the evaluation questionnaire already introduced. Organisers are expected to let participants complete the questionnaires on their own.
- Estimated duration: 10' to 15' maximum for the questionnaire completion
- Supporting materials: A printed document or an online Google Form with the questionnaire to be completed by each participant.

Evaluation Material

The list below includes the material and templates developed and provided to the pilot partners in order to support the Evaluation Trials of BigDataGrapes:

- ✔ The introductory Powerpoint presentation, which is briefly describing the BigDataGrapes project, the piloting activities, the purpose of the evaluation and why getting their feedback is very important;
- ✔ A template for producing a Powerpoint presentation per Software Demonstrator, explaining the rationale and the steps;
- ✔ Pre-recorded videos of each Software Demonstration scenario;
- ✔ User scripts for each Software Demonstration scenario, allowing the end-user to reproduce the demonstrated scenario by themselves. Inside, they should be able to find clear and detailed step-by-step guidelines that they can follow closely with minimum support on their own

computers. This document should be printed and handed-out to the participants before the evaluation questionnaire.

- A document to be printed and a shared link for the electronic version of the questionnaire to be completed by each participant (in English language, pilot partners may translate it to their native language for their end-users if necessary). Most of the questions have to be answered through a 5-point Likert scale with the lower scale to be the answer "Strongly Disagree" and the higher scale the answer "Strongly Agree". In the case that pilot partners collect hand-written answers they are obliged to transfer them into digital format, using the Google Forms link provided.

Tentative Agenda for the Evaluation Trials/Workshops

A template agenda is provided to the pilot partners and may be used as a reference, to develop their own customised versions. The table below includes an example of an end-user evaluation workshop for a BigDataGrapes Software Demonstrator.

Time	Topic	Presenter / Facilitator
Introduction: 10:00-10:15	**The BigDataGrapes Project and Piloting Activities**	Maritina Stavrakaki (AUA)
Software Demonstrator: 10:15-10:25	**Software Demonstrator: Correlations between precision agriculture and phenological data** Introducing the BDG Table and Wine Grapes Pilot Software Demonstration Scenario (introductory presentation or video)	Maritina Stavrakaki (AUA)
Hands-on Trial: 10:25-10:35	Execution of a specific scenario over the presented software tools by the end-user	Maritina Stavrakaki (AUA)
Feedback: 10:35-10:50	Providing the questionnaire on a printed version or sharing link to online questionnaire Closing	Maritina Stavrakaki (AUA)

Collecting and Reporting Results - Deadlines

The distribution of the survey to the end-users, as part of the "Pilot Basis Evaluation" will take place over two separate occasions during the evaluation trials, on M29 and M34, and are expected to run as workshops or face-to-face interviews. The first distribution of the survey to the end-users will take place a few months after the completion of the Intermediate phase, on M29, as soon as the five (5) BigDataGrapes Software Demonstrators are completed. This means that pilot partners may organise their Evaluation trials, collect feedback and report results until the end of May 2020. The second distribution will happen two months prior to the completion of the Summative phase, on M34 and pilot partners should have completed their Evaluation trials by the end of October 2020. Pilot partners are free to choose the most suitable date for the evaluation procedure to take place and are to communicate to Agroknow, AUA, and KUL at least 2 weeks before it takes place, the exact dates and venue of the Workshops, for dissemination purposes.

All partners organising an Evaluation trial are expected to take photos, in the case of a hands-on event, or screenshots, in the case of remote webinars, in order to include them in the final reports. They are also expected to ensure that the maximum number of participants have completed the questionnaires. All pilot partners will be provided with a document with pre-defined and quantified answers to be printed and distributed to the end-users as included in the Appendix. Nevertheless, the electronic version of the survey that will be available to all pilot partners is strongly suggested to be selected over the printed version. In the case that pilot partners collect hand-written answers they are obliged to transfer them into digital format, using the Google Forms link provided. It is also of high importance to protect the end-users' identity and to follow the General Data Protection Regulation (GDPR). Each end-user will be assigned an ID number before the evaluation and all data will be anonymised. Each pilot leader will keep the information that is connecting each ID to an end-user in their local storage and/or a GDPR compliant cloud platform. This data should be kept securely and only accessible to the researchers involved in the evaluation. Finally, it is also suggested that during the workshops partners use posters, flyers and other kind of promotional and dissemination materials for the project.

Online Evaluation due to the COVID-19 outbreak

All partners should organise the Evaluation online if a personal contact with the end-users is not possible to happen till the end of M29. Please report to what extend this online evaluation can be followed and what will the impact be.

In order to perform the online version of the evaluation, the BigDataGrapes project management team will provide a zoom account. This zoom account can be used by all pilot partners to carry out the evaluation interviews with the end-users, making sure there will not be any overlaps with the already scheduled monthly meetings. This account does not support multiple meetings to be carried out simultaneously. All pilot partners should contact the project management team to get the zoom login details tin order to proceed with the scheduling of their interviews as needed.

The outline of the different modules to be followed during the online evaluation trials involves online presentation sessions, including an introduction to the project, the piloting activities and a description of the Software Demonstrator logic, as well as an online hands-on trial of software and completion of the evaluation questionnaire. Therefore, each evaluation trial is expected to include again the following modules:

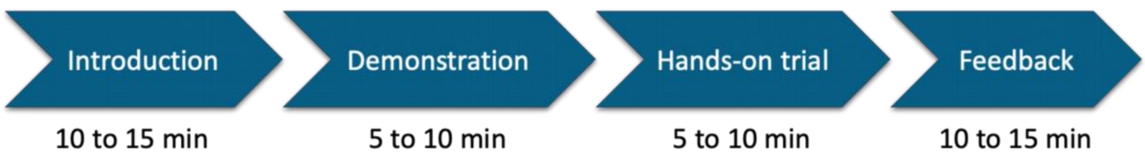


Figure A2: An overview of the online procedure

a. **Introductory/opening session**
   - **Pre-recorded video or live** introduction to the BigDataGrapes project and the piloting activities. Presentation of how the Software Demonstrators are illustrating the use of various BDG tools and components that support a complete lifecycle, from data collection and preparation, data analytics and processing, to the development of the BDG tools and components produced throughout the process.
   - Estimated duration: 10' to 15' maximum

- Supporting materials: A pre-recorded video or a set of Powerpoint slides to be used as the basis for the introductory presentation (see Annex).

b. Presentation of Software Demonstrator(s)

- **Pre-recorded video or live** explanation of the rationale behind one or more Software Demonstrators that will be presented and that are relevant to the work of the participants. Includes a presentation of the phases and steps of the workflow to be demonstrated, using a pre=recorded video or PPT slides, that explains what happens at each stage, which data sets are being used, which critical decisions are supported and which software tools are being executed.
- Estimated duration: 5' to 10' maximum per Software Demonstrator
- Supporting materials: A pre-recorded video or a set of Powerpoint slides or a pre-recorded video per Software Demonstrator explaining the rationale and the steps.

c. End-User hands-on trial

- Gives the participants the opportunity to execute a specific scenario on their own over the presented software tools. It should be based on a pre-defined script that helps participants perform specific steps and actions, similar to the demonstrated scenario. Gives them a real-life experience of the way that different software tools and services are being combined to execute the tasks within the presented scenario. Organisers are expected to closely monitor the process, both for supporting participants and for taking notes on verbal or non-verbal clues that are needed to be reported back to the project.
- Estimated duration: 5' to 10' maximum for the hands-on workshop
- Supporting materials: A Word document that includes the evaluation script with the specific steps that each participant should follow.

d. Evaluation feedback through the surveys

- This aims to give enough time to participants so that they can complete the evaluation questionnaire already introduced. Organisers are expected to let participants complete the questionnaires on their own.
- Estimated duration: 10' to 15' maximum for the questionnaire completion

Supporting materials: An online questionnaire to be completed by each participant.

APPENDIX B- BIGDATAGRAPES PILOTS' SURVEY

PILOT BASIS EVALUATION

Vineyard Information and Demographics

Please answer the following questions if you are an **industry end-user:**

Participant ID: _____ **Date:** ___/___/___

BigDataGrapes Pilot: ☐*Table and Wine Grapes* ☐*Winemaking* ☐ *Farm Management* ☐*Natural Cosmetics* ☐ *Food Protection*

[To be completed by each pilot partner during Evaluation]

Vineyard Information

Q01. What is the **total area** of the vineyard(s) of interest?

☐*Less than 1 ha* ☐*1-10 ha* ☐*10-50 ha* ☐*More than 50 ha* ☐*Not applicable*

Q02. What type of **farming system** do you practice?

☐*Conventional* ☐*Organic* ☐*Bio-dynamic* ☐*Not applicable*

Demographics

Q03. What is your **country**?

☐*Greece* ☐*France* ☐*Italy* ☐*Other*

Q04. What is your **gender**?

☐*Male* ☐*Female*

Q05. What is your **age group**?

☐*18-34* ☐*35-44* ☐*45-54* ☐*55-64* ☐*65+*

Q6. What is your **education level**?

☐*Primary* ☐*Secondary* ☐*Graduate* ☐*Post graduate*

Q07. What is the **name of your organisation** and what is **your role**?

………

……………………………………………………………………………………………………

Q08. I am generally **open to trying out new technology.**

☐*Strongly disagree* ☐*Somewhat disagree* ☐*Neutral* ☐*Somewhat agree* ☐*Strongly agree*

End-user Survey

Please answer the following questions if you are an **industry end-user: farmer, wine producer, representative from the food, cosmetic and agricultural industry or researcher:**

System Usability Scale

Instructions: For each of the following statements, mark one box that best describes your reactions to the system *today*.

	Strongly disagree				*Strongly agree*
Q1. I think that I would like to use this system frequently	1	2	3	4	5
Q2. I found the system unnecessarily complex	1	2	3	4	5
Q3. I thought the system was easy to use	1	2	3	4	5
Q4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
Q5. I found the various functions in this system were well integrated	1	2	3	4	5
Q6. I thought there was too much inconsistency in this system	1	2	3	4	5
Q7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
Q8. I found the system very cumbersome to use	1	2	3	4	5
Q9. I felt very confident using the system	1	2	3	4	5
Q10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

UTAUT Questions
Instructions: For each of the following statements, mark one box that best describes your reactions to the system.

Performance Expectancy

Statement					
PE01. I would find the system useful for my job.	*Strongly disagree*				*Strongly agree*
PE02. Using the system enables me to accomplish tasks more quickly.	*Strongly disagree*				*Strongly agree*
PE03. Using the system increases my productivity.	*Strongly disagree*				*Strongly agree*
PE04. If I use the system, I will increase my chances of getting a raise.	*Strongly disagree*				*Strongly agree*

Effort Expectancy

Statement					
EE01. My interaction with the system would be clear and understandable.	*Strongly disagree*				*Strongly agree*
EE02. It would be easy for me to become skilful at using the system.	*Strongly disagree*				*Strongly agree*
EE03. I would find the system easy to use.	*Strongly disagree*				*Strongly agree*
EE04. Learning to operate the system is easy for me.	Strongly disagree				Strongly agree

Social Influence

Statement					
SI01. People who influence my behaviour think that I should use the system.	*Strongly disagree*				*Strongly agree*
SI02. People who are important to me think that I should use the system.	*Strongly disagree*				*Strongly agree*
SI03. The senior management of this business has been helpful in the use of the system.					

	Strongly disagree				*Strongly agree*
SI04. In general, the organization has supported the use of the system.					
	Strongly disagree				*Strongly agree*
Facilitating Conditions FC01. I have the resources necessary to use the system.					
	Strongly disagree				*Strongly agree*
FC02. I have the knowledge necessary to use the system.					
	Strongly disagree				*Strongly agree*
FC03. The system is not compatible with other systems I use.					
	Strongly disagree				*Strongly agree*
FC04. A specific person (or group) is available for assistance with system difficulties.					
	Strongly disagree				*Strongly agree*

<u>NASA Task Load Index</u>

Instructions: For each of the following questions, mark one box that best describes your experience with the system. Each scale is presented as a line divided into 20 equal intervals anchored by bipolar descriptors (e g, High/Low)

Mental Demand

How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?

Very Low *Very High*

Physical Demand

How much physical activity was required? Was the task easy or demanding, slack or strenuous?

Very Low *Very High*

Temporal Demand

How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?

Very Low *Very High*

Performance

How successful were you in performing the task? How satisfied were you with your performance?

Very Low *Very High*

Effort

How hard did you have to work (mentally and physically) to accomplish your level of performance?

Very Low *Very High*

Frustration

How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?

Very Low *Very High*

The following questions were only used for Agroknow's price prediction dashboard.

Measurement of Trust

Instructions: For each of the following statements, mark one box that best describes your reactions to the system.

The prediction model is deceptive							
	Strongly disagree					*Strongly agree*	
The prediction model behaves in an underhanded manner							
	Strongly disagree					*Strongly agree*	
I am suspicious of the prediction model's outputs							
	Strongly disagree					*Strongly agree*	
I am wary of the prediction model							
	Strongly disagree					*Strongly agree*	
The prediction model's actions outputs will be harmful or injurious							
	Strongly disagree					*Strongly agree*	
I am confident in the prediction model							
	Strongly disagree					*Strongly agree*	
The prediction model is reliable							
	Strongly disagree					*Strongly agree*	
I can trust the prediction model							
	Strongly disagree					Strongly agree	
I am familiar with the prediction model							
	Strongly disagree					*Strongly agree*	

Additional Questions

Instructions: For each of the following open-ended questions, answer in your own words to describes your reactions to the system.

Feedback on trust scale questions

Which components of the visualisation influenced your answers on the previous questions?

Are there any other aspects that influenced your answers on the previous questions?

Perceived transparency

I understand why the system predicted the price as it did							
	Strongly disagree						*Strongly agree*
I understand what the system bases its prediction on							
	Strongly disagree						*Strongly agree*

Which components of the visualisation contributed to these answers?

Exit Questions

Did you understand all components in the visualisation?

INTERNAL TECHNOLOGICAL EVALUATION

Assessment Checklist

Please answer the following questions if you are a **BigDataGrapes pilot and technological partner**:

Pilot's Software Demonstrator Tasks	Status Report
Q1. Data and datasets	
Have the data sets and data sources required been identified, collected & shared?	
Which data are private and which public?	
Have those data been uploaded to the BDG platform?	
Is there a way to update those data?	
Are those data available on the website?	
Are there projected data?	
Is the data modelling completed?	
Q2. Business decisions and Competence questions	
Which are the critical business decisions to be supported?	
What are the Intelligence and data competence questions to be answered?	
Q3. Algorithm Implementations	
Is the model algorithm completed?	
Is the algorithm that fits the best identified?	
Have numerical experiments been run?	
Is the model algorithm integrated in the platform?	
Q4. Software Tool Extension & UI	
Are the visualization components' mockups designed and developed?	
Are the visualisations completed?	
Is the visualization component integrated in the platform?	
Is this Software Demonstrator an extension of SITI4farmer or FOODAKAI? Or stands alone?	
Q5. Evaluation & Dissemination activities	
Have the end-users been identified?	
Have initial feedback been collected from the end-users?	
Has a content article describing the business case of the pilot been produced?	

Global Technological Evaluation

Please answer the following questions if you are a **BigDataGrapes pilot and technological partner**:

Q6. Please select all indicators on which BigDataGrapes have an impact on:

☐ *FAIR-ness* ☐*Scalability* ☐*Resource Optimisation* ☐*Flexibility*

☐ *Consistency* ☐*Reliability* ☐*Conformity* ☐Reusability

FAIR-ness

Q7. Do the BigDataGrapes pilots…

- Help in making research data and algorithms FAIR (Findable, Accessible, Interoperable, Reusable)?

☐*Definitely not* ☐*Probably not* ☐*Neutral* ☐*Probably yes* ☐*Definitely yes*

- Support access and integrate data from various data sources and of different types (textual, SQL, RDF, images, location data, etc.)?

☐*Definitely not* ☐*Probably not* ☐*Neutral* ☐*Probably yes* ☐*Definitely yes*

Q8. Are there advantages compared to current research environments and data management practices?

☐*Definitely not* ☐*Probably not* ☐*Neutral* ☐*Probably yes* ☐*Definitely yes*

Scalability

Q9. Do the BigDataGrapes pilots allow for improved scalability in data?

☐*Definitely not* ☐*Probably not* ☐*Neutral* ☐*Probably yes* ☐*Definitely yes*

Resource Optimisation

Q10. Do the BigDataGrapes pilots enable management capabilities and resource optimisation (such as for security, compute resource management, governance and reuse)?

☐*Definitely not* ☐*Probably not* ☐*Neutral* ☐*Probably yes* ☐*Definitely yes*

Flexibility

Q11. Do the the BigDataGrapes pilots…

- Permit the addition of new data and enhanced functionality?

☐*Definitely not* ☐*Probably not* ☐*Neutral* ☐*Probably yes* ☐*Definitely yes*

- Promote the transfer of knowledge by sharing components and data?

☐*Definitely not* ☐*Probably not* ☐*Neutral* ☐*Probably yes* ☐*Definitely yes*

- Promote flexibility by the integration of open-source data and libraries?

☐*Definitely not* ☐*Probably not* ☐*Neutral* ☐*Probably yes* ☐*Definitely yes*

Consistency

Q12. Are the BigDataGrapes pilots consistent and integrated in order to support an entire data analytics pipeline?

☐*Definitely not* ☐*Probably not* ☐*Neutral* ☐*Probably yes* ☐*Definitely yes*

Q13. Do they provide a seamless end-to-end experience, to make users more productive across the whole data and analytics pipeline?

☐*Definitely not* ☐*Probably not* ☐*Neutral* ☐*Probably yes* ☐*Definitely yes*

Reliability

Q14. Are the BigDataGrapes components and tools developed for the pilots reliable enough for day-to-day use?

☐*Definitely not* ☐*Probably not* ☐*Neutral* ☐*Probably yes* ☐*Definitely yes*

Conformity

Q15. Are results from the pilots conforming to user expectations and quality standards?

☐*Definitely not* ☐*Probably not* ☐*Neutral* ☐*Probably yes* ☐*Definitely yes*

Reusability

Q16. Do the BigDataGrapes technology components support reusability? (i.e. do they have clear documentations with open access to the codes?)

☐*Definitely not* ☐*Probably not* ☐*Neutral* ☐*Probably yes* ☐*Definitely yes*