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

D2.1 - Use Cases & Technical Requirements Specification

DELIVERABLE NUMBER	D2.1
DELIVERABLE TITLE	Use Cases & Technical Requirements Specification
RESPONSIBLE AUTHOR	Panagiotis Zervas (Agroknow)

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	Panagiotis Zervas
ADDRESS	110 Pentelis Str., Marousi, GR15126, Greece
REPLY TO	pzervas@agroknow.com
PHONE	+30 210 6897 905
EU PROJECT OFFICER	Ms. Annamária Nagy
WORKPACKAGE N. | TITLE	WP2 | Grapevine-powered Industries Big Data Challenges
WORKPACKAGE LEADER	Agroknow
DELIVERABLE N. | TITLE	D2.1 | Use Cases & Technical Requirements Specification
RESPONSIBLE AUTHOR	Panagiotis Zervas (Agroknow)
REPLY TO	pzervas@agroknow.com
DOCUMENT URL	http://www.bigdatagrapes.eu/
DATE OF DELIVERY (CONTRACTUAL)	30 September 2018 (M9), 30 June 2019 (M18, Updated Version)
DATE OF DELIVERY (SUBMITTED)	28 September 2018 (M9), 28 June 2019 (M18, Updated Version)
VERSION | STATUS	2.0 | Final
NATURE	Report (R)
DISSEMINATION LEVEL	Public (PU)
AUTHORS (PARTNER)	Pythagoras Karampiperis (Agroknow), Panagiotis Zervas (Agroknow), Sotiris Konstantinidis (Agroknow), Eva Bozou (Agroknow)
CONTRIBUTORS	Milena Yankova (ONTOTEXT), Raffaele Perego (CNR), Nicola Tonellotto (CNR), Franco Maria Nardini (CNR), Stefan Scherer (GEOCLEDIAN), Arnaud Charleroy (INRA), Pascal Neveu (INRA), Aikaterini Kasimati (AUA), Maritina Stavrakaki (AUA), Simone Parisi (ABACO), Eleni Foufa (Symbeeosis), Pantelis Natskoulis (Symbeeosis)
REVIEWER	Katrien Verbert (KU Leuven)

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.1	Table of Contents	20/08/2018	Pythagoras Karampiperis (Agroknow)
0.4	Initial version	7/9/2018	Pythagoras Karampiperis (Agroknow), Panagiotis Zervas (Agroknow), Sotiris Konstantinidis (Agroknow)
0.8	Input from partners	14/9/2018	Milena Yankova (ONTOTEXT), Raffaele Perego (CNR), Nicola Tonellotto (CNR), Franco Maria Nardini (CNR), Stefan Scherer (Geocledian), Arnaud Charleroy (INRA), Pascal Neveu (INRA), Aikaterini Kasimati (AUA), Maritina Stavrakaki (AUA), Simone Parisi (ABACO), Eleni Foufa (Symbeeosis)
0.9	Internal Review	21/09/2018	Katrien Verbert (KU Leuven)
1.0	Partners review, final comments and edits	28/09/2018	Pythagoras Karampiperis (Agroknow), Panagiotis Zervas (Agroknow), Sotiris Konstantinidis (Agroknow)
1.5	Pre-final draft for 2nd version	6/6/2019	Panagiotis Zervas (Agroknow), Eva Bozou (Agroknow)
2.0	Updated version	28/6/2019	Panagiotis Zervas (Agroknow), Eva Bozou (Agroknow)

PARTICIPANTS		CONTACT
Agroknow IKE (Agroknow, Greece)		Panagiotis Zervas Email: pzervas@agroknow.com
Ontotext AD (ONTOTEXT, Bulgaria)		Todor Primov Email: todor.primov@ontotext.com
Consiglio Nazionale DelleRicherche (CNR, Italy)		Raffaele Perego Email: raffaele.perego@isti.cnr.it
Katholieke Universiteit Leuven (KULeuven, Belgium)		Katrien Verbert Email: katrien.verbert@cs.kuleuven.be
Geocledian GmbH (GEOCLEDIAN Germany)		Stefan Scherer Email: stefan.scherer@geocledian.com
Institut National de la Recherché Agronomique (INRA, France)		Pascal Neveu Email: pascal.neveu@inra.fr
Agricultural University of Athens (AUA, Greece)		Katerina Biniari Email: kbiniari@aua.gr
Abaco SpA (ABACO, Italy)		Simone Parisi Email: s.parisi@abacogroup.eu
SYMBEEOSIS LONG LIVE LIFE S.A. (Symbeeosis, Greece)		Konstantinos Gardikis Email: c.gardikis@gmail.com

EXECUTIVE SUMMARY

The deliverable D2.1, "Use Cases & Technical Requirements Specification", aims to give the outline, specify and present in detail the use cases that will be examined and undertaken during the project lifetime. It is a report documenting the use cases and the different scenarios and hypotheses that derive from them, which are directly linked and related to the BigDataGrapes pilots. More specifically, this deliverable provides the methodology that was followed by the project partners in order to define and categorize the use cases, the interpretation of the use cases with respect to their technical and infrastructural requirements, since each use case elicits to specific software requirements in terms of functionality, expected performance and required equipment.

The document is structured as follows. Chapter 1 serves as an introduction to the deliverable whereas Chapter 2 provides an overview of the use cases with details on the methodology which was followed in order to define them, their categorization and analysis as well as their importance on the domain. Each use case is separated in four sections: the scenario behind the use case, the real-life data problem to be addressed, the approach that is currently used to address the data problem, the scenario hypothesis, which practically presents a new approach to the data problem at hand thanks to the BigDataGrapes Project, the related data and their description, the technical requirements of the use case and the related pilot(s) that relate directly to the use case. Finally, Chapter 3 contains the conclusions regarding the use cases.

TABLE OF CONTENTS

LIST OF TABLES

1 INTRODUCTION

BigDataGrapes aims to explore realistic, applicable Big Data challenges in grapevine-powered industries, and as a result the early involvement of the relevant stakeholders in order to elaborate on the envisaged use cases has been defined as critical. The data problems of the targeted grapevine-powered industries require cross-sector technical solutions for the entire lifecycle of the Big Data Value Chain. While for some of the data problems entailed in the use cases there are established technologies and processes to be adopted, others require novel methodologies to ensure that the real-world demands posed by the industries are met. To this end, BigDataGrapes will produce a complete, integrated Big Data solution that responds to the different core challenges by adapting and extending existing standards and tools or, where necessary, advancing the state-of-the-art with outcomes of basic research in the relevant fields.

The driver for the project's developments is the Data Challenges faced by these industries, which have already been collected, refined and analyzed in the context of WP2, Grapevine-powered Industry Big Data Challenges (and will continue to do so) by solidifying the BigDataGrapes use cases, record and systematically analyze data assets, identify the pragmatic Big Data needs that need to be covered, and, ultimately, produce the architecture of the BigDataGrapes solution. WP2 in general aims to:

- Produce a concrete specification of the use cases where data challenges are of major significance for the relevant stakeholders;
- Elicit the technical requirements for serving these use cases, taking into account functional, operative and performance demands.

More specifically, the objectives of T2.1 "Use Cases and Requirements" are (i) the mapping of the various data problems that exist within the domain, (ii) the elicitation of the technical requirements in order to address the data problem, and (iii) the definition of the hypotheses that need to be examined in order to pave the way for experimental testing. The work of this task and its outcome, deliverable D2.1, is the main driver for the entire project since they directly linked to other activities covering a large spectrum of the project, namely activities related to WP8 and the BigDataGrapes pilots as well as to the technical WPs, which will make use of the data defined and used within the use cases and their instantiations, namely the BigDataGrapes pilots.

2 USE CASES

2.1 METHODOLOGY

The partners defined, identified and documented use cases mirroring the data challenges they face and their expectations and foreseen benefits from the application of the project's technological assets, in terms of operational, revenue and expansion gains. These are both technical and research problems that all partners would like to tackle, and they constitute the key point and are requested in the context of WP2. These requirements are expected to be based on both the partners experience as well as input from end-users.

The first step was to identify and map the different data problems that exist within the different domains. If the problem is important for the domain, then it is worth to further explore it as a use case. The second step was to examine whether there are data for a specific Use Case. In case of insufficient or no data at all, the intermediate step was to examine there would be enough time and equipment to generate the data in question. The third step was to examine if the use case is feasible and if there is the necessary experimentation space in order to carry out the related pilot(s) to verify and evaluate the use case.

2.2 USE CASES CATEGORIZATION

Three overarching (3) use cases and seven (7) relevant scenarios, in which the use cases are further divided, have been identified so far under WP2, "T2.1-Use Cases & Technical Requirements" (Table 1). Instantiations of these use cases are the pilots that have been defined in D8.1.

2.2.1 Use Case A - Earth Observation Data Anomaly Detection and Classification.

The purpose of this use case is to develop models that differentiate between Earth Observation data issues and anomalies. This is for triggering warnings to farmers and insurances concerning farm management practices or damage events. This use case is a cross-pilot use case since it is going to be addressed, tested and evaluated by all pilots.

2.2.2 Use Case B - Yield & Quality Prediction

The purpose of this use case is to leverage historical earth observation data combined with additional relevant information from the field to make educated guesses about yield and wine quality, based also on the expertise and know-how of the vine growers.

2.2.3 Use Case C - Farm Management

This use case would take care of optimizing the farm practices and of management. This would mean modelling climate, sunlight exposure, soil quality, slope and topography to predict the vine specific needs considering how different cultivation and vineyard management techniques affect grape quality and quantity.

Table 1: Use Cases and Scenarios

Use Cases (Generic)	Use Case Scenarios
A. Data Anomaly Detection & Classification	A. Earth Observation Data Anomaly Detection & Classification
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)
C. Farm Management	C1. Optimization of Farm Practices in the Vineyard C2. Management Zones Delineation for Vineyards

2.3 USE CASES ANALYSIS

Each use case is separated in seven sections, namely (a) the scenario behind the use case, (b) the real-life data problem to be addressed, (c) the approach that is currently used to address the data problem, (d) the scenario hypothesis, which practically presents a new approach to the data problem at hand thanks to the BigDataGrapes Project, (e) the related data and their description, (f) the technical requirements of the use case and (g) the related pilot(s) related directly to the use case described in D8.1.

Depending on the experimentation feasibility with respect to the required availability of data, use cases have been further divided to two sets. The first one refers to use cases that will be piloted during the first project period. The second one refers to use cases that need preparatory actions to support the required data spectrum and will be planned for the second round of BigDataGrapes pilots.

The use case entitled (A) Data Anomaly Detection & Classification has one relevant scenario named "Earth Observation Data Anomaly Detection & Classification" (Table 2).

Table 2: Earth Observation Data Anomaly Detection and Classification Scenario

Use Case	(A) Data Anomaly Detection and Classification
Scenario	*Earth Observation Data Anomaly Detection and Classification*
Real-life Problem	In order to make efficient use of Earth Observation (EO) data for Farm Management applications it is crucial to be able to differentiate between data issues and anomalies. This is not a trivial thing. This is a prerequisite to be able to provide warnings to farmers about Management practices. Anomalies detection is possible through the detection of deviations between Expectation and Observation. Inputs that can support this are: Static Heterogeneity of the field (Management Zones) & Typical patterns of expected crop development for the observed environmental conditions; Classification of anomalies should be able to differentiate between Data errors (clouds, shadows, atmospheric disturbances) & Farm Management related issues (Pests, diseases, vegetation stress through missing water or fertilizer or weather related damage).
Current Approach	Mostly manual work

Scenario Hypothesis	• **HYP1**: We are able to detect anomalies in EO data with the support of Management Zones & Typical patterns of expected crop development • **GOALS1**: ○ Find out if and how we can detect data anomalies in EO data. (what kind of data anomalies (e.g. clouds, cloud shadows, atmospheric disturbances (e.g. fire smoke), vegetation vitality decrease due to water stress, nutrient deficit, pest, disease, damage)? ○ Which spatial resolution is necessary for which data anomaly? ○ Which benefit do Farm Management Zones bring for this? ○ Which benefit do we gain from the comparison of fields with expected patterns of crop development (e.g. by comparing fields with other (reference) fields in the area) • **HYP2**: We are able to classify detected anomalies into data issues and farm management related issues • **GOALS2**: ○ Find out if and how we can classify anomalies into data issues & farm management related issues. (Understand what kind of farm management related issues are observable in EO data; ○ Develop methods to differentiate between data anomalies and farm management related issues) • **Potential methods**: ○ Time series analysis (e.g. outlier detection after curve smoothing); Data issues like clouds or shadows exhibit characteristic features (e.g. NDVI drops) in specific vegetation indexes and the multispectral images themselves while a decreased vegetation vitality shows other features ○ Single Image analysis (e.g. analysis of spectral information or indices) ○ Compare spatial patterns in images with expected patterns (Management Zones) ○ Compare time series of fields with reference time series (e.g. by generating a reference curve through aggregating all similar fields in an area and searching for deviations)	
Related Data/Description	Sentinel-2	Sentinel-2A/B MSI visible & NIR bands
	Landsat-8	LS-8 OLI visible & NIR bands
Technical Requirements	• Develop a data pre-processing component able to classify between EO data issues and anomalies. This component will rely on machine learning/deep learning libraries in order to perform the classification task. • Develop a data pre-processing component able to automatically send warning message to the specialist actors in response to certain events such that data anomaly detection.	
Related Pilot(s)	***Farm Management Pilot- ABACO***	

The use case under the generic name (B) Prediction has three (3) relevant scenarios named B1- Yield Prediction, B2- Predicting Biological Efficacy, B3- Crop Quality Prediction. More specifically, the third scenario is divided into two subcategories B3.1 Crop Quality Prediction for Optimizing Post Harvest Treatments of Table Grapes and B3.2 Crop Quality Prediction for Optimizing Winemaking (Tables 3,4,5,6).

Table 3: Quality/Yield management/prediction Scenario

<table>
<tr><th>Use Case</th><th colspan="2">(B) Prediction</th></tr>
<tr><td>Scenario</td><td colspan="2">B1- Quality/Yield management/ prediction</td></tr>
<tr><td>Real-life Problem</td><td colspan="2">Ever since the first adoptions of precision viticulture, the need for accurate yield predictions has become obvious. Yield estimations can help the growers optimize variable rate applications, the timing of harvest operations, as well as storage and shipping of their production. However, the difficulty of sampling and the lack of efficient methodologies become obstacles that greatly limit the growth and development of the sector.</td></tr>
<tr><td>Current Approach</td><td colspan="2">The traditional method used by table and wine grapes growers for predicting yield is based on weight measurement of bunches, an inefficient and time-consuming operation, which also fails to provide accurate estimations. More sophisticated yield prediction models have been developed based on data of temporal soil and weather patterns; however, the accuracy is still not adequate. Proximal and satellite data that can be converted into vegetation indices demonstrate high correlation to yield, but the number of factors that can affect the indices' values is far too great to be considered a reliable stand-alone data source. The knowledge of spatial patterns within a field is critical to select homogenous zones with site-specific input to better understand and predict the impact of weather, soil and landscape characteristics on spatial and temporal patterns of crop yields to enhance resource use efficiency at field level.</td></tr>
<tr><td>Scenario Hypothesis</td><td colspan="2">Ever since the first adoptions of precision viticulture, the need of accurate yield predictions has become obvious. Yield estimations can help the growers optimize variable rate applications, the timing of harvest operations, as well as storage and shipping of their production. However, the difficulty of sampling and the lack of efficient methodologies become obstacles that greatly limit the growth and development of the sector. The traditional method used by table and wine grapes growers for predicting yield is based on weight measurement of bunches, an inefficient and time-consuming operation, which also fails to provide accurate estimations. More sophisticated yield prediction models have been developed based on data of temporal soil and weather patterns; however, the accuracy is still not adequate. Proximal and satellite data that can be converted into vegetation indices demonstrate high correlation to yield, but the number of factors that can affect the indices' values is far too great to be considered a reliable stand-alone data source. The knowledge of spatial patterns within a field is critical to select homogenous zones with site-specific input to better understand and predict the impact of weather, soil and landscape characteristics on spatial and temporal patterns of crop yields to enhance resource use efficiency at field level.</td></tr>
<tr><td rowspan="6">Related Data/Description</td><td>Sentinel-2</td><td>Sentinel-2A/B MSI visible & NIR bands</td></tr>
<tr><td>Landsat-8</td><td>LS-8 OLI visible & NIR bands</td></tr>
<tr><td>Canopy sensing</td><td>Canopy sensing and vegetation indices data</td></tr>
<tr><td>Yield data</td><td>Historical yield data</td></tr>
<tr><td>Drone imagery</td><td>Aerial images</td></tr>
<tr><td>IoT stationary data</td><td>Soil moisture data, meteorological data</td></tr>
</table>

	Crop calendar	Records of crop growth stages and agricultural operations, yield and quality, soil analysis
	Eca sensing	Georeferenced soil electrical conductivity data
	Topographic data and elevation maps	Spatial data (boundaries and elevation data)
	Grape and berry mechanical properties	Lab analysis data
	Phenolic composition data	Lab analysis data
Technical Requirements	Develop a data analytics component able to support management decisions concerning the yield and quality of grape production (under different quality criteria).	
Related Pilot(s)	***Table and Wine Grapes Pilot- AUA***	

Table 4: Predicting Biological Efficacy Scenario

Use Case	(B) Prediction
Scenario	*B2- Predicting Biological Efficacy*
Real-life Problem	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 real-life challenge applies to both farmers and companies. Farmers during the wine making process produce valuable high-quality by-products that may be used in other industries. Nevertheless, a farmer using the state of the art doesn't exploit all the significant parameters and data that play an important role to the final quality and value of its products. The challenge is to be able to exploit data from diverse sources in order to predict some key quality parameters of the products and by-products that will eventually find an application in the industry. Potential buyers or companies on the other hand, perform market research and evaluations in order to choose suppliers of raw materials. Nevertheless, using the state of the art there is no efficient and economical way of knowing which one best suit a specific need, except trial and error by sampling and performing lab measurements on every raw material. Linking data such as the location of a domain, the weather conditions in the area or the cultivation methods can lead to conclusions regarding the most suitable supplier and raw material for a specific product with a specific biological function. The goal of the pilot is to prove the correlation between data from the field and the quality of extracts developed from vine materials.
Current Approach	Standard approach involves processing of by-products (e.g. grape seed) according to the availability of the material.
Scenario Hypothesis	The main purpose is to find how we can link crop location and weather conditions to the biological quality of the products. A company can then choose a list of suppliers for a specific need, just by evaluating crop location and weather conditions and thereby reaching conclusions regarding biological

	activity of by-products. A farmer on the other hand, can perform decision making by evaluating location and weather conditions on his field and thereby reaching conclusions regarding biological activity of its products. The farmer will then be able to make decisions on the commercialization of the by-products. This scenario hypothesis is aiming to create a predictive model that will correlate parameters concerning weather conditions and parameters linked with biological efficacy. The appropriate algorithms will be created that will use existing datasets and explore the relationship between them. Datasets concerning weather conditions will work as independent variables, while the datasets concerning biological efficacy will work as the dependent variables. A number of potential correlations will be generated (regression models?) between them. Once the correlations are generated, the selection process of the ideal correlation will focus on minimum complexity and error. This scenario hypothesis has the potential for increased scalability using additional weather and spatial data by choosing larger territories as points of interest. Our goal is to develop a decision support system (DSS) that nurtures users' trust. To achieve this goal, the system must be transparent, meaning it must be able to clearly communicate the prediction model with users and show differing effects of input variables on the model's output. Research had suggested that visual tools are the most efficient for these tasks.	
Related Data/Description	Vine leaf extract var. 1	Extraction using water soluble solvents
	Vine leaf extract var.1	Extraction using water soluble solvents
	Vine leaf extract var. 2	Extraction using water soluble solvents
	Vine leaf extract var.2	Extraction using water soluble solvents
	Sentinel-2	Sentinel-2A/B MSI visible & NIR bands
	Landsat-8	LS-8 OLI visible & NIR bands
Technical Requirements	Develop a data analytics component able to support decision making	
Related Pilot(s)	***Natural Cosmetics Pilot- Symbeeosis***	

Table 5: Crop Quality Prediction for Optimizing Post Harvest Treatments of Table Grapes Scenario

Use Case	(B) Prediction
Scenario	*B3.1- Crop Quality Prediction for Optimizing Post Harvest Treatments of Table Grapes*
Real-life Problem	Vineyards demonstrate high levels of variability of both yield and quality of the production. Selective harvesting is an example of targeted management, in which split picking of fruit at harvest is performed according to different yield and quality parameters of management zones within the field. Strategies such as selective picking may be highly profitable for grape growers; however, predictions of high accuracy are a requirement for this approach to be effective. A tool that can provide highly accurate information on crop quality can help the growers to optimize harvest, storage and processing of table grapes.

Current Approach	Currently, there is limited use of crop quality prediction in the table grapes industry by using few parameters like sugar content and berry diameter. These parameters are usually used for determining only the time of harvest on the field using berry refractometers and calipers while these measurements are conducted only on specific plants and as a result there is no information on table grapes quality for the whole field.	
Scenario Hypothesis	The table grapes growers are in need of a system that will help them 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 supermarkets. 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.	
Related Data/Description	IoT stationary data	Soil moisture data, meteorological data
	Canopy sensing	Canopy sensing and vegetation indices data
	Yield data	Historical yield data
	Drone imagery	Aerial images
	Crop Calendar	Records of crop growth stages and agricultural operations, yield and quality, soil analysis
	Eca sensing	Georeferenced soil electrical conductivity data
	Topographic data and elevation maps	Spatial data (boundaries and elevation data)
	Grape and berry mechanical properties	Lab analysis data
	Sentinel-2	Sentinel-2A/B MSI visible & NIR bands
	Landsat-8	LS-8 OLI visible & NIR bands
	Phenolic composition data	Lab analysis data
Technical Requirements	Develop a data analytics component able to support crop quality prediction	
Related Pilot	***Table and Wine Grapes Pilot- AUA***	

Table 6: Crop Quality Prediction for Optimizing Winemaking Scenario

Use Case	(B) Prediction
Scenario	*B3.2- Crop Quality Prediction for Optimizing Winemaking*
Real-life Problem	Wine making needs knowledge on grape quality at harvest. Different sugar content in wine grapes can produce wine with different characteristics. Moreover, some quality parameters like the concentration of nitrogen in wine grapes can affect the vinification process. As a result, wine makers have no idea on the quality of the wine grapes that they buy from the wine grape growers and adapt their vinification process accordingly while growers cannot obtain higher selling prices for their products due to better quality. Moreover, during wine making process, the effects of each step on the final wine quality are well known. The quality of wine is mainly due to human actions after harvest. The main current purpose is to understand how to improve grapes quality to make a good wine and which parameters drive grapes quality.

<table>
<tr><td>Current Approach</td><td colspan="2">Currently, there is limited use of crop quality prediction in wine grapes industry by using few parameters like sugar content and titratable acidity. These parameters are usually used for determining only the time of harvest on the field using berry refractometers or by collecting berry samples from few plants that take some time for providing the results. Thus, there is limited information on wine grape quality for the whole field.</td></tr>
<tr><td>Scenario Hypothesis</td><td colspan="2">The main purpose is to know the variables that have an effect on quality at harvest and then on the final product (wine).

<u>Goals:</u>

To identify the relevant leverage (vintage, yield) on a quality measurement (polyphenol content, aromas for example) to optimize wine making.
To link quality parameters with environmental data

<u>Hypothesis</u>: Which parameters (environment, genetic, …) have the main impact on grapes quality to optimize winemaking? to optimize alcoholic fermentation (fermentation strategy)? How to obtain a satisfying grapes quality at harvest to make a good wine and have a satisfying alcoholic fermentation / aroma composition?

Understand what is going to influence alcoholic fermentation / aroma production for a variety of vine (year of production, environment) to be able to have an optimized fermentation and winemaking strategy.
Maybe we need to work separately in function of the type of wine (red, white, rosé)
Start from quality parameters and see which data (inputs) impact them.

</td></tr>
<tr><td rowspan="8">Related Data/ Description</td><td>Genetic Data</td><td>Genetic profile, Morphological description, origin, etc.</td></tr>
<tr><td>Soil / Plot characteristics
[Plot management]</td><td>soil type, treatments, borders</td></tr>
<tr><td>Harvest information
[Plot management]</td><td>weight, date</td></tr>
<tr><td>Climatic data</td><td>Rainfall, temperature, radiation etc.</td></tr>
<tr><td>Laboratory analysis on must, grape and wine
[Grape and berry mechanical and chemical properties]
[Qualitative and quantitative characteristics of must]</td><td>Sugar content, alcohol, pH etc.</td></tr>
<tr><td>Alcoholic fermentation monitoring - online data
[Winemaking activities]</td><td>Co2 flow rate, fermentation kinetic…</td></tr>
<tr><td>Winemaking activities</td><td>Steps of winemaking, Bioconversion of sugar into ethanol and CO2, Monitoring of alcoholic fermentation and sugar content, yeast characteristics etc.</td></tr>
<tr><td>Sensory analysis</td><td>Expert panel of tasters’ sensory analysis (wine bitterness, astringency, phenol content, aroma etc.)</td></tr>
</table>

	Sentinel-2	Sentinel-2A/B MSI visible & NIR bands
	Landsat-8	LS-8 OLI visible & NIR bands
Technical Requirements	Develop a data analytics component able to support selling price decisions.	
Related Pilot	***Wine Making Pilot- INRA***	

The use case by the name (C) Farm Management, has two (2) scenarios C1- Optimization of Farm Practices in the Vineyard and C2- Management Zones Delineation for Vineyards (Table 7, 8).

Table 7: Optimization of Farm Practices in the Vineyard Scenario

Use Cases	(C) Farm Management	
Scenario	*C1- Optimization of Farm Practices in the Vineyard*	
Real-life Problem	Management Practices as irrigation, fertilization and phytochemicals are regularly over or underestimated with respect to the real plant needs. Especially in case of overestimation in quantities there is a negative countereffect towards the plants.	
Current Approach	Currently, cropping, irrigation levels, fertilization, spraying and pruning can only be adjusted at the vineyard block level and do not account for individual vine requirements.	
Scenario Hypothesis	For the last campaign 2018, in the two Italian vineyards pilots we will perform a data analysis crossing mid and high-resolution multispectral satellite data and best practices data from the farmer. The hypothesis and relative goals are the following: • **HYP 1**: The fertilization or phytochems spraying actions can be supported by satellite data. • **GOAL 1**: Before and after the management actions are there detectable differences in satellite data? Is an increasing of the plant health detectable afterwards? (e.g. increases in certain vegetation indexes or changes in patterns on the field (homogenization?)) • **Goal 2**: Are the actions justified by a real evident problem in the vineyard? (is an anomaly (see above) detectable before the event?) • **Goal 3:** Do the satellite data provide a benefit to the farmer when he has to decide on fertilization? Additional for the next campaign (2019) • **HYP 2**: There are interaction effects between weather, management practices and vegetation/fruit qualities that we are able to detect (what kind of effects? Methods: Machine Learning?) • **Goal 4**: Develop a model to predict where and when a fertilization is required • **Goal 5**: Provide a real Decision Support System to the farmer to improve grape quality and quantity.	
Related Data/Description	SENTEK DRILL & DROP TRISCAN	Soil moisture and temperature profiler from 10 cm to 60 cm deep
	Rain gauge Metos weather station	Rain gauge to measure precipitations occurred in a certain period
	Air temperature/humidity Metos weather station	measurements of air temperature/humidity

	IR Temperature Metos weather station	measurements of leaf / bunches IR temperatures
	Leaf moisture sensor	measurement of minutes in which leaf is humid
	Soil profile	soil data survey on texture and chemical properties
	Sentinel-2	Sentinel-2A/B MSI visible & NIR bands
	Landsat-8	LS-8 OLI visible & NIR bands
Technical Requirements	Develop a data analytics component able to support management decisions regarding the quality of the wine production.	
Related Pilot(s)	***Table and Wine Grapes Pilot- AUA, Farm Management Pilot- ABACO***	

Table 8: Management Zones Delineation for Vineyards Scenario

Use Cases	(C) Farm Management
Scenario	*C2- Management Zones Delineation for Vineyards*
Real-life Problem	Delineation of management zones has provided many advantages to the table and wine grape growers by decreasing their input costs and increasing their production value due to selective harvesting. Management zones represent subfield regions within a field with homogeneous characteristics and allow for site-specific management. Nevertheless, the determination of subfield areas is difficult due to the complex factors that can affect crop yield. The optimum number of zones to use when dividing a field may vary from year to year and are mainly functions of the following multivariate spatial and temporal attributes: yield maps, soil and topographic properties, electrical conductivity data, remote sensing, vegetation indices and weather data. Current methodologies for the delineation of management zones includes the following: 1) the use of principal component analysis (PCA) to summarize and aggregate the selected datasets, 2) the Management Zone Analyst (MZA) software using a fuzzy c-means unsupervised clustering algorithm that assigns field information into like classes or potential management zones, and 3) using multivariate geostatistical analysis, a method based on the coefficient of variation (CV) of each data. While these approaches address the complexity of delineating management zones, they do not imply linear, cost-effective and practical usage when more than two variables are introduced to the models. Spatiotemporal modelling, compared to static input, is a challenging task since it includes input dynamics as part of the problem. Additionally, there is insufficient information regarding efficient algorithms that combine further data layers over different spatial scales, in order to define the management zones based on more variables.
Current Approach	Delineation of management zones has provided many advantages to the wine grape growers by decreasing their input costs and increasing their production value due to selective harvesting. Delineation of management zones is done using satellite/drone imagery and/or soil measurements (soil type, nutrient concentration, terrain elevation). The ultimate goal is to achieve maximum differentiation among classes (zones) and minimum differentiation within each class. This enables for delineating management zones for different operations such as fertilizer application, pruning, spraying, irrigation and selective harvesting in order to reduce the infield variation or take benefit from it. However, the delineation of management zones is affected by

	numerous factors, such as the agricultural operations, demonstrating in-field temporal variability and resulting in suboptimal management strategies.	
Scenario Hypothesis	This scenario hypothesis is aiming to create a dynamic tool to delineate management zones. Algorithms will be developed (WP4) to address upcoming important developments of big data, smart farming and open source satellite data. They will aim at extending the existing methodologies using a dynamic multivariate approach and adopting algorithm designs for solving problems that address aspects of the spatiotemporal domain. The proposed method for the delineation of management zones is the use of algorithms that will convert a high dimensional input signal into a simpler low dimensional discrete signal, such that the distance and proximity relationships and topology are preserved. The algorithms will generate a number of options for delineating management zones, taking into consideration two or more variables. Once the options are generated using specific data spectrums, the selection process of the ideal option will focus primarily on the minimum dimensionality, lowest cost and minimum complexity criteria. Those options that are being characterized, as multidimensional, costly and/or complex will be eliminated from the final selection process of the optimal management zones delineation for a specific field and data spectrum. The selection process will depend on the following rules: (1) the number of management zones in which the field is divided should be feasible and dependent on the field size (2) use variable weighting factors by ranking variables according to their importance 3) supports resource optimization (e.g. for irrigation, fertilization). In addition to the above stated rules, factors such as the existing topography, as well as the minimum actuation from current status, including the complexity of equipment reinstallations and the distance from current state, should be taken into consideration for the final selection of the ideal management zones delineation. This scenario hypothesis has the potential for increased scalability using additional spatial and temporal data.	
Related Data/Description	Eca sensing	Georeferenced soil electrical conductivity data
	Topographic data and elevation maps	Spatial data (boundaries and elevation data)
	Yield data	Historical yield data
	Canopy sensing	Canopy sensing and vegetation indices data
	IoT stationary data	Soil moisture data, meteorological data
	Drone imagery	Aerial images
	Crop calendar	Records of crop growth stages and agricultural operations, yield and quality, soil analysis
	Grape and berry mechanical properties	Lab analysis data
	Phenolic composition data	Lab analysis data
	Sentinel-2	Sentinel-2A/B MSI visible & NIR bands
	Landsat-8	LS-8 OLI visible & NIR bands
Technical Requirements	Develop a data analytics component able to support management decisions concerning the delineation of management zones.	

Related Pilot(s)	***Table and Wine Grapes Pilot- AUA, Farm Management Pilot- ABACO***

3 CONCLUSIONS

This deliverable, the "Use Cases & Technical Requirements Specification", belongs to WP2, which aims to identify and solidify the BigDataGrapes use cases, record and systematically analyse data assets, identify the pragmatic Big Data needs that have to be covered, and, ultimately, produce the architecture of the BigDataGrapes solution.

As identified through this report, all use cases describe different data challenges and data problems of the various aspects of the domain, which will be addressed and evaluated through the instantiations of the Use Cases, the BigDataGrapes pilots.

The project's Use Cases and Technical Requirements presented in this report is the basic driver to an efficient definition and evaluation of the use cases, their scenarios and hypotheses of the project and it is aligned with the project vision and objectives. Moreover, it is directly linked with other activities of the project, namely the BigDataGrapes pilots as well as the technical WPs.

4 REFERENCES

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.

Arnó Satorra, J., Martínez Casasnovas, J. A., Ribes Dasi, M., & Rosell Polo, J. R. (2009). Precision viticulture. Research topics, challenges and opportunities in site-specific vineyard management. Spanish Journal of Agricultural Research, 7(4), 779-790.

Arnó, J., Rosell, J. R., Blanco, R., Ramos, M. C., & Martínez-Casasnovas, J. A. (2012). Spatial variability in grape yield and quality influenced by soil and crop nutrition characteristics. Precision Agriculture, 13(3), 393-410.

Battilani, P., Pietri, A., & Logrieco, A. (2004). Risk assessment and management in practice: ochratoxin in grapes and wine. Mycotoxins in food, 244.

Brown, M. E., Pinzon, J. E., Didan, K., Morisette, J. T., & Tucker, C. J. (2006). Evaluation of the consistency of long-term NDVI time series derived from AVHRR, SPOT-vegetation, SeaWiFS, MODIS, and Landsat ETM+ sensors. IEEE Transactions on Geoscience and Remote Sensing, 44(7), 1787-1793.

Cunha, C. R., Peres, E., Morais, R., Oliveira, A. A., Matos, S. G., Fernandes, M. A., ... & Reis, M. J. C. S. (2010). The use of mobile devices with multi-tag technologies for an overall contextualized vineyard management. Computers and Electronics in Agriculture, 73(2), 154-164.

Dunn, G. M., & Martin, S. R. (2004). Yield prediction from digital image analysis: A technique with potential for vineyard assessments prior to harvest. Australian Journal of Grape and Wine Research, 10(3), 196-198.

EFSA (European Food Safety Authority) (2017). Peer review report to the conclusion regarding the peer review of the pesticide risk assessment of the active substance trifloxystrobin. EFSA Journal 2017, 15(10):4989, 29 pp. https://doi.org/10.2903/j.efsa.2017.4989

Kitchen, N. R., Sudduth, K. A., Myers, D. B., Drummond, S. T., & Hong, S. Y. (2005). Delineating productivity zones on claypan soil fields using apparent soil electrical conductivity. Computers and Electronics in Agriculture, 46(1-3), 285-308.

Lamb, D. W., Weedon, M. M., & Bramley, R. G. V. (2004). Using remote sensing to predict grape phenolics and colour at harvest in a Cabernet Sauvignon vineyard: Timing observations against vine phenology and optimising image resolution. Australian Journal of Grape and Wine Research, 10(1), 46-54.

Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. Journal of Geophysical Research: Atmospheres, 117(D8).

Nuske, S., Achar, S., Bates, T., Narasimhan, S., & Singh, S. (2011). Yield estimation in vineyards by visual grape detection. In Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on (pp. 2352-2358). IEEE.

Pedersen, S. M., & Lind, K. M. (2017). Precision Agriculture–From Mapping to Site-Specific Application. In Precision Agriculture: Technology and Economic Perspectives (pp. 1-20). Springer, Cham.

Poulsen, M. E., Hansen, H. K., Sloth, J. J., Christensen, H. B., & Andersen, J. H. (2007). Survey of pesticide residues in table grapes: determination of processing factors, intake and risk assessment. Food additives and contaminants, 24(8), 886-895.

Somma, S., Perrone, G., & Logrieco, A. F. (2012). Diversity of black Aspergini and mycotoxin risks in grape, wine and dried vine fruits. Phytopathologia Mediterranea, 131-147.

Tan, P., Steinbach, M., Kumar, V., Potter, C., Klooster, S., & Torregrosa, A. (2001). Finding spatio-temporal patterns in earth science data. In KDD 2001 Workshop on Temporal Data Mining (Vol. 19).

Vitalif, M., Guidotti, M., Giovinazzo, R., & Cedronet, O. (1998). Determination of pesticide residues in wine by SPME and GC/MS for consumer risk assessment. Food Additives & Contaminants, 15(3), 280-287.