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

D8.4 - Integration and Operation with real-life Practices

DELIVERABLE NUMBER	D8.4
DELIVERABLE TITLE	Integration and Operation with real-life Practices
RESPONSIBLE AUTHOR	Pantelis Natskoulis (SYMBEEOSIS)

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 (AUA)
DELIVERABLE N. | TITLE	D8.4| Integration and Operation with real-life Practices
RESPONSIBLE AUTHOR	Pantelis Natskoulis (Symbeeosis)
REPLY TO	p.natskoulis@gmail.com
DOCUMENT URL	http://www.bigdatagrapes.eu/
DATE OF DELIVERY (CONTRACTUAL)	30 June 2019 (M18), 30 June 2020 (M30), 31 December 2020 (M36)
DATE OF DELIVERY (SUBMITTED)	28 June 2019 (M18), 31 July (M31), 31 December 2020 (M36)
VERSION | STATUS	3.0 | Final
NATURE	R (Report)
DISSEMINATION LEVEL	PU (Public)
AUTHORS (PARTNER)	Pantelis Natskoulis (SYMBEEOSIS)
CONTRIBUTORS	Aikaterini Kasimati (AUA), Florian Schlenz (GEOCLEDIAN), Nyi-Nyi Htun (KULeuven), Diego Rogo Garcia (KUL), Nikola Tonellotto (CNR), Vinícius Monteiro de Lira (CNR), Ioanna Polychronou (Agroknow), Giannis Stoitsis (Agroknow) Francesca Tsaropoulou (Agroknow) Mihalis Papakonstadinou (Agroknow), Iliana Giannelou (Agroknow)
REVIEWER	Aikaterini Kasimati (AUA)

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.1	Initial ToC and document structure	11/06/2019	Pantelis Natskoulis (SYMBEEOSIS)
0.4	Initial Draft Deliverable	14/06/2019	Pantelis Natskoulis (SYMBEEOSIS)
0.8	Input from partners	15/06/2019	Aikaterini Kasimati (AUA), Florian Schlenz (GEOCLEDIAN), Nyi-Nyi Htun (KULeuven), Nikola Tonellotto (CNR), Mihalis Papakonstadinou (Agroknow), Panagiotis Zervas (Agroknow)
0.9	Internal Review	17/6/2019	Aikaterini Kasimati (AUA)
1.0	Partners review, final comments and edits	28/06/2019	Pantelis Natskoulis (SYMBEEOSIS)
1.8	Input update from partners	15/06/2020	Aikaterini Kasimati (AUA), Florian Schlenz (GEOCLEDIAN), Nyi-Nyi Htun (KULeuven), Diego Rogo Garcia (KUL), Franco Maria Nardini (CNR), Vinícius Monteiro de Lira (CNR), Ioanna Polychronou (Agroknow)
2.0	Updated Version	28/06/2020	Pantelis Natskoulis (SYMBEEOSIS), Giannis Stoitsis (Agroknow) Francesca Tsaropoulou (Agroknow) Mihalis Papakonstadinou (Agroknow), Iliana Giannelou (Agroknow)
2.1	Updated TOC	1/7/2020	Pantelis Natskoulis (SYMBEEOSIS)
2.5	Updates on the data	10/12/2020	Pantelis Natskoulis (SYMBEEOSIS)
2.8	Addition of section 4 Food Protection Pilot	15/12/2020	Giannis Stoitsis (Agroknow) Francesca Tsaropoulou (Agroknow) Mihalis Papakonstadinou (Agroknow), Iliana Giannelou (Agroknow)
2.9	Internal Review	30/12/2020	Aikaterini Kasimati (AUA)
3.0	Updated Version	31/12/2020	Giannis Stoitsis (Agroknow) Francesca Tsaropoulou (Agroknow) Mihalis Papakonstadinou (Agroknow), Iliana Giannelou (Agroknow)

PARTICIPANTS		CONTACT
Agroknow IKE (Agroknow, Greece)		Nikos Manouselis Email: nikosm@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)		KatrienVerbert Email: katrien.verbert@cs.kuleuven.be
Geocledian GmbH (GEOCLEDIAN Germany)		Stefan Scherer Email: stefan.scherer@geocledian.com
Institut National de la Recherché Agronomique (INRA, France)		Pascal Neveu Email: pascal.neveu@inra.fr
Agricultural University of Athens (AUA, Greece)		Katerina Biniari Email: kbiniari@aua.gr
Abaco SpA (ABACO, Italy)		Simone Parisi Email: s.parisi@abacogroup.eu
SYMBEEOSIS EY ZHN S.A. (Symbeeosis, Greece)		Konstantinos Rodopoulos Email: rodopoulos-k@symbeeosis.com

ACRONYMS LIST

BDG	BigDataGrapes
WP	Work Package
D	Deliverable
DSS	Decision Support System
REST API	Representational State Transfer Application Program Interface
SVIs	Spectral Vegetation Indices
Vit	Vitality
NDVI	Normalized Difference Vegetation Index
NDRE1	Normalized Difference Red Edge Index (v1)
NDRE2	Normalized Difference Red Edge Index (v2)
NDRE3	Normalized Difference Red Edge Index (v3)
NDWI	Normalized Difference Water Index
SAVI	Soil Adjusted Vegetation Index
EVI2	Enhanced Vegetation Index 2
CI-RE	Chlorophyll Index - Red Edge
NCPRI	Normalized Pigment Chlorophyll Ratio Index
MAC	Maceration
UAE	Ultrasound Assisted Extraction
BA	Biological Activity
pH	Potential of Hydrogen
RI	Refractive Index
TMC	Total Microbial Count
Y&M	Yeast and Moulds count
AA1 (DPPH)	Antioxidant Activity 1 (2,2-DiPhenyl-1-PicrylHydrazyl)
AA2 (ABTS)	Antioxidant Activity 2 (2, 2'-Azino-Bis-3-ethylbenzoThiazoline-6-Sulfonic acid)
TPC	Total Phenolic Content
TFC	Total Flavonoids Content
WD	Weather Data

EXECUTIVE SUMMARY

The deliverable D8.4 “Integration with existing real-life Practices” showcases the application of BigDataGrapes technologies in data-intensive and critical operations related to the Natural Cosmetics Pilot and the Food Protection Pilot.

More specifically in the case of Natural Cosmetics Pilot, the ultimate goal was to prepare a software platform with the form of a dashboard that will expose the required functionality to practitioners in the grapevines and the end-users of relative cosmetic industries, during realistic operations and processes of these stakeholders. The dashboard uses the results of predictive analysis over data samples managed by expert users and showcases the ability to help decision-making by end-users based on a small subset of exhibits in comparison to the amount that a human should manually check. In this final version of the deliverable the scope is to showcase the initial hypothesis that BigDataGrapes (BDG) software stack could serve for the prediction of the Biological Activity (BA) parameters by incorporating their correlation analysis to Satellite-based Spectral Vegetation Indices (SVIs), while, as an additional feature, to incorporate also Weather Data (WD) for correlation analysis with BA parameters. This final version of the deliverable presents the development of the dashboard and its visualisation developed during the lifetime of the project, and its enhancement with additional input data and visualisation outputs, according to end-user needs, to address their critical decisions in natural cosmetic industries.

D8.4, “Integration with existing real-life Practices”, is based on the piloting plan of the Natural Cosmetics Pilot (SYMBEEOSIS) and the BA data collected from samples all around Greece, with GEOCLEDIAN providing the SVIs datasets, SYMBEEOSIS also collecting the WD from meteorological stations of Institute for Environmental Research (National Observatory of Athens), CNR undertaking the data correlation analysis, Ontotext the data modelling, Agroknow the data management and their appropriate transformation for uploading to the software stack, and KU Leuven the visualisation of the dashboard.

In the case of Food Protection pilot, the goal was to deliver to the food safety and quality assurance (FSQA) expert an online platform, namely FOODAKAI, that a) can monitor risks associated with any supplier, any ingredient or any product, b) can be customized to serve everyone in your safety, quality & compliance teams and c) will reduce by 50% the time devoted to food risk monitoring & assessment tasks. The FOODAKAI platform uses the Big Data software stack developed in Big Data Grapes to collect large amount of food safety incidents, to process this data, to enrich and to build prediction model that can help the FSQA experts to prevent food safety incidents in the food supply chain. D8.4 “Integration with existing real-life Practices” presents how the FOODAKAI platform was integrated into the real-life practices of the FSQA departments of food companies and how it can help them to move from reaction to prevention using the food safety analytics and risk predictions.

TABLE OF CONTENTS

1 Introduction 9

2 Natural Cosmetics Pilot 11

2.1 Specific Goals 11

2.2 Technological Guidance 12

2.3 Measurements 13

2.4 Envisaged Outcome 13

3 Data, Datasets and Use Case Scenario 15

3.1 Data and Datasets 15

3.2 Use Case Scenario - Predicting Biological Efficacy 16

3.3 Data Analytics And Processing 16

3.3.1 Satellite Data 17

3.3.2 Weather Data 21

3.3.3 Biological Activity Data 21

3.3.4 Data Correlation Analysis 22

3.3.5 Software Stack, Data Transformation and Uploading 25

3.4 Visualization 30

4 Food Protection Pilot 33

4.1 Current Process for preventing the incidents 33

4.2 Benefits from the automated process 34

4.3 Difference between current Risk Assessment approach and Global Predictions approach 34

4.4 Benefits in cost 35

4.5 A real-life prediction use case 36

4.6 Feedback from Food Companies about the benefits of the automated process 39

5 Conclusions 41

6 References 42

7 Appendix 43

7.1 Correlation analysis results for Maceration 43

7.2 Correlation analysis results for Ultrasound Assisted Extraction 45

LIST OF TABLES

Table 1: Vineyards chosen for sample collection (Correspondent file: https://tinyurl.com/y2y5fwld) 12
Table 2: Measurements of biological efficacy of developed extracts –Data description.................................... 13
Table 3: Natural Cosmetics Pilot Data and Datasets... 15
Table 4: SVIs and relative description.. 17
Table 5: Weather data and relative description.. 21
Table 6: Wineries of samples included in correlation analysis 22
Table 7: Most correlated SVI for each BA parameter after MAC. 23
Table 8: Most correlated SVI for each BA parameter after UAE..23
Table 9: Most correlated BA parameter for each SVI after MAC..24
Table 10: Most correlated BA parameter for each SVI after UAE. 24
Table 11: Problems of the current process and benefits from an automated risk prediction process 34
Table 12: Return of Investment tool developed to measure the benefits in cost...35

LIST OF FIGURES

Figure 1: Dispersion of samples across the Greek territory (Correspondent file: https://tinyurl.com/y4scyhed) 12
Figure 2: Data gathering timeline for Natural Cosmetics Pilot...16
Figure 3: A Sentinel-2 RGB image for 13th June 2019 for a parcel of the Pavlidis estate, Greece. Also shown are all available Landsat 8 and Sentinel-2 image acquisitions for 2017 - 2019...19
Figure 4: A Sentinel-2 NDVI time series for 2017 - 2019 for a parcel of the Pavlidis estate, Greece showing also the NDVI image for 13th June 2019. 19
Figure 5: A Sentinel-2 NDWI time series for 2017 - 2019 for a parcel of the Pavlidis estate, Greece showing also the NDVI image for 13th June 2019. This NDWI is known to be related to leaf water content............................20
Figure 6: A Sentinel-2 CI-RE time series for 2017 - 2019 for a parcel of the Pavlidis estate, Greece showing also the NDVI image for 13th June 2019. The Chlorophyll Index-Red Edge (CI-RE) is reported to be highly correlated with canopy Chlorophyll and Nitrogen contents..20
Figure 7: An abstract view of the BDG stack components used in Symbeeosis data flows................................25
Figure 8: The initial view of the UI tool for the dataset upload for the Symbeeosis case....................................26
Figure 9: The screen responsible for the matching of the uploaded xls schema with the Symbeeosis one 26
Figure 10: The edit/finalization screen of the uploaded data...27
Figure 11: A sample of the RDF generated for Symbeeosis 27
Figure 12: Swagger documentation of the API endpoint for the dataset upload...28
Figure 13: Swagger documentation of the API endpoint for distinct record upload...28
Figure 14: Swagger documentation of the API endpoint for the field metadata/description 28
Figure 15: Swagger documentation of the API endpoint for the rdfization of SYMBEEOSIS data 29
Figure 16: A sample SPARQL query on the SYMBEEOSIS data in GraphDB..29
Figure 17: The dashboard configuration interface of Grapevine By-Products Biological Efficacy Predictor. 31
Figure 18: Correlation output showing the analysis of input variables and their correlation with B.A parameter of interest (on the left side), as well the descending order of parcels according to inputs for biological efficacy of leaves samples (on the right side). 31
Figure 19: View Details" output for the selected parcel showing all relative information according to the end-user's final decision ". 32
Figure 20: Example of a food recall that could be prevented using the Global Predictions Dashboard 36
Figure 21: Increase in the incidents for sesame predicted by the Global Predictions Dashboard........................37
Figure 22: Prediction of hazards that will likely increase within the next months...37
Figure 23: Predictions of risks for sesame seeds 38
Figure 24: Products and suppliers that may be affected by the predicted risks..38
Figure 25: Validation of the expected benefits that the automated system can provide 39

Figure 26: Validation of the expected benefits that the automated system can provide in terms of saving time 40
Figure 27: Validation of the most important problems that the FSQA experts have when it comes to risk prevention........ 40
Figure 28: Correlation analysis for time frame "from the begging of the year until March" for MAC........43
Figure 29: Correlation analysis for time frame "from the begging of the year until April" for MAC........43
Figure 30: Correlation analysis for time frame "from the begging of the year until May" for MAC 44
Figure 31: Correlation analysis for time frame "from the begging of the year until June" for MAC 44
Figure 32: Correlation analysis for time frame "from the begging of the year until March" for UAE........45
Figure 33: Correlation analysis for time frame "from the begging of the year until April" for UAE........45
Figure 34: Correlation analysis for time frame "from the begging of the year until May" for UAE 46
Figure 35: Correlation analysis for time frame "from the begging of the year until June" for UAE 46

1 INTRODUCTION

The final version of Deliverable 8.4 "Integration and Operation with real-life Practices" aims to deliver a dashboard, targeting industry-level decision makers (natural cosmetics industry end-users) and practitioners (grapevine farming end-users), that incorporates the appropriate functionalities of the BigDataGrapes software stack used in the relevant piloting session "Natural Cosmetics Pilot" and the "Food Protection Pilot".

The Deliverable 8.4 reports the findings and the developmental stages that lead to the refinement of the dashboard. According to the Natural Cosmetics Pilot there is a need in extracting the most out of the incoming to the industry raw materials for both economic and environmental reasons. On the other hand, wine-making procedures produce a lot of by-products that may have a significant biological value if there are adequate data concerning farm management. For this purpose, a total of sixteen (16) grapes' leaves samples of different origin were collected from 15 farms located all over the Greek territory. The samples were subjected to the relevant laboratory analyses for assessment of their BA and compatibility for natural cosmetic production. At the same time, and during a specific period related to sampling, SVIs from satellite scanning and WD from meteorological stations were collected for all sampled vineyards and were correlated with the BA parameters.

Following the initial correlation analysis, the building of the models supports the DSS implementation. The dashboard that exposes the required functionality to the practitioners and cosmetic industry's end-users was developed and used during realistic operations and processes of these stakeholders. For this purpose, the software demonstrator of the dashboard presents the results of predictive analysis over data samples managed by expert users, showcasing their ability to inform decision-making based on a small subset of exhibits in comparison to the amount that a human should manually check. The demonstrator went under a narrow evaluation, as presented in the relative deliverable, D8.5, by interested end-users through a survey enhancing the dashboard refinement, so that the appropriate usability and performance, will finally meet their needs within everyday real-time practices in the grapevine industry. The results of the finalised automated system were compared by end-users with those of a full manual analysis to indicate the precision of the system and quantify the benefits in time, effort and cost.

In addition to the Natural Cosmetics pilot, deliverable D8.4 focuses on the application of the BigDataGrapes technologies in data-intensive and critical operations that are performed by the Food Safety and Quality Assurance (FSQA) experts in the Food industry in order to prevent food safety incidents. The FOODAKAI global incidents predictions dashboard that was developed by Agroknow using the Big Data Grapes software stack was used during realistic operations and processes of the FSQA departments. As already presented in deliverable D8.3, the Global Predictions Dashboard presents the results of predictive analysis to FSQA experts, showcasing their ability to inform decision making based on the risk predictions for any raw material and ingredient including the ones used in the grapevine industry. This deliverable focuses on the comparison of the current processes used to prevent food safety incidents with the one that is supported by the FOODAKAI platform. The feedback regarding the benefits in time and cost from FSQA experts working in a food company is also presented in the deliverable.

The document is structured as follows: Chapter 1 serves as an introduction to the deliverable whereas Chapter 2 provides an overview of the Natural Cosmetic Pilot, containing important relevant information, in order to describe the importance of the pilot activities and the methodology and materials that will be used. The structure of presentation includes apart from the Specific Goals, the Technological Guidance, the Measurements made, and the Envisage Outcome of the Pilot. In the Chapter 3 the Data, Datasets and the Use Case Scenario served are presented, as well as the Data Analytics and Processing. The later (Subchapter 3.3) is further divided to the components of importance for the final goal of the dashboard creation, including the use of Satellite Data, Weather Data, BA Analysis, Data Correlation Analysis, and the Software Stack, Data

Transformation and Uploading. In the same Chapter 3, Subchapter 3.4 is dedicated to the Visualisation process with the presentation of the dashboard's final version for the end-users. Chapter 4 presents the results of applying the predictions dashboard in the case of the Food Protection Pilot. Finally, Chapter 5 is a summation of the Conclusions raised by the Deliverable.

2 Natural Cosmetics Pilot

When quality managers and chemists are working at the quality testing lab of natural cosmetic companies, measure the various properties of vine leaves extract samples from a variety of suppliers (grapevine farmers), to find the lots, suppliers, varieties and geographical locations of the grapevine by-products, that presents highest quality, consistent compliance, and the desired pharmaceutical effect[1]. According to the Natural Cosmetics Pilot there is a need in extracting the most out of the incoming to the industry raw materials. 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. Bioactive compounds from winery by-products have disclosed interesting health promoting activities both in vitro and in vivo. If properly recovered, they show a wide range of potential and remunerative applications in many industrial sectors, including cosmetics, pharmaceuticals, biomaterials and food. In fact, winemaking by-products are outstanding sources of oil, phenolic compounds and dietary fibre and possess numerous health benefits and multifunctional characteristics, such as antioxidant, colouring, antimicrobial and texturizing properties[2].

2.1 Specific Goals

The collected data from the natural cosmetics pilot 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 was to face the challenge: "how data from the field can be linked to the biological efficacy of final products". This serves the main purpose of the present deliverable D8.4, to create a dashboard targeting industry-level decision makers and practitioners that incorporates the appropriate functionalities of the BigDataGrapes software stack used in the relevant piloting activities. The dashboard use benefits the cosmetic industry by giving the opportunity to choose from 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 parameters on his field and thereby reaching conclusions, not only about its primary product, but also regarding biological activity of by-products. The farmer will then be able to make decisions on the commercialization of both products (e.g., grapes, wine) and by-products (e.g., vine leaves, bines, grape seeds).

The goal was 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 has suggested that visual tools are the most efficient for these tasks, and thus BDG will use interactive visualisations through a user-friendly dashboard to help the decision-making process. For this purpose, the dashboard was showcased to end-users through a software demonstrator, called for the present Pilot "Grapevine By-Products Biological Efficacy Predictor", and their feedback aimed the refinement and evaluation of the dashboard to meet industry and relative end-users needs.

The Natural Cosmetics Pilot's Software Demonstrator showcased 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-users 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 (public/open 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 integrates intelligence data derived from satellites (with potential to include also data from meteorological stations) for the targeted vineyards, laboratory analyses on the biological activity of two different type extracts from leaves, and finally helps 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. The performance of the software demonstrator has been assessed and appraised from relative end-users selected mainly from cosmetic industry (8), but also from research organisations (2) and grapevine practitioners (1). The end-user's feedback on usefulness of the selected modules, ease of handling, successfulness of implementation, and visualisation lay-out, as well suggestions on improvements that could be employed to future versions of Natural Cosmetics dashboard, are presented with Deliverable D8.5

2.2 Technological Guidance

During project's lifetime, fifteen regions of the Greek territory have been chosen for sample collection, i.e., vine leaves. The dispersion and origin of the samples is shown in the map presented at Figure 1, where the samples of Agiorgitiko are pictured in green and the samples of Mandilaria in red marks, while in Table 1 is presented a list of the vineyards chosen for sampling with their representing variety and location. The collected leaves were dried to a draught and airiness place by avoiding direct exposure to sun and sent to the relevant pilot partner (Symbeeosis) where extraction and laboratory analyses took place.

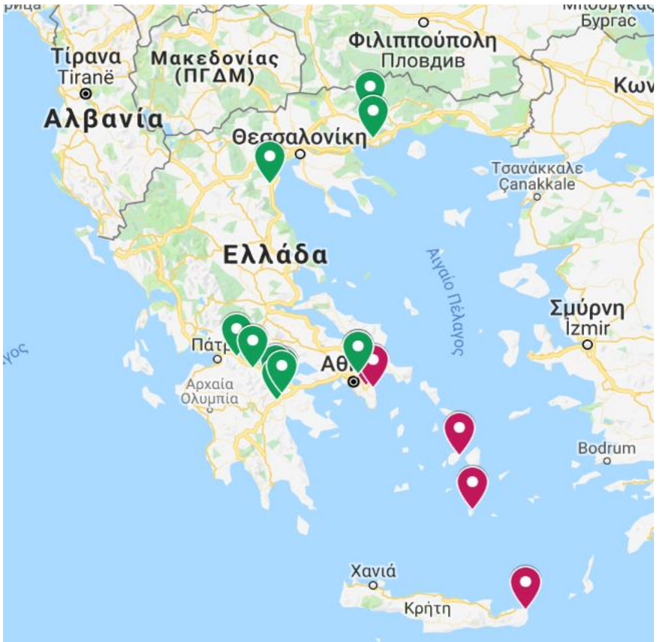


Figure 1: Dispersion of samples across the Greek territory (Correspondent file: https://tinyurl.com/y4scyhed)

Table 1: Vineyards chosen for sample collection (Correspondent file: https://tinyurl.com/y2y5fwld)

#	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 "DEMETER"	Mandilaria	Attica	Lykovrisi
15	Hellenic Agricultural Organization "DEMETER"	Agiorgitiko	Attica	Lykovrisi
16	Palyvou Estate	Agiorgitiko	Peloponnese	Nemea

2.3 Measurements

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 the Industrial Park of Markopoulo Mesogaias in Greece. At the laboratory were conducted extractions under the two different methods and the following measurements of biological activity were made: pH, RI, TPC, TFC, Total Microbial Count, Yeasts & Moulds, DPPH & ABTS assay, as pointed also in Table 2. (Correspondent file: https://docs.google.com/spreadsheets/d/1Lu7GDd-VuSrXrGuNO-wOmQfrqHZlljJi0fxe8qaSZYc/edit?usp=sharing).

Table 2: Measurements of biological efficacy of developed extracts –Data description

Name	Dataset Description	Provenance	Format	Volume
pH of vine leaf extracts	Measurement of pH	Laboratory equipment	xls	MB
Refractive index of vine leaf extracts	Measurement of Brix%	Laboratory equipment	xls	MB
Total microbial count of vine leaf extracts	Measurement of TMC	Laboratory equipment	xls	MB
Yeasts and Moulds of vine leaf extracts	Measurement of Y&M	Laboratory equipment	xls	MB
Antioxidant Activity (AA) of vine leaf extracts	DPPH & ABTS assay	Laboratory equipment	xls	MB
Total phenolic content of vine leaf extracts	Measurement of TPC	Laboratory equipment	xls	MB
Total flavonoid content of vine leaf extracts	Measurement of TFC	Laboratory equipment	xls	MB

2.4 Envisaged Outcome

We wanted to examine how the biological efficacy (in terms of BA parameters described above) depends on the location of the vineyard, the climatic conditions, the agriculture practices followed, the extraction method used, and the variety of the grape. As a final goal, we want to create a predictive tool of biological efficacy based on the location, weather and management of a certain vineyard. Bioactive compounds found in wine making by-products such as vine leaves possess multifunctional characteristics and show a wide range of potential and remunerative applications, concerning health promoting activities[3-4]. Nevertheless, the quality of these by-products and more specifically their biological efficacy can vary depending on multiple parameters, such as the origin of the sample, the recovery process and more[5-7]. 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. Finally, 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. The goal was the creation of a user-friendly dashboard to help the decision-making process of industry end-users and practitioners.

3 Data, Datasets and Use Case Scenario

3.1 Data and Datasets

Data considered for the Natural Cosmetics Pilot ranges from measurements from historical in vitro and in vivo experiments performed from the cosmetic industry (SYMBEEOSIS S.A.), to data about the conditions at actual vineyards from where the leaves are coming. During the lifetime of the BDG project samples from the referring parcels (vineyards) were analysed in order to test their compliance as cosmetic raw materials and assess their BA parameters linked with the quality of the final product. Additionally, GEOCLEDIAN has collected vegetation indices data from satellites Sentinel2 and Landsat 8 in order to test the hypothesis whether the location and field management are correlated with the BA parameters measured in the laboratory. In addition, weather data from meteorological stations near the selected parcels were collected for these 3 years in order to investigate how BA parameters could correlate with weather conditions at vineyards of interest. The relative Data and Datasets information are presented in Table 3.

Table 3: 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 (SVIs)	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, txt	MB

Figure 2 presents the Pilot's data gathering timeline.

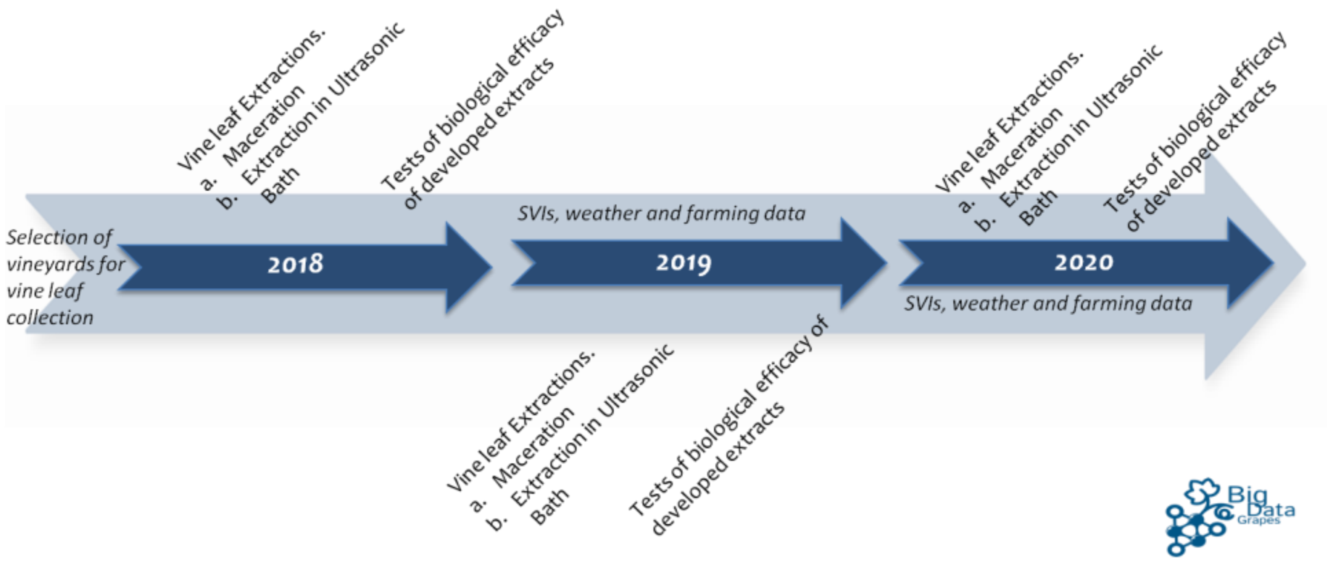


Figure 2: Data gathering timeline for Natural Cosmetics Pilot

3.2 Use Case Scenario - Predicting Biological Efficacy

The scenario presumes that precision farming and control of parameters linked to the quality of grapevine may also result in 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 scenario hypothesis is aiming to create a prediction model capable of correlating SVIs and WD to parameters linked with biological efficacy. The appropriate algorithms created will use the existing datasets and explore the relationship between them. Datasets concerning SVIs (or weather) work as independent variables, while the datasets concerning biological efficacy will work as the dependent variables. A number of potential correlations were generated between them and the selection process of the ideal correlation will focus on minimum complexity and error. The scenario hypothesis has the potential for increased scalability using additional weather and spatial data by choosing larger territories as points of interest. Finally, the cosmetic industry end-users can choose from the list of suppliers for a specific need, just by consulting the dashboard about the crop location and related SVIs (or WD) and thereby reach conclusions regarding biological activity of incoming by-products. A farmer, on the other hand, can perform decision-making by evaluating location and SVIs (or WD) on his field and thereby reaches conclusions regarding biological activity of its products. The farmer will then be able to make decisions on the commercialization of the by-products.

3.3 Data Analytics And Processing

During lifetime of BDG project, three (3) data categories have been incorporated to the Pilot: SVIs by applying the GEOCLEDIAN's Ag|knowledge tool that exploits open-source satellite data and extract information regarding the vegetation indexes on the location of each crop, WD from meteorological stations and Data related to vine leaves quality of two (2) different grape varieties (Agiorgitiko and Mandilaria), according to their phytochemical profile and BA, after extraction by UAE and MAC. The collected data from the Natural Cosmetics Pilot provided the necessary information for the evaluation of the quality of each sample, by linking the BA parameters with the SVIs and WD of the vineyard of origin. Finally, the correlation analyses pointed out which information should be taken into consideration for building the models that support the DSS.

3.3.1 Satellite Data

GEOCLEDIAN' s Cloud Processing Platform provides the field monitoring service Ag|knowledge that allows the automatic crop monitoring for fields with multispectral products. The Ag|knowledge is a REST API allowing easy access and integration of satellite remote sensing data & analytics into agricultural applications. The relative web link is: https://sites.google.com/site/geocledian/home/product-overview.
The API provides access to field monitoring products for registered parcels (i.e., fields or parts of land). The data for each parcel are immediately updated as soon as new measurements are available. The available data products that are relevant to the Natural Cosmetics Pilot include visible images (True colour images, RGB) of the parcels, and vegetation indexes (vit, NDVI, NDRE1, NDRE2, NDRE3, NDWI, SAVI, EVI2, CIRE & NCPRI) that are related to vegetation properties like e.g., chlorophyll, nitrogen or vegetation water content, as well as vegetation variations maps, that show the variation of the vegetation status. For all of these products time series and a full history of the last 5 years are available. For the experiments, only the data collected from the satellite sentinel2 were used but the platform can also access Landsat 8 satellite data. Collectively, the SVIs taken into account among their descriptions for the Pilot are presented in Table 4.

Table 4: SVIs and relative description

Name	Description
Vitality (**Vit**)	Vitality is based on the NDVI but optimised for visualization. It is a valuable quantitative vegetation monitoring tool used as an indicator for the vitality of a crop in particular for the live green vegetation.
Normalized Difference Vegetation Index (**NDVI**)	Quantifies vegetation by measuring the difference between near-infrared (vegetation strongly reflects) and red light (vegetation absorbs). Overall, NDVI is a standardized way to measure healthy vegetation, although has the disadvantage to saturate at high leaf area levels and therefore shows limited variation in dense fields with high biomass.
Normalized Difference Red Edge Index (v1) (**NDRE1**)	Substitution of NDVI's red band with NDRE's red edge band (730nm) provides a measurement that is not as strongly absorbed by just the topmost layers of leaves. NDRE can give better insight into permanent or later stage crops since it's able to measure further down into the canopy and thus provides more sensitivity in vegetation with high leaf areas.
Normalized Difference Red Edge Index (v2) (**NDRE2**)	Substitution of NDVI's red band with NDRE's red edge band (700nm)
Normalized Difference Red Edge Index (v3) (**NDRE3**)	Substitution of NDVI's red band with NDRE's red edge band (740nm)
Normalized Difference Water Index (**NDWI**)	NDWI is less susceptible to atmospheric scattering than NDVI but does not remove completely the background soil reflectance effects, similar to NDVI. Because the information about vegetation canopies contained in the SWIR channel is very different from that contained in the VIS channel, NDWI is considered as an independent vegetation index. It presents enhanced sensitivity to vegetation water content & water stress.

Soil Adjusted Vegetation Index (**SAVI**)	The index minimizes soil brightness influences from spectral vegetation indices involving red and near-infrared (NIR) wavelengths. It is interesting in sparse vegetation canopies or early growing stages.
Enhanced Vegetation Index 2 (**EVI2**)	The enhanced vegetation index (EVI) is vegetation index designed to enhance the vegetation signal with improved sensitivity in high biomass regions and improved vegetation monitoring through a de-coupling of the canopy background signal and a reduction in atmosphere influences.
Chlorophyll Index - Red Edge (**CI-RE**)	Is used to calculate the total chlorophyll content of the leaves. The CIgreen and CIred-edge values are sensitive to small variations in the chlorophyll content and consistent across most species. Apart from the very high Chlorophyll also presents Nitrogen sensitivity and thus canopy Chlorophyll & Nitrogen contents can be derived from this index.
Normalized Pigment Chlorophyll Ratio Index **(NCPRI)**	The Normalized Pigment Chlorophyll Ratio Index (NPCRI) is a numerical indicator that is associated with the chlorophyll and nitrogen content and can find applications in precision agriculture. Crops with a low Nitrogen content can have a high carotenoid to chlorophyll ratio. Using the red and blue spectral bands, NPCRI can capture the information needed to quantify chlorophyll and Nitrogen.

All SVIs were considered for four-time frame aggregations:

1. from the begging of the year until March
2. from the begging of the year until April
3. from the begging of the year until May
4. from the begging of the year until June

For each time frame, the average of the observation values was computed inside the time frame.
For each time series index point six different aggregate information were computed by the following aggregation functions:

1. max
2. min
3. mean
4. standard deviation (StD)
5. count
6. sum

Throughout the pilot's duration, GEOCLEDIAN collected and processed the described satellite data from all sites. Visible images and Spectral Vegetation Index Maps were produced, and the data automatically were provided via API in near real-time. For every parcel's geometry registered in our system is delivered the time series of satellite images together with time series statistics on all the vegetation indexes. The data can be visualized with visualization components of the accessible platform by the client as it is presented in the following Figures 3-6.

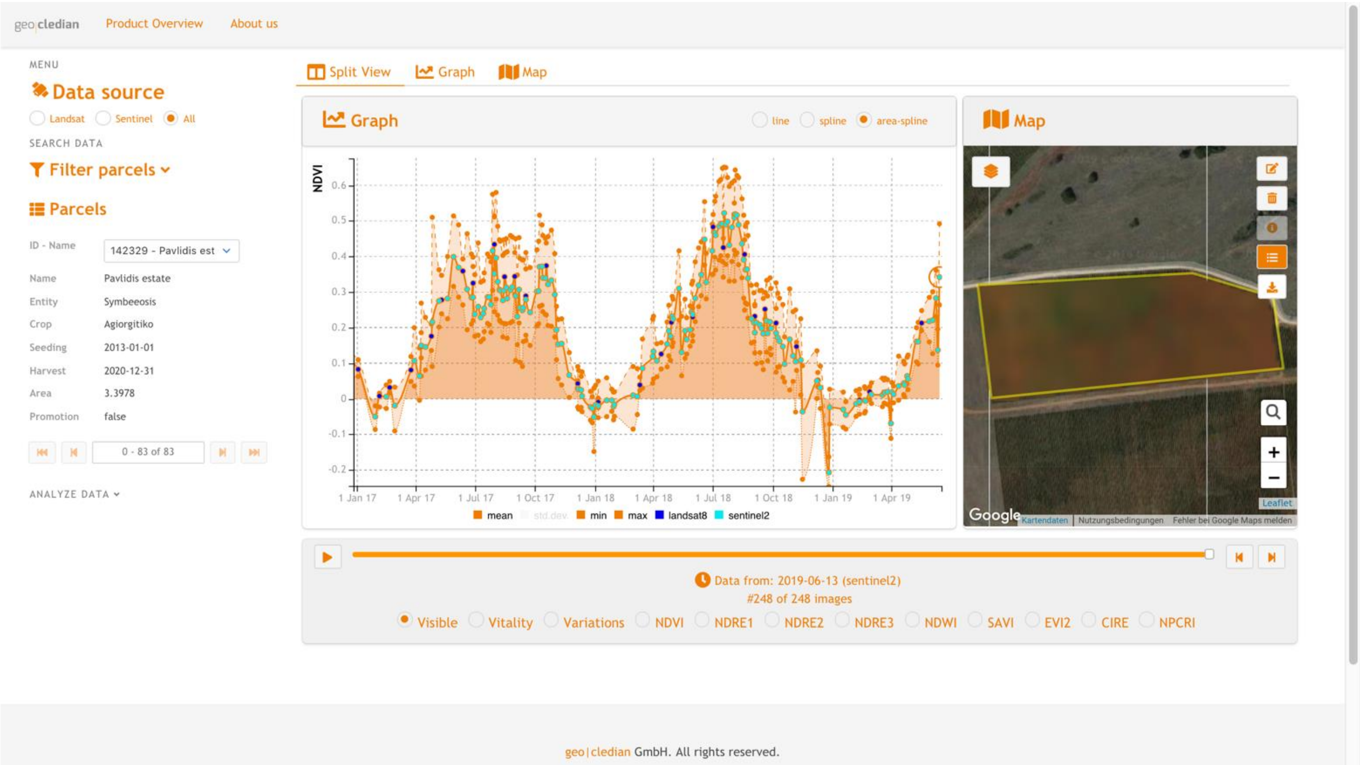


Figure 3: A Sentinel-2 RGB image for 13th June 2019 for a parcel of the Pavlidis estate, Greece. Also shown are all available Landsat 8 and Sentinel-2 image acquisitions for 2017 - 2019.

Figure 4: A Sentinel-2 NDVI time series for 2017 - 2019 for a parcel of the Pavlidis estate, Greece showing also the NDVI image for 13th June 2019.

Figure 5: A Sentinel-2 NDWI time series for 2017 - 2019 for a parcel of the Pavlidis estate, Greece showing also the NDVI image for 13th June 2019. This NDWI is known to be related to leaf water content.

Figure 6: A Sentinel-2 CI-RE time series for 2017 - 2019 for a parcel of the Pavlidis estate, Greece showing also the NDVI image for 13th June 2019. The Chlorophyll Index-Red Edge (CI-RE) is reported to be highly correlated with canopy Chlorophyll and Nitrogen contents.

3.3.2 Weather Data

The Institute of Environmental Research and Sustainable Development of The National Observatory of Athens (NOA) has developed and operates a network of automated weather stations across Greece (>350 stations) for monitoring of weather conditions with the aim to support not only the research needs (weather monitoring and analysis, weather forecast skill evaluation) but also the needs of various communities of the production sector (agriculture, constructions, leisure and tourism, etc.). The stations network is the denser network of automated stations over Greece and recorded measurements at 10 min intervals are automatically transferred to a server at NOA premises, dedicated for data archiving and quality control. Therefore, meteorological data are readily accessible by any interesting party for further processing and use. The Institute of Environmental Research and Sustainable Development provided to SYMBEEOSIS all necessary weather data from the nearest stations to the Natural Cosmetic Pilot's grapevine parcels. For demonstration purposes, only SVIs data were presented, but the platform can also access WD, perform the same processing and modelling, and present an extra visualisation page in the same platform. Collectively, the measurements of weather stations and their descriptions for the Natural Cosmetic Pilot are presented in Table 5, while the access link is https://drive.google.com/drive/folders/1QpcmeYOcE4Kz9YL3Yej2zKXWsH1jYrAo?usp=sharing.

Table 5: Weather data and relative description

Name (Symbol)	Description (measure)
Date	The date of measurement (day/month/year)
Time (t)	The local time of measurement (hour: minute)
Altitude (h)	The altitude of meteorological station (meters)
Temperature (temp_out)	The temperature of atmosphere within 10 min time frame (hour: minute)
Relative humidity (out_hum)	The relative humidity of atmosphere The Rh% within 10 min time frame (%)
Atmospheric pressure (bar)	The pressure of atmosphere (bar)
Rainfall (rain)	The height of rain in the 10 min time frame (mm/10min)
Rainfall/h (rain_rate)	The strength of rain (mm^3/h)
Speed of wind (wind_speed)	The mean speed wind within the 10 min frame (km/h)
Direction of wind (wind_dir)	The direction of wind relative to geographic cardinal directions (N, S, E, W)
High speed of wind (hi_speed)	The higher speed of wind within the 10 min frame (km/h/10min)

3.3.3 Biological Activity Data

Some BA parameters, as presented also in Table 2, refer to compliance tests (pH, RI, TMC, Y&M), meaning that samples with measurements outside the accepted ranges are discarded from the production process. Regarding the dataset used for the correlation analyses, TMC and Y&M were left out since they give as output a qualitative response. More important for the analyses were the BA parameters TPC and TFC, reporting the phenolic and flavonoids content of the sample, and AA1 (ABTS) and AA2 (DPPH), reporting AA of the sample through free radicals' scavenging potential. All these four BA parameters are crucial indicators for the BA of the extracts. Concluding, the BA parameters used for the analyses of the present formative deliverable were:

1. pH
2. RI (%)
3. AA1 DPPH (μg/mL trolox)
4. AA2 ABTS (μg/mL trolox)
5. TPC (μg/mL gallic acid)
6. TFC (μg/mL quercetin)

Sampling measurements can be accessed at the following link: https://tinyurl.com/y5tnykgr

3.3.4 Data Correlation Analysis

The samples used for the correlation analysis have been specified in Table 6 and can be found on the URL: https://docs.google.com/spreadsheets/d/1_kRsyd-bgfZHFi3XhS-g7qxJIp2qkI-7PtNYMMo69Qc.
The column Sample_Id corresponds to the identification of the sample under study, while the column Parcel_id links the sample to the parcel representation in the GEOCLEDIAN system.

Table 6: Wineries of samples included in correlation analysis

#	Sample ID	Vineyard	Variety	Region	City	Parcel ID
1	I.A.1	RIRA Vineyards	Agiorgitiko	Peloponnese	Aigio	142330
2	I.A.2	Semeli Wines	Agiorgitiko	Peloponnese	Nemea	143347
3	I.A.3	Tetramythos Wines	Agiorgitiko	Peloponnese	Ano Diakopto	196815
4	I.M.4	Moraitis Winery	Mandilaria	Aegean	Paros	142337
5	I.A.5	Pavlidis Estate	Agiorgitiko	Northern Greece	Drama	142329
6	I.M.6	Vassaltis Vineyards	Mandilaria	Aegean	Santorini	143373
7	I.A.7	Strofilia Estate Winery	Agiorgitiko	Peloponnese	Stimfalia	142332
8	I.A.8	Biblia Chora Estate	Agiorgitiko	Northern Greece	Kavala	142340
9	I.M.9	Toplou Winery	Mandilaria	Crete	Sitia	142338
10	I.A.10	Skouras Domaine	Agiorgitiko	Peloponnese	Argos	142336
11	I.M.11	Aoton Winery	Mandilaria	Attica	Peania	142339
12	I.A.12	Papagiannoulis Winery	Agiorgitiko	Northern Greece	Katerini	142335
13	I.A.13	Papagiannakos Domaine	Agiorgitiko	Attica	Markopoulo	142341
14	I.A.14	Hellenic Agricultural Organization "DEMETER" 1	Agiorgitiko	Attica	Likovrisi	143363
15	I.M.15	Hellenic Agricultural Organization "DEMETER" 2	Mandilaria	Attica	Likovrisi	143364
16	I.A.16	Palyvou Estate	Agiorgitiko	Peloponnese	Nemea	143365

According to the datasets available, correlation analysis assigned as “observed variables” 6 × 2 vectors (BA parameters × maceration & ultrasound measurements) of 16 components (parcel ids) and as “variate variables” 10 × 6 × 4 = 240 vectors (SVIs × aggregation functions × time frames) of 16 components (parcel ids). To perform the correlation analysis, a Pearson product-moment correlation coefficient was applied. The coefficient returns a value between -1 and 1 that represents the limits of correlation from a full negative correlation to a full positive correlation. A value of zero (0) means no correlation, while values that should be interpreted are often values below -0.5 or above +0.5 indicating a notable correlation. Values in-between those -0.5 and +0.5 suggest a less notable correlation. The results of the correlation analyses due to their high extent are removed in Deliverable 8.4 Appendix at the end of the document and are presented per time frame at Figures 20-23 regarding MAC and at Figures 24-27 regarding UAE of Appendix Chapter 5. Hereafter, at Tables 7-10 are presented only the most correlated SVIs with BA parameters.

Table 7: Most correlated SVI for each BA parameter after MAC.

BA parameter	Most corr. SVI	Highest Corr. Value	Time Frame	Aggregation
pH	evi2	-0.831473	interval from the beginning of the year until April	min value
TPC	ndre3	0.690595	interval from the beginning of the year until March	sum value
TFC	vit	0.786551	interval from the beginning of the year until June	min value
RI	npcri	-0.444023	interval from the beginning of the year until March	sum value
AA1	npcri	-0.813996	interval from the beginning of the year until March	sum value
AA2	vit	-0.769895	interval from the beginning of the year until June	sum value

Table 8: Most correlated SVI for each BA parameter after UAE.

BA parameter	Most corr. SVI	Highest Corr. Value	Time Frame	Aggregation
pH	ndre3	-0.620414	interval from the beginning of the year until April	sum value
TPC	ndre3	0.758299	interval from the beginning of the year until March	mean value
TFC	evi2	0.869101	interval from the beginning of the year until June	min value
RI	ndre3	-0.492838	interval from the beginning of the year until April	sum value
AA1	ndre3	-0.944947	interval from the beginning of the year until April	max value
AA2	ndre3	-0.854049	interval from the beginning of the year until March	mean value

Table 9: Most correlated BA parameter for each SVI after MAC.

SVI	Most corr. BA parameter	Highest Corr. Value	Time Frame	Aggregation
vitality	TFC	0.786551	interval from the beginning of the year until June	min value
savi	TPC	0.393863	interval from the beginning of the year until June	min value
npcri	AA1	0.813996	interval from the beginning of the year until March	sum value
ndwi	pH	0.818572	interval from the beginning of the year until June	mean value
ndvi	AA1	0.566751	interval from the beginning of the year until April	min value
ndre3	TPC	0.690595	interval from the beginning of the year until March	sum value
ndre2	AA2	0.243516	interval from the beginning of the year until March	std value
ndre1	AA1	0.557829	interval from the beginning of the year until May	min value
evi2	pH	0.831473	interval from the beginning of the year until April	min value
cire	pH	0.817427	interval from the beginning of the year until April	mean value

Table 10: Most correlated BA parameter for each SVI after UAE.

SVI	Most corr. BA parameter	Highest Corr. Value	Time Frame	Aggregation
vitality	AA1	-0.737992	interval from the beginning of the year until April	sum value
savi	AA1	-0.883041	interval from the beginning of the year until June	mean value
npcri	RI	-0.350118	interval from the beginning of the year until March	std value
ndwi	AA1	-0.653773	interval from the beginning of the year until June	sum value
ndvi	TPC	0.403468	interval from the beginning of the year until May	min value
ndre3	AA1	-0.944947	interval from the beginning of the year until April	max value
ndre2	AA2	-0.532659	interval from the beginning of the year until April	min value
evi2	TFC	0.869101	interval from the beginning of the year until June	min value
cire	TFC	0.868015	interval from the beginning of the year until June	mean value

According to the correlation results there is observed a constantly high correlation between ndwi and TFC, a fact that can be incorporated to the DSS of the dashboard. In addition, there can also be observed high correlations between the AA1 with ndre and ncpri, AA2 with ndre and vit, and TFC with ndwi, cire and evi2. Datasets could also strengthen up with repetition of sampling the upcoming years so that a furure version of the dashboard will consider all the information available in order to produce a more reliable DSS for the end-users.

3.3.5 Software Stack, Data Transformation and Uploading

The BDG stack, as abstractly described in D2.3, is employed to serve the desired outcomes of the Natural Cosmetics pilot. In the context of this specific pilot the following components are required:

- a UI tool needed for the dataset upload,
- API endpoints responsible for the storage and discovery of field specific data and metadata,
- MongoDB as storage engine for the metadata information,
- API endpoints responsible for the upload of the xlsx files, or records following the SYMBEEOSIS data schema,
- Elasticsearch for the storage of the actual data of the provided datasets,
- API endpoints responsible with the transformation of the upload data into RDF, following the work done in Data Modelling for SYMBEEOSIS,
- command line tools for the rdfization of the provided data,
- GraphDB for the storage of the RDF data and their semantic enrichment,
- GEOCLEDIAN's service for the extraction of the satellite image processing data concerning the SYMBEEOSIS fields,
- Correlation scripts responsible for the correlation of the semantically enriched data and the outcomes of the image processing.

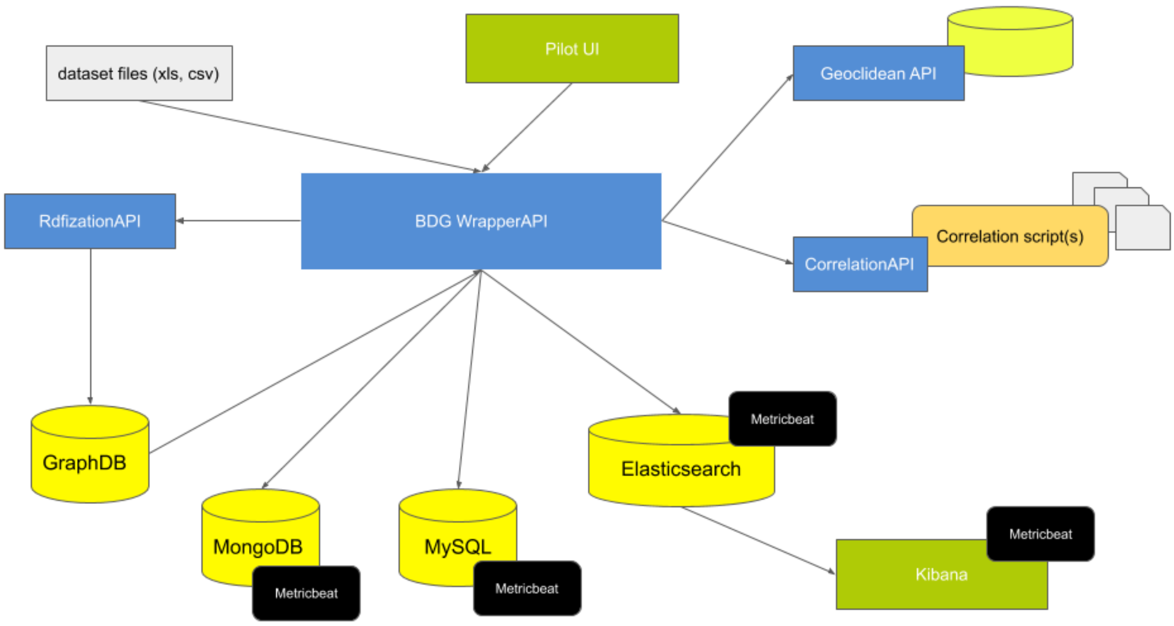


Figure 7: An abstract view of the BDG stack components used in Symbeeosis data flows

Figure 7 shows an abstract overview of the required BDG components for SYMBEEOSIS's data flows throughout the stack. In the rest of this section, we further describe each of the employed components of the stack, grouped by their usage, providing screenshots for each.

For the upload of the datasets provided by SYMBEEOSIS, there are 2 available components that can be used. In the BDG stack there is a UI tool for each pilot along with API endpoints responsible for the actual storage of the provided data.

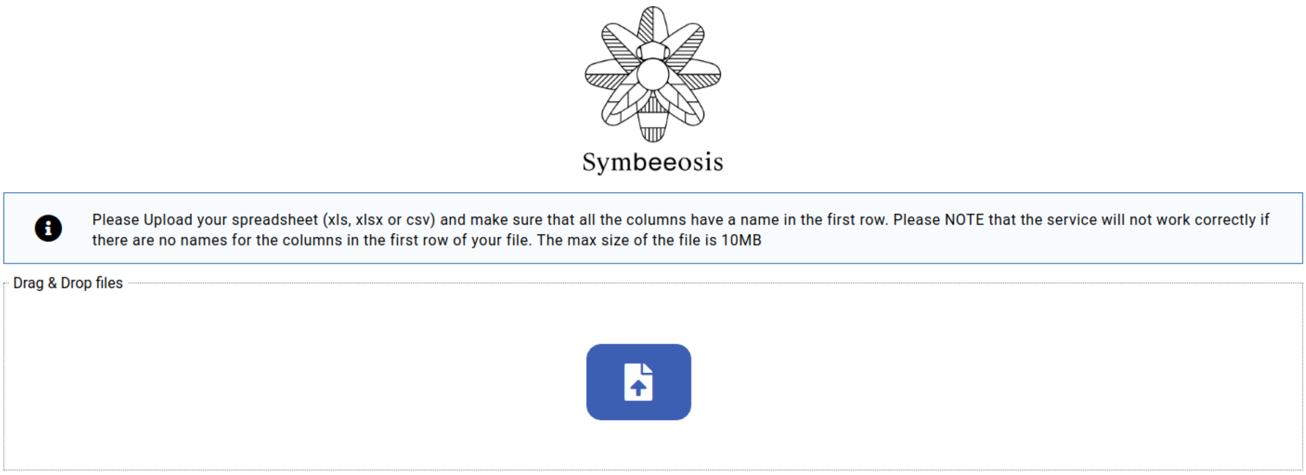


Figure 8: The initial view of the UI tool for the dataset upload for the Symbeeosis case

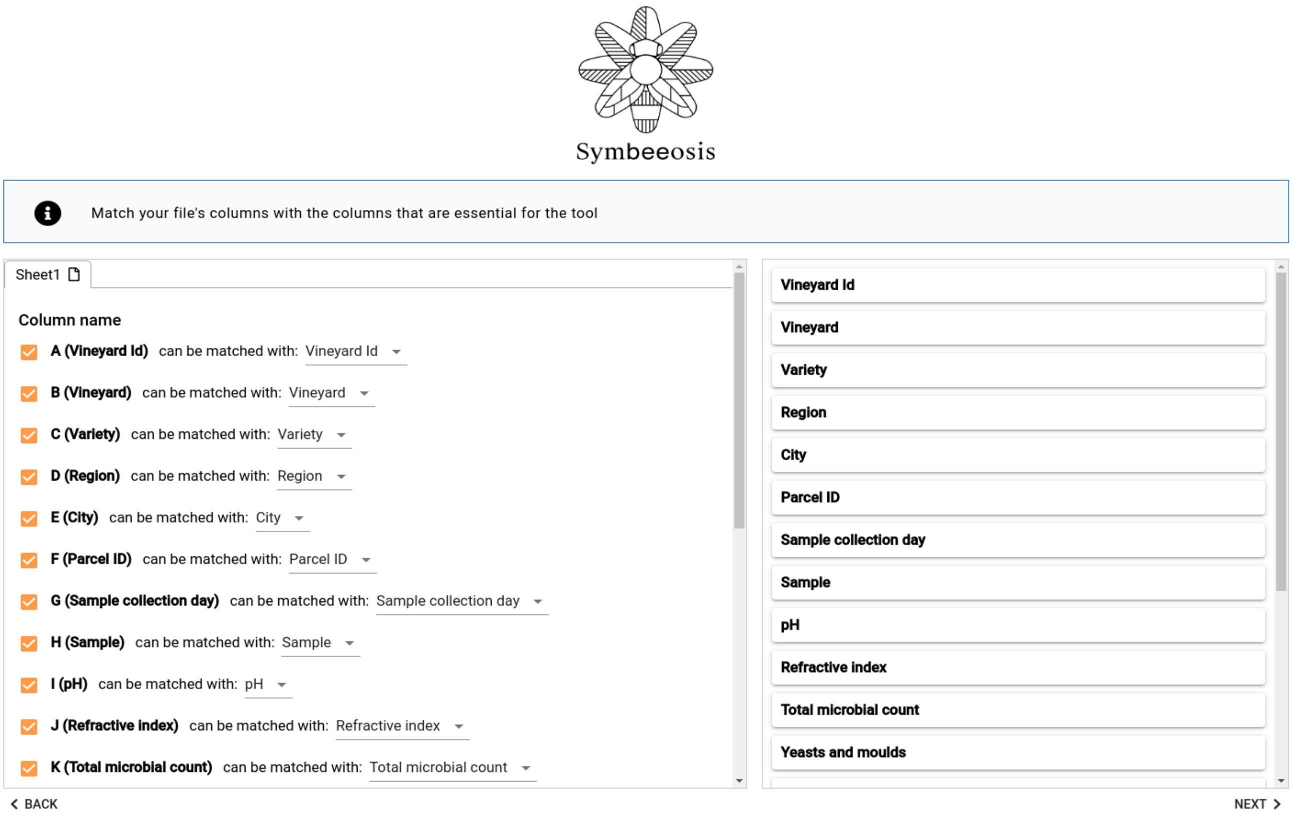
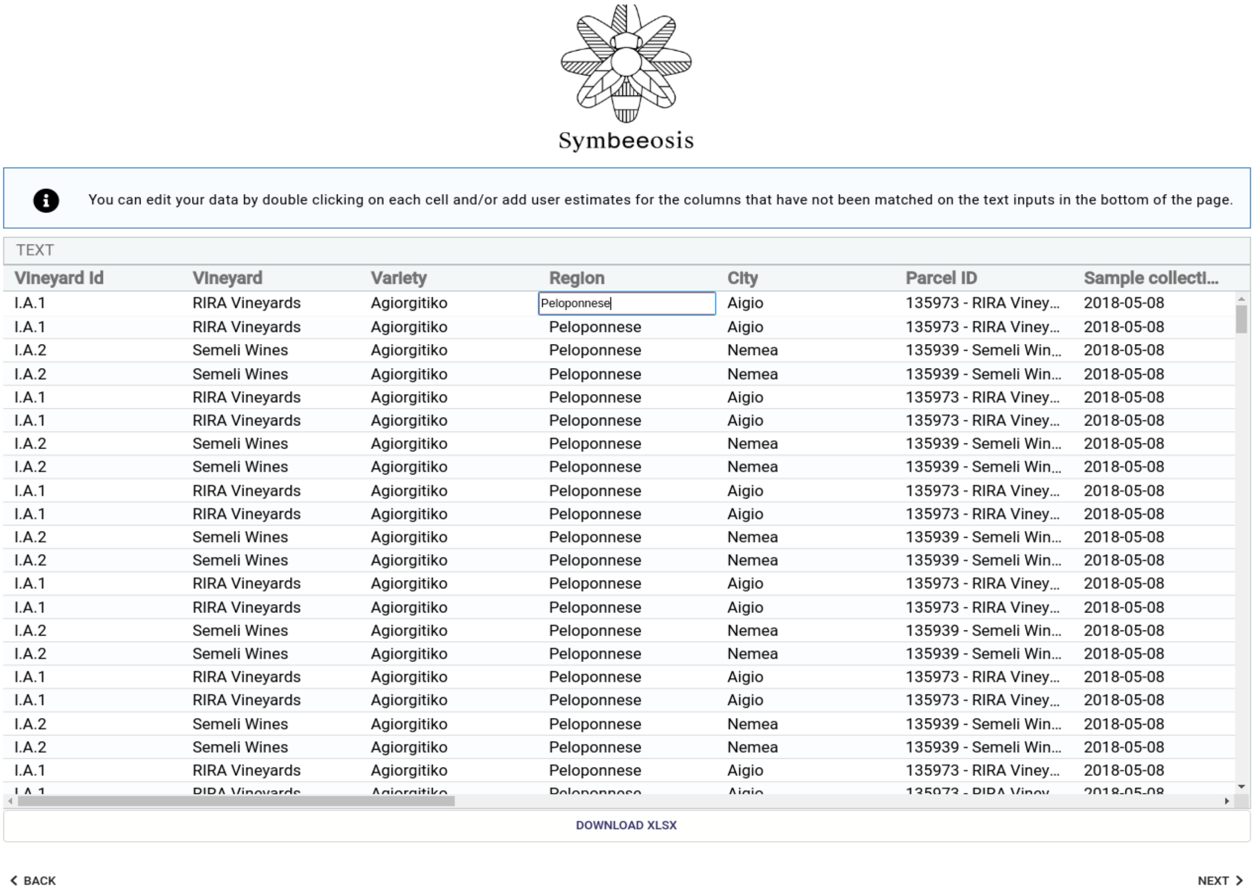


Figure 9: The screen responsible for the matching of the uploaded xls schema with the Symbeeosis one



Vineyard Id	Vineyard	Variety	Region	City	Parcel ID	Sample collecti...
I.A.1	RIRA Vineyards	Agiorgitiko	Peloponnese	Aigio	135973 - RIRA Viney...	2018-05-08
I.A.1	RIRA Vineyards	Agiorgitiko	Peloponnese	Aigio	135973 - RIRA Viney...	2018-05-08
I.A.2	Semeli Wines	Agiorgitiko	Peloponnese	Nemea	135939 - Semeli Win...	2018-05-08
I.A.2	Semeli Wines	Agiorgitiko	Peloponnese	Nemea	135939 - Semeli Win...	2018-05-08
I.A.1	RIRA Vineyards	Agiorgitiko	Peloponnese	Aigio	135973 - RIRA Viney...	2018-05-08
I.A.1	RIRA Vineyards	Agiorgitiko	Peloponnese	Aigio	135973 - RIRA Viney...	2018-05-08
I.A.2	Semeli Wines	Agiorgitiko	Peloponnese	Nemea	135939 - Semeli Win...	2018-05-08
I.A.2	Semeli Wines	Agiorgitiko	Peloponnese	Nemea	135939 - Semeli Win...	2018-05-08
I.A.1	RIRA Vineyards	Agiorgitiko	Peloponnese	Aigio	135973 - RIRA Viney...	2018-05-08
I.A.1	RIRA Vineyards	Agiorgitiko	Peloponnese	Aigio	135973 - RIRA Viney...	2018-05-08
I.A.2	Semeli Wines	Agiorgitiko	Peloponnese	Nemea	135939 - Semeli Win...	2018-05-08
I.A.2	Semeli Wines	Agiorgitiko	Peloponnese	Nemea	135939 - Semeli Win...	2018-05-08
I.A.1	RIRA Vineyards	Agiorgitiko	Peloponnese	Aigio	135973 - RIRA Viney...	2018-05-08
I.A.1	RIRA Vineyards	Agiorgitiko	Peloponnese	Aigio	135973 - RIRA Viney...	2018-05-08
I.A.2	Semeli Wines	Agiorgitiko	Peloponnese	Nemea	135939 - Semeli Win...	2018-05-08
I.A.2	Semeli Wines	Agiorgitiko	Peloponnese	Nemea	135939 - Semeli Win...	2018-05-08
I.A.1	RIRA Vineyards	Agiorgitiko	Peloponnese	Aigio	135973 - RIRA Viney...	2018-05-08
I.A.1	RIRA Vineyards	Agiorgitiko	Peloponnese	Aigio	135973 - RIRA Viney...	2018-05-08
I.A.2	Semeli Wines	Agiorgitiko	Peloponnese	Nemea	135939 - Semeli Win...	2018-05-08
I.A.2	Semeli Wines	Agiorgitiko	Peloponnese	Nemea	135939 - Semeli Win...	2018-05-08
I.A.1	RIRA Vineyards	Agiorgitiko	Peloponnese	Aigio	135973 - RIRA Viney...	2018-05-08

Figure 10: The edit/finalization screen of the uploaded data

Figures 8-10 show the necessary steps for the dataset upload using the UI tool. Initially the user uploads her xlsx files, in the next screen she can match the provided file's headers with the specifics of the SYMBEEOSIS 's BDG schema. In the final screen depicted in Figure 10, the user can make changes to the uploaded dataset and on the click of the "Next" button, the dataset is sent to the Rdfization service (the source code is available here) to be converted into RDF following the outcomes of the Data Modelling work and stored into BDG's GraphDB instance. A sample of the generated RDF is presented in Figure 11.


```
- <rdf:Description rdf:about="http://dev.bigdatagrapes.eu/data/cosmetics/T101">
    <rdf:type rdf:resource="http://purl.org/linked-data/cube#Obesrvation"/>
    <grapeVariety rdf:resource="http://dev.bigdatagrapes.eu/grapeVariety/Agiorgitiko"/>
    <antioxidantActivityDPPHTrolox rdf:datatype="http://www.w3.org/2001/XMLSchema#double">4.31</antioxidantActivityDPPHTrolox>
    <dataSet rdf:resource="http://dev.bigdatagrapes.eu/data/cosmetics"/>
    <tfcQuercetin rdf:datatype="http://www.w3.org/2001/XMLSchema#double">12.24</tfcQuercetin>
    <totalMicrobialCount rdf:resource="http://dev.bigdatagrapes.eu/microbialCount/LT10"/>
    <date rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2018-05-08T02:00:00.000+02:00</date>
    <pH rdf:datatype="http://www.w3.org/2001/XMLSchema#double">5.89</pH>
    <tpcGallicAcid rdf:datatype="http://www.w3.org/2001/XMLSchema#double">5.96</tpcGallicAcid>
    <yeastsAndMoulds rdf:resource="http://dev.bigdatagrapes.eu/microbialCount/LT10"/>
    <refractiveIndex rdf:datatype="http://www.w3.org/2001/XMLSchema#double">18.93</refractiveIndex>
    <antioxidantActivityABTSTrolox rdf:datatype="http://www.w3.org/2001/XMLSchema#double">12.38</antioxidantActivityABTSTrolox>
  </rdf:Description>
```


Figure 11: A sample of the RDF generated for Symbeeosis

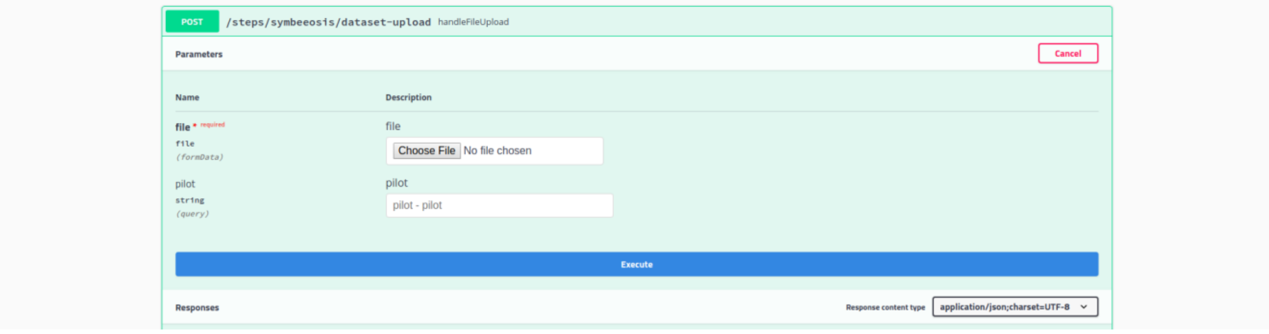


Figure 12: Swagger documentation of the API endpoint for the dataset upload

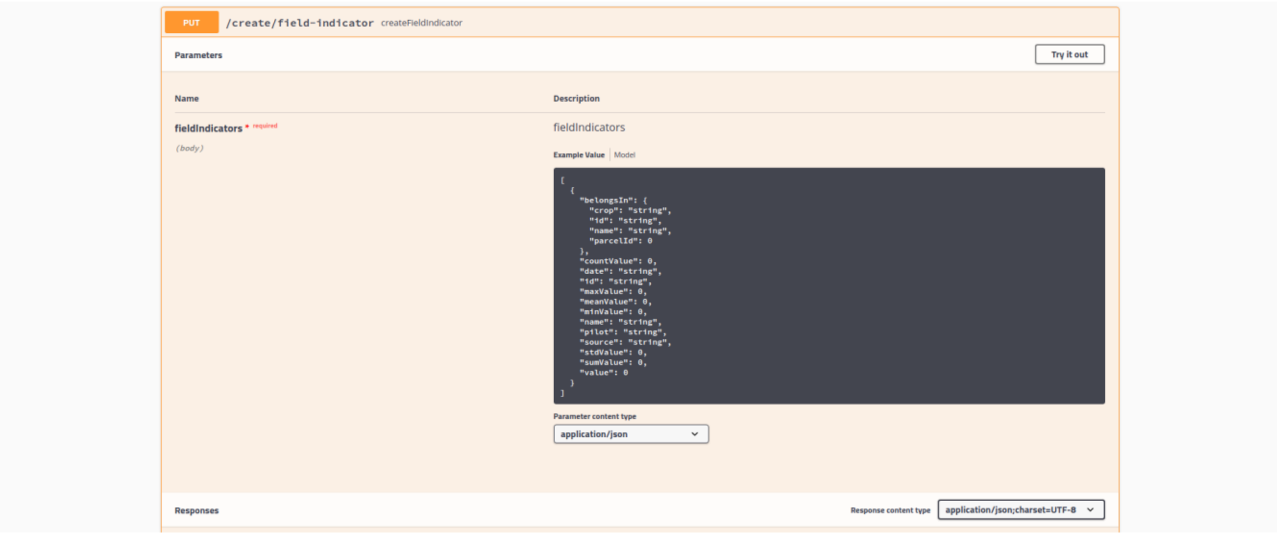


Figure 13: Swagger documentation of the API endpoint for distinct record upload

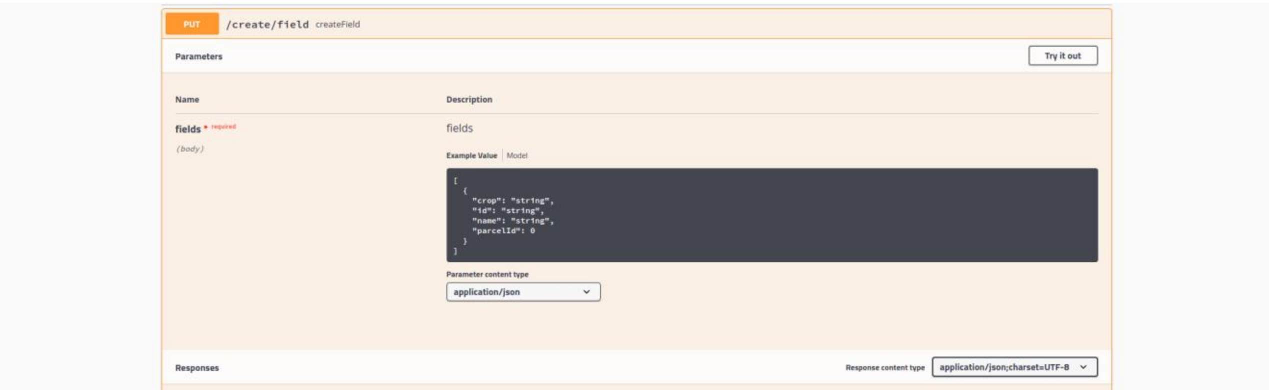


Figure 14: Swagger documentation of the API endpoint for the field metadata/description

Figures 12-14 show the API endpoints responsible for the dataset upload per pilot, the storage of distinct field indicators and the respective metadata information on fields. The actual data in the datasets along with the field indicators uploaded are stored into Elasticsearch whereas the metadata information on the fields is stored into MongoDB.

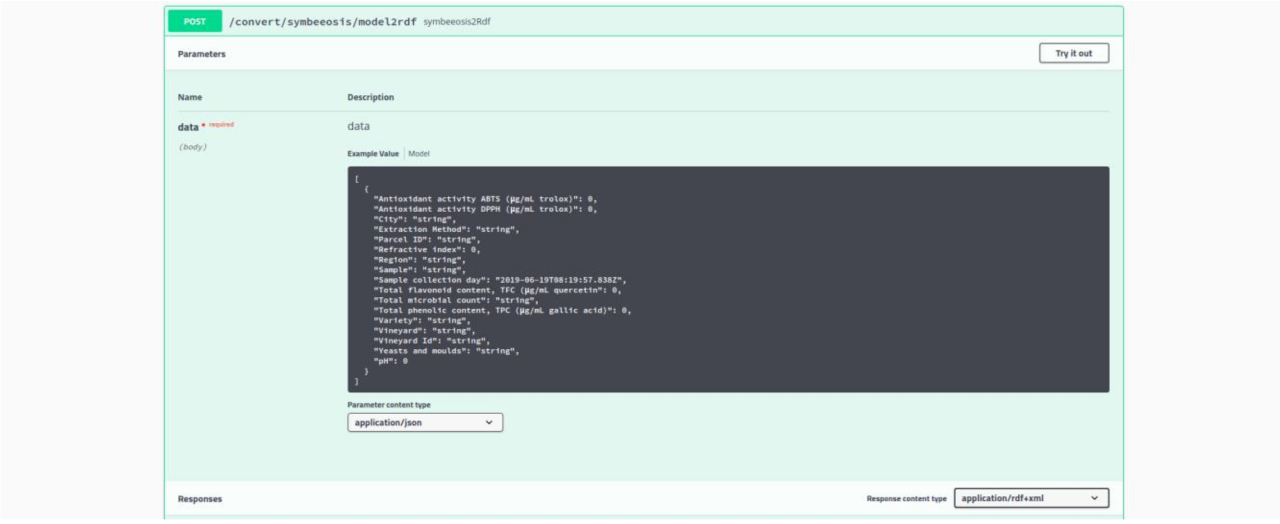


Figure 15: Swagger documentation of the API endpoint for the rdfization of SYMBEEOSIS data

Figure 15 shows the Swagger documentation of the API endpoint responsible for the conversion of the provided data into RDF/XML and their storage into GraphDB. This can also be achieved using the command line tools of the BDG stack (source code is available here).

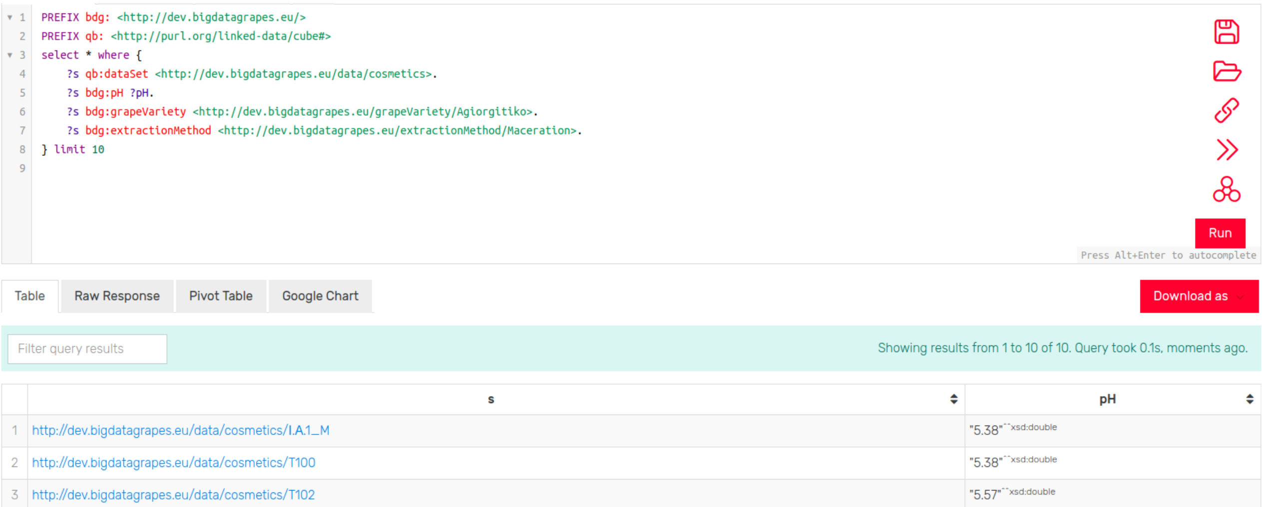


Figure 16: A sample SPARQL query on the SYMBEEOSIS data in GraphDB

In Figure 16 we present a sample SPARQL query concerning Symbeeosis's data that are stored into BDG's GraphDB instance. In this specific case we search for pH measurements of laboratory tests performed on Agiorgitiko using maceration as an extraction method on data provided in the context of the Natural Cosmetics use case.

After the storage into GraphDB, along with the respective semantic enrichment of the provided datasets, the data is extracted from the knowledge graph and stored back into BDG's Elasticsearch instance using the respective Apache Nifi dataflows and API endpoints. The same is also done for the extraction of the satellite image processing data that reside in GEOCLEDIAN's service and concerns SYMBEEOSIS 's fields.

For the correlation of the provided datasets to happen, the next step of the workflow is the extraction of all the stored data that concern SYMBEEOSIS and their conversion into tabular data. After this step is completed a request to the wrapper API endpoint that triggers this correlation happens, that once completed responds with a zip file containing all the generated histograms.

3.4 VISUALIZATION

In D5.3 of WP5, partner KU Leuven demonstrated a trust-aware DSS that uses visualization techniques to explain the influence of input (predictor) variables on prediction outcomes. Research has shown that prediction models currently employed in agricultural DSS remain opaque to users and hidden behind the software. This black-box nature can often lead to users not trusting the system's decisions especially when the system fails to provide meaningful explanations[8]. Previous work has expressed that explaining a model's predictions is an important approach for earning users' trust. Visualization is a powerful technique to address this problem and can effectively communicate uncertainty emerging from both data and prediction models[9]. For demonstration purposes a simple version of decision support system is presented after correlation modelling under the components described in 3.3.4 section of the present deliverable. The dashboard visualisation is temporarily hosted at a KU Leuven's server: http://picasso.experiments.cs.kuleuven.be:3620/, while source code has been published at the Github repository of BigDataGrapes project: https://github.com/BigDataGrapes-EU/grapevine-biological-efficacy.

The dataset used for the visualization platform is based on the datasets processed by partner CNR (3.3.4 Data Correlation Analysis) which resulted to 12 vectors of "observed variables" (BA parameters × maceration & ultrasound measurements) and 240 vectors of "variate variables" (SVIs × aggregation functions × time frames) of 16 components (parcel ids). The visualization can clearly illustrate multidimensional data and the influence of each input variable by allowing us to assess the parcel on their by-products' quality. The objective of the dashboard visualization is to help users understand the contribution of each of the features of a correlation analysis around a local data point. In the correlation bars, on the left side of the screen, we can see which attributes have the higher correlation for the target variable value. On the left side of the screen and after selecting the SVIs of interest (the one with the higher correlation for the time frame of interest) are presented the proposed parcels for grapevine leaves procurement, showing first the most promising one in terms of expected BA parameters and followed by the rest parcels in descending order. This will help end-users understand the influence of the different variables of the correlation and make their decision according to the outputs. The Cosmetic Pilot Dashboard will be a collection of tools that provides easy access to diverse visual components to display detailed data, and thus providing a unified display to the decision-maker for interaction and exploration. The dashboard interface created for demonstration is named "Grapevine By-Products Biological Efficacy Predictor" and is shown in Figure 17, the output components showing the distributions of input variables and the predicted BA parameters are shown as descending order bars of correlation in Figure 18, and all relative information for parcel of choice in Figure 19. The demonstrator of Natural Cosmetic Pilot called "Grapevine By-Products Biological Efficacy Predictor" is temporarily available at:
http://picasso.experiments.cs.kuleuven.be:3620/.
The demonstrator simulates a real-world scenario where one would measure vine leaves' extracts chemical properties and input the data in a configuration step. After that, the data could be visualized by selecting the appropriate inputs which by returning relative outputs will help the end-user to make decisions about the cosmetics production process.

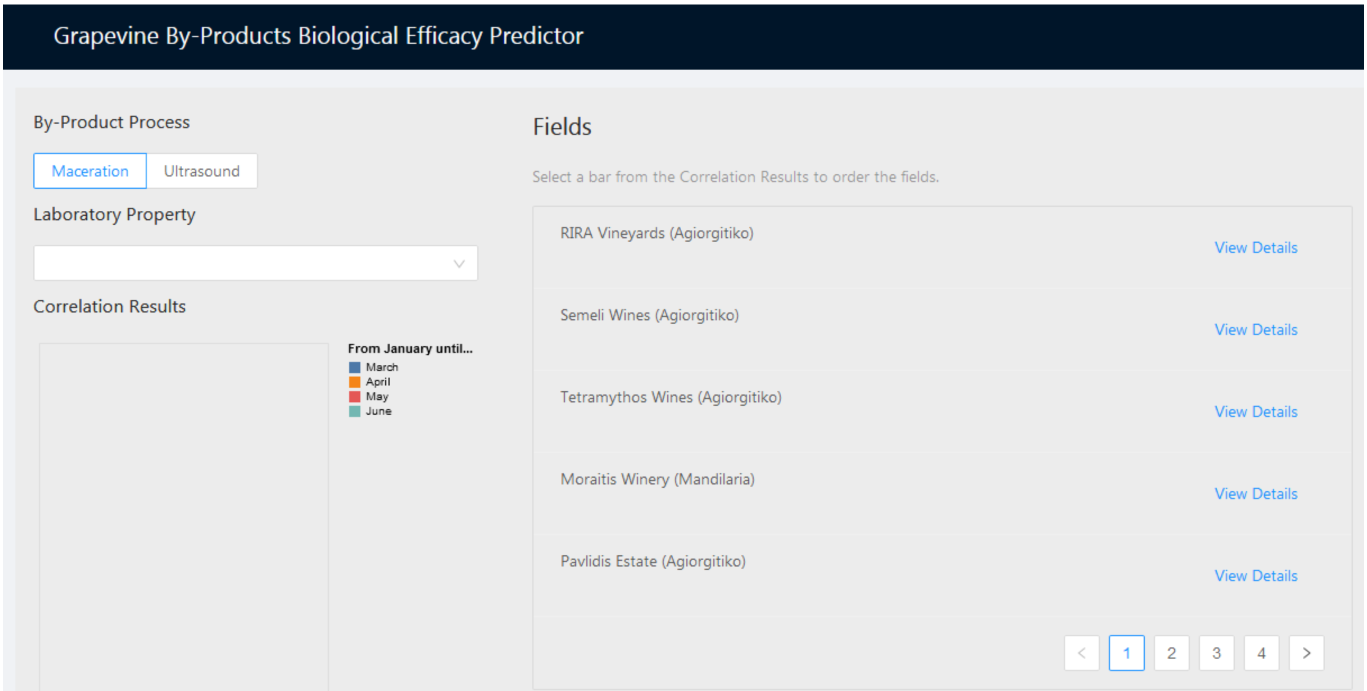


Figure 17: The dashboard configuration interface of Grapevine By-Products Biological Efficacy Predictor.

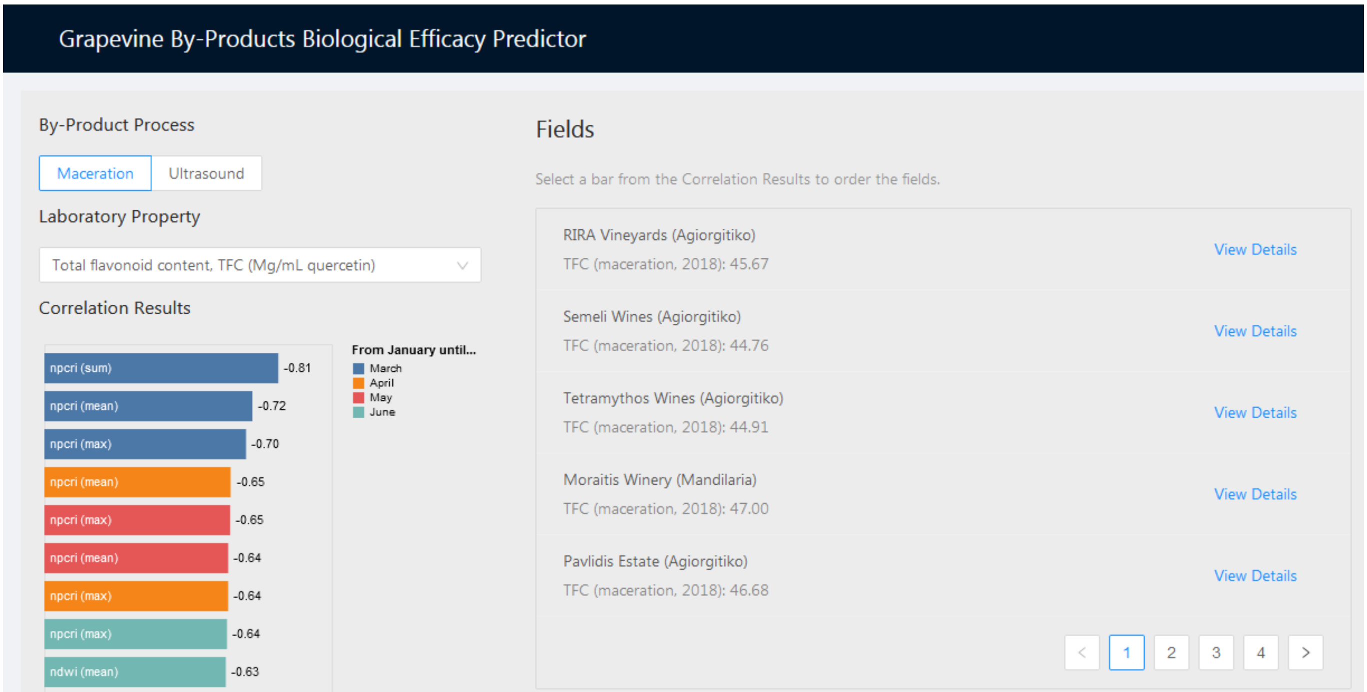


Figure 18: Correlation output showing the analysis of input variables and their correlation with B.A parameter of interest (on the left side), as well the descending order of parcels according to inputs for biological efficacy of leaves samples (on the right side).

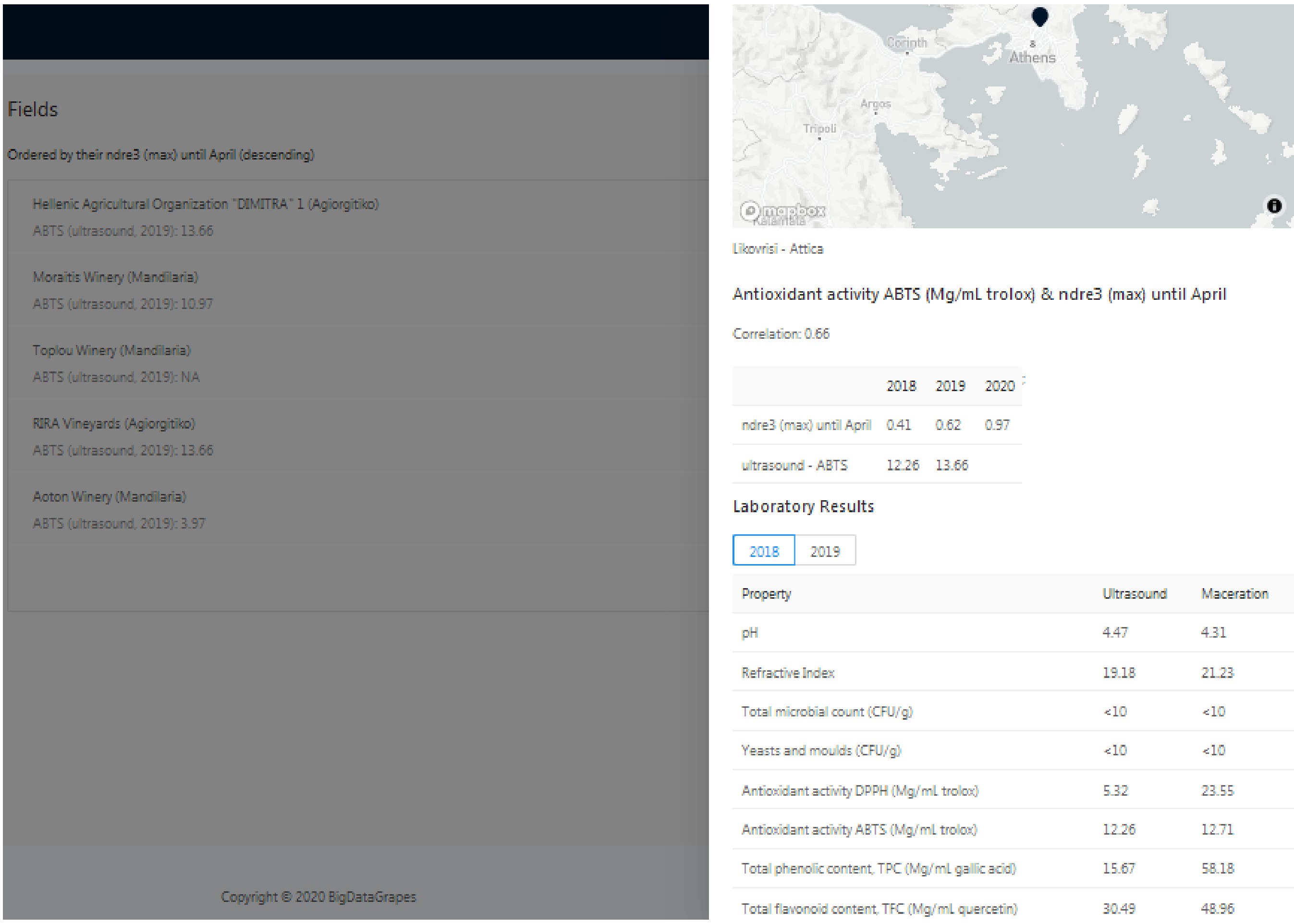


Figure 19: View Details" output for the selected parcel showing all relative information according to the end-user's final decision ".

4 Food Protection Pilot

This section of the deliverable D8.4 focuses on the application of the BigDataGrapes technologies in data-intensive and critical operations that are performed by the Food Safety and Quality Assurance (FSQA) experts in the Food industry in order to prevent food safety incidents. The FOODAKAI global incidents predictions dashboard that was developed by Agroknow using the Big Data Grapes software stack and was used during realistic operations and processes of the FSQA departments. As already presented in deliverable D8.3, the Global Predictions Dashboard presents the results of predictive analysis to FSQA experts, showcasing their ability to inform decision making based on the risk predictions for any raw material and ingredient including the ones used in the grapevine industry. This section of D8.4 focuses on the comparison of the current processes used to prevent food safety incidents with the one that is supported by the FOODAKAI platform.

4.1 Current Process for preventing the incidents

Currently the FSQA experts try to minimize the recalls in their company and the job that they need to do is to improve the two most important verification activities, namely laboratory testing and supplier verification. Their activities focus on how to make:

- the lab tests more effective.
- the audits more effective.

After working with 13 FSQA experts we identified the following pains and problems in the current process of preventing food safety incidents.

1. They still have product recalls due to an issue (hazard) that could be predicted
This highly affects the KPI that the Quality and Food Safety Groups have for the recalls (internal and external). Every time that they have a recall, they need to spend extra time for the management of recall and a significant amount of money spent when a recall happens.

2. Lab testing plan is not 100% effective. Large amount of money spent in Lab Testing that is not focused on emerging issues that can lead to a recall
This is due to the fact that the lab test plan is set up once per year for all the ingredients and is not frequently updated. The lab test does not use the information of new (emerging) risks and increasing risks (future evolution of the risk). Currently, our MQLs try to update the assessment spreadsheets by adding manually the information of new incidents that they identify using a horizon scanning solution that is not linked to their assessment tool.

3. Audits are not 100% effective. Large amounts of money spent in Audits but there are still issues that are missed and lead to a recall.
There are two important parts in audits: a) preparation of audits and b) the execution of audits. Preparation for audits is based on risk assessment of the suppliers which includes:

- risk assessment of ingredients that the suppliers is using or directly supplying to our customer and
- assessment of the supplier's food safety system, certificates and history of audits

The risk assessment of the suppliers is currently done using spreadsheets, a lot of manual work that needs time and files cannot be easily updated to highlight the emerging risks in the supply chain. Most of our customers are not updating the assessment files with incidents from the global supply chain (horizon scanning).

4.2 Benefits from the automated process

What would make the FSQA experts working in the food industry happier and help them to get the prevention job done?

- **To limit the manual work** that they are currently doing in order to identify the emerging risks that are affecting their ingredients, products and suppliers.
- **To focus their verification activities**, lab tests and audits, to the riskiest points and not just to any point of the supply chain.
- **To prevent product recalls** that happen due to inefficiency of the verification activities.

The following table presents how the Global Prediction Services can help the FSQA to get the prediction job done.

Table 11: Problems of the current process and benefits from an automated risk prediction process

PROBLEM – PAIN OF CURRENT PROCESS	BENEFIT	GLOBAL PREDICTION SERVICE
Lab testing plan is not effective.	Focus your lab test plan on the emerging risks that affect their ingredients. Limit the manual work of integrating all new incidents in your risk assessment tools.	Identify which are the ingredients with emerging risks. Identify which hazards will likely to increase in the ingredients that are at risk.
Audits are not effective	Focus your audits to suppliers that are affected by emerging risks. Limit the manual work of risk assessment for suppliers.	Identify which suppliers are affected by the emerging risks.
We still have product recalls due to an issue (hazard) that could be predicted	Prevent the recalls by having mode effective lab testing plan and audits	Identify which products are affected by the emerging risks.

4.3 Difference between current Risk Assessment approach and Global Predictions approach

Risk assessment provides an estimation of the risk based **on what we know until today.** It cannot provide information about how the risk will evolve in an ingredient. Global predictions add-on provides an estimation on how the risk will evolve. What will happen with the risks for a specific ingredient? Will they increase or decrease? Which safety parameter will increase? With the Risk Assessment package, you can identify the emerging risks but you cannot say how the risks will evolve. ***So, the Global Predictions dashboard is providing answer to "HOW THE EMERGING RISKS WILL EVOLVE".*** The emerging risk is defined by EFSA as: "A risk resulting from a newly identified hazard to which a significant exposure may occur, or from an unexpected new or increased significant exposure and/or susceptibility to a known hazard."

4.4 Benefits in cost

To measure the benefits in cost, we developed a Return of Investment (ROI) tool that focuses on the efforts that currently a food company devotes in order to identify and predict risks in their food products. The ROI tool takes into account the size of the FSQA team and the time that is devoted to specific tasks such as:

- Monitoring of food safety incidents published by the National Authorities all around the world
- Risk assessment of suppliers
- Risk assessment of ingredients
- Create reports for the hazards and potential risks in their foods

The interactive tool presented below was developed to estimate the benefit in cost.

Table 12: Return of Investment tool developed to measure the benefits in cost

A. EFFORTS DEVOTED TO RISK ASSESSMENT	Enter your data in the grey cells
Staff Efficiency: Risk assessment activities are labor-intensive, requiring experts to continuously monitor the global food safety incidents, to estimate the risk for suppliers and ingredients both using internal testing data and global data, and to create reports for the risks that need to be shared with all your team. Automation of risk monitoring and assessment can yield significant business impact. Use this ROI calculator to estimate your organization's potential impact to the bottom line.	
1. How many people are working on a full-time basis to prevent the internal and external recalls? (FTE)	20
2. Which % of their time goes to monitoring global food safety incidents (recalls, import refusals, food safety news)	5.00%
3. Which % of their time goes to risk assessment of your suppliers	5.00%
4. Which % of their time goes to risk assessment of your ingredients and raw materials?	10.00%
5. Which % of their time goes to create the reports for hazards and risks of your ingredients and suppliers?	5.00%
6. Which is the average annual cost of an expert working in your team? (in €)	€80,000.00
TOTAL ANNUAL COST OF EFFORTS TO ASSESS RISK	€400,000.00
Our promise is to reduce at least by 50% the time devoted to risk assessment	€200,000.00
Annual cost of FOODAKAI	€60,000.00
OUR GUARANTEE: give you back your investment if we will not help you to reduce efforts for risk assessment at least by 50%	
SAVINGS: This number shows the savings that can be found by using a digital solution that can reduce the time that your team needs to perform risk assessment	€140,000.00
B. PREVENTING INCIDENTS	

1. How many internal recalls (as a result of a food business's internal testing and/or auditing) does your company have every year on average?	1000
2. How many food recalls (consumer, trade, withdraw from the market) does your company have every year on average?	5
3. Which is the average cost of the internal recall? ($)	€1,000.00
4. Which is the average cost of the external recall or withdraw (consumer, trade, withdraw from the market)? ($)	€500,000.00
TOTAL COST OF INCIDENTS FOR YOUR COMPANY	**€3,500,000.00**

The ROI tool has two main parts a) the one that focuses on the savings and b) the one that focuses on the cost of the recalls that can be prevented.
In the above example of the ROI tool, we have applied the ROI calculation for a real-life scenario. More specifically, we applied it for a company that has a team of 20 FSQA and we showed to the food company that using a platform like FOODAKAI they can save at least **€140.000**. Using the average number of internal and external recalls that the company has the tool shows that the cost of the recalls is **€3.5M.**

4.5 A REAL-LIFE PREDICTION USE CASE

The following use cases were validated with the predictions that are estimated by the Global Predictions Dashboard.
USE CASE A: Company that is using sesame seeds in a number of finished products e.g., bakery products ready to eat, noodles meal. The sesame seeds are affected by an emerging risk that caused more than 140 recalls within the last two months.
Companies that this case applies to: Very large International Food Manufacturers, Very large international and local retailers.
Relevant recall (3/11/2020)

Knorr Wereldgerechten Chinese Beef Shanghai
- MRDR: 68339551
- THT: **alle houdbaarheidsdata tot 03/2022**
- EAN Consumenteneenheid (CE): 8710522736241
- EAN Handelseenheid (HE): 8710522736357

Knorr Wereldgerechten Chinese Beef Shanghai
- MRDR: 67873458
- THT: **alle houdbaarheidsdata tot 03/2022**
- EAN Consumenteneenheid (CE): 8714100247440
- EAN Handelseenheid (HE): 8714100108758

Figure 20: Example of a food recall that could be prevented using the Global Predictions Dashboard

From October 2020 the Global Predictions dashboard highlighted that sesame seeds is and will be at risk as shown in the following screen.

Incidents for sesame seeds

ACTUAL AND PREDICTED INCIDENTS FOR SESAME SEEDS

Ingredient	Current year's incidents	Next year's incidents	Tendency
sesame seeds	94	227	141 %

Figure 21: Increase in the incidents for sesame predicted by the Global Predictions Dashboard

It highlights that there is an emerging risk for ethylene oxide and predicts that there will be a further increase of the risk for ethylene oxide.

Hazards likely to increase for sesame seeds

HAZARDS LIKELY TO INCREASE

Hazard	Current year's incidents	Next year's incidents	Tendency
unauthorised substance ethylene oxide	54	3868	7063 %
unauthorised ingredient (fraud)	54	75	39 %
salmonella	39	75	92 %

Figure 22: Prediction of hazards that will likely increase within the next months

Figure 23: Predictions of risks for sesame seeds

The Global Predictions Dashboard highlights the finished product and supplier that will be affected by this emerging risk.

Your finished products that may be affected

PRODUCTS AND SUPPLIERS LIKELY TO BE AFFECTED

Product	Hazard	Risk
Honey sesame bar	unauthorised ingredient (fraud)	6.59
Noodles	unauthorised substance ethylene oxide	9.46
SUPPLIER A	unauthorised substance ethylene oxide	9.46

Figure 24: Products and suppliers that may be affected by the predicted risks

Using this very important information for the emerging risk, the FSQA department of the company will immediately:

- Include this parameter in the lab test plan of the ingredient to make sure that the there is no such chemical hazard in the ingredient
- Ask suppliers who are using this ingredient to provide certificate of analysis for the specific parameter
- Plan remote audits for suppliers that are affected by this emerging risk

Benefit

Using the information about this predicted emerging risk the company will make more effective the lab testing plan and audits and **finally will prevent a recall of its products.**

4.6 Feedback from Food Companies about the benefits of the automated process

During the BigDataGrapes project, we conducted sessions with the FSQA experts of food companies to get their feedback about the benefits of using an automated system that can help them in preventing the food safety incidents. We summarize the feedback from the experts of a food company that has a FSQA team of 13 people in the following diagrams.

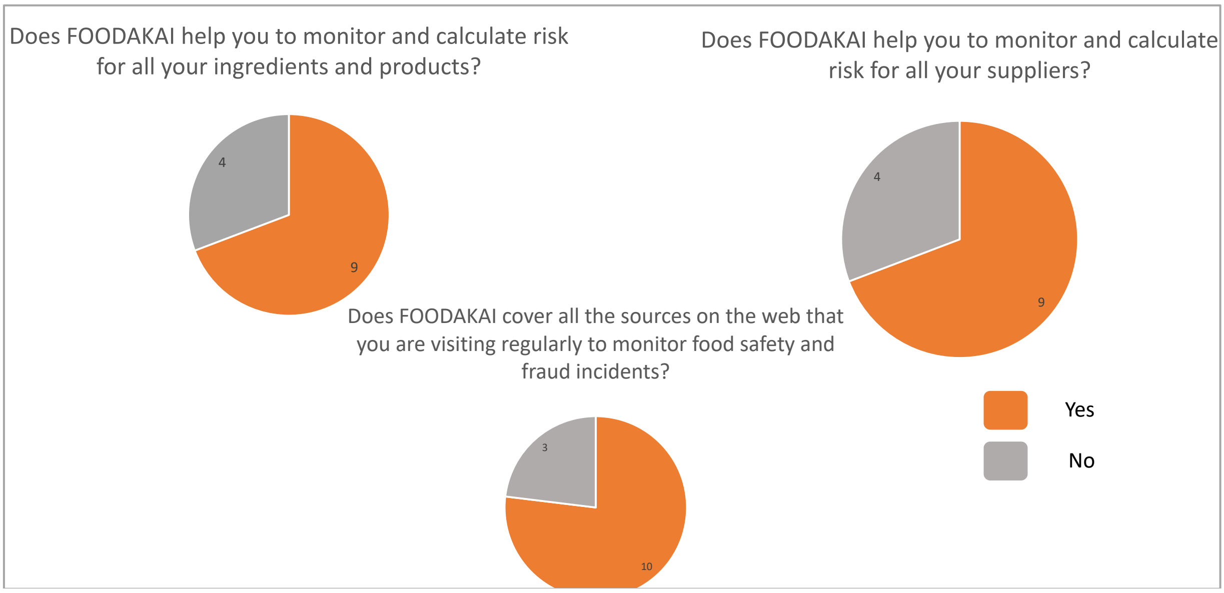


Figure 25: Validation of the expected benefits that the automated system can provide

As presented in Figure 25, the majority of the experts agree that an automated system can help them in the tasks that they perform for risk prevention.

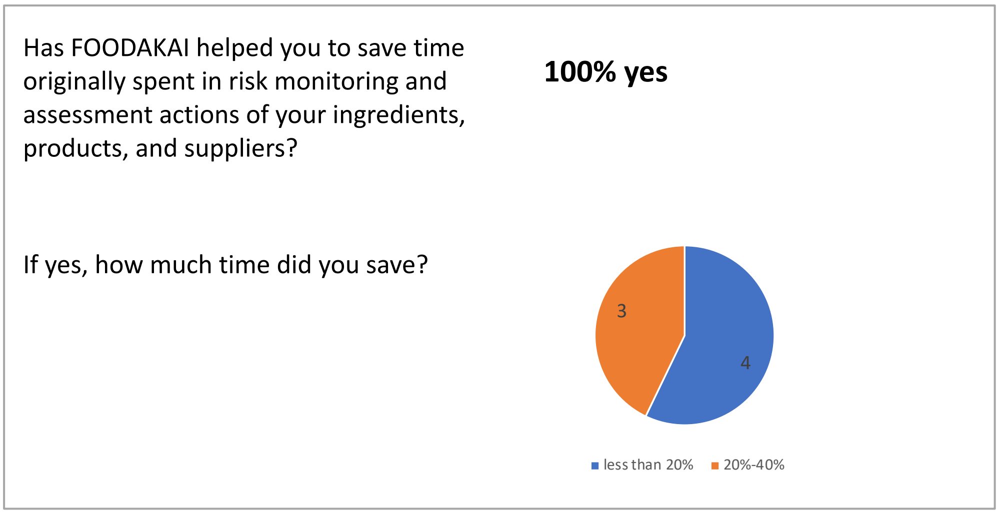


Figure 26: Validation of the expected benefits that the automated system can provide in terms of saving time

As one can see in the feedback of experts regarding time that can be saved using an automated system like the FOODAKAI predictions dashboard, all the experts agree that it can save time and most of them mentioned that 20-40% of their time spent in risk prevention activities can be saved.

Figure 27: Validation of the most important problems that the FSQA experts have when it comes to risk prevention

A very important aspect that was validated with the food company was the problems that the FSQA experts face when it comes to risk prevention. As we can see the majority of the experts mentioned that the most important problem is how to move from reaction to prevention to be able to predict risks before they happen.

5 CONCLUSIONS

The main goal of Deliverable 8.4 "Integration and Operation with real-life Practices" was the development of a dashboard, targeting industry-level decision makers (natural cosmetics industry end-user) and practitioners (grapevine farming end-user), that will incorporate the appropriate functionalities of the BigDataGrapes software stack used in the relevant piloting session "Natural Cosmetics Pilot" and the "Food Protection Pilot". The data input and the fulfilment of the pilot trials provide all necessary data for the assessment of BDG components and refine the pilot itself by the subsequent iterations after also the end-user's evaluations. The collected data from the natural cosmetics pilot provide the necessary information for the evaluation of the quality of each sample, linked with the special characteristics of the vineyard of origin. In addition, the same data is analysed for their correlation and will be incorporated to models used to build the DSS of the dashboard. The models using as input data SVIs of vineyards, WD and BA parameters of grape leaves extracts help decision making about the incoming raw materials of the cosmetics industry as well as the practitioners of grape production for the quality of their by-products through a friendly to the end-user dashboard, the so-called demonstrator.

The performance of the Natural Cosmetics Pilot software demonstrator has been assessed and appraised from 15 relative end-users selected mainly from cosmetic industry and food industry, but also from research organisations and grapevine practitioners, as presented in deliverable D8.5. The FOODAKAI predictions dashboard has also been evaluated by 25 users, in the context of the project assessment activities presented in D8.5. These included food safety and quality assurance experts, food scientists and business experts from the industry and academia. They confirmed the potential of the tool not only as a valuable risk assessment and prediction tool in the context of the day to day operations of a company in the food chain, but also as a tool to support food science research and academic activities. In this task we also focused on the comparison of the current processes used by the experts with the automated one that is supported by the predictive dashboards. Based on the feedback of the experts it is concluded that the important tasks that are currently performed manually such as the monitoring and management of data, can be automated with the help of the BigDataGrapes software stack.

The feedback from the assessment of both demonstrators confirms that the selected modules address important real-world user needs and their implementation offers an effective user experience, including useful visualisations. The suggestions for improvements that the evaluation participants provided could be employed to improve the demonstrators in their next versions.

6 REFERENCES

1. Amarowicz *et al.* (2008). Grapevine leaves as a source of natural antioxidants. *Polish Journal of Food Nutrition and Science* 58, pp. 73-78.
2. Fiume, M.M. (2012). Safety Assessment of *Vitis Vinifera* (Grape)-Derived Ingredients. Cosmetic Ingredient Review Expert Panel 2012. Pp. 1-35.
3. Letsiou *et al.* (2017). Skin Protective Effects of *Nannochloropsis gaditana* Extract on H2O2-Stressed Human Dermal Fibroblasts. *Frontiers in Marine Science* 4, pp. 221.
4. 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.
5. 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.
6. 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.
7. Katalinic *et al.* (2009). Insight in the phenolic composition and antioxidative properties of *Vitis vinifera* leaves extracts. *Polish Journal of Food Nutrition and Science* 1, pp. 7-15.
8. Rossi, V. *et al.* (2014). Addressing the implementation problem in agricultural decision support systems: The example of vite.net®. *Computers and Electronics in Agriculture* 100, pp. 88-99.
9. Wachowiak, M.P. *et al.* (2017). Visual analytics and remote sensing imagery to support community-based research for precision agriculture in emerging areas. *Computers and Electronics in Agriculture* 143, pp. 149-164.

7 Appendix

7.1 Correlation analysis results for Maceration

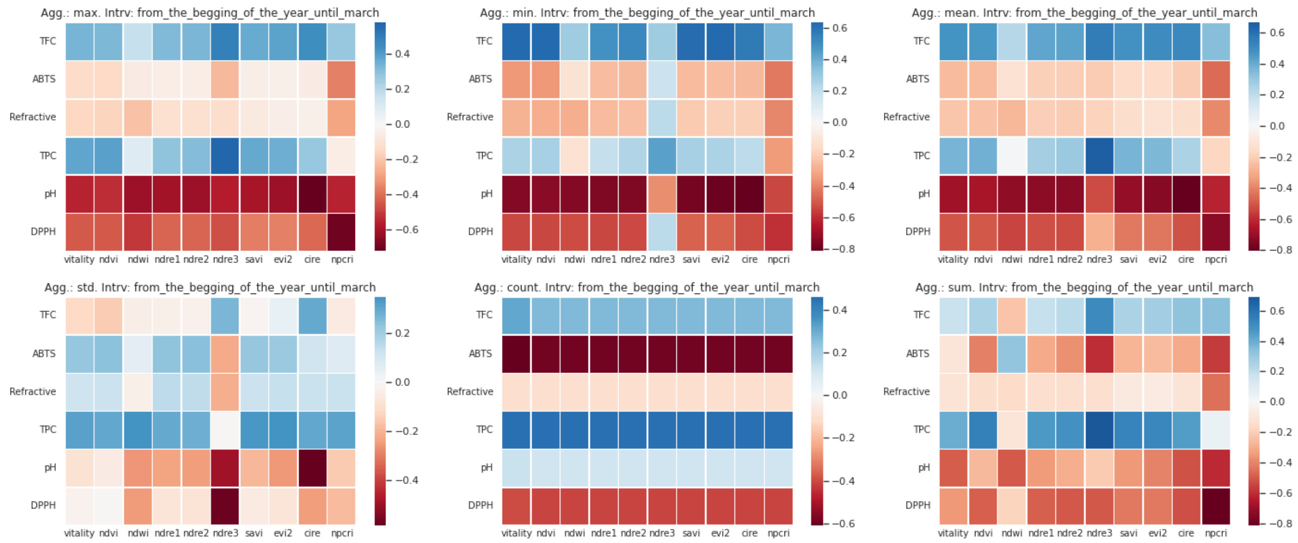


Figure 28: Correlation analysis for time frame "from the begging of the year until March" for MAC

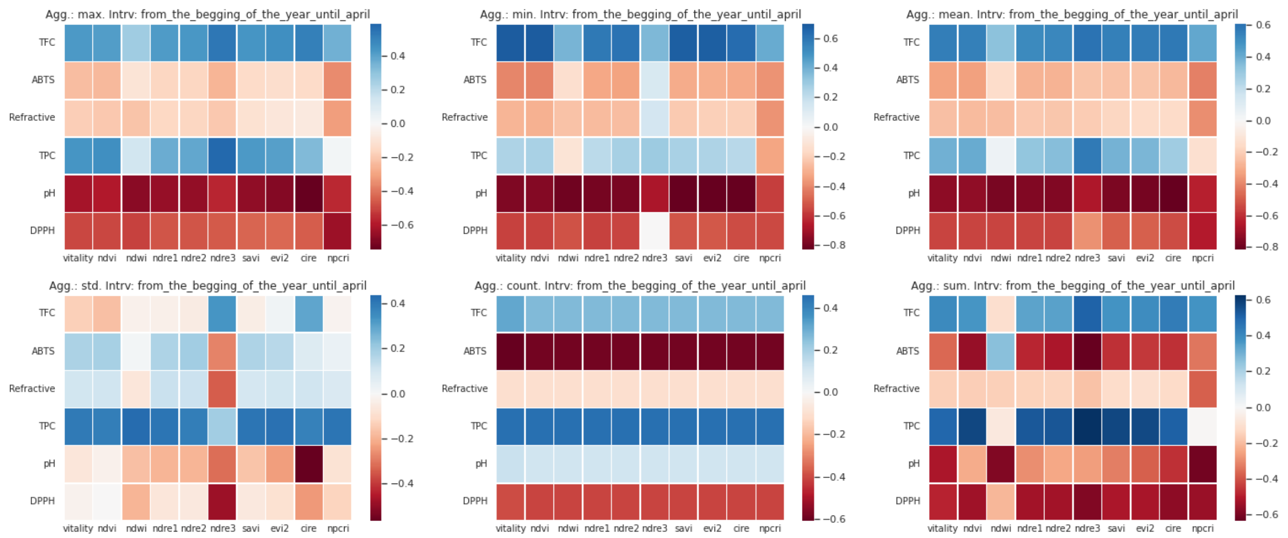


Figure 29: Correlation analysis for time frame "from the begging of the year until April" for MAC

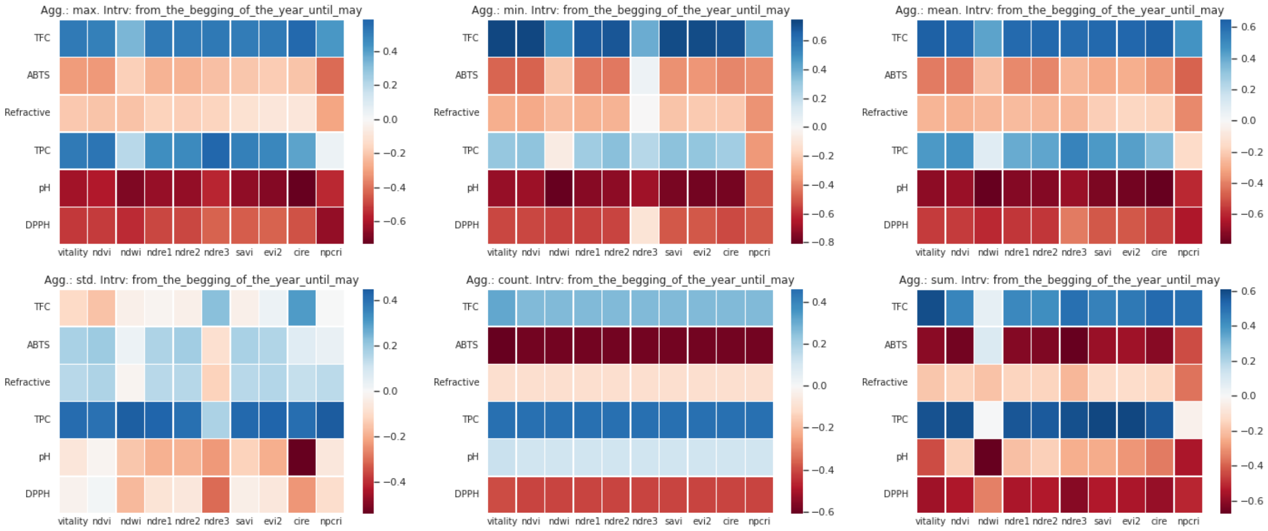


Figure 30: Correlation analysis for time frame "from the begging of the year until May" for MAC

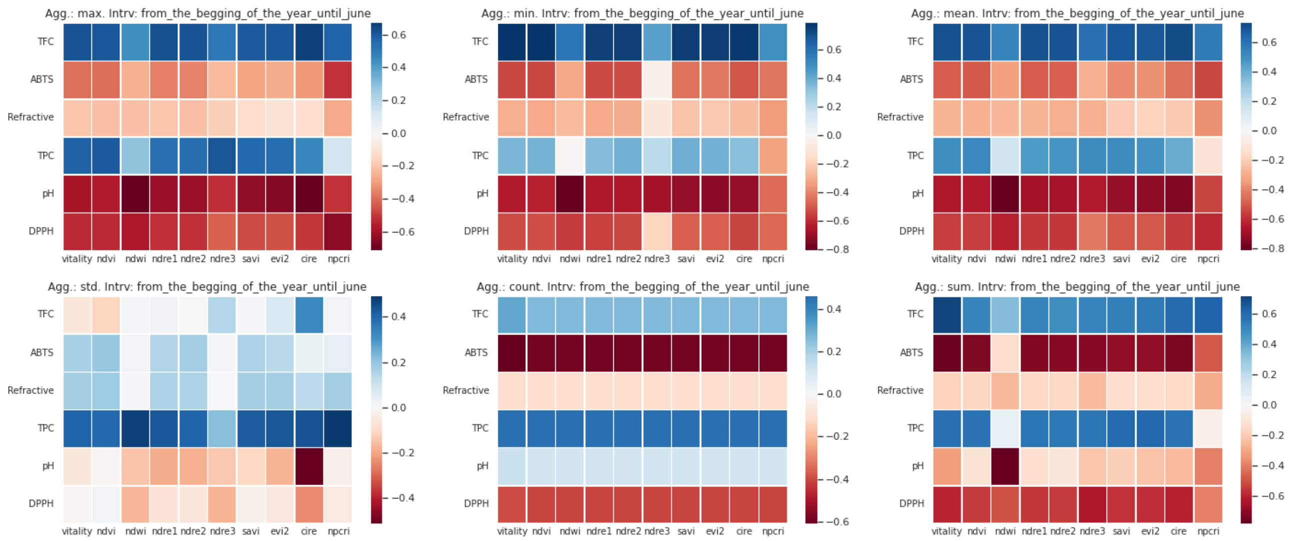


Figure 31: Correlation analysis for time frame "from the begging of the year until June" for MAC

7.2 CORRELATION ANALYSIS RESULTS FOR ULTRASOUND ASSISTED EXTRACTION

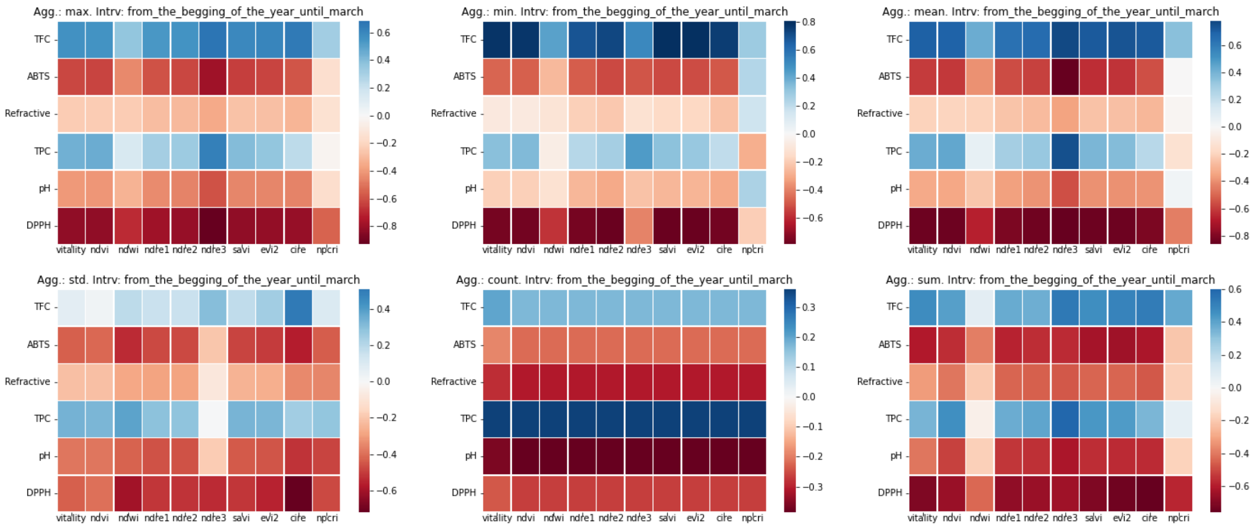


Figure 32: Correlation analysis for time frame "from the begging of the year until March" for UAE

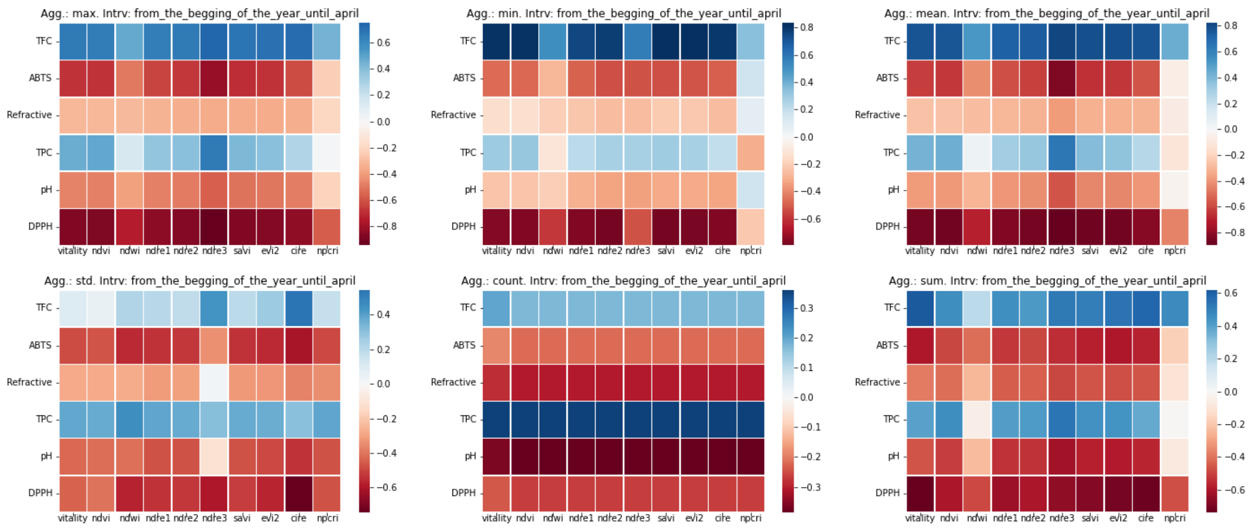


Figure 33: Correlation analysis for time frame "from the begging of the year until April" for UAE

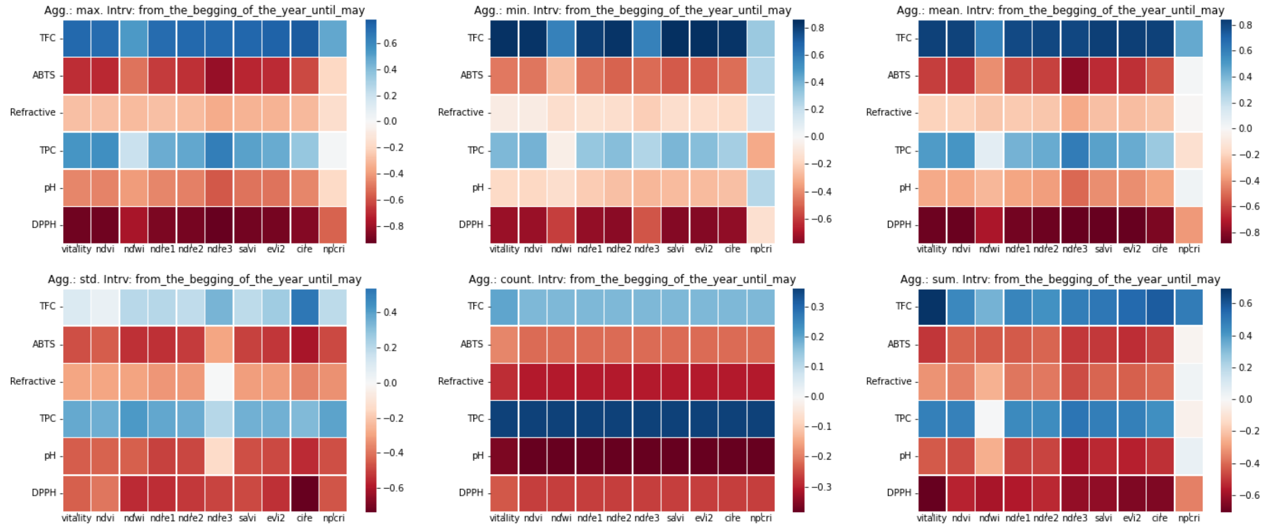


Figure 34: Correlation analysis for time frame "from the begging of the year until May" for UAE

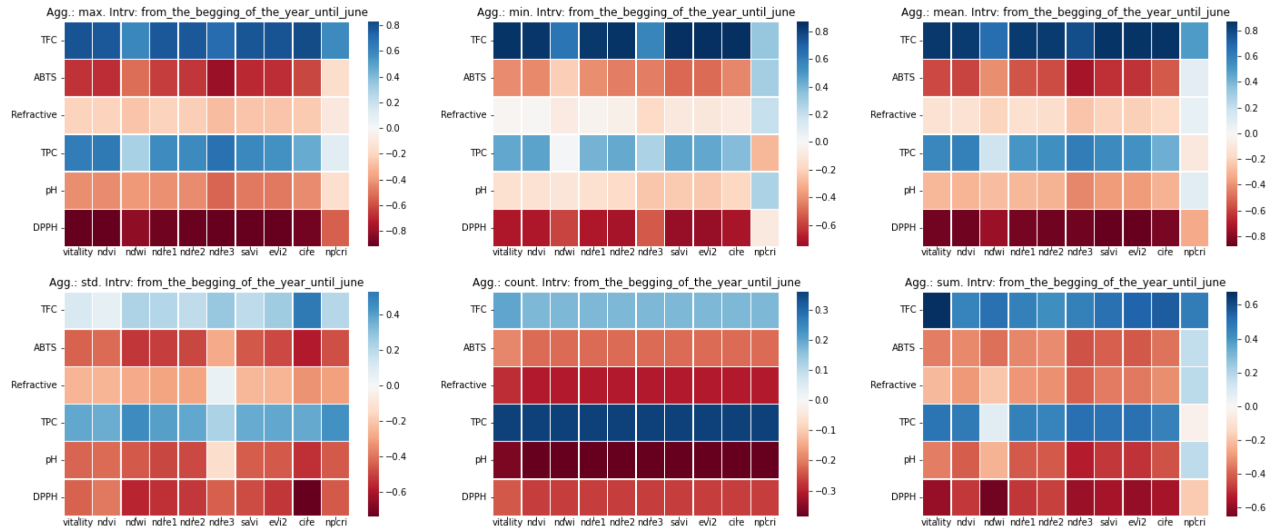


Figure 35: Correlation analysis for time frame "from the begging of the year until June" for UAE