Ref. Ares(2019)4019693 - 25/06/2019

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	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	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)
DATE OF DELIVERY (SUBMITTED)	28 June 2019 (M18)
VERSION | STATUS	1.0 | Final
NATURE	R (Report)
DISSEMINATION LEVEL	PU (Public)
AUTHORS (PARTNER)	Pantelis Natskoulis (Symbeeosis)
CONTRIBUTORS	Aikaterini Kasimati (AUA), Florian Schlenz (GEOCLEDIAN), Nyi-Nyi Htun (KULeuven), Nikola Tonellotto (CNR), Panagiotis Zervas (Agroknow), Mihalis Papakonstadinou (Agroknow)
REVIEWER	Aikaterini Kasimati (AUA), Vasileios Psiroukis (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)

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)		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 LONG LIVE LIFE S.A. (Symbeeosis, Greece)		Kostas Gardikis Email: c.gardikis@gmail.com

ACRONYMS LIST

BDG	BigDataGrapes
WP	Work Package
D	Deliverable
DSS	Decision Support System
REST API	Representational State Transfer Application Program Interface
SVIs	Spectral Vegetation Indices
Vit	Vitality
NDVI	Normalized Difference Vegetation Index
NDRE1	Normalized Difference Red Edge Index (v1)
NDRE2	Normalized Difference Red Edge Index (v2)
NDWI	Normalized Difference Water Index
SAVI	Soil Adjusted Vegetation Index
EVI2	Enhanced Vegetation Index 2
CI-RE	Chlorophyll Index - Red Edge
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
MTT	3-(4,5-diMethylThiazol-2-yl)-2,5-diphenylTetrazolium bromidefor
SIRT1	Sirtuin 1

EXECUTIVE SUMMARY

The deliverable D8.4 "Integration with existing real-life Practices" will showcase the application of BigDataGrapes technologies in data-intensive and critical operations related to the Natural Cosmetics Pilot. More specifically, the ultimate goal is to prepare 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 will use the results of predictive analysis over data samples managed by expert users and will showcase 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 first version of the deliverable the scope is to provide an initial assessment of the hypothesis that BigDataGrapes (BDG) software stack can serve for the prediction of the Biological Activity (BA) parameters by applying correlation analysis to Satellite-based Spectral Vegetation Indexes (SVIs) and demonstrating the future steps to build up the dashboard. Updated versions of this deliverable, including finalised dashboard, are due to M30 and M36 of the project lifetime.

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, CNR 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.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

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

The present intermediate version reports the findings so far and the remaining developmental stages that will 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) grape leave 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 were collected from satellite scanning for all sampled vineyards and were correlated with the BA parameters.

Following the initial correlation analysis, the building of the models that will support the DSS will be implemented. Finally, a dashboard that will expose the required functionality to practitioners and cosmetic industry's end-users will be developed and used during realistic operations and processes of these stakeholders. The dashboard will present 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 results of the automated system will be compared with those of a full manual analysis to indicate the precision of the system and quantify the benefits in time, effort and cost.

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 Use Case Scenario are presented, as well as the Data Analytics and Processing. This last 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, BA Analysis, Data Correlation Analysis, and the Software Stack, Data Transformation and Uploading. Finally, in the same Chapter 3, Subchapter 3.4 is dedicated to the Visualisation process using as an example the Wine-making dashboard already prepared for the relevant Pilot. Chapter 4 is a summation of the Conclusions and Future Steps of the Deliverable at the present phase.

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 will provide the necessary information for the evaluation of the quality of each sample, linked with the special characteristics of the vineyard of origin. The goal is to face the challenge: "how data from the field can be linked to the biological efficacy of final products". This will serve the main purpose of the present 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 will benefit 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 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, and thus BDG will use interactive visualisations through a user-friendly dashboard to help the decision-making process.

2.2 TECHNOLOGICAL GUIDANCE

For the first period of the project (M1-M18), 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 where dried to a draught aired place by avoiding direct exposure to sun and sent to the relevant pilot partner (Symbeeosis) where extraction and laboratory analyses took place.

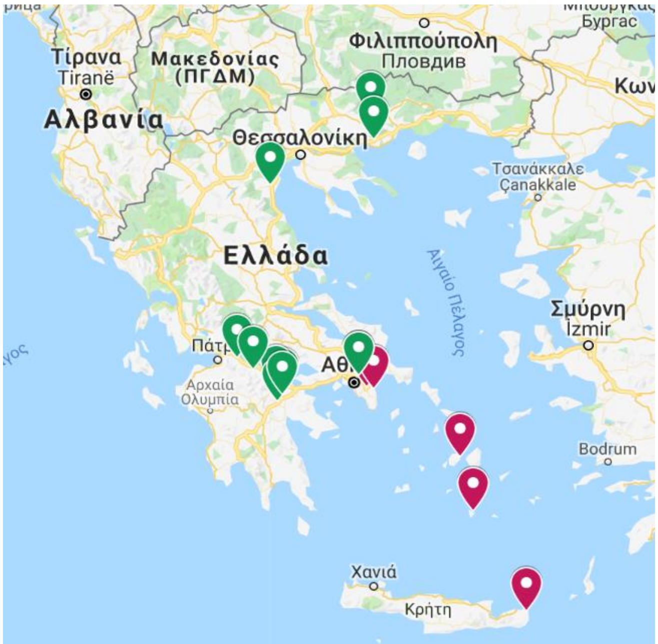


Figure 1: Dispersion of samples across the Greek territory (Correspondent file: https://tinyurl.com/y4scyhed)

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 "DIMITRA"	Mandilaria	Attica	Lykovrisi
15	Hellenic Agricultural Organization "DIMITRA"	Agiorgitiko	Attica	Lykovrisi
16	Agricultural University of Athens	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 Industrial Park of Markopoulo Mesogaias in Greece. At the laboratory conducted extractions under the two different methods and the following measurements of biological activity: pH, RI, TPC, TFC, Total Microbial Count, Yeasts & Moulds, DPPH & ABTS assay, as pointed also in Table 2. The measurement of toxicity (MTT assay) and of gene expression (target SIRT1 mRNA transcripts) on skin cells are still pending and will be contacted together with the second-year samples analyses[3]. (Correspondent file: **https://tinyurl.com/y5z8e8lf**).

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
Toxicity on skin cells	MTT assay	Laboratory equipment	xls	MB
Gene expression on skin cells	Real time PCR	Laboratory equipment	xls	MB

2.4 ENVISAGED OUTCOME

We want to examine how the biological efficacy (in terms of BA parameters described above) depends on the location of the vineyard, 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 the biological efficacy based on the location 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[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 will provide 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 will point out which information should be taken into account for building the models that will support the DSS. The goal is 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 to be 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 first year of the BDG project and for the intermediate phase of the present deliverable, 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. 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 (NDVI, NDRE1, NDRE2, NDWI, SAVI, EVI2, CIRE)	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		

Figure 2 presents the Pilot's data gathering timeline.

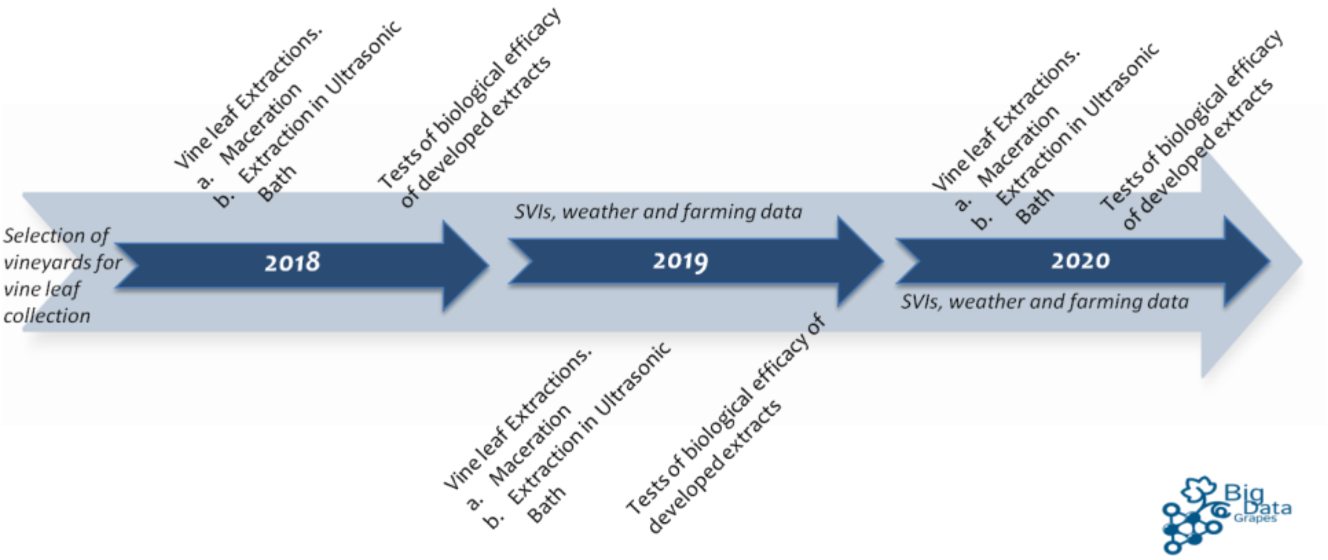


Figure 2: Data gathering timeline for Natural Cosmetics Pilot

3.2 Use Case Scenario - Predicting Biological Efficacy

The scenario presumes that precision farming and control of parameters linked to the quality of wine 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 BA data to parameters linked with biological efficacy. The appropriate algorithms created will use the existing datasets and explore the relationship between them. Datasets concerning SVIs 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 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 indices and thereby reaching conclusions regarding biological activity of incoming by-products. A farmer, on the other hand, can perform decision-making by evaluating location and SVIs indices 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.

3.3 Data Analytics And Processing

Until the present phase, two (2) data categories have been incorporated to the Pilot: SVIs by applying the GEOCLEDIAN Ag|knowledge tool that exploits open source satellite data and extract information regarding the vegetation indexes on the location of each crop, and Data related to vine leaves of two (2) different grape varieties (Agiorgitiko and Mandilaria) and their phytochemical profile and BA after extraction by UAE and MAC. The collected data from the Natural Cosmetics Pilot will provide the necessary information for the evaluation of the quality of each sample, by linking the BA parameters with the SVIs Data of the vineyard of origin. Finally, the correlation analyses of BA parameters with the SVIs will point out which information should be taken into consideration for building the models that will 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 (NDVI, NDRE1, NDRE2, NDWI, SAVI, EVI2, CIRE) 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 access also Landsat 8 satellite data. Collectively, the SVIs taken into account among their description 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 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 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 presents also Nitrogen sensitivity and thus canopy Chlorophyll & Nitrogen contents can be derived from this index.

All SVIs were considered for four-time frame aggregations:

1. from 30 days before the crop date of the sample
2. from 1/1/2018 to the sample date in the same year (different for each y)
3. from 1/1/2018 to the min of samples dates in the same year (different for each y)
4. from the time we identify the min value of y since the beginning of the sampling season

For each time frame, partner computed the average of the observation values 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 collects and processes the described satellite data of all sites. Visible images and Spectral Vegetation Index Maps will be produced, and the data will automatically be 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.

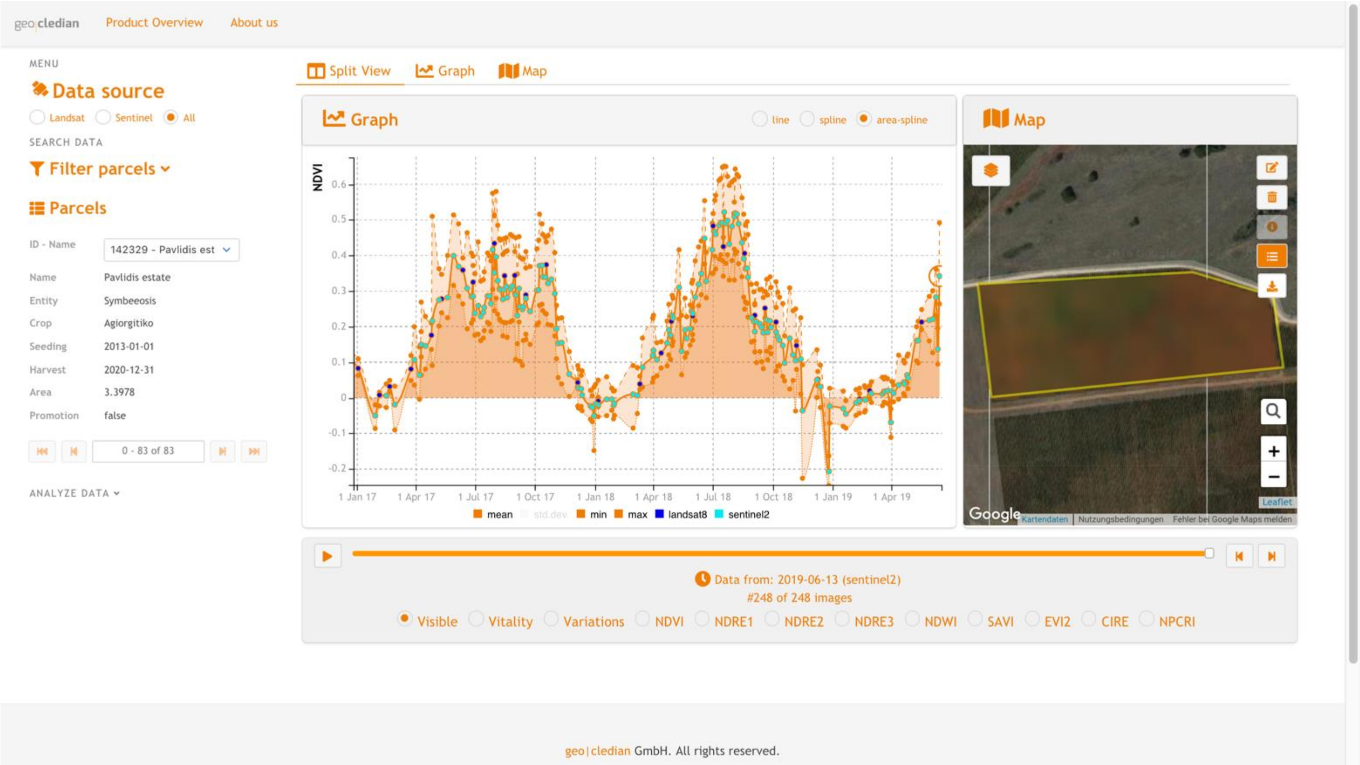


Figure 3: A Sentinel-2 RGB image for 13th June 2019 for a parcel of the Pavlidis estate, Greece. Also shown are all available Landsat 8 and Sentinel-2 image acquisitions for 2017 - 2019.

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 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 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. Finally, MMT and SIRT1 assays are relative to the anti-ageing properties[3] of the extract, but due to the high complexities regarding their analyses were left to be measured together with the 2nd year samples and will be included in the forthcoming phases of the present report. 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)

All 1st year sampling measurements can be accessed at the following link: **https://tinyurl.com/y5tnykgr**

3.3.3 Data Correlation Analysis

The samples used for the correlation analysis have been specified in Table 5 and can be found on the URL: https://docs.google.com/spreadsheets/d/1_kRsyd-bgfZHFi3XhS-g7qxJIp2qkI-7PtNYMMo69Qc.
The columns Sample_Id corresponds to the identification of the sample under study and the column Parcel_id link the sample to the parcel representation in the GEOCLEDIAN system. The samples I.A.3, I.M.6, I.A.14, I.M.15 and I.M.16 have been excluded since there is no corresponding Parcel_id to these samples while samples I.A.7 and I.A.12 have also been filtered out from the analysis since both have the same Parcel_id. All rejected samples will be included to the analysis in the next version of the Deliverable 8.4 after correcting the input errors.

Table 5: Wineries of samples included in correlation analysis

#	Sample ID	Vineyard	Variety	Region	City	Parcel ID
1	I.A.1	RIRA Vineyards	Agiorgitiko	Peloponnese	Aigio	135973
2	I.A.2	Semeli Wines	Agiorgitiko	Peloponnese	Nemea	135939
3	I.M.4	Moraitis Winery	Mandilaria	Aegean	Paros	136162
4	I.A.5	Pavlidis Estate	Agiorgitiko	Northern Greece	Drama	135947
5	I.A.8	Biblia Chora Estate	Agiorgitiko	Northern Greece	Kavala	136168
6	I.M.9	Toplou Winery	Mandilaria	Crete	Sitia	136164
7	I.A.10	Skouras Domaine	Agiorgitiko	Peloponnese	Argos	136159
8	I.M.11	Aoton	Mandilaria	Attica	Peania	136167

9	I.A.13	Winery Papagiannakos Domaine	Agiorgitiko	Attica	Markopoulo	136169

According to the datasets available, correlation analysis assigned as "observed variables" 6 × 2 vectors (BA parameters × maceration & ultrasound measurements) of 9 components (parcel ids) and as "variate variables" 8 × 6 × 4 = 192 vectors (SVIs × aggregation functions × time frames) of 9 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 correlations due to the extend are removed in Deliverable's 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 6-9 are presented only the most correlated SVIs with BA parameters.

Table 6: Most correlated SVI for each BA parameter after MAC.

BA parameter	Most corr. SVI	Highest Corr. Value	Time Frame	Aggregation
pH	savi	-0.735546	interval from min value observation date until the crop date	std value
TPC	ndwi	0.738708	from 30 days before the crop date of the sample	min value
TFC	ndwi	0.542995	from 30 days before the crop date of the sample	min value
RI	vitality	-0.592526	from min value observation date until the crop date	sum value
AA1	evi2	-0.693165	from min value observation date until the crop date	std value
AA2	ndvi	-0.650048	from min value observation date until the crop date	max value

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

BA parameter	Most corr. SVI	Highest Corr. Value	Time Frame	Aggregation

pH	ndvi	-0.758273	from min value observation date until the crop date	std value
TPC	ndwi	-0.769248	from begging of the year until min value observation date	mean value
TFC	ndwi	0.68462	from 30 days before the crop date of the sample	min value
RI	savi	0.687912	from min value observation date until the crop date	sum value
AA1	evi2	0.649713	from 30 days before the crop date of the sample	max value
AA2	evi2	-0.632314	from min value observation date until the crop date	std value

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

SVI	Most corr. BA parameter	Highest Corr. Value	Time Frame	Aggregation
vit	RI	-0.592526	interval from min value observation date until the crop date	sum value
savi	pH	-0.735546	interval from min value observation date until the crop date	std value
ndwi	TPC	0.738708	from 30 days before the crop date of the sample	min value
ndvi	AA2	-0.650048	interval from min value observation date until the crop date	max value
ndre2	TPC	0.459028	interval from min value observation date until the crop date	sum value

ndre1	pH	0.663137	interval from min value observation date until the crop date	max value
evi2	AA1	-0.693165	interval from min value observation date until the crop date	std value
cire	TPC	0.519241	interval from min value observation date until the crop date	count value

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

SVI	Most corr. BA parameter	Highest Corr. Value	Time Frame	Aggregation
vit	RI	0.665395	from 30 days before the crop date of the sample	count value
savi	RI	0.687912	from min value observation date until the crop date	sum value
ndwi	TPC	-0.769248	from the begging of the year until min value observation date	mean value
ndvi	pH	-0.758273	from min value observation date until the crop date	std value
ndre2	AA1	0.567201	from 30 days before the crop date of the sample	mean value
ndre1	AA1	-0.617066	from the begging of the year until min value observation date	std value
evi2	RI	0.664826	from minvalue observation date until the crop date	std value
cire	pH	0.698912	from minvalue observation date until the crop date	max value

According to the correlation results there is observed a constantly high correlation between ndwi and TPC, 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 evi2, AA2 with ndvi and evi2, and TFC with ndwi. As soon as datasets are completed with the missing data (e.g. rejected Parcel_Ids, forthcoming seasons samples) all the correlation analyses will be contacted again. Datasets will also strengthen up with the repetition of samplings each year so that the upcoming phases of the Deliverable 8.4 will consider all the information available in order to produce a more reliable DSS of the dashboard.

3.3.4 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 xls 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,
- Geoclidean'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.

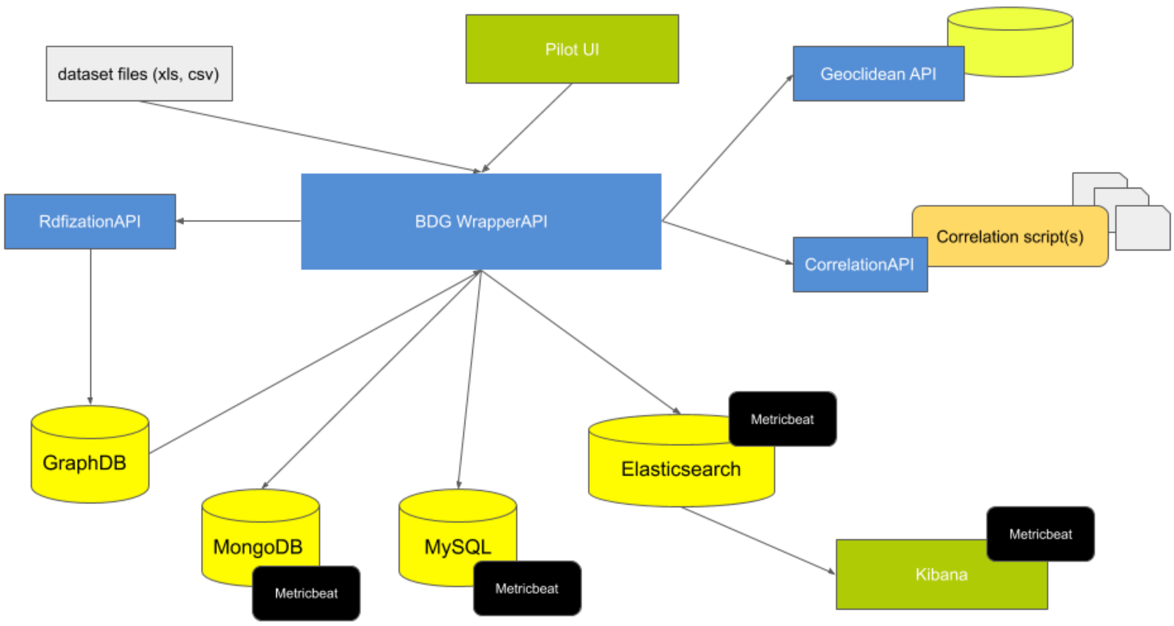


Figure 7: An abstract view of the BDG stack components used in Symbeeosis data flows

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 also 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.

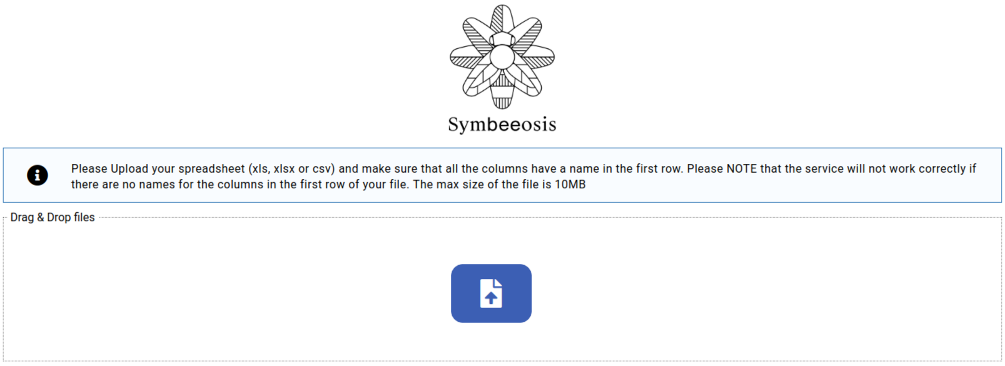


Figure 8: The initial view of the UI tool for the dataset upload for the Symbeeosis case

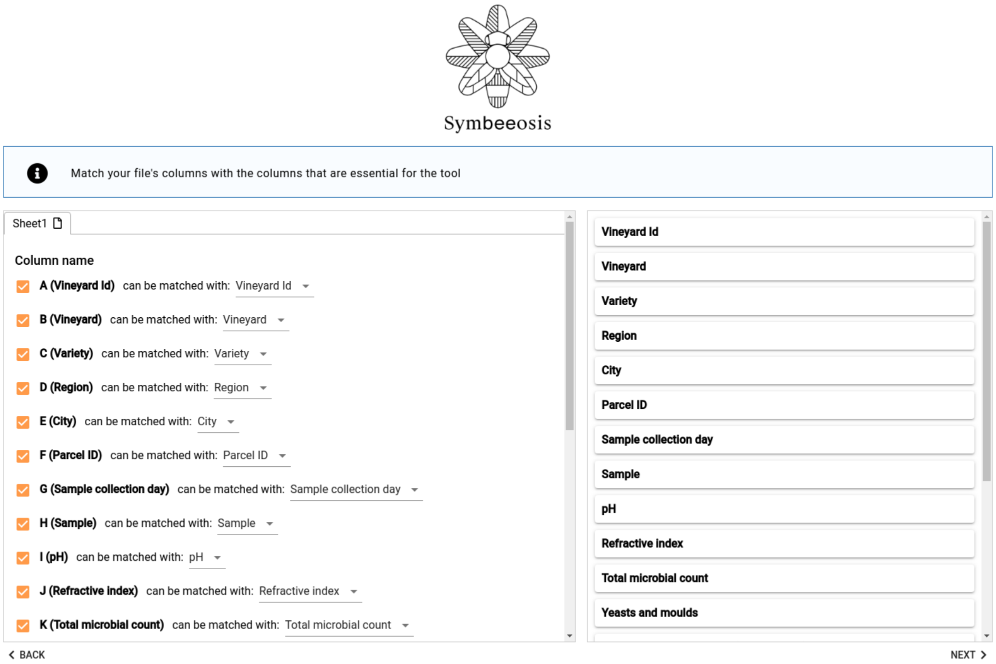
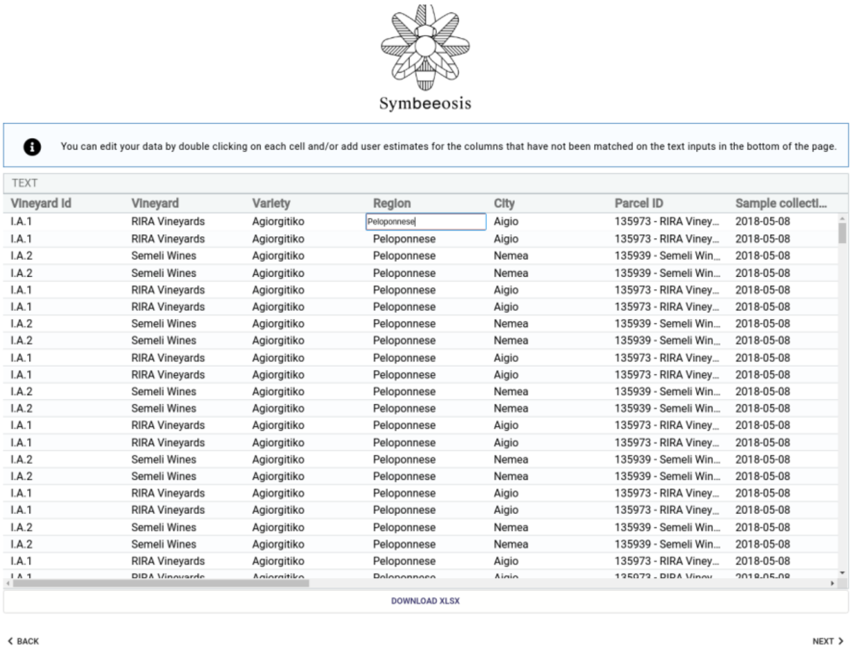


Figure 9: The screen responsible for the matching of the uploaded xls schema with the Symbeeosis one



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 xls 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

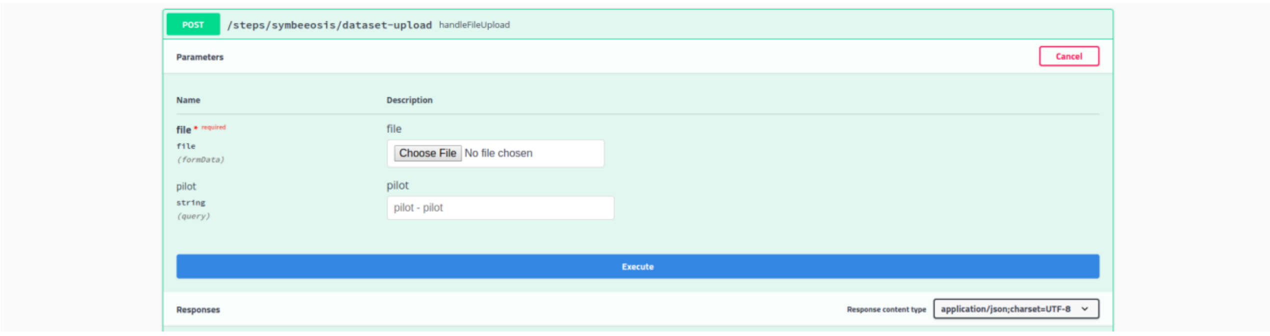


Figure 12: Swagger documentation of the API endpoint for the dataset upload

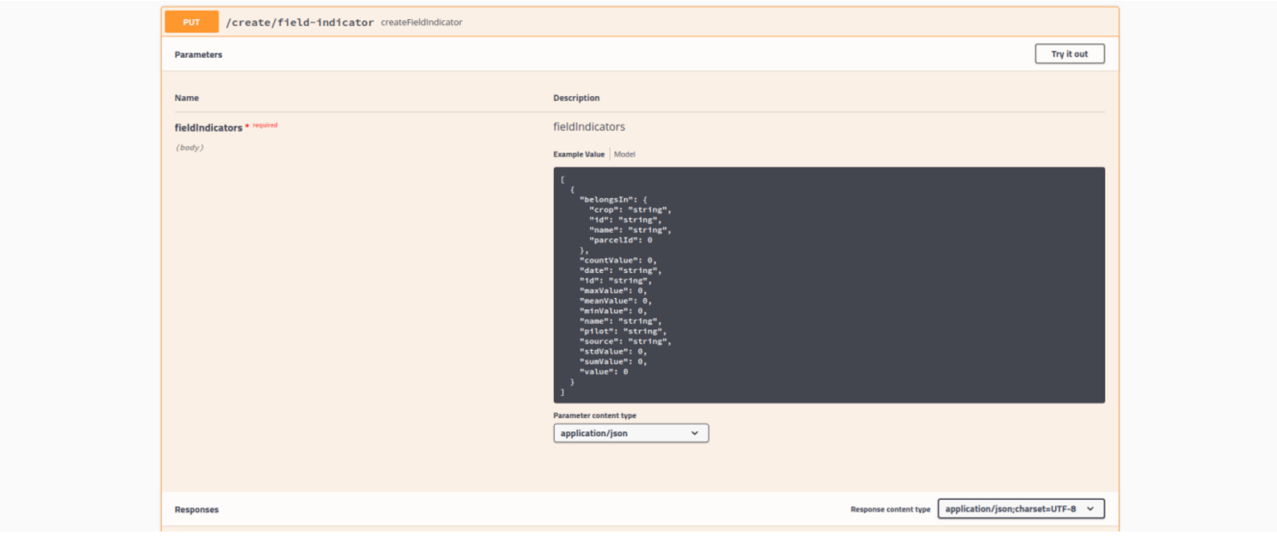


Figure 13: Swagger documentation of the API endpoint for distinct record upload

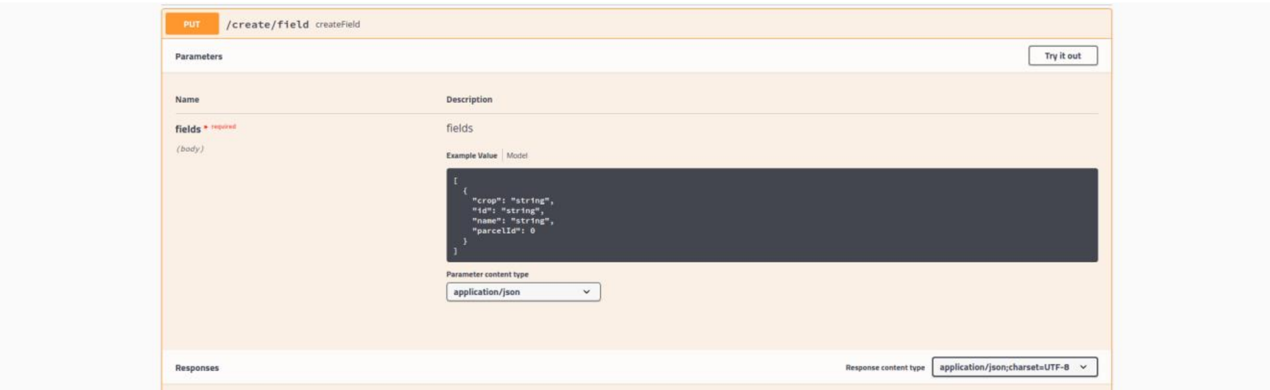


Figure 14: Swagger documentation of the API endpoint for the field metadata/description

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.

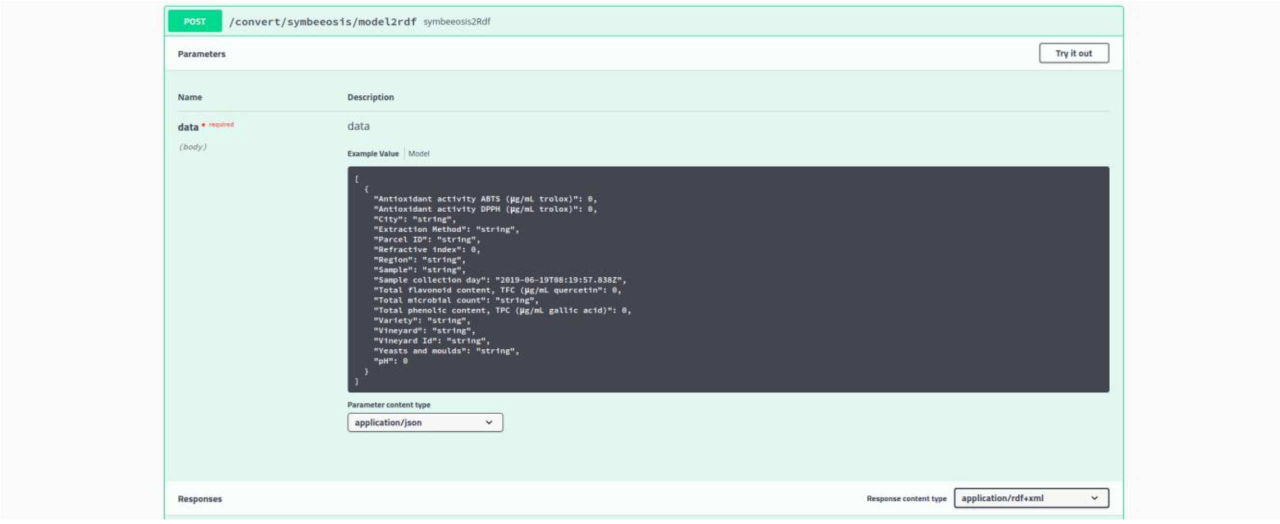


Figure 15: Swagger documentation of the API endpoint for the rdfization of Symbeeosis data

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).

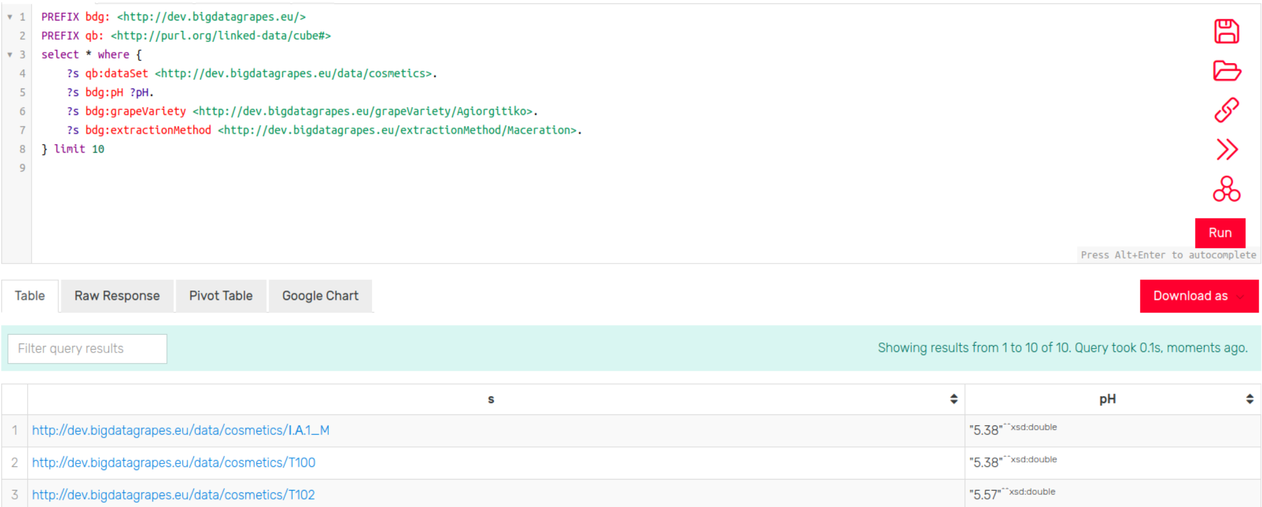


Figure 16: A sample SPARQL query on the Symbeeosis data in GraphDB

In figures 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 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 Geoclidean'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 DDS 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 and until the incorporation of Natural Cosmetics Pilot dataset to the visualization platform, hereafter is demonstrated the decision support system made from the wine quality dataset after linear regression modelling. Relevant link: **https://tinyurl.com/y5z8e8lf**.

In the same way as with Wine Making Pilot, the dataset used for the visualization platform will be based on the datasets processed by partner CNR (3.3.3 Data Correlation Analysis) which resulted to 12 vectors of "observed variables" (BA parameters × maceration & ultrasound measurements) and 192 vectors of "variate variables" (SVIs × aggregation functions × time frames) of 9 components (parcel ids). The visualization can clearly illustrate multidimensional data and the influence of each input variable by allowing us to arrange the variables accordingly. Two web components will be developed for the proposed visualization with Regression Parallel Coordinates and Regression Waterfall. The objective of the waterfall visualization is to help users understand the contribution of each of the features of a regression model around a local data point. In the parallel coordinates' visualization, we can easily see which attributes decrease the target variable value. Both components will be designed to help users understand the influence of the different variables on the prediction of the regression model while the data model will be provided to the web components in a JSON format. The Cosmetic Pilot Dashboard will be a collection of tools that provide an 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 of Wine Making Pilot is shown in Figure 17, while output interfaces showing the distributions of input variables and the predicted wine quality are shown for Regression Waterfall in Figure 18 and for Regression Parallel Coordinates in Figure 19. A demo of these components are available at: https://bigdatagrapes-eu.github.io/deliverable5.3/. The dashboard will simulate a real-world scenario where one would measure vine leafs' extracts chemical properties and input the data in a configuration step. After that, the data could be visualized by selecting any of our visualizations available, which will then help the user make decisions about the cosmetics production process.

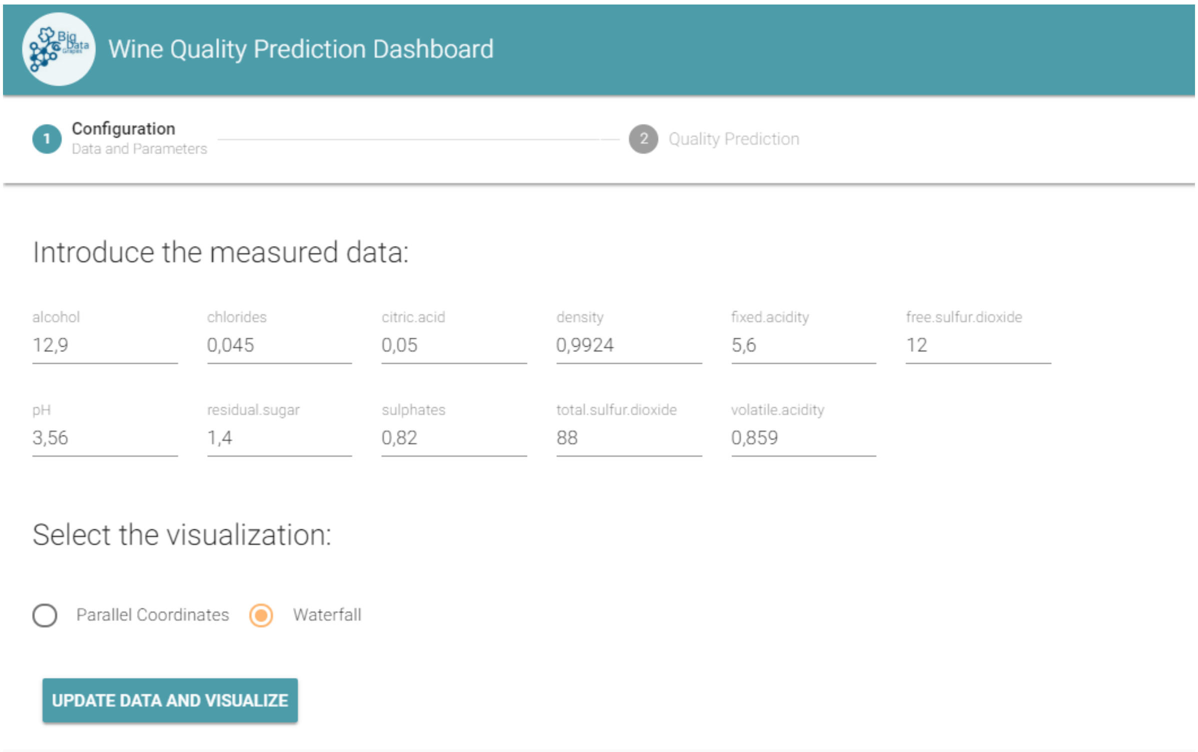


Figure 17: The dashboard configuration interface of wine quality prediction.

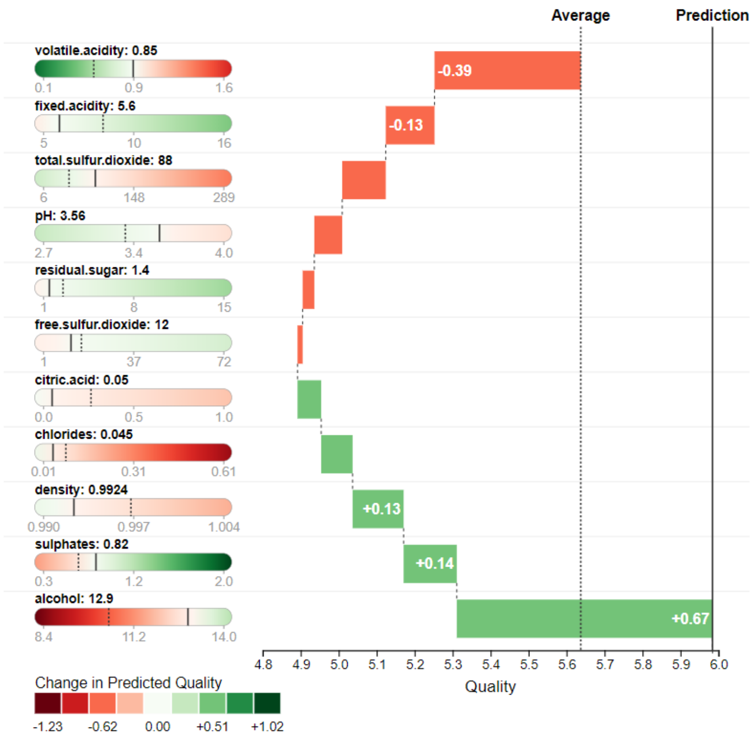


Figure 18: Waterfall plot showing the distributions of input variables and the predicted wine quality.

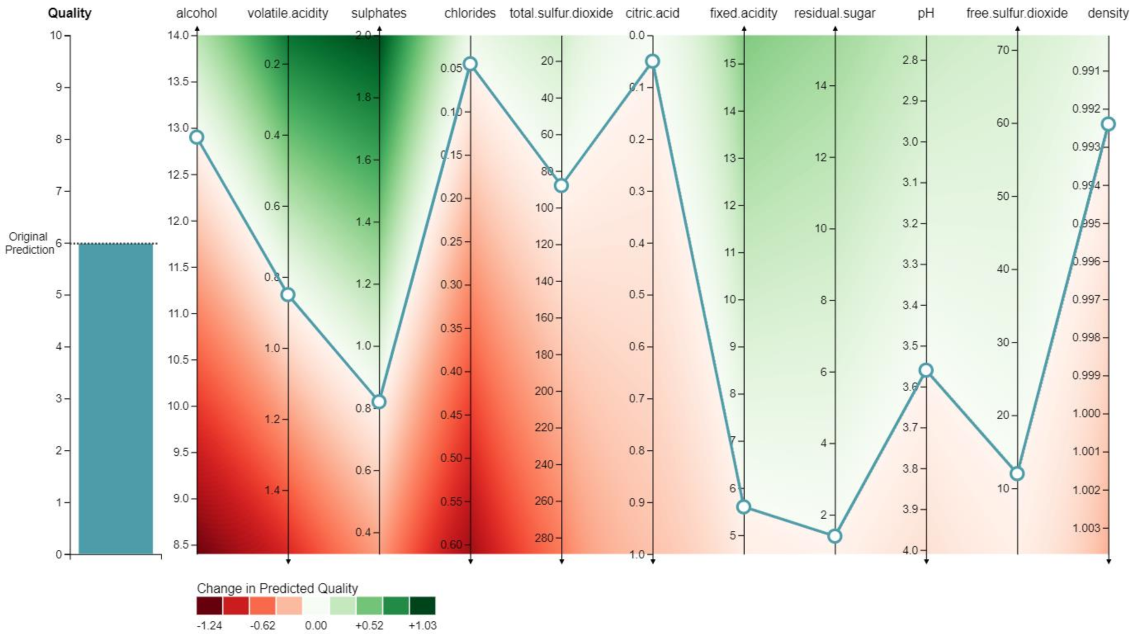


Figure 19: Parallel coordinates showing the distributions of input variables and the predicted wine quality in comparison to the baseline.

4 Conclusions & Next steps

The main goal of Deliverable 8.4 "Integration and Operation with real-life Practices" is 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". The first round of controlled pilot activities is using the first versions of newly developed BDG components. For the present form of the deliverable these components correspond to restricted piloting trials and thus are of limited complexity and extend. The upcoming data input and the fulfilment of the pilot trials will provide all necessary data for the assessment of BDG components and will refine the pilot itself by their subsequent iterations. The collected data from the natural cosmetics pilot will provide the necessary information for the evaluation of the quality of each sample, linked with the special characteristics of the vineyard of origin. In addition, the same data will be 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 and BA parameters of grape leaves extracts will 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.

Next steps of the present deliverable that will be added to its forthcoming versions are the completion of input data for the Natural Cosmetic Pilot with future seasons' sampling and measurements of BA parameters, the incorporation of remaining BA parameters of MTT assay and SIRT1 gene expression, their correlation analyses with the future seasons SVIs, and with the appropriate modelling of the development and assessment of the relevant dashboard in the same way as for the rest of Piloting dashboards of BDG.

5 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.

6 APPENDIX

6.1 CORRELATION ANALYSIS RESULTS FOR MACERATION

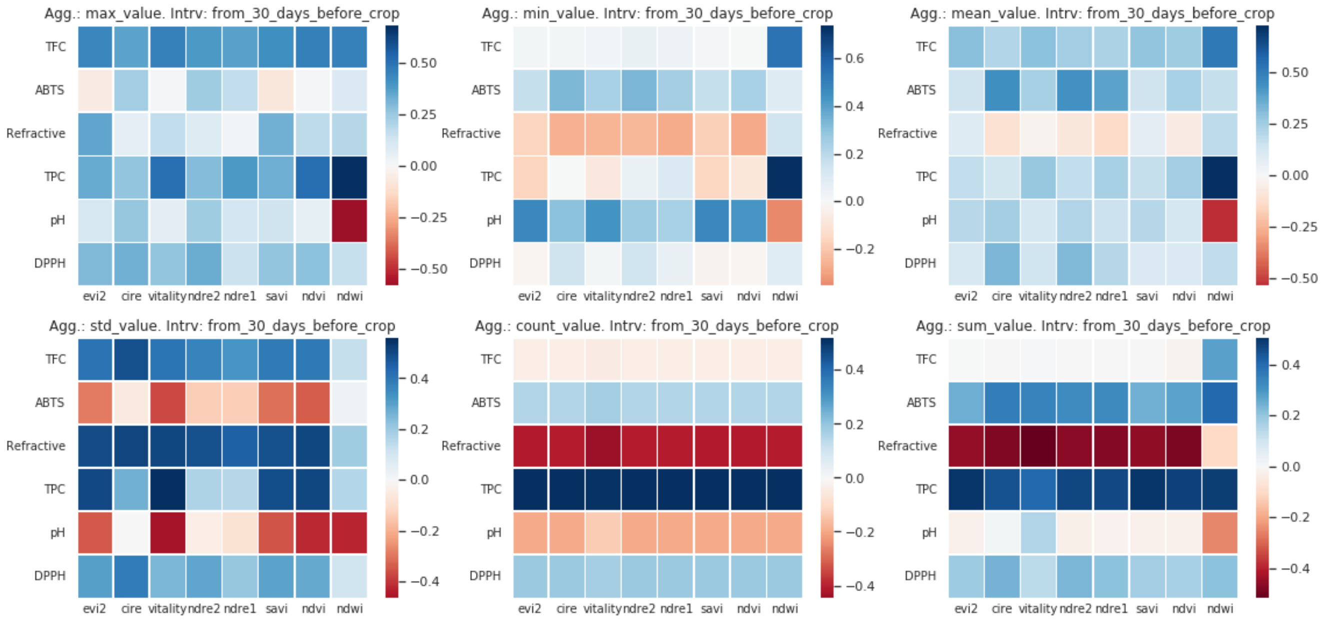


Figure 20: Correlation analysis for time frame "from 30 days before the crop date of the sample"

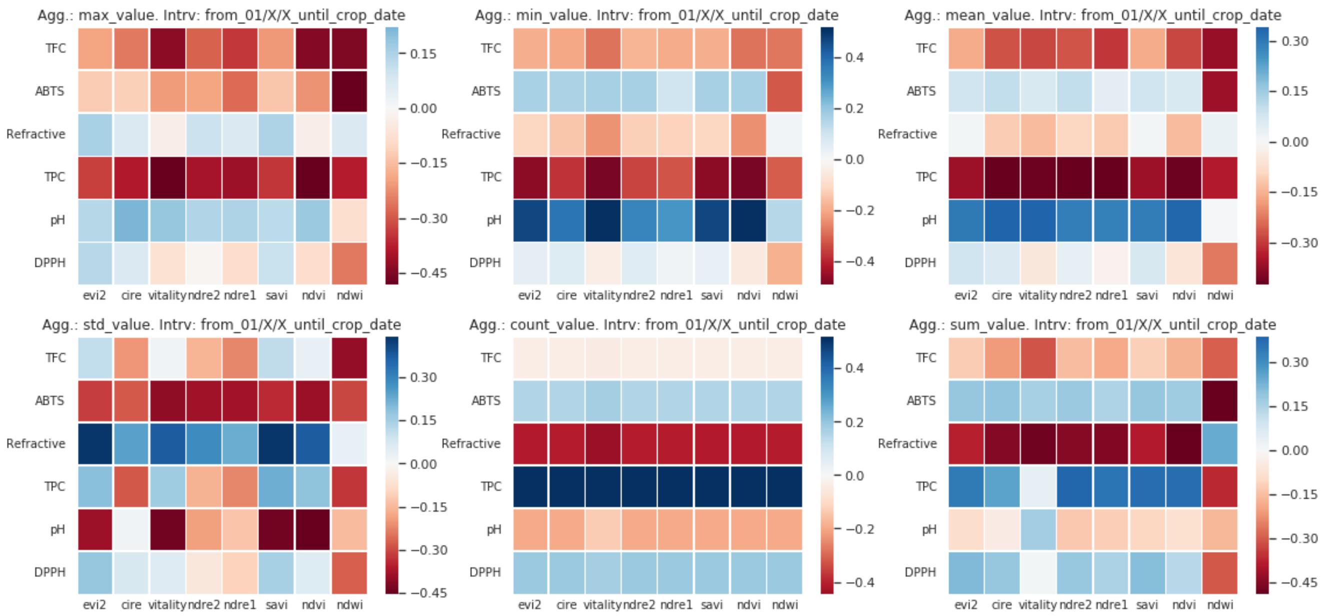


Figure 21: Correlation analysis for time frame "from 1/1/2018 to the sample date in the same year"

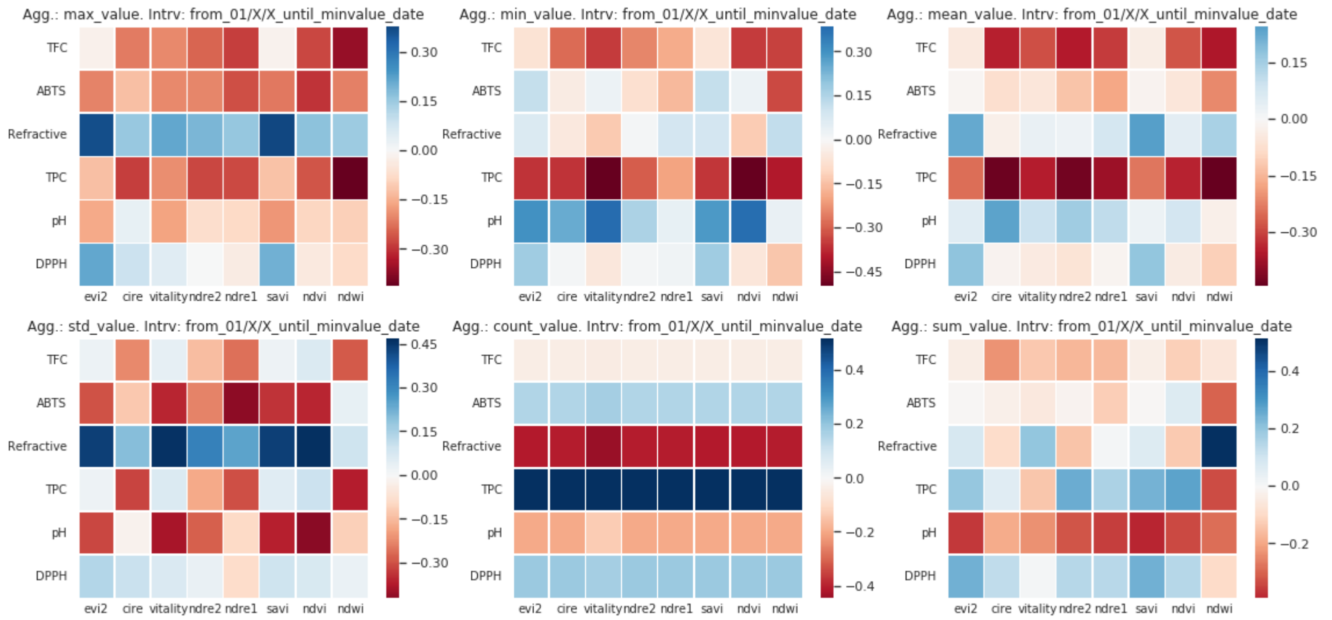


Figure 22: Correlation analysis for time frame "from 1/1/2018 to the min of samples dates in the same year"

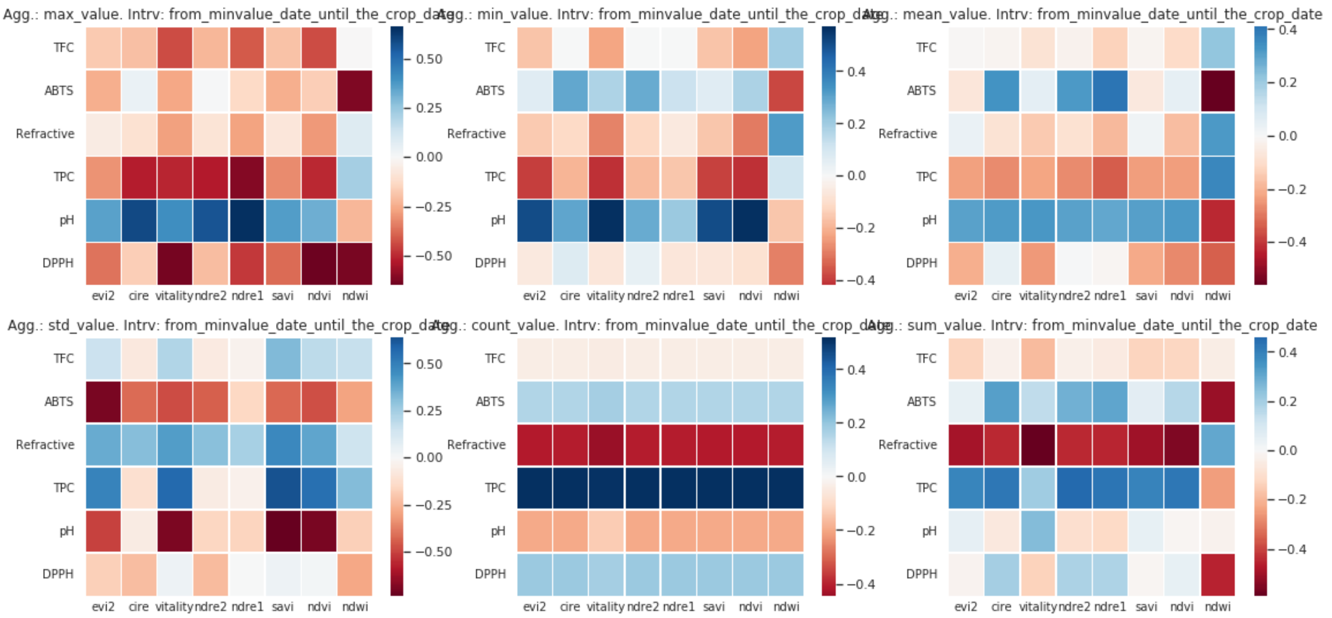


Figure 23: Correlation analysis for time frame "from the min value date until the crop date"

6.2 Correlation analysis results for Ultrasound Assisted Extraction

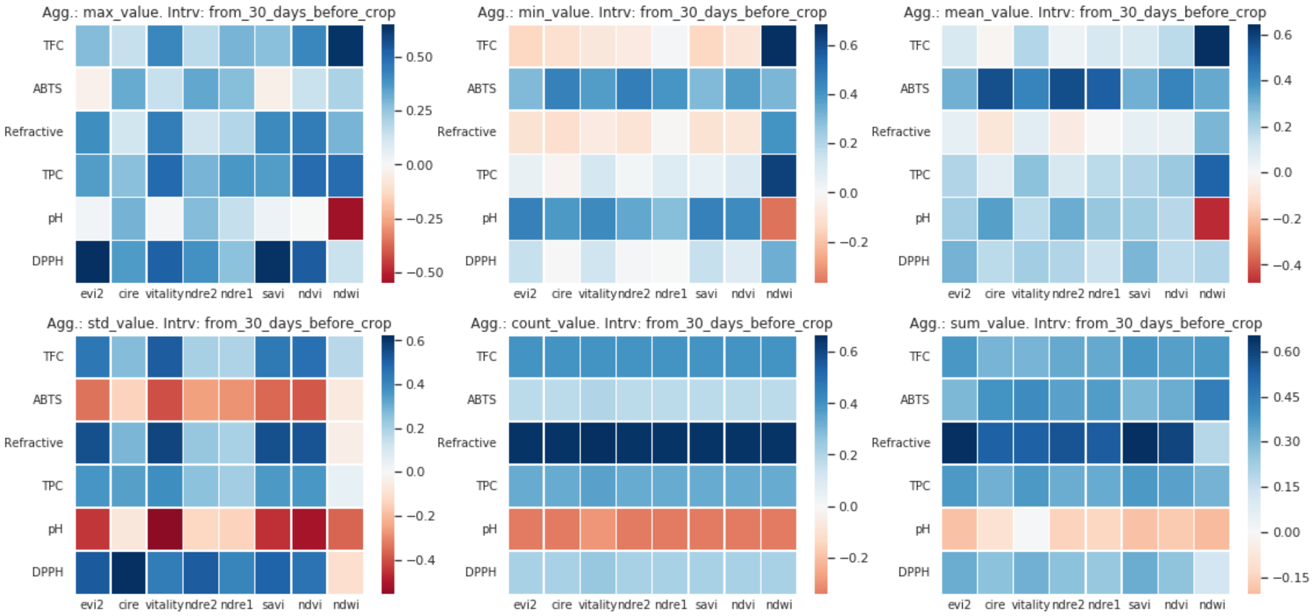


Figure 24: Correlation analysis for time frame "from 30 days before the crop date of the sample"

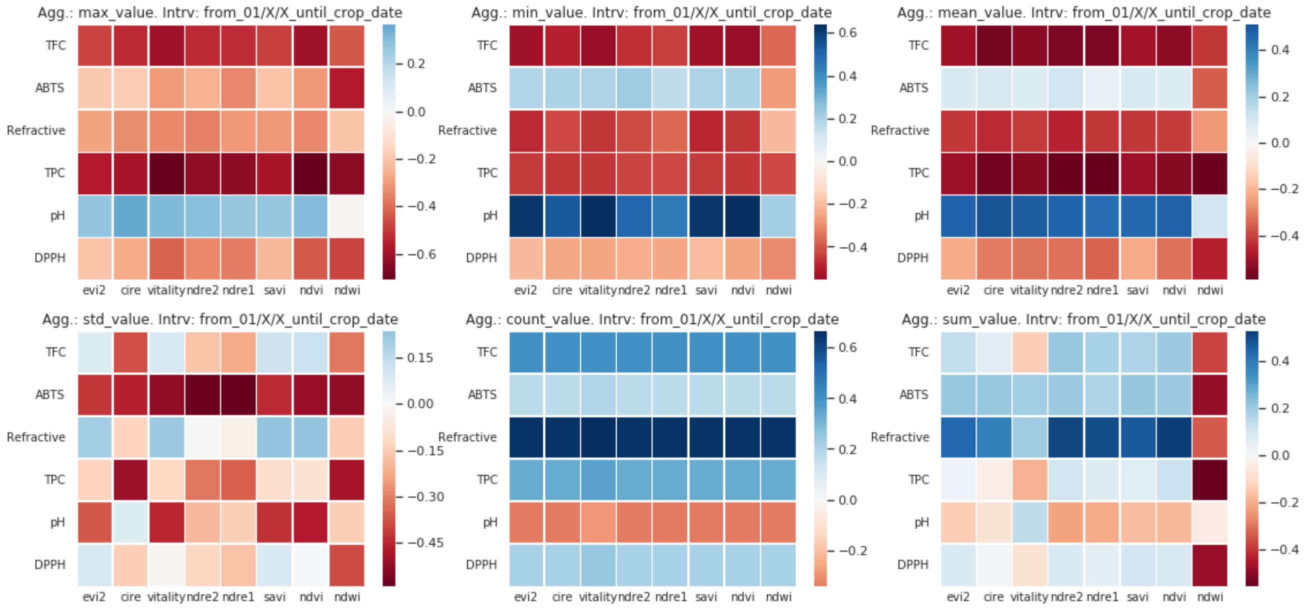


Figure 25: Correlation analysis for time frame "from 1/1/2018 to the sample date in the same year"

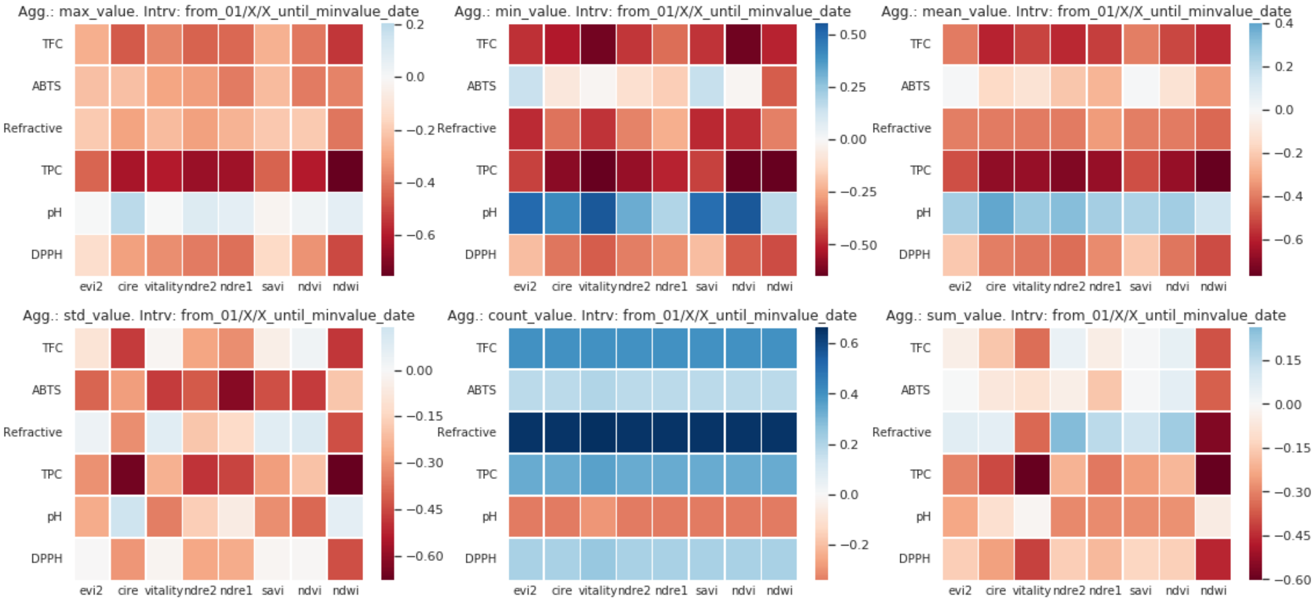


Figure 26: Correlation analysis for time frame "from 1/1/2018 to the min of samples dates in the same year"

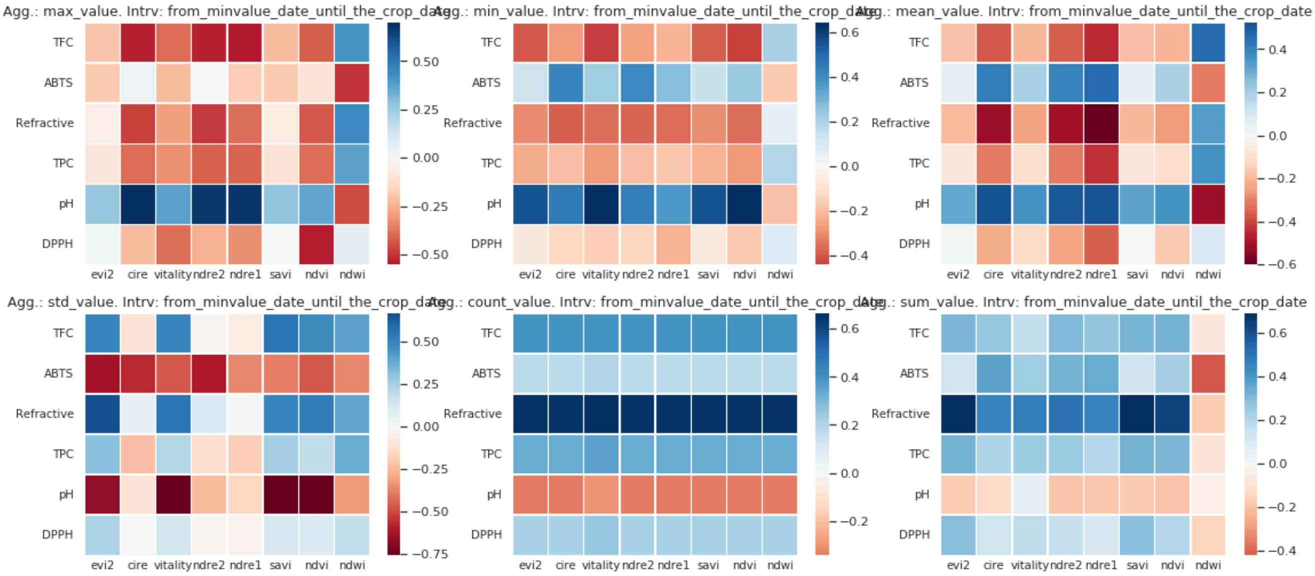


Figure 27: Correlation analysis for time frame "from the min value date until the crop date"