

Article

# A Collaborative Pilot Platform for Data Annotation and Enrichment in Viticulture

Phivos Mylonas <sup>1,\*</sup>, Yorghos Voutos <sup>1</sup> and Anastasia Sofou <sup>1,2</sup>

<sup>1</sup> Department of Informatics, Ionian University, 49132 Corfu, Greece; c16vout@ionio.gr (Y.V.); natasa.sofou@gmail.com (A.S.)

<sup>2</sup> School of Electrical & Computer Engineering, National Technical University of Athens, 15773 Athens, Greece

\* Correspondence: fmylonas@ionio.gr; Tel.: +30-26610-87756

Received: 1 March 2019; Accepted: 18 April 2019; Published: 22 April 2019



**Abstract:** It took some time indeed, but the research evolution and transformations that occurred in the smart agriculture field over the recent years tend to constitute the latter as the main topic of interest in the so-called Internet of Things (IoT) domain. Undoubtedly, our era is characterized by the mass production of huge amounts of data, information and content deriving from many different sources, mostly IoT devices and sensors, but also from environmentalists, agronomists, winemakers, or plain farmers and interested stakeholders themselves. Being an emerging field, only a small part of this rich content has been aggregated so far in digital platforms that serve as cross-domain hubs. The latter offer typically limited usability and accessibility of the actual content itself due to problems dealing with insufficient data and metadata availability, as well as their quality. Over our recent involvement within a precision viticulture environment and in an effort to make the notion of smart agriculture in the winery domain more accessible to and reusable from the general public, we introduce herein the model of an aggregation platform that provides enhanced services and enables human-computer collaboration for agricultural data annotations and enrichment. In principle, the proposed architecture goes beyond existing digital content aggregation platforms by advancing digital data through the combination of artificial intelligence automation and creative user engagement, thus facilitating its accessibility, visibility, and re-use. In particular, by using image and free text analysis methodologies for automatic metadata enrichment, in accordance to the human expertise for enrichment, it offers a cornerstone for future researchers focusing on improving the quality of digital agricultural information analysis and its presentation, thus establishing new ways for its efficient exploitation in a larger scale with benefits both for the agricultural and the consumer domains.

**Keywords:** smart viticulture; precision viticulture; collaborative platform; data aggregation; metadata analysis

## 1. Introduction

Over recent years, the smart agriculture sector has evolved tremendously towards the bridging of two separate worlds: information technologies and primitive agriculture. In this framework, smart agriculture followed the general tendency and became part of the so-called *digital evolution* in the form primarily of massive digitization, as well as annotation, activities, along with actions towards multimedia content generation from all possible sources. The latter resulted into an obvious problem of our epoch, i.e., the one of *information overload*. Initiatives to aggregate the digital content resulted in the solution of the so-called traditional digital platforms. Typically, such a traditional platform operates as a cross-domain hub, making the content accessible to its users, independently on whether

the latter are considered to be content producers (e.g., farmers or smart agriculture stakeholders) or consumers (e.g., the general public). Among the characteristics of the platform is the fact that data and information are made available for search and study. Additional applications and web services may reuse and re-purpose the data, whereas the main strength of such platforms lay in the vast number of content items they contain.

Of course, not all issues are tackled efficiently by such a traditional digital platform. In most cases, they offer limited usability and accessibility due to their insufficient data and metadata quality. As it is the case with many information technology sub-domains, there is a variety of factors that affects the quality of smart agriculture data and metadata:

- lack of structured and rich descriptive metadata;
- complex, heterogeneous, and multi-channel aggregation workflows;
- possible shortcomings in the data providing process (surpassing manual quality control of automatic metadata generation in digital repositories);

These are some of the main causes that result in poor information descriptions. This drawback highly affects the accessibility, visibility and dissemination range of the available digital content. It also limits the potential of added-value services and applications that re-use the available resources in innovative ways, also limiting the overall user experience in the process. On top of that, metadata quality improvement usually faces another important problem, i.e., the problem of scale, since improving the metadata quality coming from different sources often requires a huge amount of time, effort, and resources that aggregators cannot afford.

In this framework, metadata enrichment services through automated metadata processing and feature extraction, possibly along with crowd-sourcing annotation services, available in a centralized way through a dedicated platform can offer a remarkable opportunity for improving the metadata quality of digital smart agriculture content stored in such typical platforms, while at the same time engaging users and raising awareness about smart agriculture assets. In this work, we focus on a viticulture-friendly (Figure 1), cooperative, content sourcing, publication platform, aiming to enrich and improve metadata quality, facilitate the exploration, visibility, and promote digital content reprocessing through various smart farming applications. Our project targets various types of users from the viticulture domain with different levels of expertise and, in addition, offers enriched capabilities based on present resources deploying the platform's existing interfaces. The platform offers metadata management and organization in collections and introduces the concept of spaces, defined to enable data visualization based on various intuitive levels of aggregation. Through linked data principles and automated metadata processing services that aim to amend acquirable web resources, while improving the quality of metadata. Finally, it offers the possibility to interested stakeholders to launch ad-hoc crowd-sourcing campaigns with measurable results, thus mobilizing and engaging users to execute useful tasks for the enrichment and validation of selected smart agriculture metadata (e.g., by adding annotations such as semantic tagging, image tagging, geo-tagging, or by validating existing annotations through up-voting or down-voting within user-friendly and engaging ways).

Finally, the rest of this paper is structured as follows: Section 2 provides an overview of the smart agriculture domain by illuminating some related pioneer works, as well as illustrating the overall precision viticulture domain. Section 3 focuses on the design and architecture of the enhanced platform and its constituent parts. It also provides a detailed description of the data management approach in terms of semantic heterogeneity handling (Section 3.1), content search and aggregation workflow (Section 3.2), including the organization and management of content in terms of collections and spaces, as well as content metadata generation and enrichment methodologies adopted in the platform (Section 3.3). Finally, Section 4 concludes the work and addresses a couple of future works with respect to the platform's evolution over time.

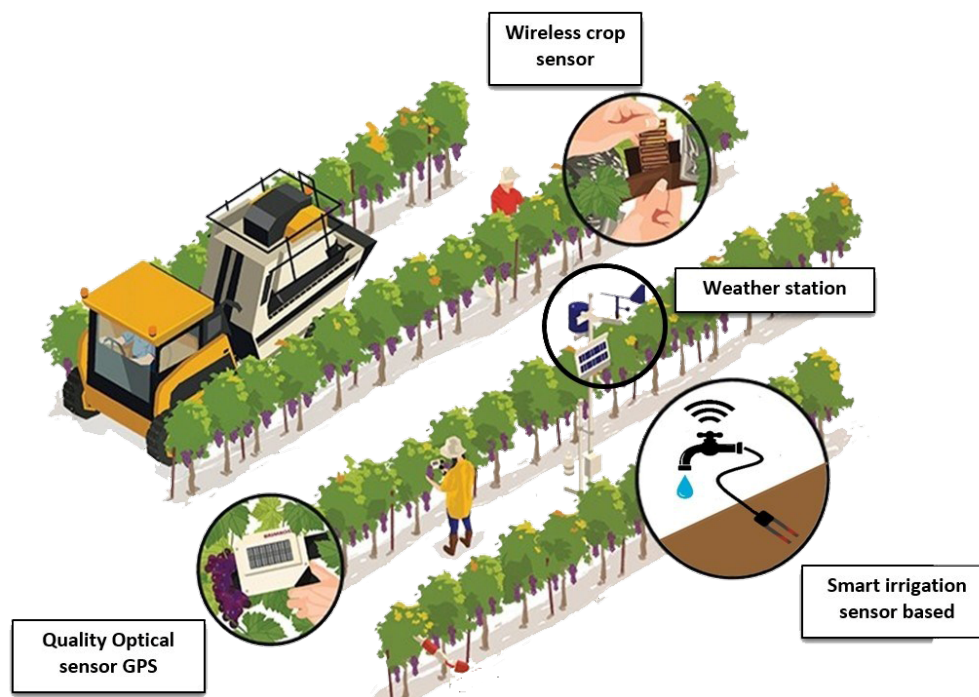


Figure 1. Precision viticulture illustration.

## 2. Smart Agriculture Overview

Over recent years, big data technologies have pushed high-performance computing in the direction of implementing machine learning (ML) for the further development of data science in multi-disciplinary agri-technologies. Thus, several ML models have been deployed within the agricultural production systems, with their main focus being on approaches dealing mostly with crop management through artificial neural networks (ANNs) [1]. Smart agriculture, in the form of machine learning to sensor data, has evolved into an automated decision making under the scope of production improvement [2]. The correlation of agricultural data is based mainly on feature extraction related to the domain of study and defines the use of proper ML tool and consequently rule setting. Consequently, ML applications in agriculture tend to require indispensable tools and methods that meet the specifications of the data model that will be deployed in a specific case study.

### 2.1. Pioneer Works on Smart Agriculture/viticulture

New and emerging technologies and methodologies, such as cloud computing and crowd-sourcing, aid today's farmers towards an overall better performance of their owned farming resources. In this framework, the notion of precision agriculture, in general, and precision viticulture—when focusing solely on vineyards—in particular, is precision farming applied to optimize the overall performance, while minimizing environmental impacts and risk [3]. In viticulture, for instance, this is accomplished by measuring local variation in factors that influence grape yield and quality (i.e., soil, topography, micro-climate, vine health) and applying appropriate viticulture management practices (i.e., trellis design, pruning, fertilizer application, irrigation, timing of harvest) [4,5]. Precision viticulture is based on the premise that high in-field variability for factors that affect vine growth and grape ripening warrants intensive management customized according to local conditions. In this manner, Morais et al. [6] studied the effect on the quality of the wine of various barrel types during aging. They developed a control protocol for monitoring the chemical composition and quality of the produce. The hardware part of the project was implemented through an array of sink nodes based on Raspberry Pi model 2, and the software related to the development of a prediction model was implemented using the *mySENSE* IoT platform. Morais et al. focused on oxygen consumption during the wine aging process, through the discovery of substantial differences between wood barrel

types. Remote sensing is a very useful tool for analyzing the impact of agro-environmental factors in the quality of vine yields. They employed satellite-derived prescription maps (PM) from aerial imagery through the analysis of intra-vineyard variability for vine color (vigor). Normalized difference vegetation index (NDVI) maps offer surveying capabilities for the optimization and further control in wine production along with field data acquisition. They presented a monitoring system that improves the distinctive capability for agronomic applications.

An application of a real-time monitoring platform in viticulture is presented in the work of Canete et al. [7]. Their research emphasized the control of structural health and ullage of wine casks on the wine aging process through detection system for possible crevices on the cask's surface. They managed to gather information from inside the aging containers with the introduction of the prototype "Smart Cork", which provides the capability to real-time monitor and resolve any issues. More specifically the sensing device was based on IoT technologies and supported by a cloud platform and a control interface for on-site warning in order to evaluate the suitability of the aging process in real time.

Insular environmental monitoring is a research subject was studied by Naumowicz et al. [8]. They implemented an IoT platform that based on deployed wireless sensor network for three years design on the scope of monitoring seabirds on Skomer Island (UK National Reserve, West Wales). Although their application was not specific to agriculture their research focused on the particular characteristics of insular areas. Their sensory arrangement consists of a real-world deployable sensor network based on battery powered sensor nodes to investigate the habitat of the Manx shearwater (*Puffinus puffinus*) seabird.

Muangprathub et al. [9] proposed an optimization system for water supply in agricultural crops. They developed a wireless sensor network infrastructure under the scope of controlling and managing data and operations remotely. Furthermore, the deployed nodes provided concrete and use full data for irrigation while increasing agricultural productivity.

Finally, to conclude this brief overview of related research literature, we should stress the fact that the herein discussed approach differs from all the aforementioned depicted in Table 1, in the sense that it combines on-site soil and atmospheric monitoring based on integrated equipment of weather instruments. Furthermore, it is worth mentioning that its gathered data are employed within a semantic structure for smart viticulture through integrated data handling on the scope of current state-of-the-art artificial intelligence [10].



**Table 1.** Overview of research works in smart agriculture.

Work	Task(s)	Method(s)	Pros	Cons	Dataset(s)
[4]	Study of spatial variability in winegrape yield	Assessment of the yield growth probability through <i>k</i> -means	Introduction of a system of zonal vineyard management	They didn't address the physiology of the grape	Field data from two areas in South Australia
[5]	Demonstrate temporal stability within-vineyard variation and winegrape quality	Geo-statistical assessment ( <i>k</i> -means) clustering and	Justify the use of zonation on yield monitoring	Requirements for large scale sampling and analysis	Field data from two areas in South Australia
[6]	Investigation of behaviour models on wine aging	Deployment of a distributed monitoring system (hardware and software	Detect differences between type of wood barrels and the different storage conditions on wine aging quality	Their model produced variations due to oxidation in aging process	Measurements of environmental parameters from two wineries along with barrel dimensions
[11]	Investigate satellite data potential to evaluate changes in vegetation characteristics on occasional acquisitions	Investigation of spatial resolution role in radiometric features of data	Concrete correlations between grapevine physiological, reproductive and qualitative indices	Weak quantitative interpretation of mapped vigour	Aerial and mid-resolution satellite multispectral images from two growing seasons (2013 and 2014)
[7]	Real time monitoring system with IoT prototyping to maintain the balance between ullage and wine	Prototype embedded system to detect possible damages to barrels and to track the level of wine inside the cask	Provide a plan for tracking and assessing the suitability of the delicate wine elaboration process	Low distinctive capability	Recordings of physical and natural characteristics from IoT devices
[10]	Preliminary ontological scheme for environmental monitoring	Spatio-temporal ontological development	Relation between environmental factors and spatio-temporal distribution	Not deployed with field data	-
[8]	Automated Insular environmental monitoring	Deployment of a WSN for wildlife monitoring	Remote deployment	Not widely applicable	-
[9]	Optimized water supply system for crops	IoT network of sensors and software infrastructure	Automatic deployment of a monitoring system	The system is not scalable	Climatic variables gathered by a WSN

## 2.2. Precision Viticulture

Precision viticulture is a specialized field of smart agriculture aiming at maximizing the oenological potential of winegrape yield. High-quality production areas demand the adoption of spatially aware management systems that assist the optimization of wine quality and yield. Additionally, technological innovations allow specialized vineyard management that promotes high quality and reduction of environmental impact. Furthermore, further advancements on the field of applied geography and information communication technologies (ICT) enhance the probability of scientific leaps in precision viticulture. Novel optimized solutions provide a plethora of tools for advanced monitoring and control in vine growth. In conclusion, smart viticulture aims at the exploitation of methods that identify high-resolution spatial variability on vine fields that provide a recommendation basis for improved quality, yield, and sustainable management.

Agricultural impact on the environment is reduced through sustainable and efficient practices derived from smart farming that permits the management of various agricultural factors along with environmental preservation. It is based on a management approach that estimates the localized needs of the vineyards, based on heterogeneous structural factors of the yield. The managerial requirements evolve around the variability on the physiological response, strongly related to grape quality, which in terms of agronomy it will satisfy the efficiency and quality of crop production. Vineyard spatial variability directs the development of specialized tools that assist monitoring and control vine growth through. These technological solutions are based on remote and proximal sensing sensors that monitor vineyard status, namely water and nutrient availability, plant health and pathogen attacks, or soil conditions. More specifically precision viticulture exploits the widest range of available sensory observations to interpret high-resolution spatial variability within the vineyard, and assist the efficient yield management efficiency in relation to quality, production, and sustainability.

The rise of machine learning and artificial intelligence has offered novel tools for vineyard management, early pest detection, growth stage and the potential quality of vine grapes estimation. Unmanned aerial vehicles equipped with red-green-blue (RGB) and near infrared (NIR) cameras are commonly used in precision viticulture to acquire image data that are used to extract color information in terms of various indices (e.g., excess green index, excess red index, color index of vegetation extraction, etc.) that are often used to segment the green vegetation regions in remote sensing images. Computer vision techniques can be applied on remote sensing images to extract various color, shape and size related features and vegetative indices, to estimate canopy features in the field, to estimate cluster morphology which not only impact the cluster architecture and compactness (leading to looser or tighter clusters), but are also considered as indicators of grape and wine quality. Machine learning techniques combined with machine vision technology has been widely used and studied in agriculture to identify and detect plants (crops and weeds). In viticulture, machine learning approaches are used typically for grape and foliage detection, classification of early-stage grape yield estimation, whereas neural networks are used to classify wine samples into different varieties, locations, and years of production. In the following two subsections, we shall investigate further research with respect to two important traits of precision viticulture, namely soil management and yield/quality estimation.

## 2.3. Utilization of Artificial Intelligence

One of the most important and studied sub-fields of the above is soil management because it heavily affects the production and outcome of the vineyard. One of the first integrated works on the subject is from Mehdizadeh et al. [12]; the authors investigated the performance of equations and soft computing approaches, implementing 16 empirical equations from temperature-based, mass transfer-based, radiation-based, and meteorological parameters-based categories. More specifically they calibrated different types of used empirical equations and evaluated their performances in estimating  $ET_0$  (reference evapotranspiration ( $\text{mm day}^{-1}$ )). Thereafter, they addressed the capabilities of soft computing approaches, including multivariate adaptive regression splines (MARS), support vector machines (SVM) and gene expression programming (GEP) and in result they conducted a

comparative analysis between empirical equations and the soft computing approaches. Their findings concluded that precise estimation of  $ET_o$  is required in many fields related to water and aquatic domain, while testing the effectiveness of the empirical equations.

Artificial intelligence provides a plethora of methodologies on modeling continuous and variable environmental factors. Feng et al. [13] developed an estimation model for daily  $ET_o$  based on temperature data that introduced two models along with extreme learning machine (ELM) and generalized regression neural network (GRNN). They gathered data from six meteorological stations from the Sichuan basin (SW China) and thereafter compared them using the ELM and GRNN models along with the temperature-based Hargreaves (HG) model. Their data sequence corresponded to a calibrated version considering FAO-56 Penman–Monteith  $ET_o$  as the benchmark, that produced two data management scenarios: (1) the models were trained and tested using the local data of each station and (2) the models were trained using the pooled data from all the stations and tested in each station, which were compared against to the well-known empirical Hargreaves models.

The development of Artificial Intelligence models for daily  $ET_o$  estimation solely with temperature data was addressed by Feng et al. [13]. More specifically they introduced two models along with ELM and GRNN in six meteorological stations from the Sichuan basin, in southwest China, and compared the proposed ELM and GRNN with the corresponding temperature-based Hargreaves (HG) model and its calibrated version considering FAO-56 Penman–Monteith  $ET_o$  as the benchmark. In relation to the implementation of the above they considered that the models' training was based on local weather station data and derived data from the all of the stations in total, compared altogether against the empirical Hargreaves models. Furthermore, the authors' research showed the great capabilities for  $ET_o$  estimation of ELM and GRNN models, mainly because they performed much better than the original models. In conclusion, Feng et al. [13] stated that ELM and GRNN provided a slight better performance than the local ELM and GRNN, they render a better fit for  $ET_o$  estimation methodology based on incomplete local meteorological data.

Precise estimation of evapotranspiration is crucial for accurate crop-water estimation by implementing machine learning (ML) techniques, such as artificial neural networks (ANNs). Specifically, the estimation of weekly reference crop evapotranspiration ( $ET_o$ ) was assessed by Patil and Deka [14] by developing an improved extreme learning machine (ELM) algorithm. The authors attempted to model the process of evapotranspiration in arid regions of India under limited data scenario through the evaluation of the capabilities of ELM to model the process of evapotranspiration and input selection. They concluded that the proposed ELM model performed better than the Hargreaves and ANN model, as the best to estimate weekly evapotranspiration at Jodhpur and Pali weather station, based less human intervention to select the optimum model parameters. In addition to that the effectiveness of using such extrinsic inputs is proved by the evaluation of suitability of using ( $ET_o$ ) values from various weather stations modeling the process of evapotranspiration.

Prediction of daily dew point temperature was studied by Mohammadi et al. [15] by proposing an extreme learning machine (ELM)-based model. The effectiveness of ELM was validated against SVM and ANN based on the case studies of daily averaged measured databases for two stations in Iran with different climate conditions by focusing on the input elements of the physical factors influencing the formation of dew, five widely available parameters, average air temperature, relative humidity, atmospheric pressure, vapor pressure, and horizontal global solar radiation. More specifically they deployed statistical indicators to provide a statistical comparison between the predicted and measured data for three modeling techniques to address that the ELM model is superior over SVM and ANN for the predictions of daily dew point temperature. In conclusion, the ELM was found better to predict the nonlinear variations of daily dew point temperature in its different ranges by estimating the probability of occurrence of predicted values compared to measured data revealed that for both stations the probability distribution of the predicted values.

In principle, it is a rather challenging task to perform accurate yield prediction, estimation and quality control [16], because such factors are affected by environmental variables. Still, many

approaches for enhanced vineyard management techniques are introduced, which involve automatic yield estimation [17–19], grape quality evaluation [20], and grapevine variety identification [21]. In a relevant approach, Ramos et al. [22] proposed a method of calculating the number of fruits on a coffee branch in the field. Their system categorized fruits as either harvestable or non-harvestable through identifying their number, weight, and maturation percentage non destructively. They were based on data obtained from mobile devices and image processing algorithms along with the development of a dedicated detection method of masking occluded from un-occluded fruits from field images. Ramos et al. produced estimation models based on assessing determination coefficients higher than 0.93 and exploiting direct measurements capable of higher coefficient rate, comparable to similar models.

Taking this a step further, the scope of a machine vision system for automated harvesting was studied by Amatya et al. [23] through branch detection used morphological features including orientation, length, and thickness of partially visible segments. The authors aimed at detecting cherry tree branches during full foliage season that support automated cherry harvesting operation using mechanical shaking of limbs. Furthermore, they deployed RGB images of cherry tree canopies using a Bayesian classifier for the identification of the segmentation branch pixels, in which they managed a classification accuracy of 89.6%. They also implemented noise filtering and grouping techniques of the same branch in a specified neighborhood, along with curve fitting method was then used to fit an equation through detected branch segments, providing overall accuracy in detecting individual branches was by 89.2%. In conclusion, their study showed promising results in detecting branches of sweet cherry trees during harvest season in the presence of full foliage and fruit on their study sample.

One of the worth-noting works in the field is the one by Radhika and M. Shashi [24], where they exploit support vector machines (SVMs) for the task of predicting the maximum daily temperature. Moreover, as a baseline, they use a multi-layer perceptron (MLP) trained through back-propagation. To accomplish the aforementioned prediction task, they rely on weather data provided by the University of Cambridge for the time period of 2003–2007. Through their experimental findings, the authors showcase that the SVM-based approach, is able to achieve significant results, thus showcasing its suitability for this particular prediction problem.

Similarly, Gill et al. [25] develop an SVM regression model, in order to predict the moisture of soil. In particular, the authors utilize both meteorological, and soil moisture data, from 11 weather stations from the Oklahoma region in the United States, for training and testing their prediction model. Finally, by comparing their proposed model against an artificial neural network (ANN), they showcase its efficiency. Huuskonen and Oksanen [26] presented a methodology in relation to remote soil sampling while presenting the results on the augmented reality platform. Their work intended to reproduce a top soil map with management zones for precision farming based on the gaming experience. They deployed a UAV on pre-destined sampling sites in order to create soil maps. The maps feed a platform with content that enhances user experience in order to assist agricultural research.

Lastly, another novel approach in agriculture was implemented by Fuentes et al. [27] in which they employed machine learning models for the categorization of cultivars. They used morpho-colorimetry and NIR spectroscopy in grapevine leaves in order to assess the main differences between cultivars, and their according levels of water stress. The researcher employed automated image analysis under the scope of morphological and color feature extraction from scanned vine leaves along with an AI-based model (ANN) for rapid, accurate, and inexpensive ampelography/cultivar classification. Fuentes et al. method proved to be highly accurate in recognizing different cultivar classes in a large sample of vine leaves.

Table 2 provides a brief comparative overview of the above fundamental research efforts. As it was the case with previous Table 1, Table 2 includes six columns: The first contains each work's bibliographic reference number; the second describes the main task the particular work attempts to tackle; the third focuses on the depicted methodology the authors propose or utilize in order to solve the particular research task at hand; the fourth column presents a representative set of positive characteristics (if applicable); the fifth column presents a representative set of negative characteristics (if applicable); finally, the sixth column provides information on the utilized dataset (if any).

**Table 2.** Overview of research works focused in the utilization of AI.

Work	Task(s)	Method(s)	Pros	Cons	Dataset(s)
[1]	Review of applied Machine Learning in agriculture	Presentation of a number of relevant papers with emphasis on key and unique features of popular Machine Learning models	Categorization of Machine Learning methodologies	Not related to decision making processes	forty (40) related articles
[12]	Equation performance of monthly mean reference evapotranspiration	Comparative analysis between empirical equations and soft computing approaches	-	-	Meteorological dataset from southwest of Asia
[13]	Modeling reference ET <sub>0</sub> through ML and techniques	Implementation of temperature based ELM and GRNN models for the estimation of daily ET <sub>0</sub>	Emphasize on the capabilities of ELM and GRNN models	Experimental data were not used	A dataset of daily meteorological variables from 6 stations of the National Climatic Centre of the China Meteorological Administration
[14]	Evaluation of capabilities of Extreme ML to model the process of evapotranspiration	Implementation of ANN1, LS-SVM1, ELM1 and Hargreaves equations in climatic data	Comparability between models	-	Weekly climatic data from two weather stations in India
[15]	Daily Dew point prediction methodology	Extreme Machine Learning	Reliable prediction techniques deployed in varied climatic conditions	Decrease of model's performance between training and testing	2555 days of measurements from two meteorological stations
[17]	Automatic yield estimation	Leaf spectroscopy and partial least squares	The model has many classification efficiencies	Data obtained from a limited number of samples	Hyperspectral images of vine leaves
[18]	Complete characterization of grape seed quality	Detection of morphological differences through image analysis	They produced an efficient evaluation tool	They employed a small distinctive sample	Images from one hundred berries from a specific cultivation
[21]	Grapevine variety identification	Automatic leaf spectroscopy classification of grapevine varieties	Their work produced robust results	Limited samples for required laboratory testing	Images from grapevines of Vitis vinifera L.
[22]	Automatic identification of harvestable and not harvestable fruits on a coffee branch	Object based classification based on a Machine Vision System & an Image Processing Algorithm	Appropriate tool for assessing harvest potential	Labour intensive procedure	Acquired and adjusted grape images from 1018 coffee branches
[23]	Branch shaking localization for automated cherry harvesting	Machine vision system for automatic cherry recognition and counting	High success rate	Time consuming procedure	Color RGB images of cherry trees from Washington, USA
[27]	Accurate& automated ampelographic measurements	Two types of ANN models	High accuracy rate	Requires complex processes	Morpho-colorimetric data obtained from vine leaves
[26]	Remote soil sampling	Drone imaging and segmentation	Novel integrated agricultural monitoring approach	Not tested in large scale	Soil data from the study area



### 3. Agricultural Data Aggregation—An Interconnected System

At this point, it should have been made obvious by the previous analysis that through the use of innovative new technologies, farmers may be supported in activities related to the management of their crops so that they can achieve higher yields and better quality while at the same time they apply the exact amounts of inputs (e.g., irrigation water, fertilizers, and pesticides) needed by their crops. This not only improves the financial benefit of farmers (thanks to the minimized production costs), but also minimizes the impact of agriculture on the environment. Within this challenging framework, the main motivation behind the creation of a viticulture-friendly platform is to utilize different data repositories in unison and promote the digital content by enhancing its accessibility and discoverability, and provide artificial intelligence based services for further analyzing available data (e.g., images of various modalities, multi-source measurements, sensor data, historical data). Thus the discussed platform can be viewed as a complete ecosystem that offers enriched services, aggregates content from multiple sources, resolves interoperability issues by performing automatic mappings in the back-end, improves metadata quality through linked data principles and automated free text and image analysis services where applicable, and mobilizes and engages its users (i.e., farmers, winemakers, stakeholders, the general public) to execute useful tasks for the enrichment and validation of selected metadata, as well as the provision and extraction of new information. In its core, it is a platform especially designed with a focus on human-computer collaboration with application in the area of viticulture. The latter offers a set of services such as content integration, management, retrieval, and curation, automatic metadata enrichment, information retrieval and extraction, and inserts the human being in the loop in order to accommodate human expertise and intelligence.

The main characteristics, as well as benefits from the introduction and utilization of such a platform may be summarized in the following, whereas a diagrammatic overview of the proposed platform is illustrated in Figure 2:

- Viticulture data are aggregated from multiple repositories and published;
- Viticulture data are translated into usable information consumable for decision-making purposes;
- Heterogeneous data are translated into the same form using a common data model;
- The data model is able to actually evolve, so as to include additional requirements, data sources and other models;
- Data can be made more accessible to a wide range of users, such as data scientists, business analysts, etc., by granting a unified access to knowledge from multiple sources, thus promoting the viti-cultural content;
- Data can be queried and queries may be asked in multiple ways that haven't been anticipated while modeling the data;
- Back-end processing can be initiated to extract useful information from translated data;
- Viticulture-related information is retrieved, stored, extracted and sustained for future use;

A more descriptive architectural flowchart of the proposed implementation is illustrated in Figure 3. The information flow starts on the left, where related content providers (e.g., user archives or even libraries of related content) and corresponding aggregators provide their content metadata descriptions to the developed platform. In principle, this first inaugural part of the process tends to be a rather difficult task, mainly due to the heterogeneity of the various available content metadata formats. Using the proposed modular system, the metadata transformation module maps and transforms the uploaded data to elements through an automatic transformation process. The outcome of this module remains a transformed version of the content metadata. Continuing, translated metadata are represented as KB expressions by exploiting the semantic enrichment module and are then stored in the semantic repository module on the upper right corner of the flowchart. It is worth noting at this point that all metadata elements are represented within the semantic repository as entities of the terminological knowledge. This knowledge has the form of an ontology, which is related to both various popular ontology metadata standards, such as Dublin Core or LIDO, as well as to more specific

terminological axioms providing details about crop varieties, categories of agricultural crops, physical properties, numerous interrelations, and other important agricultural features. Content providers and aggregators develop this type of knowledge in the form of so-called thematic terminologies. The exploitation of such knowledge types is then expected to happen within the semantic query answering module, responsible for the semantic enrichment of the available content metadata, as well as for the implementation of meaningful reasoning tasks so as to extract efficient related results.

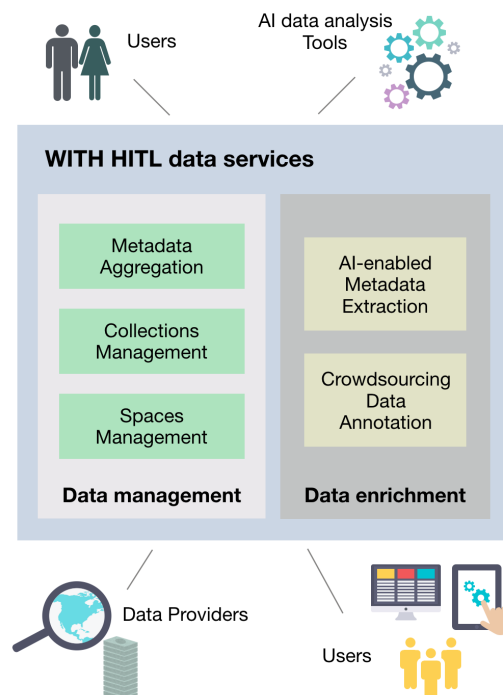


Figure 2. Platform overview.

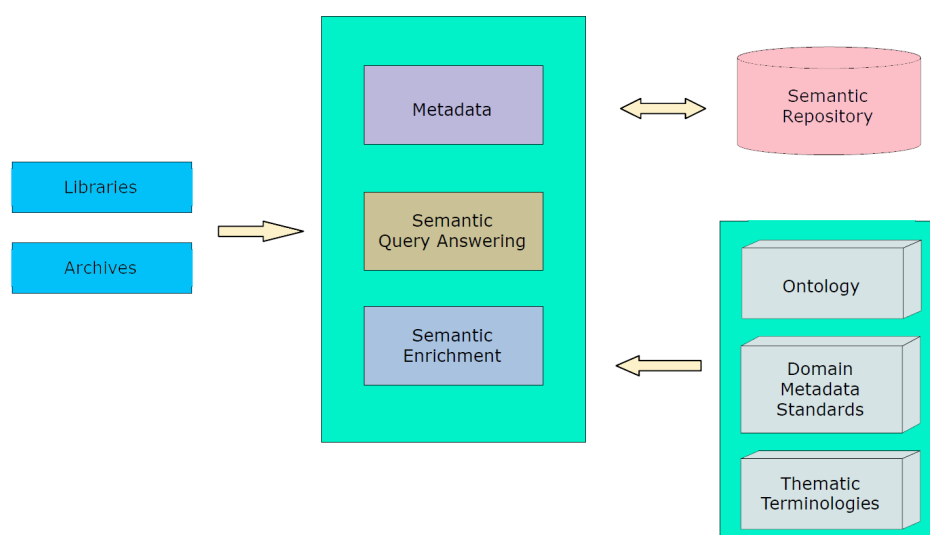


Figure 3. Platform architecture flowchart.

Our system as a whole is available to test and use online (<http://withculture.eu/#/custom/evinos>) and currently addresses numerous kinds of users from various fields of the viticulture domain that are interested in exploiting, analyzing and creatively reusing content through smart tools and applications. It serves as a multipurpose platform with varying level of functionality abstraction depending on the user's level of expertise. More specifically:

- For *end users: wine lovers, viticulture, enthusiasts, winemakers and producers etc.*, the platform serves as an outlet for the breakthrough of content and interfaces bonded to propagated repositories. Users can potentially create customizable multi-source content (e.g., searched through multiple viticulture repositories), on the scope of further enrichment of dissemination for viticulture. Customized collections will create a virtual aggregation of data related to viticulture allowing a straightforward method of presenting popular narratives in a web-based environment. Furthermore, the platform offers collaborative content creation and modification within registered groups of users.
- For *specialized professionals (from viticulture, agriculture domain, etc.)*, the platform offers the tools to synthesize and construct eclectic collections, for the enrichment, optimization, and development of specialized semantic knowledge.
- For *viticulture content holders, i.e., winemakers and producers*, it offers an easy to use content & metadata repository and management system, that can ensure interoperability with standards, best practices, and guidelines. Winemakers are able to upload their historic data from a specific vineyard, like for instance soil data, management data, meteorological information, and actual yields per season. Data can be stored in the platform's repository and then plugged into the system's machine learning models to predict the yield from the coming season, even from the early stages of growth.
- For *organizations*, the platform offers the notion of *spaces*, i.e., a novel way to organize and promote content and improve its metadata, as well as engage with users through campaigns.
- For *aggregators*, it provides the MINT (Metadata INTeroperability, <http://mint.image.ece.ntua.gr/redmine/>) [28,29], open source, web-based platform. It offers full implementation of workflows on the scope of ingestion, formal mapping, transformation, and aggregation of metadata records.
- For *developers*, it provides a back-end toolset for digital content re-usability, employing a fully functional API ([api.withculture.eu/assets/developers-lite.html](http://api.withculture.eu/assets/developers-lite.html)) that disseminates the total of the available data and services to anyone willing to utilize for further exploitation.

Above depiction of the main possible utilization of the platform by the aforementioned kinds of viticulture users' expertise depicts the expansion capabilities and potential of the proposed system, being a modular platform with varying levels of granularity. The latter may be defined as the lesser or greater extent in the description of the expected behavior of the platform, according to its functional specifications and the particular goal associated with its requirements and allows for its great impact with respect to making agriculture smart using traditional, computer-aided automation, and artificial intelligence (AI) based and Internet of things (IoT) technologies.

The basic underlying system architecture, as already illustrated in Figures 2 and 3, mainly depicts the user involvement in the data services of the platform; the latter is designed based on an innovative human-in-the-loop approach, to target the advancement of data quality and organization, to offer data services that users can employ in order to better manage, retrieve, enrich, and promote their data. Specifically, with the aid of *metadata aggregation*, *collection management*, and *spaces management* services, viticulture users can upload, collect, and organize content from several sources and also create interesting data views and presentations and store them in the platform's underlying database. Moreover, using the *AI-enabled metadata extraction* services that integrate automatic or semi-automatic annotation tools (like automatic text analysis, image annotation, etc.), users may select interesting ways to analyze the relevant content, extract knowledge and develop and store new content descriptions. Finally, by initiating and managing crowd-sourcing campaigns, users may involve other users in

recognition tasks that are difficult for machines to perform (e.g., wine variety recognition from a grape roe picture), thus further improving the available data content descriptions.

Based on the above system architecture, the platform's user involvement may be summarized in four interactive processes, as depicted in Figure 4. The *content search* and the *content management* processes that enable users to collect content and organize content, as well as the *Metadata enrichment* and the *crowd-sourcing* processes that enable users to advance content descriptions, using AI content analysis tools or human annotations, respectively. In the next subsections, we shall focus on these main processes, providing further technical details and use-case examples where applicable.



**Figure 4.** User involvement in the proposed platform.

### 3.1. Handling of Metadata Semantic Heterogeneity

As the world becomes full of sensor networks producing information, smart viticulture enters the somehow “foreign” fields of data processing, storage and retrieval at a fast pace. Through various sources of information derived from a wide data spectrum (e.g., geo-information, soil measurements, weather forecasts, chemical analyses, to mention just a few data streams), grapevine-powered stakeholders are in a position to grow an extremely rich system of integrated Internet of things data to power more informed decisions. The main challenge lies in the high heterogeneity of data types in the agricultural/viticultural sector. Agriculture makes use of various data (e.g., research, meteorological, soil, financial/economic, statistical, satellite/remote sensing, administrative, germplasm, crop experiments, and field trials, to name a few) and for each data type, there are numerous standards used. These standards are not always linked between each other or at least with the most prominent one of the sector.

In principle, all data collected from a viticultural object (i.e., a vineyard, parcel, etc.) are stored in metadata records. The records are an important aspect for documenting and maintaining the semantic interrelations among various sources of information. These types are used in querying relevant data like viticulture historical data, over long periods of time. It is worth mentioning that standardized descriptive metadata enhances the overall search experience and render reliable the retrieval techniques within or across multiple collections. Information management is assisted by descriptive, administrative, technical, and preservation metadata, which they ensure data intellectual integrity. Furthermore, the primary sector in the now has employed ICT research groups under the scope of disseminating knowledge related to smart agriculture. Inter-operability introduces itself as an important factor for management and aggregation of digital resources among proprietary data structures. As an effect, many types of software applications have evolved implementing

existing methodologies such as crosswalks, translation algorithms, metadata registries, and specialized data dictionaries.

In the field of smart viticulture, in particular, the crosswalk provides a mapping of metadata entities within various schemata. Metadata definition is highlighted as the precise interpretation for implemented the required framework. Furthermore, common elements are hardly distinguished among different schemata due to existing incompatible records. Tables of equivalent elements serve the purpose of crosswalks and, even though the equivalences may be inexact, they represent an expert's judgment on the matter of software processes. More specifically, the conceptual differences are immaterial to the software's operation that involves records encoded in the two models. Therein-after, retrieval mechanisms support querying capabilities among different data sources that wholly support the desired semantic inter-operability of viticulture schemata.

An additional important factor on the adoption of crosswalks is the support to convertibility among data formats. Apart from single point of access or cross-domain searching capabilities, the aggregation of metadata records among varied from viticulture sources (e.g., vineyards located in quite distinct geographical areas or producing different wine types) presumably creates conflicting results. In the case of automatically generated metadata, organizations or individuals that did not follow best practices or standard thesauri and controlled vocabularies in the process, provide mapping metadata elements from different schemas under the scope of cross walking. Furthermore, datatype registration and formatting address the capabilities of semantic interoperability in relation to the values included in the metadata elements (e.g., rules for recording wine types or encoding standards for dates) assessing the local authority files for adopted terminologies.

As a solution to this issue, we introduce a novel adaptation of the MINT-mapping tool that provides an aggregation mechanism on critical activities, such as:

- harvesting and aggregating metadata inputs, created under shared community standards or proprietary metadata schemata,
- migrating from providers' standard or local models to a reference model,
- transforming records among intermediate standards to pan-European (Europeana Semantic Elements and the Europeana Data Model [30,31]).

### 3.2. Content Search and Management

Content search and management is divided into two distinct, yet collaborating parts, namely *content aggregation* and *content collection and spaces management*. In the following subsections, we shall try to provide a brief overview, together with an illustrative example case of their interchangeability.

#### 3.2.1. Content Aggregation

The content aggregation task is mainly accomplished through searching in external agricultural digital repositories and libraries. In the herein proposed framework, the utilization of WITH tool allows us to mash-up APIs from different digital resources and provide a single, yet powerful, service, which gives access to a large set of heterogeneous multimedia items (e.g., images, text, different metadata schemata, videos). The tool supports a variety of different data models (e.g., EDM, LIDO, etc.) and formats (e.g., XML, JSON-LD), and thus resolves most of the aforementioned interoperability issues. The latter is achieved by integrating the tool in the proposed architecture of fully implemented workflows or the ingestion, formal mapping, transformation and aggregation of viticulture records in the back-end through the proposed open-source metadata web-based platform.



The user interface (UI) of the knowledge base platform (<http://withculture.eu/#/custom/evinos>) offers federated and faceted search services, which enable its user to apply multiple search criteria in different combinations based on the metadata of search results, navigate through the results via different presentation views and retrieve more information about items of interest (see Figure 5). One of the system's provisions is the parallel query into all of the selected repositories, based on a specific search term or phrase, in order to create facets to be used for narrowing down results. As most digital data in external repositories are accompanied with rich metadata, the user may choose from a variety of specialized filters such as rights, creator, media type and dates, resulting in the desirable objects among thousands of others. Figure 6 indicates the improved adaptability that offers the search results and facets are accessed through the API, so other developers can include the query features into their customized applications.

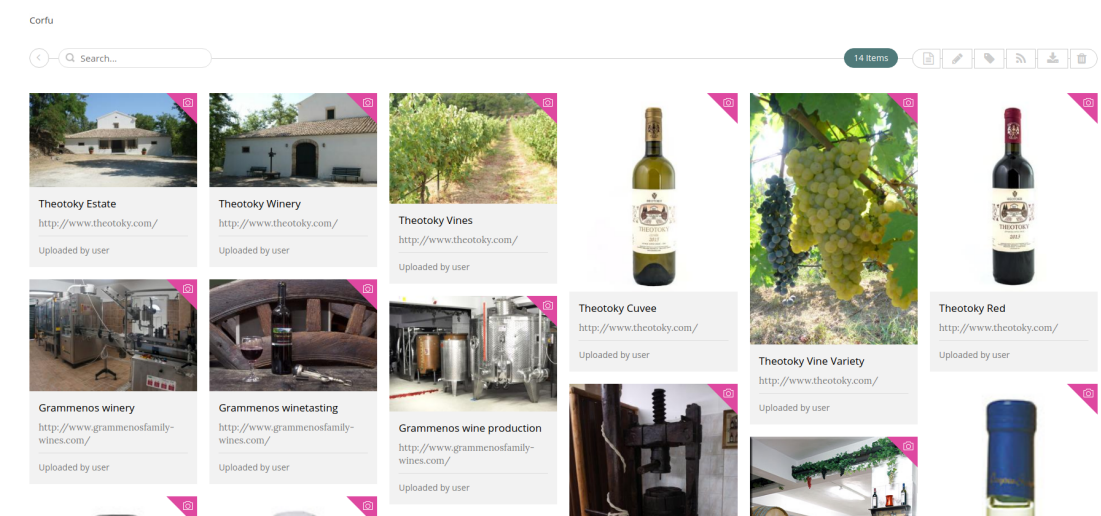


Figure 5. User data collection.

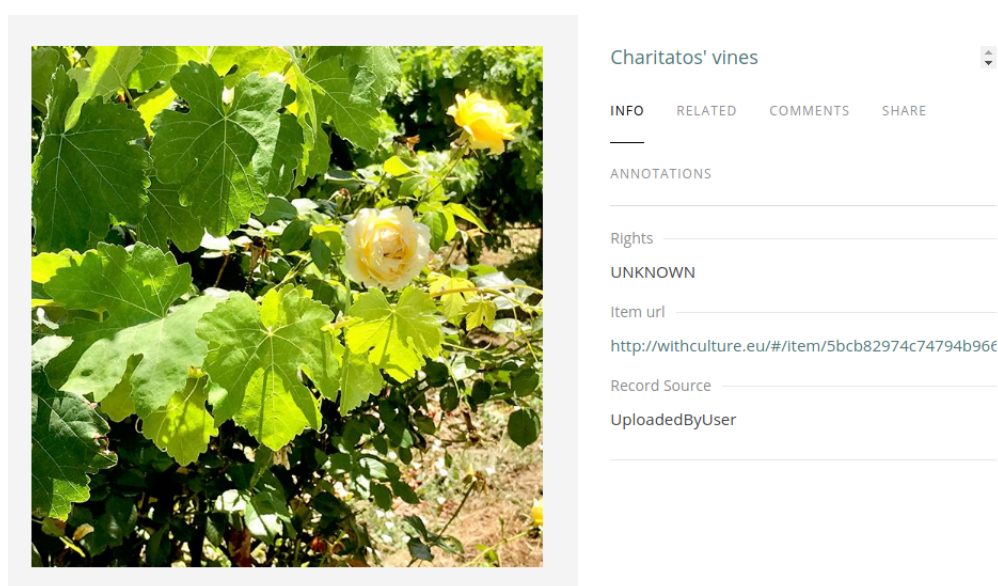


Figure 6. Selected item from the collection.

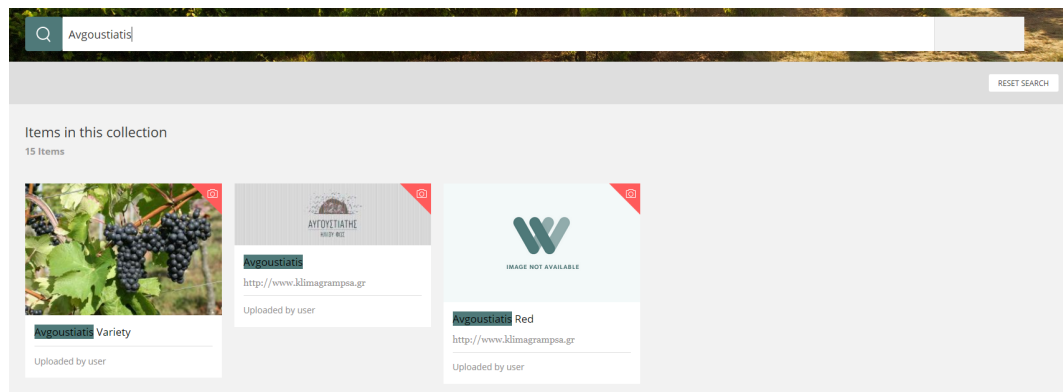
It worth mentioning that the required data model is based on the imported records following the paradigm of major repositories. More particularly, the model demands extensions for fulfilling compatibility issues among various data models. The proposed platform is compliant with Europeana data model (EDM) [30,31] and supports two-way transformation capabilities. Furthermore, any type of data amendment is possible to be implemented through its two-way transformation capabilities. Regarding the matter of its application programming interface, we introduced a common support framework on serialization types (JSON, XML, and RDF). This enhanced share-ability and export-ability of data among similar platforms worldwide.

### 3.2.2. Collections and Spaces Management

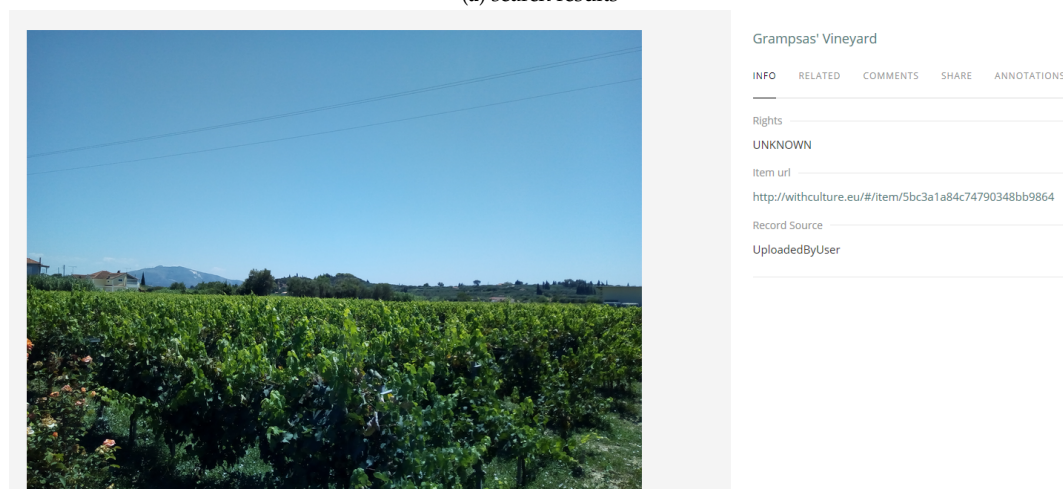
As a next step, platform users are able to create collections that become available under their specific profile. Data management is implemented through an internal data model that assesses record imports and transformations in the platform, under the scope of collection enrichment through user inputs. Particularly the system keeps intact the users' original records in the form of an accessible backup and while allowing the amendment of data into their custom collections. Along with collections, platform users may organize their content, which was either uploaded or collected from external repositories, into exhibitions that provide enhanced and more playful visualization features.

The platform also provides thematic content organization in the so-called *spaces*. Spaces is a concept that corresponds to specific, access-based views of stored data, enables the organization of digital content in different thematic categories and views, and allows different ways of interaction with the end-user (e.g., users are able to re-use the provided content, comment and share). This added feature enables interested wine professionals to design and host custom web spaces in order to promote their content and engage with users. Furthermore, devoted expansions of content holder serve the purpose of visualizing and disseminating on the web digital collections, exhibitions and stories, in addition to content access and use capabilities made to the public by the platform. Alternatively, linked open data (LOD) are published by the means of specific structures, namely the *spaces*. For the sake of efficient organization, content access within a space (collections and exhibitions, etc.) is limited to specified user groups. The scope of the search engine in each space can be customized as well, for example to exclude some sources or, to only search for specific type of sources, such as pictures, etc.. Individual customization of the individual front-end (e.g., descriptive texts, images, CSS) is also available within each space.

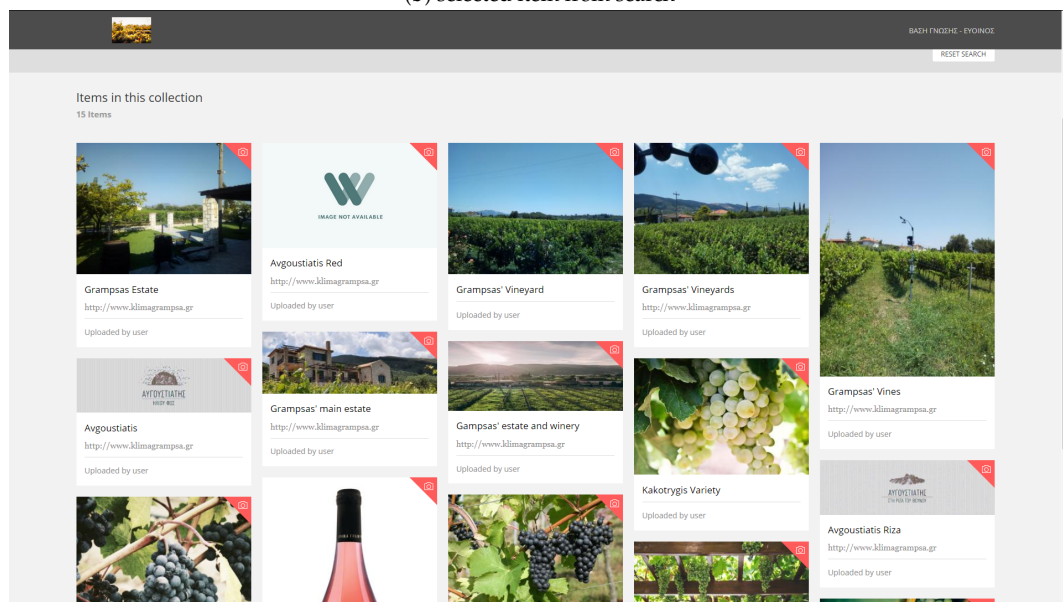
As a case in point, let's introduce Yorghos, a wine enthusiast fascinated by Greek wines of the Ionian Islands region, who wants to explore and discover relevant viticulture artifacts and build, organize and contextualize his own related library. The first thing for Yorghos to do within the platform is to create a user profile and to initialize an empty collection, where all the assets will be saved in the following. Let us assume that Yorghos owns content from Greek wineries, so he uploads and curates it by providing some initial metadata, such as name, description, and rights. To enrich his collection and add more assets, he uses the platform's search service to look for traditional Greek wine types of the Ionian Islands geographical region, like "Avgoustiatis" and "Moschato". Yorghos collects assets from worldwide repositories, such as the Digital Public Library of America, Europeana, and the European Wine Database (<http://www.eu-vitis.de/index.php>), and further filters the results by setting the preferable rights and mime type illustrated in Figure 7a. From a set of 15 items returned containing "Avgoustiatis", he opens the ones he is interested in, as shown in Figure 7b, reviews related images, inspects the accompanying metadata and collects the items he wants. Collected items are automatically transformed into the platform's data model, an excerpt of which, is shown in Algorithm 1. Yorghos may then observe and manage his collection shown in Figure 7c.



(a) search results



(b) selected item from search



(c) collection

Figure 7. User data management.

**Algorithm 1** Platform's datamodel.

---

```

"descriptiveData": {
  "label": "Greek wine type",
  "description": "This image depicts a Greek wine type
from the Ionian Islands geographical region",
  "keywords": [
    "Greek",
    "wine",
    "avgoustiatis",
    "winery"
  ],
  "isShownAt": "http://www.europeana.eu/api/ANgfDzTpW",
  "isShownBy": "https://www.uvinum.co.uk/zakynthos-wine/avgoustiatis-in-the-mountains-root-2015",
  "rdfType": "http://www.europeana.eu/schemas/edm/ProvidedCHO",
  "country": "Greece",
  "dclanguage": "English",
  "dctype": "scanned image",
  "dcrights": "Public Domain",
  "dctermspatial": "Zakynthos, 1999",
  "dcformat": "jpg"
}

```

---

### 3.3. Metadata Generation and Enrichment

As a means to both incorporate and interlink viticulture data that in principle come from a plethora of different sources (e.g., sensors, IoT devices), we utilize the platform's characteristics. In particular, the viticulture information that is collected and aggregated in the platform initially contains the original metadata that are provided directly by the various smart viticulture information sources. These are stored in the respective representation of each item using the described data model. To allow enrichment of these metadata with additional, either manually or automatically generated, metadata, the developed platform supports the use of annotations.

To facilitate annotation creation, retrieval, management, and interoperability, the platform includes a thesauri manager that is responsible for importing, through an offline process, the Linked Data vocabularies and datasets that make up the pool of potential annotation resources. The thesauri manager converts the imported vocabularies from their source format (e.g., SKOS thesauri, OWL ontologies, N-triples datasets) to a common model, stores them in the platform's thesauri database and indexes them to allow fast search and retrieval. In all the above vocabularies and datasets, each resource is always accompanied by one or more textual labels, possibly in more than one language. These labels provide textual representations for the specific resource and are used for indexing the resources and facilitating look-up.

The herein utilized annotation model is based on W3C's Web Annotation Model [32], which is a structured model and format to enable annotations to be shared and reused across different hardware and software platforms. In brief, an annotation consists of an id, a list of annotators, a body, a target, and a list of scores. An annotation may be generated either automatically by a content analysis software, a web-service etc., or manually by a human annotator. Thus, the list of annotators contains all relevant information about the origins of each annotation. The core part of the annotation is its body which identifies the relevant Linked Data resource. The target of an annotation identifies the item, or the part of an item, which the particular annotation relates to the body resource.

Based on the above annotation model, the platform provides a number of tools for the manual and automatic generation of annotations. The latter are based on natural language processing, string matching, dictionary look-up tools, as well as three supervised learning models, namely k-nearest neighbors, linear and random forest regression to analyze the existing textual metadata of records in order to detect in them occurrences of named entities and terms that belong to the underlying set of vocabulary and dataset resources; an indicative illustration of the latter is included in the Appendix A of the manuscript. In the case of the manual annotation process, the user has to directly choose a resource from the underlying thesauri database and add it as annotation for a



particular item. The user starts typing in a keyword and the user interface assists him through an auto-complete functionality (shown in Figure 8), which consults the underlying thesauri index, and suggests to the user resources having textually matching labels. The user can restrict the scope of auto-complete suggestions by selecting only specific vocabularies.

An important part of the manual annotation tools forms the geo-tagging tool, which allows the generation of annotations with a geo-URI in their body, so as to represent geographical locations. To generate such an annotation, the user of the platform may click on the desired points of the provided map. In conclusion, the automatic annotation tools are aimed to provide textual and visual metadata analysis. Textual metadata are the main descriptive metadata of the items, namely their title and description fields. Textual item data can also be used to automatically generate annotations through the dictionary look-up tool that is implemented in the platform. Beyond the scope of the current publication, the deployment of further services is implemented by the extraction of textual and image related metadata that assist the enforcement of automated tools on image annotation tasks.

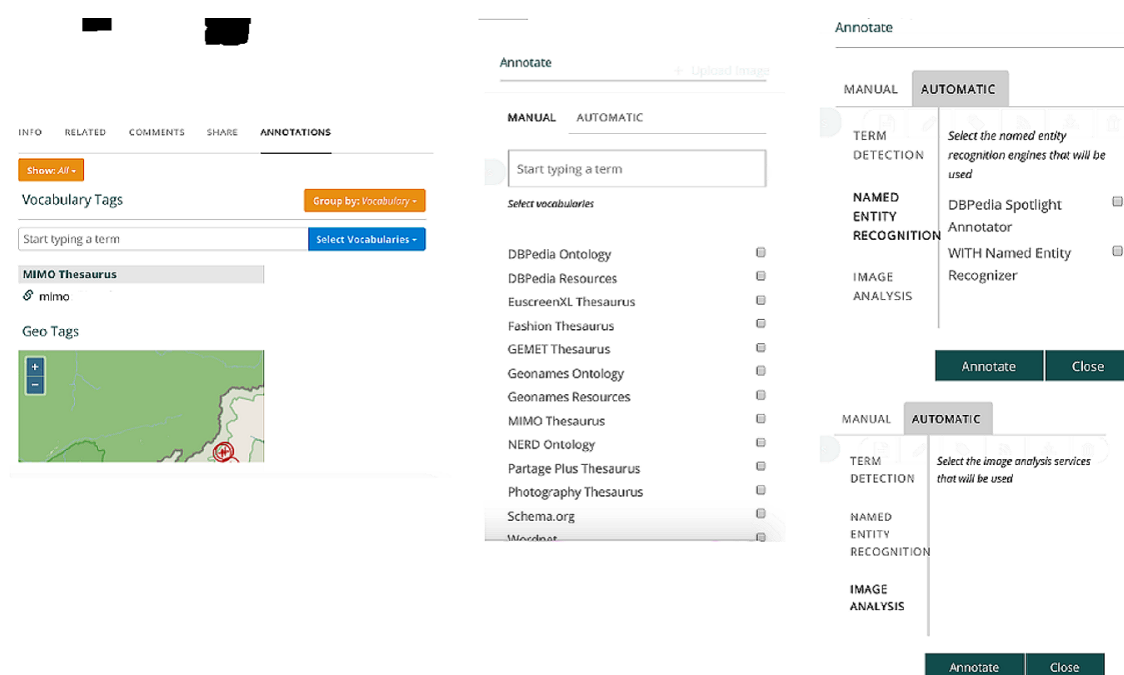


Figure 8. Manual annotation paradigm.

### 3.4. Case Study: The Zakynthian Variety Selection

The herein discussed platform has been successfully utilized in farms across the Ionian Islands. Among the first venues involved was a large winery in Zakynthos, Greece. The installation of the environmental monitoring system in Zakynthos focuses on the basic characteristics of the cultivation that is found in the “Klima Grampsa” estate on the island. The entire unit consists of a vineyard and a wine production station. In addition, the company manufactures typical Zakynthian varieties of wine, such as Avgoustiatis and Goustolidi, maintaining complete production characteristics. The cultivation is located in a lowland area surrounded by farmland and residential environments in south-west Zakynthos island and the field of study refers to the wider area of the Gramsa estate near the village Lagopodo (see Figure 9) belonging to the municipal unit of Artemision.



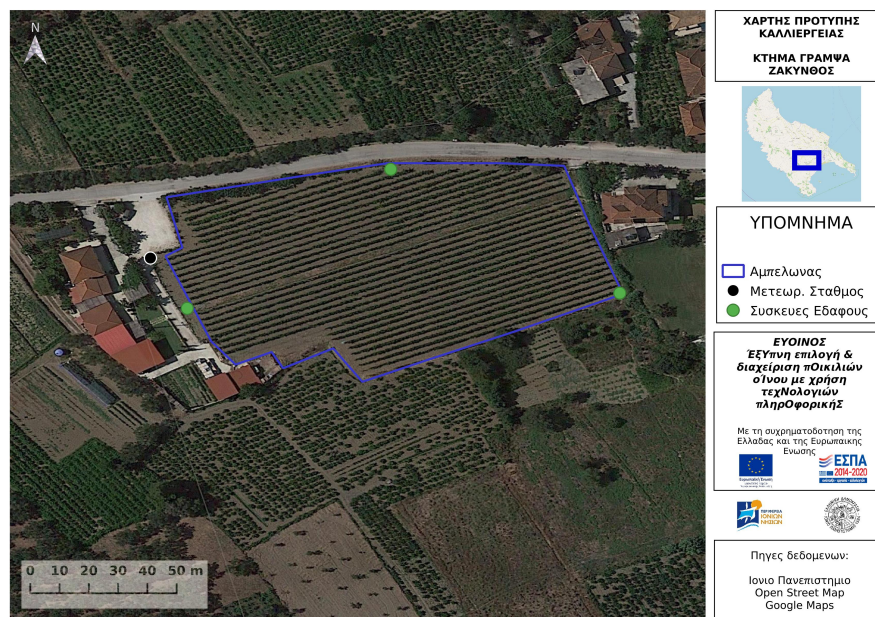


Figure 9. Map of Klima Grampsas installation.

The vineyard owners immediately embraced the initiative and contributed practically and theoretically to the project, aiming to increase conversions and product revenue for their clients via upgraded quality and optimized production. They used the platform to collect, organize, analyze, and publicize winery data on behalf of their winery products. By looking at the interests and behaviors of the people engaged most with their varieties, they were able to feed these insights back into the overall company planning to inform their strategy on how to best reach these consumers.

As a first step, the owners provided multimedia content (mostly photos and textual data) of their land, premises, plants and products (see Figure 10). This formed the so-called *Grampsas collection* in the knowledge base.

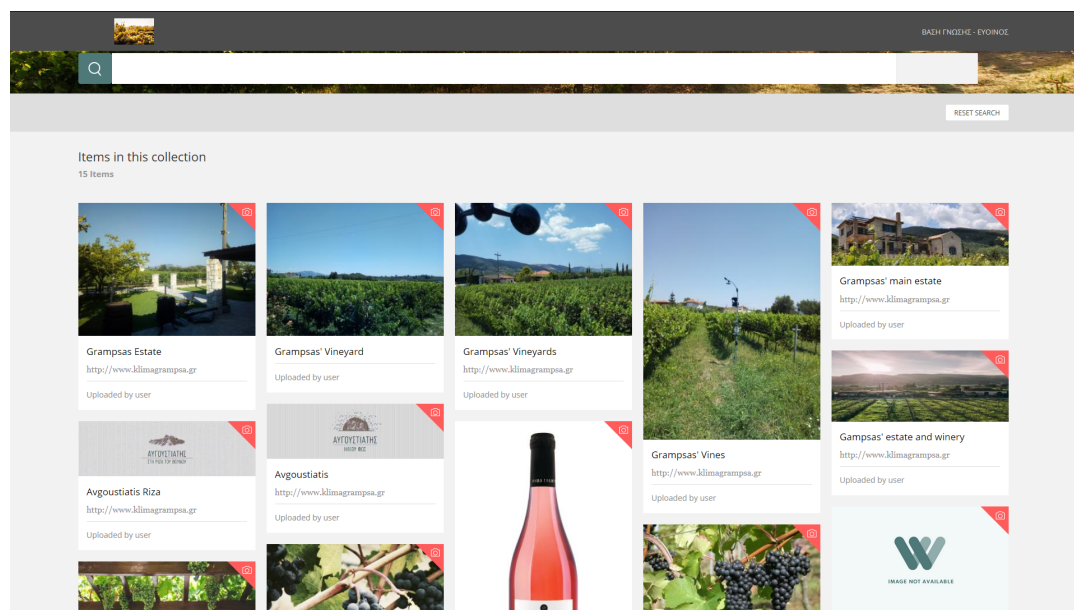


Figure 10. Grampsas collection.

As a second step, detailed information about the individual items of the collection were provided. These included descriptions of the produced varieties, such as Avgoustiatis, Goustolidi, and Kakotrygis, as well as the respective wines produced by each variety (e.g., *Avgoustiatis Red*, see Figure 11).

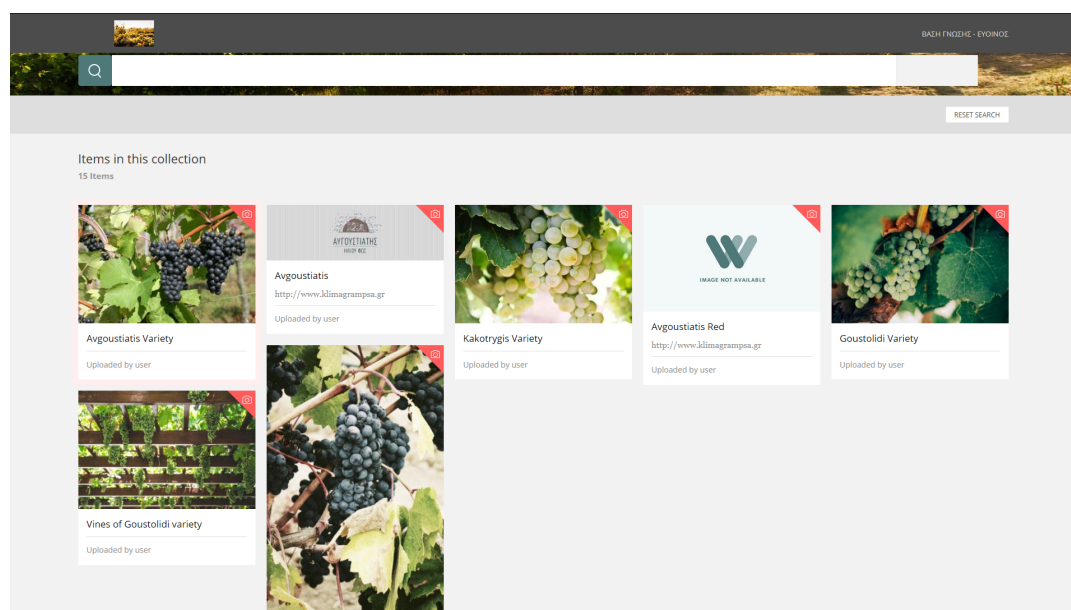
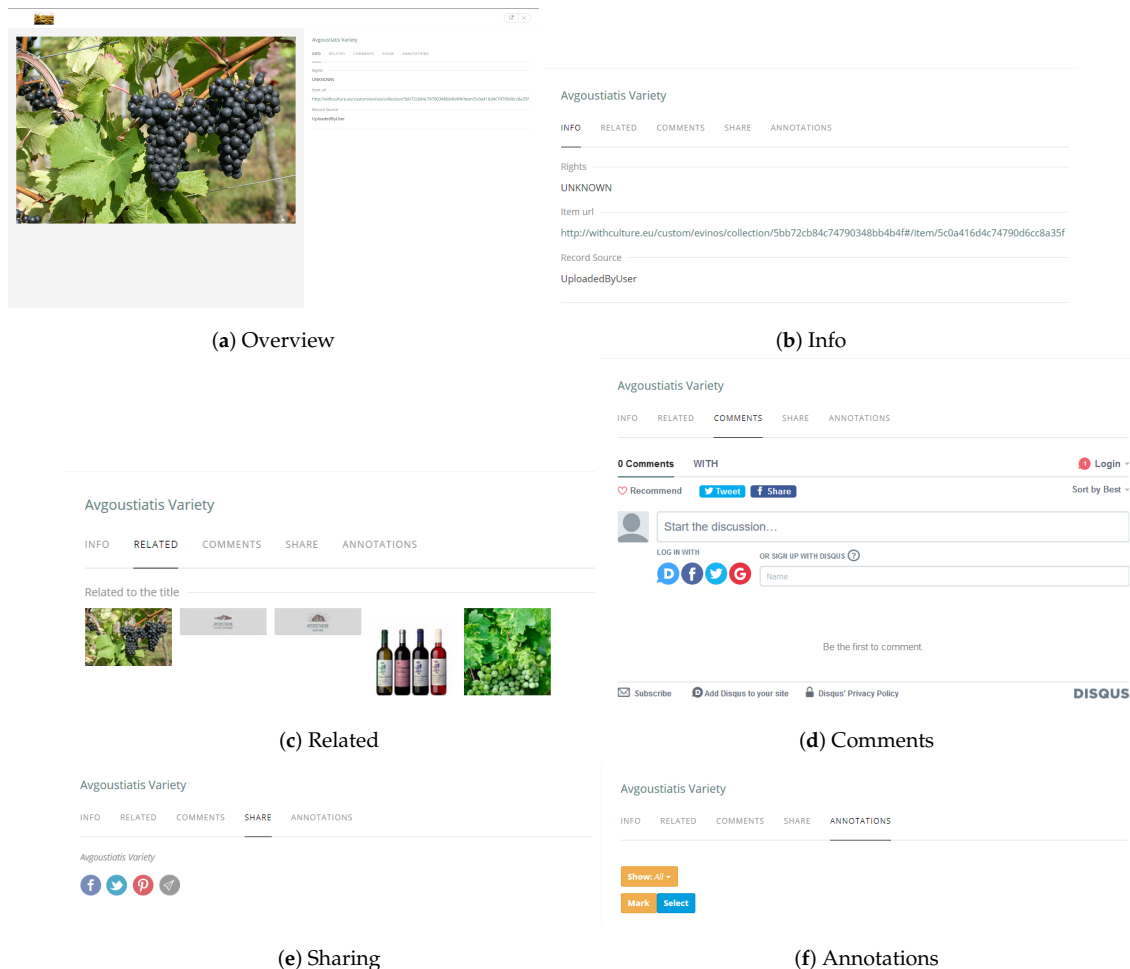


Figure 11. Grampas knowledge base item collection.

Initially the vineyard owner provided a recent high resolution photo of the item under examination. So, for instance in the case of the Avgoustiatis variety, Figure 12a depicts its overview. Continuing, the entries of the knowledge base were filled up with data. The *Info* tab (Figure 12b) contains information about the rights of the multimedia item uploaded, the item's Uniform Resource Locator (URL) and the actual knowledge base record's source. The latter is typically the user who uploaded the actual content item. The second tab *Related* presents similar to the uploaded content items that are already included in the knowledge base. In the case of a photo, given the single uploaded image, it retrieves similar images from the knowledge base and depicts them. The platform is also able to suggest tags that may be attached to the original uploaded image (Figure 12c). The *Comments* tab (Figure 12d) enables the public to participate in the process because it provides a fully functional comments section based on the popular Disqus (<https://disqus.com/>) service. Disqus is a networked community platform used by hundreds of thousands of sites all over the world. By incorporating the Disqus functionality, the platform gained a feature-rich comment system complete with social network integration, advanced administration, and moderation options.

Tab *Sharing* (Figure 12e) deals with social media integration of the platform and allows its visitors to quickly associate, post, or comment about the knowledge base items on their favorite social media sites. This is of crucial importance because social media is nowadays integrated in almost every aspect of the people's everyday life, and at the same time forms a great way to leverage the power of social media to increase awareness of the owners of the vineyards marketing efforts. Lastly, tab *Annotations* (Figure 12f) offers the ability to the content uploader to provide important annotations describing the actual content uploaded by either inserting them manually during the process or by allowing the platform to process them and add them in a semi-automated way during an offline post-processing step. Finally, it is also worth mentioning that at this point, after the initial stage of content uploaded mainly by the vineyard owners has been concluded, any interested stakeholder may now use the platform to add new, edit or delete existing information; the latter two actions are subject to the administrator's approval, of course. This allows for a great user participatory approach to smart viticulture, actively involving all stakeholders in the process to help ensure the results meet their needs and are rather exploitable.



**Figure 12.** Knowledge base (KB) item walk-through.

As a final, not yet implemented, step, vineyard owners will be able to use machine-learning tools that predict the likelihood of a customer to engage with a specific wine variety product, and automatically adjusts the target customers with those findings. The overall feedback we are getting so far from them is that they were able to significantly increase conversions using the platform's technology.

#### 4. Discussion and Conclusions

In this paper, we presented a collaboration platform based on linked data and machine principles with unique features, such as automatic metadata enrichment and services, specifically aimed for the viticulture domain. We provided a detailed description of the workflow and the user engagement in the platform and illustrated its basic principles, alongside with a respective real-life case study paradigm. The platform forms an evolving, flexible, and interconnected integrated information system. It follows a modular design and implementation that allows it to efficiently adapt over time over numerous distinct tasks. For instance, new knowledge base repositories may be aggregated, new knowledge base spaces may be created and new platform features and services may constantly be designed, deployed and utilized within the proposed architecture. It was designed and developed aiming at strengthening cooperation of the interested stakeholders within the viticulture domain, initially within the Region of Ionian Islands and beyond.

The herein proposed part of the platform aims to propose innovative solutions for the improvement of smart viticulture/agriculture services, their efficient management, including all

key stakeholders, by providing incentives for the role of vineyard owners and consumers of viticulture products. Moreover, it aims at involving the wisdom of the crowd, in order to ensure wider dissemination of the achieved results. This, of course, remains a work in progress for our research team, in the sense that among our future goals are the extension of the current platform to include data from other than viticulture application sub-domains of the general smart agriculture domain, as well as the introduction of enhanced features for the already implemented viticulture one.

Among the technical improvements of the platform that we plan to implement in the near future lies an automated tag extraction service, based on specifically trained deep convolutional neural networks (CNNs), to be included in the tabular knowledge base item approach (Figure 12a). This service will be used to provide automated tags that will be fed to users for verification and validation and then will be included in the training set. In addition, we plan to exploit the provided user annotations of knowledge base items for the training of new, or improving the existing, ML algorithms, towards an enhanced promotion of digital agriculture data from various sources. Analytics of user-generated data or data simply derived from private sources, coupled together with huge amounts of publicly available data from large-scale organizations (e.g., the Hellenic National Meteorological Service) will allow the utilization of new types and combinations of data that were never available before. For instance, data on farming production costs (i.e., seeds, fertilizer, labor and yield levels) for various crops collected directly from farmers, combined with spatial, weather and commodity datasets, as well as insights like cost curves, may be successfully analyzed by machine learning algorithms, so as to predict market prices of agricultural products with a reliable degree of accuracy.

So, finally, with respect to the marketing part of our initiative, we plan to initiate deeper wine customer connections to drive better marketing results for the vineyard owners by utilizing existing, or developing from scratch, analytics solutions. The goal is to obtain reliable data from large numbers of real-world people, process them in a single framework, where we may identify peoples' insights, what actions should be implemented and towards which audience, how it is performing, and make changes almost in real-time. Taking this a step further, we intend to transform the system into an integrated agriculture analytics platform used to predict the acceptance of products by the market and their economic performance in economic terms, as well as forecast crop yields with increased accuracy. We believe that the data revolution currently happening in agriculture has huge implications for both society and related business models, and the impact will be colossal in the years to come. Consequently, our ultimate goal is to develop an integrated platform infrastructure for further assessment and utilization in the directions of data analysis in smart agriculture, supported by artificial intelligence tools and methodologies. Additionally, the exploitation of statistical processing methods and even semantic analysis (automatic metadata deduction), will lead our research to the implementation of a unified platform for analytics, smarter marketing and better results in the agricultural domain.

**Author Contributions:** P.M., Y.V. conceived of the idea, designed the platform, implemented the individual parts and tailored the concluding remarks, P.M., Y.V., and A.S. drafted the initial manuscript and P.M., Y.V. and A.S. revised the final manuscript.

**Funding:** This research was funded by the European Union and Greece (Partnership Agreement for the Development Framework 2014–2020) under the Regional Operational Programme Ionian Islands 2014–2020, project title: “Smart vine variety selection and management using ICT - EYOINOS”, project number: 5007310.

**Conflicts of Interest:** The authors declare no conflict of interest.



## Appendix A

```

In [1]: import pandas as pd
import numpy as np
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
import math

In [2]: df_raw = pd.read_csv('dataset.txt', sep='|', low_memory='False',
    parse_dates=['Date_Time'])

In [3]: df_raw.columns = [x.strip().replace(' ', '_') for x in df_raw.columns]

In [4]: df_raw.describe(include='all').T
Out[4]:
```

	count	unique	top	freq	first	last	mean	std	min	25%	50%	75%	max
Date_Time	1008	1008	2019-02-26 09:30:00	1	2019-02-24 18:40:00	2019-03-03 18:30:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Temp	1008	NaN	NaN	NaN	NaN	NaN	8.90081	36.3097	-3.61	1.7075	7.665	14.4025	113
Humidity	1008	74	92.0%	54	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Barometer	1008	NaN	NaN	NaN	NaN	NaN	1031.96	4.33417	1023.13	1028.75	1031.88	1035.77	103
Pressure	1008	46	29.86	50	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
WindSpeed	1008	NaN	NaN	NaN	NaN	NaN	1.7619	2.4179	0	0	0	3.2	9.6
WindDirection	448	NaN	NaN	NaN	NaN	NaN	139.47	111.785	0	45	90	247.5	337
WindGust	1008	NaN	NaN	NaN	NaN	NaN	5.95873	6.39589	0	0	3.2	11.2	25.6
WindGustDir	701	NaN	NaN	NaN	NaN	NaN	148.288	107.098	0	68	113	248	338
Rain	1008	NaN	NaN	NaN	NaN	NaN	0	0	0	0	0	0	0
Rainrate	1008	NaN	NaN	NaN	NaN	NaN	0	0	0	0	0	0	0
Dewpoint	1008	NaN	NaN	NaN	NaN	NaN	1.66936	4.18039	-5.32	-2.13	1.4	5.33	11.1
Windchill	1008	NaN	NaN	NaN	NaN	NaN	7.77203	6.92719	-3.61	1.7075	7.61	14.39	19.6
Radiation	1008	NaN	NaN	NaN	NaN	NaN	179.656	253.838	0	0	0	393	821
UV	1008	NaN	NaN	NaN	NaN	NaN	0.642857	1.0705	0	0	0	1.1	3.6

```

In [5]: print("number of columns/features: " + str(df_raw.shape[1]))

```

Figure A1. Summarized algorithm report - Excerpt 1.

```

number of columns/features: 15

In [6]: print("number of rows/measurements: " + str(df_raw.shape[0]))
number of rows/measurements: 1008

In [7]: df_raw.dtypes
Out[7]:
```

Date_Time	datetime64[ns]
Temp	float64
Humidity	object
Barometer	float64
Pressure	object
WindSpeed	float64
WindDirection	float64
WindGust	float64
WindGustDir	float64
Rain	float64
Rainrate	float64
Dewpoint	float64
Windchill	float64
Radiation	float64
UV	float64
dtype:	object

Figure A2. Summarized algorithm report - Excerpt 2.



```

In [8]: print("Percentage of null values in the dataset based on each feature")
Percentage of null values in the dataset based on each feature

In [9]: df_raw.isnull().sum().sort_index()/len(df_raw)

Out[9]: Barometer      0.000000
Date_Time      0.000000
Dewpoint      0.000000
Humidity      0.000000
Pressure      0.000000
Radiation      0.000000
Rain      0.000000
Rainrate      0.000000
Temp      0.000000
UV      0.000000
WindDirection  0.555556
WindGust      0.000000
WindGustDir    0.304563
WindSpeed      0.000000
Windchill      0.000000
dtype: float64

In [10]: df_raw.fillna(df_raw.mean(), inplace=True)

In [11]: df_raw.drop(["Date_Time", "Windchill", "Pressure"], axis=1, inplace=True)

In [12]: df_raw['Humidity'] = df_raw['Humidity'].str.rstrip('%').astype('float') / 100.0

In [13]: X, y = df_raw.drop("Temp", axis=1), df_raw.Temp
X_test, X_train = X[:300], X[300:]
y_test, y_train = y[:300], y[300:]
X_test.shape, X_train.shape, y_test.shape, y_train.shape

#X_val, X_train = X_train_val[], X_train_val[]
#y_val, y_train = y_train_val[], y_train_val[]

Out[13]: ((300, 11), (708, 11), (300,), (708,))

In [22]: m = RandomForestRegressor(n_jobs=-1)

In [36]: m.fit(X_train, y_train)

```

Figure A3. Summarized algorithm report - Excerpt 3.

```

Out[36]: RandomForestRegressor(bootstrap=True, criterion='mse', max_depth=None,
max_features='auto', max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, n_estimators=10, n_jobs=-1,
oob_score=False, random_state=None, verbose=0, warm_start=False)

In [37]: m.score(X_train, y_train)

Out[37]: 0.9324235524560729

In [38]: m.score(X_test, y_test)

Out[38]: 0.8983705875252752

```

Figure A4. Summarized algorithm report - Excerpt 4.

## References

1. Liakos, K.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine Learning in Agriculture: A Review. *Sensors* **2018**, *18*, 2674. [[CrossRef](#)] [[PubMed](#)]
2. Wolfert, S.; Verdouw, L.G.C.; Bogaardt, M.J. Big Data in Smart Farming—A review. *Agric. Syst.* **2017**, *153*, 69–80. [[CrossRef](#)]
3. Proffitt, T.; Bramley, R.; Lamb, D.; Winter, E. *Precision Viticulture: A New Era in Vineyard Management and Wine Production*; Winetitles: Ashford, UK, 2006.
4. Bramley, R.G.V.; Hamilton, R.P. Understanding variability in winegrape production systems: 1. Within vineyard variation in yield over several vintages. *Aust. J. Grape Wine Res.* **2004**, *10*, 32–45. [[CrossRef](#)]
5. Bramley, R.G.V. Understanding variability in winegrape production systems 2. Within vineyard variation in quality over several vintages. *Aust. J. Grape Wine Res.* **2005**, *11*, 33–42. [[CrossRef](#)]

6. Morais, R.; Peres, E.; Boaventura-Cunha, J.; Mendes, J.; Cosme, F.; Nunes, F.M. Distributed monitoring system for precision enology of the Tawny Port wine aging process. *Comput. Electron. Agric.* **2018**, *145*, 92–104. [[CrossRef](#)]
7. Canete, E.; Chen, J.; Martin, C.; Rubio, B. Smart Winery: A Real-Time Monitoring System for Structural Health and Ullage in Fino Style Wine Casks. *Sensors* **2018**, *18*, 803. [[CrossRef](#)]
8. Naumowicz, T.; Freeman, R.; Kirk, H.; Dean, B.; Calsyn, M.; Liers, A.; Braendle, A.; Guilford, T.; Schiller, J. Wireless Sensor Network for habitat monitoring on Skomer Island. In Proceedings of the 35th Annual IEEE Conference on Local Computer Networks, LCN 2010, Denver, CO, USA, 10–14 October 2010; pp. 2882–2888.
9. Muangprathub, J.; Boonnam, N.; Kajornkasirat, S.; Lekbangpong, N.; Wanichsombat, A.; Nillaor, P. IoT and agriculture data analysis for smart farm. *Comput. Electron. Agric.* **2019**, *156*, 467–474. [[CrossRef](#)]
10. Voutos, Y.; Mylonas, P. A semantic data model for sensory spatio-temporal environmental concepts. In Proceedings of the 22nd Panhellenic Conference on Informatics, Athens, Greece, 29 November–1 December 2018.
11. Borgogno-Mondino, E.; Lessio, A.; Tarricone, L.; Novello, V.; de Palma, L. A comparison between multispectral aerial and satellite imagery in precision viticulture. *Precis. Agric.* **2018**, *19*, 195–217. [[CrossRef](#)]
12. Mehdizadeh, S.; Behmanesh, J.; Khalili, K. Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. *Comput. Electron. Agric.* **2017**, *139*, 103–114. [[CrossRef](#)]
13. Feng, Y.; Peng, Y.; Cui, N.; Gong, D.; Zhang, K. Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data. *Comput. Electron. Agric.* **2017**, *136*, 71–78. [[CrossRef](#)]
14. Patil, A.P.; Deka, P.C. An extreme learning machine approach for modeling evapotranspiration using extrinsic inputs. *Comput. Electron. Agric.* **2016**, *121*, 385–392. [[CrossRef](#)]
15. Mohammadi, K.; Shamsirband, S.; Motamedi, S.; Petkovic, D.; Hashim, R.; Gocic, M. Extreme learning machine based prediction of daily dew point temperature. *Comput. Electron. Agric.* **2015**, *117*, 214–225. [[CrossRef](#)]
16. Clingeffer, P.; Petrie, P.; Dunn, G.; Martin, S.; Krstic, M.; Welsh, M. *Final Report to Grape and Wine Research & Development Corporation: Crop Control for Consistent Supply of Quality Winegrapes*; CSIRO Division of Horticulture: Canberra, Australia, 2005.
17. Diago, M.; Correa, C.; Millan, B.; Barreiro, P.; Valero, C.; Tardaguila, J. Grapevine Yield and Leaf Area Estimation Using Supervised Classification Methodology on RGB Images Taken under Field Conditions. *Sensors* **2012**, *12*, 16988–17006. [[CrossRef](#)] [[PubMed](#)]
18. Rodriguez-Pulido, F.; Gomez-Robledo, L.; Melgosa, M.; Gordillo, B.; Gonzalez-Mireta, M.; Heredia, F. Ripeness estimation of grape berries and seed by image analysis. *Comput. Electron. Agric.* **2012**, *82*, 128–133. [[CrossRef](#)]
19. Usha, K.; Singh, B. Potential applications of remote sensing in horticulture—A review. *Sci. Hortic.* **2013**, *153*, 71–83. [[CrossRef](#)]
20. Zhang, B.; Huang, W.; Li, J.; Zhao, C.; Fan, S.; Wu, J.; Liu, C. Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review. *Food Res. Int.* **2014**, *62*, 326–343. [[CrossRef](#)]
21. Diago, M.; Fernandes, A.; Millan, B.; Tardaguila, J.; Melo-Pinto, P. Identification of grapevine varieties using leaf spectroscopy and partial least squares. *Comput. Electron. Agric.* **2013**, *99*, 7–13. [[CrossRef](#)]
22. Ramos, P.; Prieto, F.; Montoya, E.; Oliveros, C. Automatic fruit count on coffee branches using computer vision. *Comput. Electron. Agric.* **2017**, *137*, 9–22. [[CrossRef](#)]
23. Amatya, S.; Karkee, M.; Zhang, Q.; Whiting, M.D. Automated Detection of Branch Shaking Locations for Robotic Cherry Harvesting Using Machine Vision. *Robotics* **2017**, *6*, 31. [[CrossRef](#)]
24. Radhika, Y.; Shashi, M. Atmospheric Temperature Prediction using Support Vector Machines. *Int. J. Comput. Theory Eng.* **2009**, *1*, 55–58. [[CrossRef](#)]
25. Gill, M.K.; Asefa, T.; Kemblowski, M.W.; McKee, M. Soil moisture prediction using Support Vector Machines. *JAWRA J. Am. Water Resour. Assoc.* **2006**, *42*, 1033–1046. [[CrossRef](#)]
26. Huuskonen, J.; Oksanen, T. Soil sampling with drones and augmented reality in precision agriculture. *Comput. Electron. Agric.* **2018**, *154*, 25–35. [[CrossRef](#)]

27. Fuentes, S.; Hernandez-Montes, E.; Escalona, J.; Bota, J.; Viejo, C.G.; Poblete-Echeverria, C.; Tongson, E.; Medrano, H. Automated grapevine cultivar classification based on machine learning using leaf morpho-colorimetry, fractal dimension and near-infrared spectroscopy parameters. *Comput. Electron. Agric.* **2018**, *151*, 311–318. [[CrossRef](#)]
28. Drosopoulos, N.; Tzouvaras, V.; Simou, N.; Christaki, A.; Stabenau, A.; Pardalis, K.; Xenikoudakis, F.; Kollias, S. A Metadata Interoperability Platform. In *Museums and the Web 2012 (MW2012)*; Museums and the Web: San Diego, CA, USA, 2012.
29. Kollia, I.; Tzouvaras, V.; Drosopoulos, N.; Stamou, G. A systemic approach for effective semantic access to cultural content. *Semant. Web* **2012**, *3*, 65–80.
30. Isaac, A.; Clayphan, R. Europeana Data Primer. 2010. Available online: [https://pro.europeana.eu/files/Europeana\\_Professional/Share\\_your\\_data/Technical\\_requirements/EDM\\_Documentation/EDM\\_Primer\\_130714.pdf](https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf) (accessed on 21 April 2019).
31. Isaac, A.; Charles, V. Europeana Data Model Definition. 2017. Available online: [https://pro.europeana.eu/files/Europeana\\_Professional/Share\\_your\\_data/Technical\\_requirements/EDM\\_Documentation/EDM\\_Definition\\_v5.2.7\\_042016.pdf](https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Definition_v5.2.7_042016.pdf) (accessed on 21 April 2019).
32. Sanderson, R.; Ciccicarese, P.; Young, B. Web Annotation Data Model, 2017. Available online: <https://www.w3.org/TR/annotation-model/> (accessed on 21 April 2019).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).