# A generic ontological network for Agri-food experiment integration – Application to viticulture and winemaking

Aunur Rofiq Muljarto, Jean-Michel Salmon, Brigitte Charnomordic, Patrice Buche, Anne Tireau, Pascal Neveu

# A generic ontological network for agri-food experiment integration - Application to viticulture and winemaking

Aunur Rofiq Muljarto[a,d], Jean-Michel Salmon[b], Brigitte Charnomordic[a], Patrice Buche[c,e], Anne Tireau[a], Pascal Neveu[a]

[a] *INRA, MISTEA Joint Research Unit, UMR729, F-34060 Montpellier, France*
[b] *INRA, Unité Expérimentale de Pech Rouge, UE0999, F-11430 Gruissan, France*
[c] *INRA, IATE Joint Research Unit, UMR1208, F-34060 Montpellier, France*
[d] *Dept. of Agroindustrial Technology, Brawijaya University, Malang 65145, Indonesia*
[e] *INRIA GraphiK, LIRMM, UMR5506, F-34095 Montpellier, France*

**Abstract**

This paper presents an ontological approach of observational data integration across complementary sub-domains, i.e., agriculture production and food processing, with an application to viticulture and winemaking process. The two main steps in this approach are i) to integrate preexisting ontologies to create a so-called ontology network and ii) to populate the ontology network with actual experimental data from different sources. The Agri-Food Experiment Ontology (AFEO), a new ontology network was developed based on two existing ontology resources, i.e., AEO (Ontology for Agriculture Experiment) and OFPE (Ontology for Food Processing Experiment). It contains 136 concepts which cover various viticultural practices, and winemaking products and operations. AFEO was used to guide the data integration of two different data sources, i.e., viticulture experimental data stored in a relational database and winemaking experimental data stored in Microsoft Excel files. Two potential uses by researchers of viticulture-winemaking integrated data using AFEO are shown. The first one is about wine traceability and the second one is related to the influence of grape varieties, irrigation practices, and different winemak-

*Email address:* `patrice.buche@inra.fr` (Patrice Buche)

*Preprint submitted to Computer and Electronics in Agriculture* *February 8, 2017*

ing methods on GSH concentration in wine. Those examples show that data integration guided by an ontology network can provide researchers with the information necessary to address extended research questions, illustrated in the paper by viticulture and winemaking processes.

*Keywords:* ontology network, data integration, viticulture, winemaking

## 1. Introduction

Research in agri-food and related domains dealing with sustainability is evolving more and more extensively in recent years to be more integrative, collaborative, and interdisciplinary [1]. These tendencies consider the agri-food domain as an interconnected system with various entities and complex relationships among them [2]. Consequently, data sources over the whole food chain are becoming available and can be combined to address new questions. For example, to answer a question or to test a hypothesis about the effects of different viticulture treatments on wine quality, researchers need to access various data sources from a set of winemaking experiments and viticulture practices. Data analysis could then provide better information and understanding of what actually happened during a set of experiments and give the possibility of acquiring new knowledge all along the agri-food chain.

However, to access and to incorporate various data sources, researchers have to deal with some obstacles. Data are commonly stored in scattered places that sometimes make it difficult to combine them. Moreover, data are very diverse in terms of formats, naming, storage mechanisms, and query or retrieval mechanisms. The heterogeneity of scientific data may come from many factors, such as (i) most scientific data are collected distinctively based on independent research projects; (ii) the data structures are frequently selected according to the collection methods (e.g., to make data easier to record) or the format is suggested by analysis tools, instead of standard data representations (e.g., relational database schema); and (iii) the terms and concepts used to label data are not standardized, neither within nor across scientific disciplines and research groups [3]. The difficulty of organizing available data and knowledge in a unified way not only limits research productivity but also reduces data traceability [4].

Research experiments are commonly divided into some sub-domains, such as agriculture production, post-harvest, and food transformation process. Even though the explicit relation between them is clearly understandable, each of these has different objectives, scopes, and circumstances. For instance, the agricultural experiments are normally performed in the fields, in which external factors have a significant contribution while food processing experiments are generally carried out in the laboratory with more controllable environments. From a practical point of view, they require different methods for collecting and organizing observational data that would lead to differences in the data format or structure and in the way the data are stored. The heterogeneity also occurs due to the vast scope of agri-food domains, ranging from plant cultivation up to the final processed food product. Each discipline uses its own knowledge expression, terms, concepts and semantic relations that might cause some difficulties to share the observational data.

Studies in the last two decades have shown that ontologies represent a flexible way to link the information contained in heterogeneous data sources within or across domains [4, 5]. Ontologies also provide the common concepts for data integration, thus opening the possibility to draw more comprehensive conclusions and to view data from different perspectives. Ontologies also allow certain types of automated reasoning to be performed in order to fulfil some specific requirements. The capabilities of data and knowledge sharing as well as reasoning will help to develop more advanced Information Systems able to manage heterogeneous data sources and to design platforms for more collaborative and accurate scientific data analysis.

The contribution of this work is to provide a method to prepare and to integrate data sources prior to further analysis in order to answer complex questions that require access to various agri-food scientific data sources. This method works out in viticulture and winemaking to solve complex questions using heterogeneous experimental data sources. To achieve this purpose, we developed the Agri-Food Experiment Ontology (AFEO), a new ontology network resource, based on two ontology resources, i.e., AEO[1] (Ontology for Agriculture Experiment), which is also an original contribution of this paper, and OFPE[2] (Ontology for Food Processing Experiment). The AEO and OFPE are ontologies that have been developed separately in research laboratories as generic knowledge representations of two respective sub-domains, i.e., agricultural production and food transformation process. By following the NeOn [6] methodology, we integrated these two ontologies into an ontology network, in order to facilitate data integration across these complementary sub-domains. An ontology network [6] is a new ontology engineering concept, which allows ontology re-use and avoids custom-building new ontologies from scratch.

Although in this paper the proposed ontology is specialized and tested for viticulture and winemaking experiments, the core elements of AFEO are fairly generic and might be adapted to other food products. Furthermore, the ontological definitions (concepts and relations) can be used to impose and preserve a logical structure for new types of scientific data, that may appear due to new sensors, protocols and analyses.

This paper is structured as follows: Section 2 gives a brief survey on recent

[1]http://agroportal.lirmm.fr/ontologies/AEO
[2]http://agroportal.lirmm.fr/ontologies/OFPE

work related to scientific data integration and agri-food ontologies; Section 3 describes the global agri-food experiment ontology design process, introduces the two ontology resources used, their integration and specialization to viticulture and winemaking experiments; Section 4 presents how the new proposed ontology network is instantiated to integrate two different data sources: Section 5 is dedicated to discussion and a potential use in the domain of viticulture and winemaking; Finally, in Section 6, conclusions are drawn and further work outlined.

## 2. State of the art

The works on ontologies have increased recently not only in computer science but also in various domains, including agri-food related domains. This is driven by the need to communicate and share knowledge with common understanding. This section is divided into two subsections: i) scientific data integration with respect to an ontological approach and ii) current work on agri-food ontologies.

### *2.1. Scientific data organization and integration*

Over the recent years researchers have faced significant problems to manage scientific data due to their increasing volume and complexities. This causes researchers to spend quite a bit of time to manage and integrate scientific data rather than to directly focus on their analysis [7] /* BRIGITTE: AJOUT REF PHENOME */. Scientific data are generally collected from measurements directly linked to real-world phenomena [8]. Within the agri-food domain, cross-disciplinary scientific data are required to explore complex and temporal aspects of food quality and the impact of practices and operations.

The need for a more adaptable mechanism to organize scientific data has been addressed in the literature, both in non-ontological based approaches such as, LabKey Server [9] and SciPort [10]; and in ontological based approaches such as proposed in [11], [7], [12], [3], [13] and [14]. Some of these approaches are targeted to specific scientific domains while the others are developed to be more generic and extensible. The non-ontological approach mostly relies on data models, such as database or XML schema, where attributes and relationships of domain concepts are captured in standardized structures. The ontological approach has some additional advantages in terms of data interoperability and knowledge reasoning [15, 16, 17, 18].

To the best of our knowledge, we did not find an ontological approach to represent scientific experiments as well as observational data which fulfill all our needs. An ontology of scientific experiments, the EXPO [19], has been proposed to formalize the generic concepts of experimental design, methodology and result representation. However, this ontology does not provide a clear explanation about scientific data representation. Neither does this ontology describe how to manage a set of experiments in which several interrelated experiments have to be conducted in a given order. The Extensible Observation Ontology (OBOE) has been developed to serve as a formal and generic conceptual framework for describing the semantics of observational data sets (i.e., data sets consisting of observations and measurements) [7]. The basic concepts of the observational model consists of five classes and six properties [3]. This ontology can be applied to various types of observations. Nevertheless, it is more suited to representing scientific data which are generated from measurement by sensors. It does not provide other types of observational data such as expert judgements or results from a calculation procedure.

### *2.2. Agri-food ontologies*

In order to fulfill the needs of representing complex knowledge, sharing common understanding within or across scientific domains and reusing domain knowledge in agri-food sectors, some ontologies have been proposed, either very general or more specific. Some researchers focused on building ontologies to conceptually model agricultural practices related to crop productions, such as, hilly citrus production ontology [20], precision agriculture ontology [21], crop-pest ontology [22] and potato ontology [23]. Meanwhile, some works concentrating on food taxonomy have also been carried out, such as FOODS (Food-Oriented Ontology-Driven System) [24] and wine classification [25].

The valid definition of concepts and a complete taxonomy for agricultural practices and food processing are important to model agri-food experiments. Most of the above ontologies refer to the terms or concepts listed in AGROVOC. AGROVOC is a multi-lingual vocabulary developed by the Food and Agriculture Organization (FAO) of the United Nations and covers divers areas including food, nutrition, agriculture, fisheries, forestry, environment etc [26]. Some ontologies extend these concepts in order to accommodate their data sources according to the target domains, such as shown in [20],[22], and [23].

Another part of the agri-food domain that recently attracted the attention of researchers is related to food traceability. It is driven by a growing interest in developing systems for the food supply chain. Some essential works in this domain have been carried out, such as The Food Track and Trace Ontology (FTTO) [27] and TraceALL [28]. Even though the FTTO is dedicated to food traceability management, this ontology provides representative food concepts that also become valuable resources for food processing experiments. This ontology includes four core concepts: actor, food product, process and service product.

The ontologies mentioned above have proven useful in their own domain. However those ontologies mostly cover one sub-domain, for example only agricultural practices or only food processing. The question of data integration across complementary sub-domains often has not been discussed yet in the Agri-food sector.

## 3. The global Agri-food experiment ontology design approach

In this section, we explain how we built an ontology network (AFEO), starting from two complementary existing ontologies (AEO and OFPE), in order to integrate heterogeneous data sources, both from field experiments and from food processing processes. To design AFEO ontology, we have followed the NeOn methodology [6]. NeOn provides methodological support for constructing an ontology network (AFEO in our case), when different pre-existing resources must be re-engineered in order to be reused for a new purpose. NeOn permitted us to design iteratively the AFEO ontology network based on some of the nine scenarios proposed in the methodology which cover commonly occurring situations. The section is divided into three parts. The first part defines the requirements associated with the desired result, the AFEO ontology network. The second part presents the scope of both selected ontology resources (AEO and OFPE) as inputs of the ontology integration process. The third part describes the ontological transformations used to integrate AEO and OFPE ontologies into AFEO in order to fulfill the requirements. Main OWL 2 (Web Ontology Language, second edition) ontological constructs used in this paper are presented in [29].

### *3.1. AFEO ontology requirements*

Following NeOn scenario 1 (From specifications to implementation), we established the ontology requirements (called OR in the rest of the paper). AFEO must

- model a global food chain linking the biomass production process to the biomass transformation process (OR1),

- represent experimental observations all along the integrated process (OR2),
- be able to be specialized for a given food chain which is, in this paper, viticulture and winemaking (OR3).

### *3.2. Description of the initial ontologies*

Following NeOn scenario 3 (Reusing ontological resources), AEO and OFPE ontological resources must be inspected in order to determine if they are good candidates to fulfill the AFEO requirements and to analyze their granularity. AEO and OFPE ontologies were built separately for different intended communities. AEO serves as a representation of expert knowledge in the field of agricultural practices while OFPE was developed to represent the knowledge related to food processing experiments. It is the reason why they have been selected as their union should satisfy OR1 requirement presented in section 3.1. The present section describes both ontologies, in terms of their concepts and relations, in order to determine if some knowledge is missing to fulfill AFEO requirements.

#### *3.2.1. Agricultural Experiments Ontology (AEO)*

Table 1: Some important object properties of AEO

| Property name | Inverse | Descriptions |
|---|---|---|
| includes | included | relation among cultivation locations (FarmArea, CultivationArea and its subclasses) |
| applies | appliedTo | relation between classes under AgriActivity and some classes under ObservedAgriEntity. For instance, an instance of Fertilizing is applied to an instance of Plot or SubPlot |
| hasPart | partOf | relation between Plant class and Organ class |
| hasLocation | locationOf | relation between AgriExperiment and CultivationArea |
| involves | involvedIn | relation between AgriExperiment and AgriActivity as well as ObservedAgriEntity. |

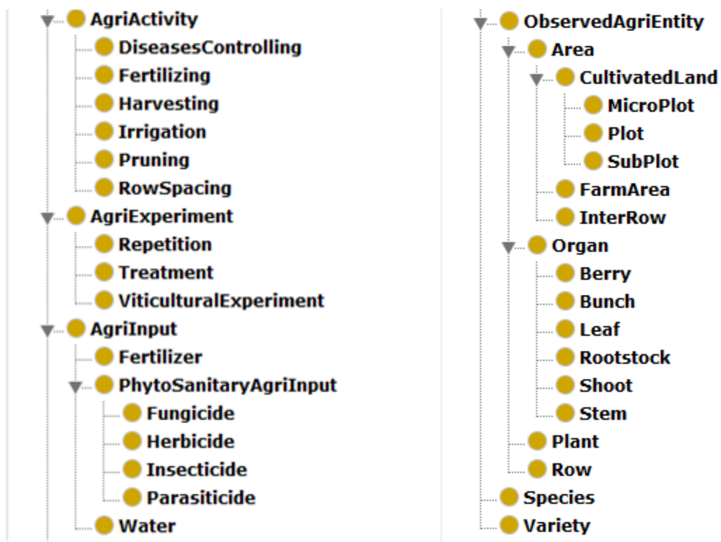


Figure 1: Concept hierarchies in AEO.

AEO is an ontology aimed to represent objects related to agricultural practices. It is based on three main concepts, *AgriExperiment*, *ObservedAgriEntity*, and *AgriActivity* (Figure 1). *AgriExperiment* is a concept that represents experiments carried out by researchers. *ObservedAgriEntity* represents objects that are observed and followed during field experiments while the *AgriActivity* represents common activities or operations performed during experiments. Some important semantic relations have been defined between those concepts as shown in Table 1. Object property *includes* plays an important role in this ontology. It is used to represent the relation between spatial locations, e.g., *FarmArea* a concept that represents the area of agricultural land used for farming. This type of relation also applies to *Plot*, *Sub-plot* and *Micro-plot* which are subclasses of *CultivatedArea*. Each plot may be composed of several subplots and each subplot has two or more micro-plots. By defining *includes* object property as a transitive object property, we can easily navigate between locations. This object property also represents the multi-scale aspect of locations that is useful for tracking the origin of a particular product.

### 3.2.2. Ontology for Food Processing Experiments (OFPE)

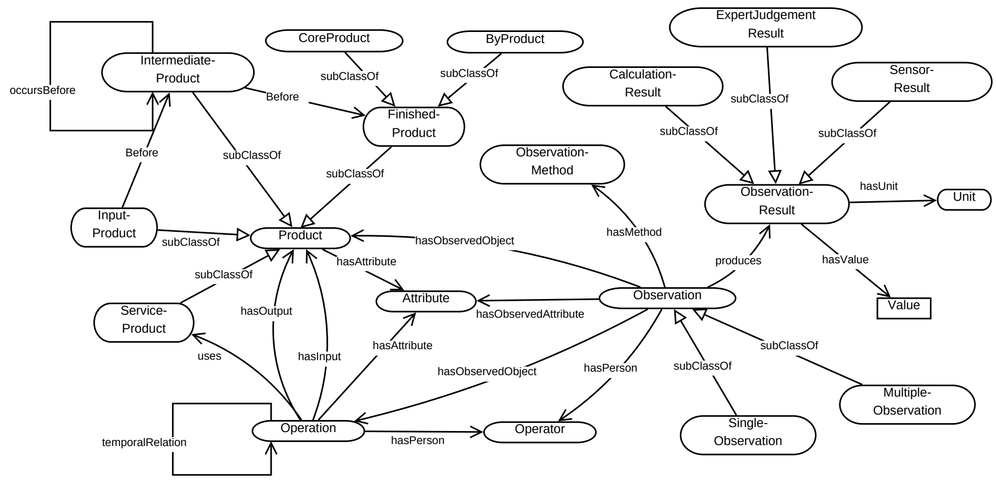


Figure 2: Core concepts and relations in OFPE.

OFPE (called Onto-FP in [30]) is an ontology dedicated to food processing experiments, where raw materials are transformed into final products. It includes different classes that represent products and activities during food transformation processes, which can be classified into four main concepts, i.e., *Product*, *Operation*, *Attribute*, and *Observation*. The core concepts and relations of OFPE are presented in Figure 2.

The *Product* concept represents different types of material existing during the food transformations. As shown in Figure 2, four subclasses under this class are *InputProduct*, *IntermediateProduct*, *FinishedProduct*, and *ServiceProduct*. The first three subclasses are representation of product stages which occur during a set of transformation processes. Starting with *InputProduct*, representing any product harvested from farm that act as raw material, followed by *IntermediateProduct* that models semi or unfinished products during process flow, and terminated by *FinishedProduct* that represents end products or final

Table 2: Temporal relations between operations

| Relations | Inverse | Examples from winemaking |
|---|---|---|
| Before*(A,B)* | After*(B,A)* | Crushing is performed after destemming |
| Meets*(A,B)* | MetBy*(B,A)* | Draining is started after maceration |
| Overlaps*(A,B)* | OverlappedBy*(B,A)* | Malolactic fermentation can be started before alcoholic fermentation is finished |
| Starts*(A,B)* | StartedBy*(B,A)* | Alcoholic fermentation starts when Maceration is started |
| Finishes*(A,B)* | FinishedBy*(B,A)* | Sulfitation finished malolactic fermentation |
| During*(A,B)* | Contains*(B,A)* | Alcoholic fermentation occurs during maceration |
| Equal*(A,B)* | | Extraction of ethanol is started and finished at the same time as alcoholic fermentation |

outputs of food transformation process. The *ServiceProduct* represents all materials used by operations to transform particular product during process flow. The *Operation* class conceptualizes the knowledge related to the transformation activities. The concept of *Attribute* represents all features or qualities belonging to Product or Operation classes.

*Observation* is an abstract model of activity where an instance of Attribute is measured. This class has an important role in representing measurement values of attributes collected from different experiments. Observation can be a single observation (*SingleObservation*) that means one time only measurement or a set of observation (*MultipleObservation*) where multiple measurements of a particular attribute (of product or operation) are needed. An instance of observation links to an instance of product or operation and to one or more instances of attribute through object property *hasObservedObject* and *hasObservedAttribute.* This observation instance also has method which is represented by *ObservationMethod* class and produces result depicted by *ObservationResult* that holds the values of measurement. In food processing the observation results come in various forms, such as values directly recorded by sensors, values from expert judgement or values from a particular calculation. To accommodate those forms, this ontology provides three subclasses under *ObservationResult*, including *SensorResult*, *ExpertJudgementResult*, and *CalculationResult* (see Figure 2).

Relations among main classes in this ontology can be grouped into two parts, i.e., product transformation and temporal aspects of operations. Product transformation can be described either by using a transitive object property *Before* (or its inverse *After*) or object property *hasInput* and *hasOutput*. The *Before* links between different products, in particular between classes *InputProduct*, *IntermediateProduct*, and *FinishedProduct* (see Figure 2). To describe temporal aspects of operations, this ontology uses the concepts defined in the Ontology of Time proposed in [31]. Examples of use of these temporal relations are presented in Table 2.

### *3.3. AEO and OFPE re-engineering for integration*

After inspection, the two selected ontology resources presented in Section 3.2.1 and Section 3.2.2 have been considered valid by domain experts for their respective intended domains, i.e. biomass production and food processing. However, when both are directly merged, the result is not entirely suitable to support the objective of this work which focuses on agri-food experimental data integration. Indeed, OR1 requirement is not completely fulfilled because the linkage between production and transformation is not done. Moreover, AEO does not permit to represent experimental observations, which is required in OR2. Last, specialization to vineyard (resp. winemaking) (OR3) is not present in AEO (resp. OFPE) ontology.

Therefore, we used scenario 6 (Reusing, Merging and Re-engineering Ontological Resources) of the NeOn methodology. Firstly, AEO and OFPE ontologies have been linked. Secondly, they have been re-engineered to generate AFEO, a new ontological network.

**OR1 fulfillment** AEO and OFPE cover two different scientific disciplines which are closely related. AEO covers the ontology terms related to entities and activities occurring during field agricultural experiments, starting at planting and ending at harvesting. OFPE contains various products and operations as well as their attributes. It can be used to build semantic relations representing a set of food processing experiments where several products and operations are interrelated in a specific order. Normally, a particular set is started by receiving raw materials as an input (represented by *InputProduct* class in OFPE) and transformed by some operations to become a processed food. Most raw materials in food processing are agricultural products that come from the field. They are the result of harvesting activity (represented by the *Harvesting* class in AEO). Based on that description, the object property *hasOutput*, part of OFPE, can be used to link between *Harvesting* class part of the AEO, and *InputProduct* part of OFPE. More formally, the domain of the object property *hasOutput*, which is the *Operation* class in OFPE is enlarged to the alternative *Harvesting* class of AEO. *Berry* and *Bunch*, which correspond to the *InputProduct*, are considered as equivalent classes in AEO and OFPE (see Table 3). Moreover, the object property *hasOutput* is defined with the cardinality one to many, because in some cases one or more instances of *InputProduct* come from the same instance of *Harvesting*. As *AgriActivity* and *Operation* can be considered as equivalent classes in AEO and OFPE (see in Table 3), *Before* or more generally all temporal relations presented in table 2, part of OFPE, can be used to link between *AgriActivity* class, part of the AEO, and *Operation*, part of OFPE. More formally, the domain of the object property *Before* or more generally all temporal relations presented in table 2, which is the *Operation* class in OFPE is enlarged to the alternative *AgriActivity* class of AEO. By using these formal extended semantic relations, the ontology linking between AEO and OFPE can be built and OR1 requirement is now fulfilled. Figure 4 shows with an example how these extended semantic relations are applied.

Table 3: Class mapping between AEO and OFPE.

| AEO class | OFPE class | Mapping |
|---|---|---|
| Berry | Berry | owl:equivalentClass |
| Bunch | Bunch | owl:equivalentClass |
| AgriActivity | Operation | owl:equivalentClass |

**OR2 fulfillment** The next step after linking AEO and OFPE is re-engineering this new ontology network. Indeed, as presented in Section 3.2.1, AEO does not provide important concepts and properties with regard to observations, measurement data representation, as well as observed attributes. Contrary, the OFPE contains a specific part that is dedicated to accommodate observation activities. In order to fulfill OR2 requirement, we re-engineered the new ontology network by the following steps which implement scenario 8 (Restructuring ontological resources). Firstly, we re-used the semantic relation *hasAttribute* (existing in OFPE) to link the *ObservedAgriEntity* and *AgriActivity* classes in AEO to the *Attribute* class. Formally, the domain of the object property *hasAttribute*, which is the *Product* class in OFPE is enlarged to the *ObservedAgriEntity* and *AgriActivity* classes of AEO. Secondly, we re-used semantic relations regarding the observation which exist in OFPE. In OFPE, the *Observation* class links to two classes, i.e. *Product* and *Operation* by using the semantic relation *hasObservedObject* (see Figure 2). The *ObservedAgriEntity* and *AgriActivity* classes in AEO have a similar role to *Product* and *Operation* which represent objects of observation in OFPE. Thus, formally, we enlarged the range of the *hasObservedObject* object property to the *ObservedAgriEntity* and *AgriActivity* in AEO.

**OR3 fulfillment** As stated in OR3, AFEO has to be applied to a specific domain, in this paper, viticulture and winemaking experimentation. However, from the inspection of Section 3.2.1 and Section 3.2.2, it can be seen that both ontology resources (AEO and OFPE) do not include any specific terms of viticulture and winemaking experiments. Therefore, AFEO needs to be restructured. In this case, the scenarios 2 (Reusing and re-engineering non-ontological resources) and 8 (Restructuring ontological resources) of the NeOn methodology were performed to get a new version of AFEO. The list of specific terms related to viticulture and winemaking was provided by domain experts. These terms were formalized as new concepts to specialize AFEO (Ontology specialization activity of scenario 8). AFEO preserves all the classes from AEO (see Figure 1) which provides generic classes that mostly also exist in viticulture. Some specific classes have been added, such as CanopyManagement (sub-class of AgriActivity), GreenHarvesting (sub-class of Pruning), Vineyard (sub-class of CultivatedArea), GrapeVariety (sub-class of Genotype) and Grapevine (sub-class of Plant). Figure 3 shows the new required specific classes under the *Product* and the *Operation* classes for winemaking.

To illustrate the final structure of AFEO[3], we present a part of the whole ontology in Figure 4. This diagram contains different viticulture entity and activity classes, winemaking specific classes of product and operation, and their semantic relations.

Another interesting point illustrated in Figure 4 refers to the practical application of complex temporal relations between operations. For example the *Crushing* is performed before the *AlcoholicFermentation*, the *PumpingOver* has to be done during the *Maceration*, etc.. By using these semantic relations, a particular sequence or itinerary of operations (experiments) can be investigated. The ability of finding a specific sequence of products and operations as well as linking to agricultural activities and entities is a strong point of AFEO ontology

[3]http://agroportal.lirmm.fr/ontologies/AFEO

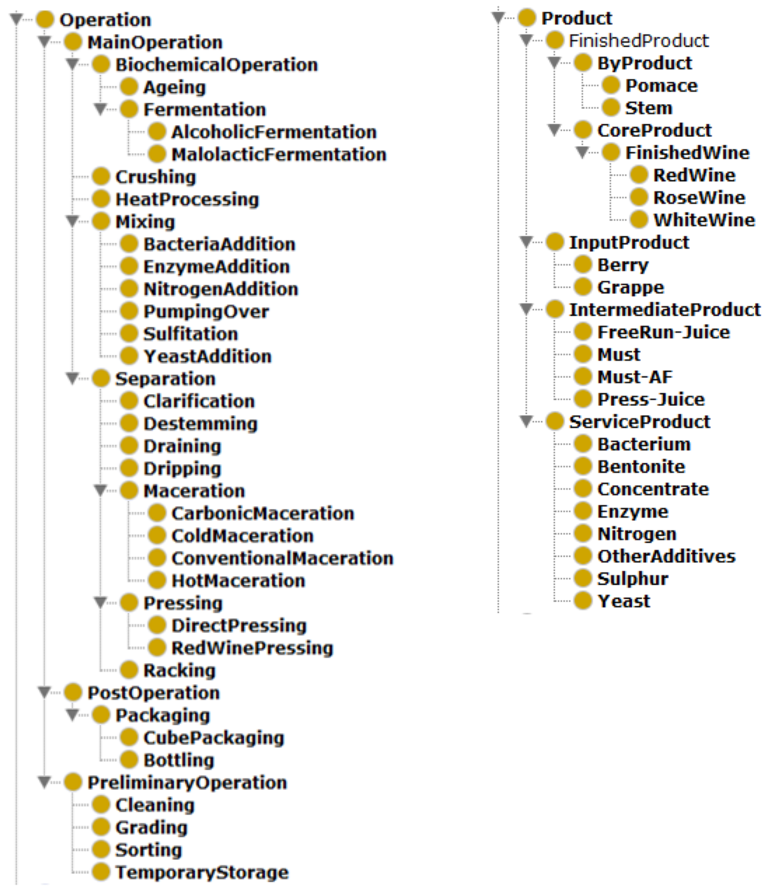


Figure 3: Operation and product specialization of the AFEO for winemaking experiment.

network because it gives a reliable way to build a connection between observational data belonging to interrelated experiments. AFEO [4] (resp. AEO [5] and OFPE [6]) ontology is available on AgroPortal, derived from BioPortal [32] for agronomical ontologies.

[4]http://agroportal.lirmm.fr/ontologies/AFEO
[5]http://agroportal.lirmm.fr/ontologies/AEO
[6]http://agroportal.lirmm.fr/ontologies/OFPE

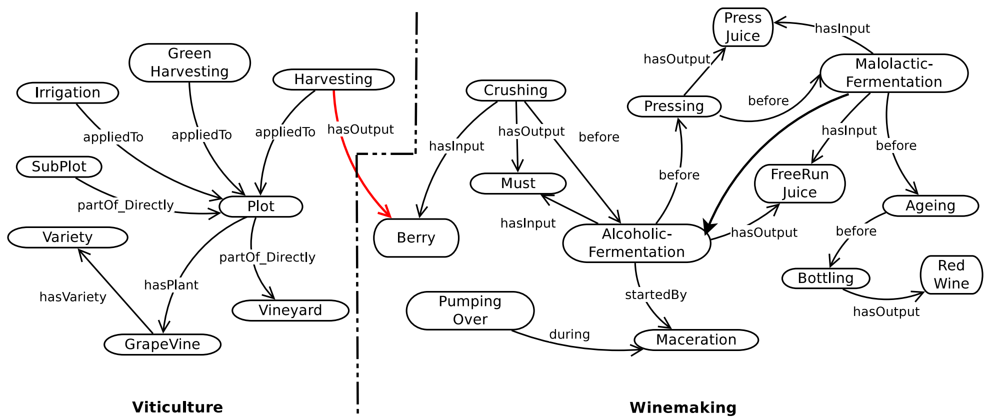


Figure 4: A part of the AFEO that represents viticulture and winemaking integration.

## 4. Ontology population with viticulture and winemaking experimental data

In a domain specific ontology development stage, an ontology instantiation (population) is an important step. The main task is to semantically annotate all related resources based on the proposed ontology. AFEO, result from Section 3, can be seen as an agri-food experiment knowledge layer that acts as a container for viticulture and winemaking experimental data integration. From a practical point of view, this ontology network needs to be instantiated with actual experimental data so that specific data queries can be performed to support further statistical analysis. This section presents how AFEO has been instantiated in order to integrate the experimental data from various stages of agri-food experiments. The first part describes two different data sources used in this work and the second part explains the principles of the ontology population step.

### *4.1. Data sources*

Data sources used for this work can be divided into two groups, i.e., viticulture and winemaking experimental data. Both are temporal data, however they are different in terms of structure and storage formats.

The viticulture experimental data were collected during several years of viticultural experiments from different vineyards in the south of France. These data contain information about vineyards, plots, sub-plots, grapevines, different treatments of experiment (for example irrigation, non irrigation, green harvesting, non-green harvesting, etc.), and a huge number of observations regarding the grape characteristics during the growing season. They are stored in a single relational database linked to *Silex-VitiOeno Pilotype*, a web application that manages viticulture experimental data.

The winemaking experimental data contain information about different stages of the winemaking process, different wine products as well as a big amount of observational data about operations and products characteristics during process flow. These data were collected annually from the *Unité Expérimentale de Pech Rouge*, France, where the winemaking experiments were conducted. Unlike the viticulture experimental data, these data are recorded in the form of Microsoft Excel files. To the current date, more than 500 Microsoft Excel files have been collected. These files can be classified into two groups: i.e,

- a group that represents different operations during the winemaking process. Each file consists of a single sheet that contains a table of operations used during a set of winemaking experiments, started from harvesting to bottling. The table also contains the attributes of each operation;
- a group that records observational data concerning the different product features. Each file consists of three sheets, where each sheet represents observational results of some attributes for a particular product stage, such as grape, must, must after alcoholic fermentation, and finished wine.

### *4.2. Ontology instantiation process*

Both data sources have to be transformed into uniform format based on AFEO ontology network. Consequently, the Resource Description Framework

(RDF) format has been chosen to represent these data. The RDF is a standard model for data interchange and metadata processing [33]. It has features that facilitate data linking even if the underlying schema differ. The RDF can be seen as a graph data model that uses URIs to name two nodes as well as the relationship between them. Due to the different source data formats, we developed two specific scripts using the Python language as illustrated in Figure 5. The first script creates instances by performing some queries to the viticulture database and transforming them to RDF format. The second script is specialized for making instances from the Microsoft Excel winemaking experimental data files. Both of them were developed guided by the AFEO ontology network. RDF data are stored in a triplestore, a specific database for storing and retrieving RDF triples through semantic queries expressed in SPARQL.

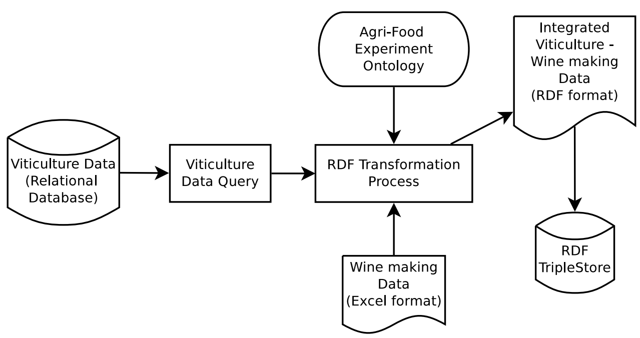


Figure 5: Data population processing steps.

## 5. Results and discussion

The RDF data produced in Section 4.2 are viticulture-winemaking integrated data that contain objects and activities/operations along with observational results from various experiments. These data can be used in many ways by querying them according to the user requirements. This section presents two examples on how this integrated data can help viticulture and winemaking researchers to find some related data prior a statistical analysis. The first part describes a practical use related to wine traceability and the second part explains how to extract and analyze interrelated observational data between viticulture practices and wine qualities.

### 5.1. Wine traceability

Wine is a food product whose value is generally determined by three groups of characteristics, i.e., appearance, flavour, and aroma. In the recent years, wine consumers are more and more oriented to consider detailed information about those characteristics as well as the overall process from the grape to the bottle [34]. The information is used to confirm the wine identity (authenticity) and to ensure it is produced transparently in order to avoid fraud and toxicological issues [35]. The viticulture-winemaking integrated data generated from Section 4.2 can be queried in order to get such traceability information, giving valuable knowledge about the entire process from viticulture practices through selected winemaking operations, until finished product delivery.

By using the RDF integrated data, specific queries allow to trace products or operations during various experiments. It can be done backwardly, from finished wine as a final product to the plot or vineyard where the grapevines are planted, or forwardly. The following lines is an example of SPARQL query to trace the origin of wine products.

**Query 1**.

```
SELECT ?w ?t ?l ?nv ?na
WHERE
{
 ?o ofpe:After ?h .
 ?w rdf:type ofpe:FinishedWine .
```

```
 ?o ofpe:hasOutput ?w
 ?h aeo:appliedTo ?l
 ?l aeo:included ?p
 ?p rdf:type aeo:Plot
 ?p aeo:hasVariety ?v
 ?v rdfs:label ?nv
 ?l aeo:included ?a
 ?a rdf:type aeo:Vineyard
 ?a rdfs:label ?na
}
```

| W (FinishedWine) | T (BerryGrape) | L (SubPlot) | NV (GrapeVariety) | NA (Vineyard) |
|---|---|---|---|---|
| wt_12001 | g_12001 | PR-Cha-i1 | Chardonnay | Pech Rouge |
| wn_12001 | g_12001 | PR-Cha-i1 | Chardonnay | Pech Rouge |
| wn_12002 | g_12002 | PR-Cha-i0 | Chardonnay | Pech Rouge |
| wt_12002 | g_12002 | PR-Cha-i0 | Chardonnay | Pech Rouge |
| wt_12003 | g_12003 | RIE-Gre-i0 | Grenache | Rieux |
| wn_12003 | g_12003 | RIE-Gre-i0 | Grenache | Rieux |
| wn_12021 | g_12021 | StSAU-Cha-i1 | Chardonnay | Saint Sauveur |
| wt_12021 | g_12021 | StSAU-Cha-i1 | Chardonnay | Saint Sauveur |
| wn_12020 | g_12020 | StSAU-Cha-i0 | Chardonnay | Saint Sauveur |

Figure 6: List of finished wine instances and their origin

Figure 6 shows a screen shot of the query result which lists instances of finished wine, their input products, sub-plots and vineyards where the grapes are planted, and their grape varieties. In Query 1, the *After* object property is used in order to trace all existing products during winemaking process. By using the *After*, which is defined as a transitive property, this SPARQL query performs an inference to determine link between *BerryGrape* instance ($t$) as an input product and *FinishedWine* instance ($w$) as the final product, whatever the number of intermediate products and associated operations. Figure 6 highlights some information about wine traceability that might help researchers to know global information about the origin of finished wine products. For instance, we can see that, from the same berry grape lot, can be produced two different finished wine products: wt_12001 and wn_12001 are finished wine instances that have the same input product, i.e., g_12001. This indicates that probably there were different treatments during the winemaking process, that may affect the characteristics of the final product.

### *5.2. Influence of viticulture practices on grape and wine qualities*

As said above, wine quality can be analysed by three groups of characteristics. Each group can be further detailed by various attributes. For example, the appearance is determined by the color attribute along with other visual features such as clarity and fluidity, while the wine flavour has some basic attributes, such as sweetness, acidity, bitterness, fruit flavour, etc. These attributes are influenced by various interrelated factors, both viticulture practices (e.g., grape variety, location, vintage, etc.) and winemaking selected operations (e.g., crushing method, maceration, pressing, yeasts added, etc.).

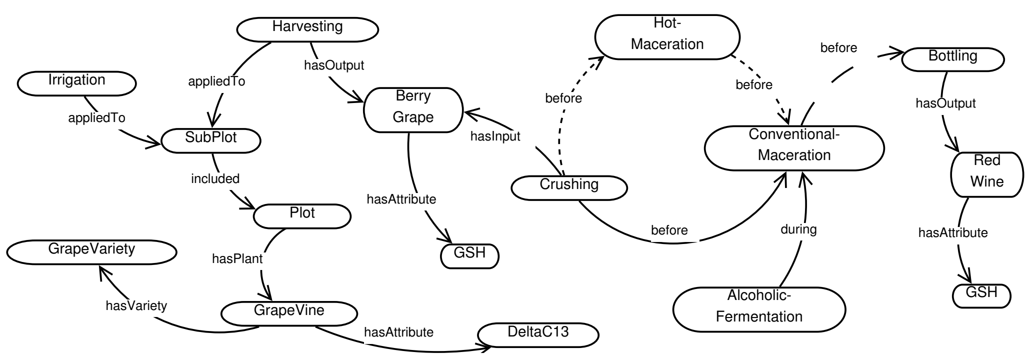


Figure 7: Semantic relations linking GSH concentration in grape and red wine.

To illustrate the interest of viticulture-winemaking integrated data, we now present one set of experiments that has been conducted in the *Pilotype* project, funded by OSEO innovation and the Languedoc-Roussillon regional council. This project includes multi-site experiments located in the south of France. The same experimental design was set up in seven sites across the Languedoc Roussillon region in order to test for the effects of vine stress status and winemaking protocol on grape potential and wine quality in contrasted environmental conditions. We discuss here a subset of these experiments. Three grape varieties are being studied, i.e. Merlot, Syrah and Cabernet Sauvignon; and two different winemaking methods, i.e., with an innovative hot pre-fermentation maceration and without it (classical method). In the experimental plots, the wine water status has been monitored by different means: on-line sap flow sensors, predawn leaf water potential and DeltaC13 at harvest time. The last one, which is a global indicator of the water stress experienced by the vine, will be used in the following. Many quality parameters resulting from physico-chemical analyses are available to assess grape composition at harvest time, must composition and finished wine composition. Among them, let us take the example of the glutathione (GSH) concentrations in wine and berry grape. GSH, a natural tripeptide found in grapes and wine, is a powerful antioxidant that protects wines from oxidation and loss of aroma or flavor. Glutathione levels fluctuate during production, as the compound can be absorbed by yeast and then released after fermentation. If final GSH levels are low in wines, they will experience faster loss of fresh varietal and fruity aromas, and poor ageing potential. By using the AFEO, this experiment can be represented semantically as visualized in Figure 7.

In order to perform a statistical analysis, some data sets need to be queried from the viticulture-winemaking integrated data, including all the *RedWine* instances and their corresponding *BerryGrape* instances along with their GSH values, all the *Subplot* and *Plot* instances as the wines origin, all the corresponding GrapeVariety instances, and all itinerary types which are used to identify what winemaking methods are used. Using the SPARQL query statements, data can be extracted from viticulture-winemaking integrated data. Based on the available data which cover two years experiments (2012 and 2013), the query provides the results as follow: there are a total of 40 instances of *RedWine* which are produced from 20 instances of the *BerryGrape*, 20 instances of the *Subplot* and 9 instances of the *Plot* as the origin of the *BerryGrape* instances. The *RedWine* instances are divided equally into two groups, 20 instances from the hot pre-fermentation maceration methods and the rest come from the classical method.

The statistical study aims to determine if the winemaking protocol has a significant influence on the GSH wine concentration. As data do not follow a Gaussian distribution (as confirmed by a Shapiro test), the Wilcoxon test on paired samples is run. It highlights a significant (p-value $<0.05$) difference in GSH concentrations depending on the winemaking protocol: The hot pre-fermentation maceration influences the wine GSH concentration and increases it.

A graphical analysis is then done to visually explore, by taking advantage of the links between viticulture and wine data, if the vine water status could have an impact on that influence. Figure 8 shows, for each of the three varieties, on the abscissa the grape GSH concentration and on the ordinate the corresponding wine GSH concentration. Circles correspond to the year 2012, and squares to the year 2013. The classical winemaking protocol is plotted in red, and the hot pre-fermentation maceration one in blue. The size of the symbols is proportional to the Delta C13 indicator: the higher Delta C13, the bigger the symbol and the less the vine was exposed to water stress. Plots visually confirm the increase of GSH in wine, for a given GSH in grapes, with the hot pre-maceration protocol, but they do not show any influence of the vine water status.

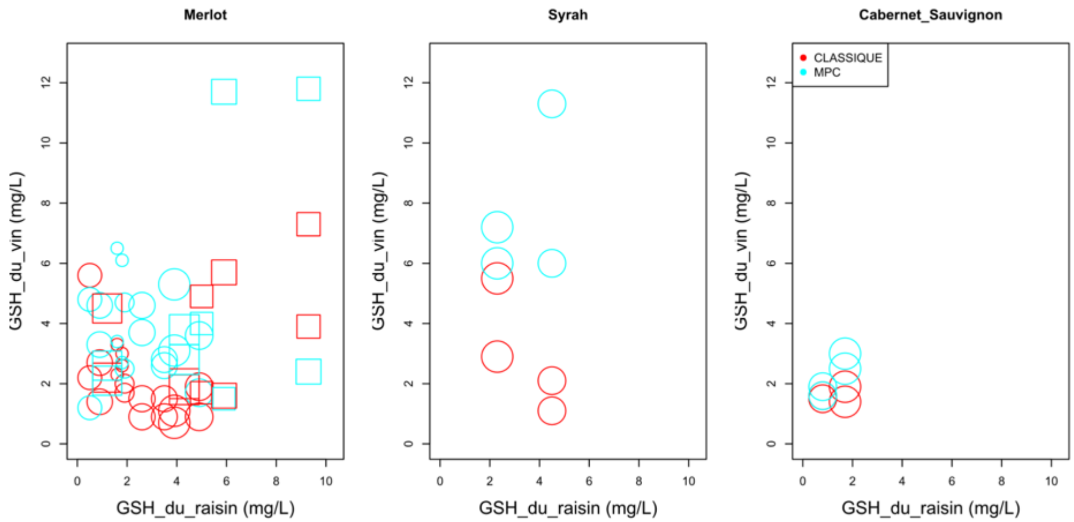


Figure 8: Plots of GSH concentration in red wine versus GSH concentration in grape.

## 6. Conclusion

This paper presents an ontology-based approach for experimental scientific data integration across complementary sub-domains, i.e., agricultural practices and food processing, with an application to viticulture and winemaking process. The two main steps in this approach are i) to develop an ontology network and ii) to populate the ontology with actual experimental data from different sources. The ontology network -Agri-Food Experiment Ontology (AFEO)- was developed based on two existing ontology resources, i.e., AEO (Ontology for Agriculture Experiment) and OFPE (Ontology for Food Processing Experiment). It contains 136 concepts which covers various viticulture practices, and winemaking products and operations. AFEO was used to guide data integration from two different data sources, i.e., viticulture experimental data which are stored as relational database and winemaking experimental data which are stored in Microsoft Excel files.

Results show the potential uses of the AFEO along with viticulture-winemaking integrated data to provide linked data to be used by researchers. Two practical uses are presented, i.e. wine traceability and the influence of grape varieties and different winemaking methods on GSH concentration. They give an idea of how data integration can support extended research questions which require data from different parts the Agri-food chain.

The development steps presented in this paper can also be viewed as an effort to decrease the gap between scientific disciplines, by allowing researchers from different disciplines to formulate their knowledge in a formal ontological form, in order to share it and to build more comprehensive and collaborative scientific researches.

AFEO focuses on viticulture and winemaking objects and experimental activities. However, as indicated in Section 5 the weather and climate has significant influence on grape quality. Therefore in the future, aligning AFEO with an ontology that represents knowledge from the meteorological domain will allow a better analysis of viticulture and winemaking experimental results. Another interesting perspective is to specialize AFEO to other food products that require linkage from agriculture practices to food transformation process.

## References

## References

[1] N. Perrot, H. De Vries, E. Lutton, H. G. J. van Mil, M. Donner, A. Tonda, S. Martin, I. Alvarez, P. Bourgine, E. van der Linden, M. Axelos, Some remarks on computational approaches towards sustainable complex agri-food systems, Trends in Food Science & Technology 48 (2016) 88–101. URL http://www.sciencedirect.com/science/article/pii/S0924224415002186

[2] J. Wolfert, C. Verdouw, C. Verloop, A. Beulens, Organizing information integration in agri-foodA method based on a service-oriented architecture and living lab approach, Computers and Electronics in Agriculture 70 (2) (2010) 389–405.

[3] S. Bowers, Scientific Workflow, Provenance, and Data Modeling Challenges and Approaches, Journal on Data Semantics 1 (1) (2012) 19–30.

[4] S. P. Gardner, Ontologies and semantic data integration, Drug Discovery Today 10 (14) (2005) 1001–1007.

[5] D. P. K. (Seedah, B. Sankaran, W. J. O'Brien, Approach to classifying freight data elements across multiple data sources, TRANSPORTATION RESEARCH RECORD 2529 (18) (2015) 56–65.

[6] M. C. Suárez-Figueroa, A. Gómez-Pérez, M. Fernández-López, The NeOn Methodology for Ontology Engineering, in: M. C. Suárez-Figueroa, A. Gómez-Pérez, E. Motta, A. Gangemi (Eds.), Ontology Engineering in a Networked World, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 9–34.

[7] J. Madin, S. Bowers, M. Schildhauer, S. Krivov, D. Pennington, F. Villa, An ontology for describing and synthesizing ecological observation data, Ecological Informatics 2 (3) (2007) 279–296.

[8] M. P. S. Shawn Bowers, Joshua S. Madin, A Conceptual Modeling Framework for Expressing Observational Data Semantics, in: Q. Li, S. Spaccapietra, E. Yu, A. Olivé (Eds.), Conceptual Modeling - ER 2008, Vol. 5231 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 41–54.

[9] E. K. Nelson, B. Piehler, J. Eckels, A. Rauch, M. Bellew, P. Hussey, S. Ramsay, C. Nathe, K. Lum, K. Krouse, D. Stearns, B. Connolly, T. Skillman, M. Igra, LabKey Server: an open source platform for scientific data integration, analysis and collaboration., BMC bioinformatics 12 (2011) 71.

[10] F. Wang, P.-E. Bourgué, G. Hackenberg, M. Wang, D. Kaltschmidt, P. Liu, SciPort: an adaptable scientific data integration platform for collaborative scientific research, Very Large Data Bases (2007) 1310–1313.

[11] B. Luduscher, K. Lin, S. Bowers, E. Jaeger-Frank, B. Brodaric, C. Baru, Managing scientific data: From data integration to scientific workflows*, Geological Society of America Special Papers 397 (2006) 109–129.

[12] S. Leonelli, Integrating data to acquire new knowledge: Three modes of integration in plant science., Studies in history and philosophy of biological and biomedical sciences 44 (4) (2013) 503–514.

[13] Y.-F. F. Li, G. Kennedy, F. Ngoran, P. Wu, J. Hunter, An ontology-centric architecture for extensible scientific data management systems, in: Future Generation Computer Systems, Vol. 29, 2013, pp. 641–653.

[14] P. Fox, D. L. McGuinness, L. Cinquini, P. West, J. Garcia, J. L. Benedict, D. Middleton, Ontology-supported scientific data frameworks: The Virtual Solar-Terrestrial Observatory experience, Computers and Geosciences 35 (4) (2009) 724–738.

[15] C. (Daraio, M. Lenzerini, C. Leporelli, P. Naggar, A. Bonaccorsi, A. Bartolucci, The advantages of an ontology-based data management approach: openness, interoperability and data quality, SCIENTOMETRICS 108 (1) (2016) 441–455.

[16] M. (Imran, R. I. M. Young, Reference ontologies for interoperability across

multiple assembly systems, INTERNATIONAL JOURNAL OF PRODUCTION RESEARCH 54 (18) (2016) 5381–5403.

[17] N. Arch-int, S. Arch-int, Semantic ontology mapping for interoperability of learning resource systems using a rule-based reasoning approach, EXPERT SYSTEMS WITH APPLICATIONS 40 (18) (2013) 7428–7443.

[18] C. Lousteau-Cazalet, B. Barakat, J. P. Belaud, P. Buche, G. Busset, B. Charnomordic, S. Dervaux, S. Destercke, J. Dibie, C. Sablayrolles, C. Vialle, A decision support system for eco-efficient biorefinery process comparison using a semantic approach, Computers and Electronics in Agriculture 127 (2016) 351–367.

[19] L. N. Soldatova, R. D. King, An ontology of scientific experiments., Journal of the Royal Society, Interface / the Royal Society 3 (11) (2006) 795–803.

[20] Y. Wang, Y. Wang, J. Wang, Y. Yuan, Z. Zhang, An ontology-based approach to integration of hilly citrus production knowledge, Computers and Electronics in Agriculture 113 (2015) 24–43.

[21] G. Song, M. Wang, X. Ying, R. Yang, B. Zhang, Study on Precision Agriculture Knowledge Presentation with Ontology, AASRI Procedia 3 (2012) 732–738.

[22] H. W. Beck, S. Kim, D. Hagan, A Crop-Pest Ontology for Extension Publications, in: EVITA 2005 Proceedings, no. July, Vila Real, 2005, pp. 1169–1176.

[23] A. J. Haverkort, J. L. Top, F. Verdenius, Organizing Data in Arable Farming: Towards an Ontology of Processing Potato, Potato Research 49 (3) (2007) 177–201.

[24] C. Snae, M. Bruckner, FOODS: A Food-Oriented Ontology-Driven System, in: 2nd IEEE International Conference on Digital Ecosystems and Technologies, Ieee, 2008, pp. 168–176.

[25] J. Graça, M. Mourao, O. Anunciação, P. Monteiro, H. S. Pinto, V. Loureiro, Ontology building process: the wine domain, in: EVITA 2005 Proceedings, no. i, Vila Real, 2005.

[26] S. Rajbhandari, J. Keizer, The AGROVOC Concept Scheme A Walk-through, Journal of Integrative Agriculture 11 (5) (2012) 694–699.

[27] T. Pizzuti, G. Mirabelli, M. A. Sanz-Bobi, F. Goméz-Gonzaléz, Food Track & Trace ontology for helping the food traceability control, Journal of Food Engineering 120 (2014) 17–30.

[28] M. Salampasis, D. Tektonidis, E. P. Kalogianni, TraceALL: a semantic web framework for food traceability systems, Journal of Systems and Information Technology 14 (4) (2012) 302–317.

[29] W. O. W. Group, OWL 2 Web Ontology Language Document Overview (Second Edition) W3C Recommendation 11 December 2012 (2012).
URL `https://www.w3.org/TR/2012/REC-owl2-overview-20121211/`

[30] A.-R. Muljarto, J.-M. Salmon, P. Neveu, B. Charnomordic, P. Buche, Ontology-Based Model for Food Transformation Processes - Application to Winemaking, in: S. Closs, R. Studer, E. Garoufallou, M.-A. Sicilia (Eds.), Metadata and Semantics Research SE - 30, Vol. 478 of Communications in Computer and Information Science, Springer International Publishing, 2014, pp. 329–343.

[31] J. R. Hobbs, F. Pan, An ontology of time for the semantic web, ACM

Transactions on Asian Language Information Processing 3 (1) (2004) 66–85.

[32] N. Noy, N. Shah, P. Whetzel, B. Dai, M. Dorf, N. Griffith, C. Jonquet, D. Rubin, M. Storey, C. Chute, M. Musen, BioPortal: ontologies and integrated data resources at the click of a mouse, Nucleic Acids Res. (1).

[33] R. Cyganiak, D. Wood, M. Lanthaler, RDF 1.1 Concepts and Abstract Syntax (2014).
URL http://www.w3.org/TR/rdf11-concepts/

[34] M. G. Cimino, F. Marcelloni, Enabling Traceability in the Wine Supply Chain, in: G. Anastasi, E. Bellini, E. Di Nitto, C. Ghezzi, L. Tanca, E. Zimeo (Eds.), Methodologies and Technologies for Networked Enterprises, Springer Berlin Heidelberg, 2012, pp. 397–412.

[35] M. Palade, M.-e. Popa, Wine traceability and authenticity  A literature review, Series F. Biotechnologies XVIII (2014) 226–233.