

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

▶ To cite this version:

Aunur Rofiq Muljarto, Jean-Michel Salmon, Brigitte Charnomordic, Patrice Buche, Anne Tireau, et al.. A generic ontological network for Agri-food experiment integration – Application to viticulture and winemaking. Computers and Electronics in Agriculture, 2017, 140, pp.433-442. 10.1016/j.compag.2017.06.020. hal-01552148

HAL Id: hal-01552148 https://hal.science/hal-01552148v1

Submitted on 16 Jun2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

^aINRA, MISTEA Joint Research Unit, UMR729, F-34060 Montpellier, France ^bINRA, Unité Expérimentale de Pech Rouge, UE0999, F-11430 Gruissan, France ^cINRA, IATE Joint Research Unit, UMR1208, F-34060 Montpellier, France

^dDept. of Agroindustrial Technology, Brawijaya University, Malang 65145, Indonesia ^eINRIA 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-

Preprint submitted to Computer and Electronics in Agriculture

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

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 1. Introduction

Research in agri-food and related domains dealing with sustainability is 2 evolving more and more extensively in recent years to be more integrative, 3 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 6 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 8 viticulture treatments on wine quality, researchers need to access various data sources from a set of winemaking experiments and viticulture practices. Data 10 analysis could then provide better information and understanding of what actu-11 ally happened during a set of experiments and give the possibility of acquiring 12 new knowledge all along the agri-food chain. 13

However, to access and to incorporate various data sources, researchers have 14 to deal with some obstacles. Data are commonly stored in scattered places 15 that sometimes make it difficult to combine them. Moreover, data are very di-16 verse in terms of formats, naming, storage mechanisms, and query or retrieval 17 mechanisms. The heterogeneity of scientific data may come from many factors, 18 such as (i) most scientific data are collected distinctively based on independent 19 research projects; (ii) the data structures are frequently selected according to 20 the collection methods (e.g., to make data easier to record) or the format is 21

suggested by analysis tools, instead of standard data representations (e.g., re-22 lational database schema); and (iii) the terms and concepts used to label data 23 are not standardized, neither within nor across scientific disciplines and research 24 groups [3]. The difficulty of organizing available data and knowledge in a unified 25 way not only limits research productivity but also reduces data traceability [4]. 26 Research experiments are commonly divided into some sub-domains, such 27 as agriculture production, post-harvest, and food transformation process. Even 28 though the explicit relation between them is clearly understandable, each of 29 these has different objectives, scopes, and circumstances. For instance, the 30 agricultural experiments are normally performed in the fields, in which external 31 factors have a significant contribution while food processing experiments are 32 generally carried out in the laboratory with more controllable environments. 33 From a practical point of view, they require different methods for collecting 34 and organizing observational data that would lead to differences in the data 35 format or structure and in the way the data are stored. The heterogeneity 36 also occurs due to the vast scope of agri-food domains, ranging from plant 37 cultivation up to the final processed food product. Each discipline uses its own 38 knowledge expression, terms, concepts and semantic relations that might cause 39 some difficulties to share the observational data. 40

Studies in the last two decades have shown that ontologies represent a flexi-41 ble way to link the information contained in heterogeneous data sources within 42 or across domains [4, 5]. Ontologies also provide the common concepts for data 43 integration, thus opening the possibility to draw more comprehensive conclu-44 sions and to view data from different perspectives. Ontologies also allow certain 45 types of automated reasoning to be performed in order to fulfil some specific 46 requirements. The capabilities of data and knowledge sharing as well as rea-47 soning will help to develop more advanced Information Systems able to manage 48

heterogeneous data sources and to design platforms for more collaborative and 49 accurate scientific data analysis. 50

The contribution of this work is to provide a method to prepare and to inte-51 grate data sources prior to further analysis in order to answer complex questions 52 that require access to various agri-food scientific data sources. This method 53 works out in viticulture and winemaking to solve complex questions using het-54 erogeneous experimental data sources. To achieve this purpose, we developed 55 the Agri-Food Experiment Ontology (AFEO), a new ontology network resource, 56 based on two ontology resources, i.e., AEO¹ (Ontology for Agriculture Experi-57 ment), which is also an original contribution of this paper, and $OFPE^2$ (Ontol-58 ogy for Food Processing Experiment). The AEO and OFPE are ontologies that 59 have been developed separately in research laboratories as generic knowledge 60 representations of two respective sub-domains, i.e., agricultural production and 61 food transformation process. By following the NeOn [6] methodology, we inte-62 grated these two ontologies into an ontology network, in order to facilitate data 63 integration across these complementary sub-domains. An ontology network [6] 64 is a new ontology engineering concept, which allows ontology re-use and avoids 65 custom-building new ontologies from scratch. 66

Although in this paper the proposed ontology is specialized and tested for 67 viticulture and winemaking experiments, the core elements of AFEO are fairly 68 generic and might be adapted to other food products. Furthermore, the onto-69 logical definitions (concepts and relations) can be used to impose and preserve 70 a logical structure for new types of scientific data, that may appear due to new 71 sensors, protocols and analyses. 72

73

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

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

work related to scientific data integration and agri-food ontologies; Section 3 74 describes the global agri-food experiment ontology design process, introduces 75 the two ontology resources used, their integration and specialization to viticul-76 ture and winemaking experiments; Section 4 presents how the new proposed 77 ontology network is instantiated to integrate two different data sources: Section 78 5 is dedicated to discussion and a potential use in the domain of viticulture 79 and winemaking; Finally, in Section 6, conclusions are drawn and further work 80 outlined. 81

82 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.

⁸⁸ 2.1. Scientific data organization and integration

Over the recent years researchers have faced significant problems to manage 89 scientific data due to their increasing volume and complexities. This causes 90 researchers to spend quite a bit of time to manage and integrate scientific data 91 rather than to directly focus on their analysis [7] /* BRIGITTE: AJOUT REF 92 PHENOME */. Scientific data are generally collected from measurements di-93 rectly linked to real-world phenomena [8]. Within the agri-food domain, cross-94 disciplinary scientific data are required to explore complex and temporal aspects 95 of food quality and the impact of practices and operations. 96

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 ¹⁰⁰ such as proposed in [11], [7], [12], [3], [13] and [14]. Some of these approaches ¹⁰¹ are targeted to specific scientific domains while the others are developed to be ¹⁰² more generic and extensible. The non-ontological approach mostly relies on data ¹⁰³ models, such as database or XML schema, where attributes and relationships ¹⁰⁴ of domain concepts are captured in standardized structures. The ontological ¹⁰⁵ approach has some additional advantages in terms of data interoperability and ¹⁰⁶ knowledge reasoning [15, 16, 17, 18].

To the best of our knowledge, we did not find an ontological approach to 107 represent scientific experiments as well as observational data which fulfill all our 108 needs. An ontology of scientific experiments, the EXPO [19], has been proposed 109 to formalize the generic concepts of experimental design, methodology and result 110 representation. However, this ontology does not provide a clear explanation 111 about scientific data representation. Neither does this ontology describe how to 112 manage a set of experiments in which several interrelated experiments have to 113 be conducted in a given order. The Extensible Observation Ontology (OBOE) 114 has been developed to serve as a formal and generic conceptual framework for 115 describing the semantics of observational data sets (i.e., data sets consisting 116 of observations and measurements) [7]. The basic concepts of the observational 117 model consists of five classes and six properties [3]. This ontology can be applied 118 to various types of observations. Nevertheless, it is more suited to representing 119 scientific data which are generated from measurement by sensors. It does not 120 provide other types of observational data such as expert judgements or results 121 from a calculation procedure. 122

123 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 133 practices and food processing are important to model agri-food experiments. 134 Most of the above ontologies refer to the terms or concepts listed in AGROVOC. 135 AGROVOC is a multi-lingual vocabulary developed by the Food and Agricul-136 ture Organization (FAO) of the United Nations and covers divers areas includ-137 ing food, nutrition, agriculture, fisheries, forestry, environment etc [26]. Some 138 ontologies extend these concepts in order to accommodate their data sources 139 according to the target domains, such as shown in [20], [22], and [23]. 140

Another part of the agri-food domain that recently attracted the attention 141 of researchers is related to food traceability. It is driven by a growing interest 142 in developing systems for the food supply chain. Some essential works in this 143 domain have been carried out, such as The Food Track and Trace Ontology 144 (FTTO) [27] and TraceALL [28]. Even though the FTTO is dedicated to food 145 traceability management, this ontology provides representative food concepts 146 that also become valuable resources for food processing experiments. This on-147 tology includes four core concepts: actor, food product, process and service 148 product. 149

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 154 sector.

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

In this section, we explain how we built an ontology network (AFEO), start-156 ing from two complementary existing ontologies (AEO and OFPE), in order 157 to integrate heterogeneous data sources, both from field experiments and from 158 food processing processes. To design AFEO ontology, we have followed the 159 NeOn methodology [6]. NeOn provides methodological support for constructing 160 an ontology network (AFEO in our case), when different pre-existing resources 161 must be re-engineered in order to be reused for a new purpose. NeOn per-162 mitted us to design iteratively the AFEO ontology network based on some of 163 the nine scenarios proposed in the methodology which cover commonly occur-164 ring situations. The section is divided into three parts. The first part defines 165 the requirements associated with the desired result, the AFEO ontology net-166 work. The second part presents the scope of both selected ontology resources 167 (AEO and OFPE) as inputs of the ontology integration process. The third part 168 describes the ontological transformations used to integrate AEO and OFPE on-169 tologies into AFEO in order to fulfill the requirements. Main OWL 2 (Web 170 Ontology Language, second edition) ontological constructs used in this paper 171 are presented in [29]. 172

173 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),

180

181

• be able to be specialized for a given food chain which is, in this paper, viticulture and winemaking (OR3).

182 3.2. Description of the initial ontologies

Following NeOn scenario 3 (Reusing ontological resources), AEO and OFPE 183 ontological resources must be inspected in order to determine if they are good 184 candidates to fulfill the AFEO requirements and to analyze their granularity. 185 AEO and OFPE ontologies were built separately for different intended com-186 munities. AEO serves as a representation of expert knowledge in the field of 187 agricultural practices while OFPE was developed to represent the knowledge 188 related to food processing experiments. It is the reason why they have been 189 selected as their union should satisfy OR1 requirement presented in section 3.1. 190 The present section describes both ontologies, in terms of their concepts and 191 relations, in order to determine if some knowledge is missing to fulfill AFEO 192 requirements. 193

¹⁹⁴ 3.2.1. Agricultural Experiments Ontology (AEO)

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

Table 1: Some important object properties of AEO



Figure 1: Concept hierarchies in AEO.

AEO is an ontology aimed to represent objects related to agricultural prac-195 tices. It is based on three main concepts, AgriExperiment, ObservedAgriEntity, 196 and AgriActivity (Figure 1). AgriExperiment is a concept that represents ex-197 periments carried out by researchers. ObservedAgriEntity represents objects 198 that are observed and followed during field experiments while the AgriActiv-199 ity represents common activities or operations performed during experiments. 200 Some important semantic relations have been defined between those concepts 201 as shown in Table 1. Object property *includes* plays an important role in this 202 ontology. It is used to represent the relation between spatial locations, e.g., Far-203 *mArea* a concept that represents the area of agricultural land used for farming. 204 This type of relation also applies to *Plot*, *Sub-plot* and *Micro-plot* which are sub-205 classes of CultivatedArea. Each plot may be composed of several subplots and 206 each subplot has two or more micro-plots. By defining *includes* object property 207 as a transitive object property, we can easily navigate between locations. This 208 object property also represents the multi-scale aspect of locations that is useful 209

- ²¹⁰ for tracking the origin of a particular product.
- 211 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 218 the food transformations. As shown in Figure 2, four subclasses under this 219 class are InputProduct, IntermediateProduct, FinishedProduct, and ServiceProd-220 uct. The first three subclasses are representation of product stages which occur 221 during a set of transformation processes. Starting with InputProduct, repre-222 senting any product harvested from farm that act as raw material, followed by 223 IntermediateProduct that models semi or unfinished products during process 224 flow, and terminated by *FinishedProduct* that represents end products or final 225

Table 2: Temporal relations between operations						
Relations	Inverse	Examples from winemaking				
Before(A,B)	$\operatorname{After}(B,A)$	Crushing is performed after destem-				
		ming				
Meets(A,B)	MetBy(B,A)	Draining is started after maceration				
$\operatorname{Overlaps}(A,B)$	OverlappedBy(B,A)	Malolactic fermentation can be started				
		before alcoholic fermentation is finished				
$\operatorname{Starts}(A,B)$	$\operatorname{StartedBy}(B, A)$	Alcoholic fermentation starts when				
		Maceration is started				
$\operatorname{Finishes}(A,B)$	$\operatorname{FinishedBy}(B,A)$	Sulfitation finished malolactic fermen-				
		tation				
$\operatorname{During}(A,B)$	$\operatorname{Contains}(B,A)$	Alcoholic fermentation occurs during				
		maceration				
$\operatorname{Equal}(A,B)$		Extraction of ethanol is started and fin-				
		ished at the same time as alcoholic fer-				
		mentation				

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 231 measured. This class has an important role in representing measurement values 232 of attributes collected from different experiments. Observation can be a single 233 observation (SingleObservation) that means one time only measurement or a set 234 of observation (MultipleObservation) where multiple measurements of a partic-235 ular attribute (of product or operation) are needed. An instance of observation 236 links to an instance of product or operation and to one or more instances of at-237 tribute through object property hasObservedObject and hasObservedAttribute. 238 This observation instance also has method which is represented by Observa-239 tionMethod class and produces result depicted by ObservationResult that holds 240 the values of measurement. In food processing the observation results come 241 in various forms, such as values directly recorded by sensors, values from ex-242

pert judgement or values from a particular calculation. To accommodate those 243 forms, this ontology provides three subclasses under ObservationResult, includ-244 ing SensorResult, ExpertJudgementResult, and CalculationResult (see Figure 2). 245 Relations among main classes in this ontology can be grouped into two parts, 246 i.e., product transformation and temporal aspects of operations. Product trans-247 formation can be described either by using a transitive object property Before 248 (or its inverse After) or object property hasInput and hasOutput. The Before 249 links between different products, in particular between classes InputProduct, In-250 termediateProduct, and FinishedProduct (see Figure 2). To describe temporal 251 aspects of operations, this ontology uses the concepts defined in the Ontology of 252 Time proposed in [31]. Examples of use of these temporal relations are presented 253 in Table 2. 254

255 3.3. AEO and OFPE re-engineering for integration

After inspection, the two selected ontology resources presented in Section 3.2.1 256 and Section 3.2.2 have been considered valid by domain experts for their respec-257 tive intended domains, i.e. biomass production and food processing. However, 258 when both are directly merged, the result is not entirely suitable to support the 259 objective of this work which focuses on agri-food experimental data integration. 260 Indeed, OR1 requirement is not completely fulfilled because the linkage between 261 production and transformation is not done. Moreover, AEO does not permit 262 to represent experimental observations, which is required in OR2. Last, spe-263 cialization to vineyard (resp. winemaking) (OR3) is not present in AEO (resp. 264 OFPE) ontology. 265

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 270 which are closely related. AEO covers the ontology terms related to entities and 271 activities occurring during field agricultural experiments, starting at planting 272 and ending at harvesting. OFPE contains various products and operations as 273 well as their attributes. It can be used to build semantic relations representing 274 a set of food processing experiments where several products and operations are 275 interrelated in a specific order. Normally, a particular set is started by receiving 276 raw materials as an input (represented by *InputProduct* class in OFPE) and 277 transformed by some operations to become a processed food. Most raw materials 278 in food processing are agricultural products that come from the field. They are 279 the result of harvesting activity (represented by the *Harvesting* class in AEO). 280 Based on that description, the object property has Output, part of OFPE, can 281 be used to link between Harvesting class part of the AEO, and InputProduct 282 part of OFPE. More formally, the domain of the object property hasOutput, 283 which is the *Operation* class in OFPE is enlarged to the alternative *Harvesting* 284 class of AEO. Berry and Bunch, which correspond to the InputProduct, are 285 considered as equivalent classes in AEO and OFPE (see Table 3). Moreover, 286 the object property *hasOutput* is defined with the cardinality one to many, 287 because in some cases one or more instances of InputProduct come from the same 288 instance of Harvesting. As AgriActivity and Operation can be considered as 289 equivalent classes in AEO and OFPE (see in Table 3), *Before* or more generally 290 all temporal relations presented in table 2, part of OFPE, can be used to link 291 between AgriActivity class, part of the AEO, and Operation, part of OFPE. 292 More formally, the domain of the object property *Before* or more generally all 293 temporal relations presented in table 2, which is the Operation class in OFPE 294 is enlarged to the alternative AgriActivity class of AEO. By using these formal 295 extended semantic relations, the ontology linking between AEO and OFPE can 296

- ²⁹⁷ be built and OR1 requirement is now fulfilled. Figure 4 shows with an example
- ²⁹⁸ how these extended semantic relations are applied.

II O						
AEO class	OFPE class	Mapping				
Berry	Berry	owl:equivalentClass				
Bunch	Bunch	owl:equivalentClass				
AgriActivity	Operation	owl:equivalentClass				

Table 3: Class mapping between AEO and OFPE.

OR2 fulfillment The next step after linking AEO and OFPE is re-engineering 299 this new ontology network. Indeed, as presented in Section 3.2.1, AEO does not 300 provide important concepts and properties with regard to observations, mea-301 surement data representation, as well as observed attributes. Contrary, the 302 OFPE contains a specific part that is dedicated to accommodate observation 303 activities. In order to fulfill OR2 requirement, we re-engineered the new ontol-304 ogy network by the following steps which implement scenario 8 (Restructuring 305 ontological resources). Firstly, we re-used the semantic relation hasAttribute 306 (existing in OFPE) to link the ObservedAgriEntity and AgriActivity classes in 307 AEO to the Attribute class. Formally, the domain of the object property hasAt-308 tribute, which is the Product class in OFPE is enlarged to the ObservedAgriEn-309 tity and AgriActivity classes of AEO. Secondly, we re-used semantic relations 310 regarding the observation which exist in OFPE. In OFPE, the Observation class 311 links to two classes, i.e. *Product* and *Operation* by using the semantic relation 312 hasObservedObject (see Figure 2). The ObservedAgriEntity and AgriActivity 313 classes in AEO have a similar role to Product and Operation which represent 314 objects of observation in OFPE. Thus, formally, we enlarged the range of the 315 hasObservedObject object property to the ObservedAgriEntity and AgriActivity 316 in AEO. 317

OR3 fulfilment 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 320 ontology resources (AEO and OFPE) do not include any specific terms of viti-321 culture and winemaking experiments. Therefore, AFEO needs to be restruc-322 tured. In this case, the scenarios 2 (Reusing and re-engineering non-ontological 323 resources) and 8 (Restructuring ontological resources) of the NeOn methodol-324 ogy were performed to get a new version of AFEO. The list of specific terms 325 related to viticulture and winemaking was provided by domain experts. These 326 terms were formalized as new concepts to specialize AFEO (Ontology special-327 ization activity of scenario 8). AFEO preserves all the classes from AEO (see 328 Figure 1) which provides generic classes that mostly also exist in viticulture. 329 Some specific classes have been added, such as CanopyManagement (sub-class 330 of AgriActivity), GreenHarvesting (sub-class of Pruning), Vinevard (sub-class 331 of CultivatedArea), GrapeVariety (sub-class of Genotype) and Grapevine (sub-332 class of Plant). Figure 3 shows the new required specific classes under the 333 *Product* and the *Operation* classes for winemaking. 334

To illustrate the final structure of AFEO³, 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

³http://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.

⁴http://agroportal.limm.fr/ontologies/AFEO

⁵http://agroportal.lirmm.fr/ontologies/AEO

⁶http://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 experimen tal data

In a domain specific ontology development stage, an ontology instantiation 352 (population) is an important step. The main task is to semantically annotate all 353 related resources based on the proposed ontology. AFEO, result from Section 3, 354 can be seen as an agri-food experiment knowledge layer that acts as a container 355 for viticulture and winemaking experimental data integration. From a practical 356 point of view, this ontology network needs to be instantiated with actual exper-357 imental data so that specific data queries can be performed to support further 358 statistical analysis. This section presents how AFEO has been instantiated in 359 order to integrate the experimental data from various stages of agri-food exper-360 iments. The first part describes two different data sources used in this work and 361 the second part explains the principles of the ontology population step. 362

363 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 viti-367 cultural experiments from different vineyards in the south of France. These 368 data contain information about vineyards, plots, sub-plots, grapevines, different 369 treatments of experiment (for example irrigation, non irrigation, green harvest-370 ing, non-green harvesting, etc.), and a huge number of observations regarding 371 the grape characteristics during the growing season. They are stored in a single 372 relational database linked to Silex-VitiOeno Pilotype, a web application that 373 manages viticulture experimental data. 374

The winemaking experimental data contain information about different stages 375 of the winemaking process, different wine products as well as a big amount of 376 observational data about operations and products characteristics during process 377 flow. These data were collected annually from the Unité Expérimentale de Pech 378 *Rouge*, France, where the winemaking experiments were conducted. Unlike the 379 viticulture experimental data, these data are recorded in the form of Microsoft 380 Excel files. To the current date, more than 500 Microsoft Excel files have been 381 collected. These files can be classified into two groups: i.e. 382

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.

391 4.2. Ontology instantiation process

³⁹² Both data sources have to be transformed into uniform format based on ³⁹³ AFEO ontology network. Consequently, the Resource Description Framework

(RDF) format has been chosen to represent these data. The RDF is a standard 394 model for data interchange and metadata processing [33]. It has features that 395 facilitate data linking even if the underlying schema differ. The RDF can be 396 seen as a graph data model that uses URIs to name two nodes as well as the 397 relationship between them. Due to the different source data formats, we devel-398 oped two specific scripts using the Python language as illustrated in Figure 5. 399 The first script creates instances by performing some queries to the viticulture 400 database and transforming them to RDF format. The second script is special-401 ized for making instances from the Microsoft Excel winemaking experimental 402 data files. Both of them were developed guided by the AFEO ontology net-403 work. RDF data are stored in a triplestore, a specific database for storing and 404 retrieving RDF triples through semantic queries expressed in SPARQL. 405



Figure 5: Data population processing steps.

406 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.

⁴¹⁶ 5.1. Wine traceability

Wine is a food product whose value is generally determined by three groups 417 of characteristics, i.e., appearance, flavour, and aroma. In the recent years, 418 wine consumers are more and more oriented to consider detailed information 419 about those characteristics as well as the overall process from the grape to the 420 bottle [34]. The information is used to confirm the wine identity (authenticity) 421 and to ensure it is produced transparently in order to avoid fraud and toxico-422 logical issues [35]. The viticulture-winemaking integrated data generated from 423 Section 4.2 can be queried in order to get such traceability information, giving 424 valuable knowledge about the entire process from viticulture practices through 425 selected winemaking operations, until finished product delivery. 426

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.

432 Query 1.

- 433 SELECT ?w ?t ?l ?nv ?na
- 434 WHERE
- 435 {
- $_{\rm 436}$?o ofpe:After ?h .
- 437 ?w rdf:type ofpe:FinishedWine .

- 438 ?o ofpe:hasOutput ?w
- 439 ?h aeo:appliedTo ?l
- 440 ?l aeo:included ?p
- 441 ?p rdf:type aeo:Plot
- 442 ?p aeo:hasVariety ?v
- 443 ?v rdfs:label ?nv
- 444 ?l aeo:included ?a
- ⁴⁴⁵ ?a rdf:type aeo:Vineyard
- 446 ?a rdfs:label ?na
- 447 }

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 fin-448 ished wine, their input products, sub-plots and vineyards where the grapes are 449 planted, and their grape varieties. In Query 1, the After object property is 450 used in order to trace all existing products during winemaking process. By 451 using the After, which is defined as a transitive property, this SPARQL query 452 performs an inference to determine link between BerryGrape instance (t) as an 453 input product and *FinishedWine* instance (w) as the final product, whatever the 454 number of intermediate products and associated operations. Figure 6 highlights 455 some information about wine traceability that might help researchers to know 456 global information about the origin of finished wine products. For instance, 457

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.

⁴⁶³ 5.2. Influence of viticulture practices on grape and wine qualities

As said above, wine quality can be analysed by three groups of characteris-464 tics. Each group can be further detailed by various attributes. For example, the 465 appearance is determined by the color attribute along with other visual features 466 such as clarity and fluidity, while the wine flavour has some basic attributes, 467 such as sweetness, acidity, bitterness, fruit flavour, etc. These attributes are 468 influenced by various interrelated factors, both viticulture practices (e.g., grape 469 variety, location, vintage, etc.) and winemaking selected operations (e.g., crush-470 ing method, maceration, pressing, yeasts added, etc.). 471



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 Rous-476 sillon region in order to test for the effects of vine stress status and winemaking 477 protocol on grape potential and wine quality in contrasted environmental con-478 ditions. We discuss here a subset of these experiments. Three grape varieties 479 are being studied, i.e. Merlot, Syrah and Cabernet Sauvignon; and two different 480 winemaking methods, i.e., with an innovative hot pre-fermentation maceration 481 and without it (classical method). In the experimental plots, the wine water 482 status has been monitored by different means: on-line sap flow sensors, pre-483 dawn leaf water potential and DeltaC13 at harvest time. The last one, which 484 is a global indicator of the water stress experienced by the vine, will be used in 485 the following. Many quality parameters resulting from physico-chemical analy-486 ses are available to assess grape composition at harvest time, must composition 487 and finished wine composition. Among them, let us take the example of the 488 glutathione (GSH) concentrations in wine and berry grape. GSH, a natural tri-489 peptide found in grapes and wine, is a powerful antioxidant that protects wines 490 from oxidation and loss of aroma or flavor. Glutathione levels fluctuate during 491 production, as the compound can be absorbed by yeast and then released after 492 fermentation. If final GSH levels are low in wines, they will experience faster 493 loss of fresh varietal and fruity aromas, and poor ageing potential. By using 494 the AFEO, this experiment can be represented semantically as visualized in 495 Figure 7. 496

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 503 the available data which cover two years experiments (2012 and 2013), the query 504 provides the results as follow: there are a total of 40 instances of *RedWine* which 505 are produced from 20 instances of the BerryGrape, 20 instances of the Subplot 506 and 9 instances of the *Plot* as the origin of the *BerryGrape* instances. The 507 RedWine instances are divided equally into two groups, 20 instances from the 508 hot pre-fermentation maceration methods and the rest come from the classical 509 method. 510

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 prefermentation maceration influences the wine GSH concentration and increases it.

A graphical analysis is then done to visually explore, by taking advantage of 518 the links between viticulture and wine data, if the vine water status could have 519 an impact on that influence. Figure 8 shows, for each of the three varieties, on 520 the abscissa the grape GSH concentration and on the ordinate the corresponding 521 wine GSH concentration. Circles correspond to the year 2012, and squares to 522 the year 2013. The classical winemaking protocol is plotted in red, and the hot 523 pre-fermentation maceration one in blue. The size of the symbols is proportional 524 to the Delta C13 indicator: the higher Delta C13, the bigger the symbol and the 525 less the vine was exposed to water stress. Plots visually confirm the increase of 526 GSH in wine, for a given GSH in grapes, with the hot pre-maceration protocol, 527 but they do not show any influence of the vine water status. 528



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

529 6. Conclusion

This paper presents an ontology-based approach for experimental scientific 530 data integration across complementary sub-domains, i.e., agricultural practices 531 and food processing, with an application to viticulture and winemaking pro-532 cess. The two main steps in this approach are i) to develop an ontology network 533 and ii) to populate the ontology with actual experimental data from differ-534 ent sources. The ontology network -Agri-Food Experiment Ontology (AFEO)-535 was developed based on two existing ontology resources, i.e., AEO (Ontology 536 for Agriculture Experiment) and OFPE (Ontology for Food Processing Experi-537 ment). It contains 136 concepts which covers various viticulture practices, and 538 winemaking products and operations. AFEO was used to guide data integra-539 tion from two different data sources, i.e., viticulture experimental data which 540 are stored as relational database and winemaking experimental data which are 541 stored in Microsoft Excel files. 542



Results show the potential uses of the AFEO along with viticulture-winemaking $% \mathcal{A}$

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.

561 References

562 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 agrifood systems, Trends in Food Science & Technology 48 (2016) 88–101.
URL http://www.sciencedirect.com/science/article/pii/
S0924224415002186

- 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.
- ⁵⁷⁷ [5] D. P. K. (Seedah, B. Sankaran, W. J. O'Brien, Approach to classifying
 ⁵⁷⁸ freight data elements across multiple data sources, TRANSPORTATION
 ⁵⁷⁹ 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.
- 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. Ram-593 say, C. Nathe, K. Lum, K. Krouse, D. Stearns, B. Connolly, T. Skillman, 594 M. Igra, LabKey Server: an open source platform for scientific data inte-595 gration, analysis and collaboration., BMC bioinformatics 12 (2011) 71. 596 [10] F. Wang, P.-E. Bourgué, G. Hackenberg, M. Wang, D. Kaltschmidt, P. Liu, 597 SciPort: an adaptable scientific data integration platform for collaborative 598 scientific research, Very Large Data Bases (2007) 1310–1313. 599 [11] B. Luduscher, K. Lin, S. Bowers, E. Jaeger-Frank, B. Brodaric, C. Baru, 600 Managing scientific data: From data integration to scientific workflows*, 601 Geological Society of America Special Papers 397 (2006) 109–129. 602 [12] S. Leonelli, Integrating data to acquire new knowledge: Three modes of 603 integration in plant science., Studies in history and philosophy of biological 604 and biomedical sciences 44(4)(2013)503-514. 605 [13] Y.-F. F. Li, G. Kennedy, F. Ngoran, P. Wu, J. Hunter, An ontology-centric 606 architecture for extensible scientific data management systems, in: Future 607 Generation Computer Systems, Vol. 29, 2013, pp. 641-653. 608 [14] P. Fox, D. L. McGuinness, L. Cinquini, P. West, J. Garcia, J. L. Benedict, 609 D. Middleton, Ontology-supported scientific data frameworks: The Vir-610 tual Solar-Terrestrial Observatory experience, Computers and Geosciences 611 35 (4) (2009) 724–738. 612 [15] C. (Daraio, M. Lenzerini, C. Leporelli, P. Naggar, A. Bonaccorsi, A. Bar-613 tolucci, The advantages of an ontology-based data management approach: 614
- openness, interoperability and data quality, SCIENTOMETRICS 108 (1) (2016) 441–455.
- 617 [16] M. (Imran, R. I. M. Young, Reference ontologies for interoperability across

- multiple assembly systems, INTERNATIONAL JOURNAL OF PRODUC TION 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 Agri culture Knowledge Presentation with Ontology, AASRI Procedia 3 (2012)
 732–738.
- [22] H. W. Beck, S. Kim, D. Hagan, A Crop-Pest Ontology for Extension Publi cations, 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.

- 642 [24] C. Snae, M. Bruckner, FOODS: A Food-Oriented Ontology-Driven Sys-
- tem, 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).
- 658 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).
- ⁶⁷³ 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.
- ⁶⁷⁸ [35] M. Palade, M.-e. Popa, Wine traceability and authenticity A literature
 ⁶⁷⁹ review, Series F. Biotechnologies XVIII (2014) 226–233.