



HAL
open science

Recent progresses in data and knowledge integration for decision support in agri-food chains

Sophie Aubin, Pierre Bisquert, Patrice Buche, Juliette Dibie-Barthelemy, Liliana Ibanescu, Clement Jonquet, Catherine Roussey

► **To cite this version:**

Sophie Aubin, Pierre Bisquert, Patrice Buche, Juliette Dibie-Barthelemy, Liliana Ibanescu, et al.. Recent progresses in data and knowledge integration for decision support in agri-food chains. IC 2019 - 30es Journées francophones d'Ingénierie des Connaissances, AFIA, Jul 2019, Toulouse, France. pp.43-59. <hal-02284538>

HAL Id: hal-02284538

<https://hal.science/hal-02284538v1>

Submitted on 11 Sep 2019

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.



HAL Authorization

Recent progresses in data and knowledge integration for decision support in agri-food chains

Sophie Aubin¹, Pierre Bisquert², Patrice Buche², Juliette Dibie³,
Liliana Ibanescu³, Clément Jonquet⁴, Catherine Roussey⁵

¹IST INRA, Angers, France
sophie.aubin@inra.fr

²IATE INRA, INRIA GraphiK, Montpellier University, France
{pierre.bisquert, patrice.buche}@inra.fr

³UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005 Paris, France
{dibie, liliana.ibanescu}@agroparistech.fr

⁴LIRMM, Université de Montpellier & CNRS, France
jonquet@lirmm.fr

⁵Université Clermont Auvergne, Irstea, UR TSCF, Centre de Clermont-Ferrand,
9 avenue Blaise Pascal CS 20085, F-63178 Aubière, France
catherine.roussey@irstea.fr

Abstract : Research in Agri-food and related fields dealing with sustainability has undergone important changes in the past years and tends to be more integrative, collaborative, and interdisciplinary. In this article, we present an excerpt of recent and ongoing projects in the French community proposing data integration guided by an ontology for collective decision support in the Agri-food domain. The use of ontologies for primary production and secondary production, i.e. transformation of primary production into food products, are addressed. An example of collective decision support system for food packaging selection which relies on scientific data annotated thanks to ontologies and food chain stakeholders' preferences is described. French initiatives at the international level to share agri-food ontologies are presented.

Key-words : Ontology, data integration, flexible querying, computational social choice, argumentation

1 Introduction

Research in Agri-food and related fields dealing with sustainability has undergone important changes in the past years and tends to be more integrative, collaborative, and interdisciplinary (Perrot et al., 2016). The Agri-food domain is considered as an interconnected system with various entities and complex relationships among them (Wolfert et al., 2010). More and more numerous data coming from heterogeneous sources cover the whole food chain and can be combined to address new questions. For instance, to test a hypothesis about the effects of different viticulture treatments on wine quality, researchers need to access and analyze various data sources at different scales, from the field to the bottle. Data integration is not so easy and researchers have to face several issues. Data are commonly stored in scattered places and their formats, naming, storage and query or retrieval mechanisms are very diverse. The heterogeneity of scientific data may come from many factors, such as (i) they are collected separately based on independent research projects with very specific targets and aims; (ii) the data structure frequently depends upon the collection method (e.g., to make data easier to record) or is function of future analysis, instead of using standard data representation (e.g., relational database schema); (iii) the terms and concepts used to annotate data are not standardized, neither within nor across scientific disciplines and research groups (Bowers, 2012). The difficulties in organizing data and knowledge in a unified way do not only limit research productivity but also reduce data traceability (Gardner, 2005). Research experiments are commonly divided into sub-domains, such as agricultural production, post-harvest, and food transformation. Even though the explicit links between those sub-domains may be easy to explicit and understand, each of them has different objectives, scopes and circumstances. For instance, agricultural experiments are usually conducted in the plots where environmental factors are difficult to control, while food processing experiments are generally carried out in

laboratories with controlled environmental conditions. From a practical point of view, both experiments require different methods for collecting and organizing observational data that yield differences in data format, structure and storage. The heterogeneity also occurs due to the vast scope of Agri-food sub-domains. For instance, each discipline and sub-domain involved in Agri-food domains uses its own way to express knowledge, terms, concepts and semantic relationships, which makes the scientific data sharing harder.

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 (Gardner, 2005; Seedah et al., 2015; Doan et al., 2012). An ontology defines a set of primitives to model a domain of interest: classes, attributes (or properties) and relations between members of the classes (Guarino et al., 2009). Ontology is used to create and/or reuse standardized vocabularies and to index data sources with those vocabularies in order to allow data source interoperability. It opens 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. These features will help to develop advanced Information Systems able to manage heterogeneous data sources and to design platforms for more collaborative scientific data analysis and decision support to food chains.

In this article, we present an excerpt of recent and ongoing projects in the French community proposing data integration guided by an ontology for collective decision support in the Agri-food domain. Section 2 addresses the use of ontologies for primary production. Section 3 presents the use of ontologies for secondary production, i.e. transformation of primary production into food products. An example of decision support system for food packaging selection which relies on scientific data annotated thanks to ontologies will be described in the Section 4. Section 5 presents French initiatives at the international level to share agri-food ontologies. Finally, we conclude in Section 6.

2 Data integration guided by ontology for primary production

Precision agriculture is nowadays getting more and more attention in Europe due to the development of remote sensors and wireless sensors. Precision agriculture can be summarized as putting the right amount of matter (e.g., water, nutriment or pesticide) at the right time and at the right place. The goal is to reduce the quantity of matter that are lost in the soil or the air. Precision agriculture is based on accurate observations. These observations come from various types of sensors such as farmers (i.e., human sensors), remote sensors or sensors in the field. These observations could be quite complex. Some precise protocols should be followed to perform an accurate observation.

In this section we focus on crop observations, that are observations made on cultivated plots. These observations can be for instance the crop growth stage, the presence of pests in the cultivated plot, observations of pedo climatic conditions (e.g., soil moisture, quantity of rain, outdoor temperature). These observations help identifying potential risks like water stress. For a long time, agronomists have developed models (e.g., simulation, decision tree) to evaluate these risks and to help the design of Decision Support Systems (DSS) consuming crop observations. DSS may activate different types of actions: send alerts to farmers' smartphone in order to advise them to consolidate the risk evaluation by observing the presence of the problems on the plot. DSS may also automatically activate an equipment to solve the problem (e.g., irrigate a plot if the water stress risk is high).

Due to the fact that all these observations come from various kinds of sensors and have many different types, the storage of these observations suffers from heterogeneity issues. One solution is to use ontologies to store all these observations. The focus of Section 2 is to present the advantages of using a well-known ontology design pattern (Property and FeatureOfInterest)

to store in a similar way all types of observations. In Section 2.3 we present two uses cases developed at Irstea, a French Research Institute for Agriculture.

2.1 Ontologies related to Observations

From our point of view, there are two types of observations in primary production: (1) observations made by sensor located in the cultivated plot, such as soil moisture probes or agricultural weather stations; (2) observations made by farmers or any crop experts following an observation protocol. Ontologies for storing sensor measurements were very well studied in the literature. We can cite for example Semantic Sensor Network (SSN) (Compton et al., 2012) and Smart Appliances REference For Environment (SAREF4ENVI) (ETSI, 2017). Concerning human observations, the Extensible Observation Ontology (OBOE) (Madin et al., 2007) is the only one dedicated to environmental scientific observations. Note that the ontologies dedicated to the description of experiments are not in the scope of this section.

The SSN ontology is used as the core of the crop observation model and more precisely a new module, call SOSA.

SSN is an ontology developed by the Semantic Sensor Network Incubator Group (SSN-XG). The final report of the original SSN was published on the site of World Wide Web Consortium (W3C) on 28th June 2011. SSN ontology describes sensors and their measurements (Compton et al., 2012). SSN ontology defines a *Sensor* as “*Device, agent (including humans), or software (i.e. simulation) involved in, or implementing, a Procedure. Sensors respond to a Stimulus, e.g., a change in the environment, or Input data composed of the Results of prior Observations, and generate a Result...*” .

The joint W3C and OGC Spatial Data on the Web Working Group has developed a new version of the SSN including a module called SOSA (Sensor, Observation, Sampler and Actuator) (Haller et al., 2017). This module replaces the SSO (Stimulus Sensor Observation) Pattern. To describe sensing acts, SOSA provides *sosa:Sensor* (e.g. a thermometer) that make *sosa:Observation* (e.g. a measurement) about *sosa:ObservableProperty* (e.g. temperature). The *sosa:ObservableProperty* (e.g. temperature) is a property of a *sosa:FeatureOfInterest* (e.g. the outdoor air, the air of a specific room, a human body, an oven). Moreover, to identify the specimen where the sensing act was performed, SOSA presents *sosa:Sampler* that makes *sosa:Sampling* of some *sosa:FeatureOfInterest* to produce a *sosa:Sample*. The *sosa:Sample* that corresponds to the *sosa:FeatureOfInterest* air is the volume of air around the weather station. This design pattern is useful to describe precisely the sensor measurement. For example our weather station interface presents its temperature measurement with the label “temperature”. Everybody will understand that the weather station measures the outdoor temperature of the air. But other thermometer may be used in agricultural equipments: soil, building or cattle. So the precision of the SSN/SOSA design pattern enables to clarify the sensor measurements and it is the core of the model detailed in the next section.

2.2 Crop Observation Model based on SSN/SOSA

To build our model of crop observation based on SSN/SOSA we needed others ontologies presented in Table 1. The acronyms given in the second column of Table 1 are used in Figure 1.

TABLE 1 - the list of ontologies used to model crop observation

Ontology name	acronym	authors	reference
Semantic Sensor Network	sosa	W3C	(Haller et al, 2017)
Time	time	W3C	(Hobbs & Pan, 2004)
GeoSparql	geo	OGC	(Battle & Kolas, 2012)
Prov ontology	prov	W3C	(Lebo et al., 2013)
Simple Knowledge Organisation System	skos	W3C	
Quantities, Units, Dimensions and data Types	qudt	QUDT.org member of W3C	(Hodgson et al., 2014)

As shown in Figure 1, the crop observation denoted **P25Maize20170101** is an instance of the class *sosa:Observation*. This observation is related to a cultivated plot, denoted **P25** in Figure 1. A cultivated plot is modeled as an instance of *sosa:Sample* because a cultivated plot at a specific time is indeed a specific sample of a crop (e.g. maize). Therefore the crop is an instance of *sosa:FeatureOfInterest* which has as many samples as cultivated plots of the crop exist. The plot is also a geographic object. So the representation of a cultivated plot is an instance of both *sosa:Sample* and of *geo:Feature* classes.

The properties used to describe the observation are:

- *sosa:hasFeatureOfInterest* that links an instance of *sosa:Observation* to an instance of *sosa:Sample*. The instance of a sample may be one specific cultivated plot or a set of plots depending on the use case.
- *sosa:isSampleOf* that links the instance representing the plot to an instance of *skos:Concept* representing the type of crop (e.g. maize). This skos instance may come from any thesaurus describing crop like Agrovoc (<http://aims.fao.org/fr/agrovoc>).
- *sosa:observedProperty* that links an instance of *sosa:Observation* to an instance of *skos:Concept* representing the crop growth stage, denoted **v2** in Figure 1. This skos instance may come from any thesaurus. For example we can reuse the CROP ontology (http://agroportal.lirmm.fr/ontologies/CO_715) that contains the BBCH crop growth stage, a scale for a uniform coding of growth stages.
- *sosa:phenomenonTime* that links an instance of *sosa:Observation* to an instance of *time:Instant* or *time:Interval*. In the case of an observation of crop growth stage the time entity is an interval that represent the period when the cultivated plot reach the growth stage. Note that the interval is described by several time entity like the beginning instant as presented in Figure 1.
- *sosa:resultTime* that links an instance of *sosa:Observation* to an *xsd:dateTime* value that represents the day when the observation was done.
- *sosa:hasSimpleResult* is an attribute that contains a percentage. This percentage may express different ratio depending of the use case. It may be the percentage of plants that reach the given growth stage on the number of plants in the plot. Note that a plot reaches a given growth stage when 50% of the plants of the plot reach this growth stage. Other interpretation of the percentage may be the ratio between the number of plots that has reached the crop growth stage on the total number of plots in the sample.
- *sosa:hasResult* that links an instance of *sosa:Observation* to an instance of *sosa:Result*. This instance has two attributes: one indicates the unit of the measurement and the other the value.

- *sosa:madeBySensor* that links an instance of *sosa:Observation* to an instance of *sosa:Sensor*. This instance may be a person or a device depending of the use case.

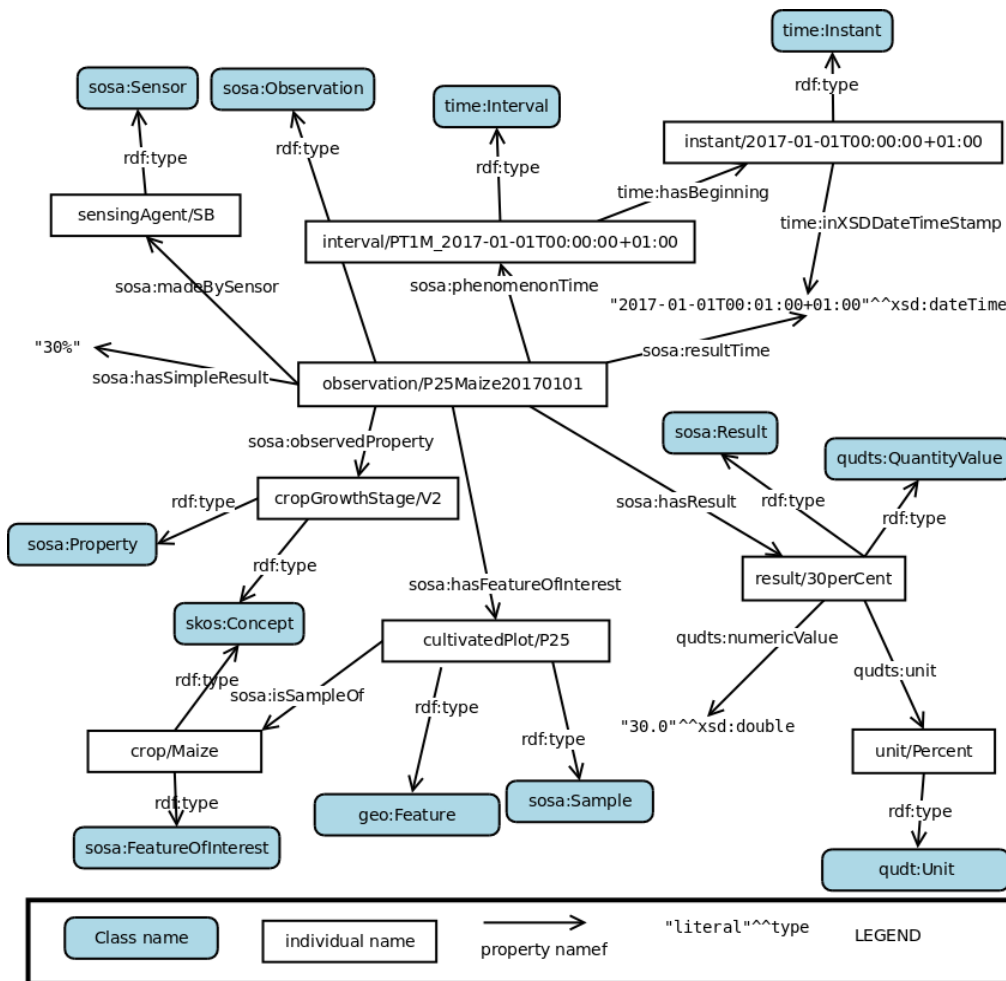


FIGURE 1 - an example of crop observation based on SSN/SOSA.

2.3 Two Use Case Descriptions in Primary Production

The crop model observation described in the previous section is implemented in two use cases: the design of an archive of crop observations and the design of a context aware system to automate irrigation.

The first use case is part of the French ANR D2KAB project. The goal of this project is to build a search engine dedicated to agricultural alert bulletins called “Bulletin de Santé du Végétal”. An alert bulletin is dedicated to a specific area (e.g. a French administrative unit) and a specific crop category (e.g. cereals). A bulletin contains some human observations: crop growth stage and pest presence. It also contains some evaluation of risk about pest attacks. Note that a presence of a pest on a cultivated plot becomes an attack only if the crop has reached a defined growth stage and if the presence of the pest is significant. Some protocols of pest observation are defined to evaluate the pest attack. Otherwise the pest has no impact on the crop production.

A bulletin does not contain an aggregation of all observations about one crop, but it contains few useful observations selected by specialists, the authors of the bulletin. One of the goals of the D2KAB project is to extract from the text content of the bulletin the crop growth

stage observations. Results are published in (Roussey & Abderrahmani Ghorfi 2018). The outcome of this project is that the annotated corpus becomes a spatio temporal archive of crop observations for France and may be queried for different purposes.

The second use case designs a context-aware system to automate maize irrigation. The context of a cultivated plot is acquired by a wireless sensor network located on this plot. This context contains some automatic evaluation of crop growth stage based on temperature measurements. It also contains some soil moisture measurements and pluviometer measurements. In order to take an irrigation decision, the context aware system stores different types of observations: raw sensor measurements, spatio-temporal aggregation of measurements, qualitative data inferred from the comparison of the aggregated value and a threshold. A decision process starts from some simple quantitative measurements (e.g., soil moisture at 30 cm depth is 42 cbar) and finishes by a qualitative data that summarize the crop situation (e.g., crop water stress is high). Results are published in (Poveda et al, 2018) and confirm that the crop water stress depend on its growth stage and the amount of water in the soil.

3. Data integration guided by ontology for secondary production

One of the most significant current discussion in food production is to formulate food products with high qualities such as nutritional requirements or sensory acceptable by consumers as well as low environmental impact. This supposes to build a decision support tool combining data and knowledge from different domains in food science (e.g. nutrition, sensory and perception, eco-design, microbiology, biochemistry, process engineering) with data and knowledge in environmental analysis.

Besides the different domains of the available data, several other issues need to be addressed to be able to help experts in their decision process such as data heterogeneity, their different scale, their different purposes for which they have been collected and that can be complementary or even contradictory, their temporal evolution through the different unit operation. A major problem is the lack of cross domain studies and no clearly identified methodology about how to combine, aggregate and integrate data and knowledge in order to perform a multi criteria analysis in food production.

When assessing the quality of food product, it is important to have knowledge about product properties all along its production process, about different process parameters and the environmental impact of the whole production process. This section presents two use cases of data integration using ontologies in food production: @Web and PO .

3.1 @Web

@Web (Buche et al., 2013b) is a Web application designed to collect, integrate and query experimental data extracted from scientific documents found on the Web. @Web implements a complete workflow to manage experimental data: extraction and semantic annotation of data from scientific documents, data source reliability assessment and uniform querying of the collected data stored in a database opened on the Web.

@Web relies on an Ontological and Terminological Resource (OTR) (Buche et al., 2013a) which guides scientific data semantic annotation and querying. An OTR associates a terminological component to an ontology in order to establish a clear distinction between the linguistic expressions in different languages (i.e. the term) and the notion it denotes (i.e. the concept) (Roche et al., 2009; McCrae et al., 2011; Cimiano et al., 2011). For instance, English terms “Ethylene vinyl alcohol” and “EVOH” and the French term “Ethylène alcool vinylique” denote the same symbolic concept Ethylene_vinyl_alcohol. The @Web OTR is designed to

model scientific experiments. It is composed of two layers: a generic one and a specific one dedicated to a given application domain. The @Web OTR allows n-ary relations relevant for a given application domain to be defined. Those n-ary relations are used to annotate data in tables. In (Buche et al., 2013a) the @Web OTR for the Food packaging domain is presented (see Figure 2 for an example in this domain). In (Lousteau-Cazalet et al., 2016) the generic part of the @Web OTR was reused and its specific part was defined for the biorefinery process.

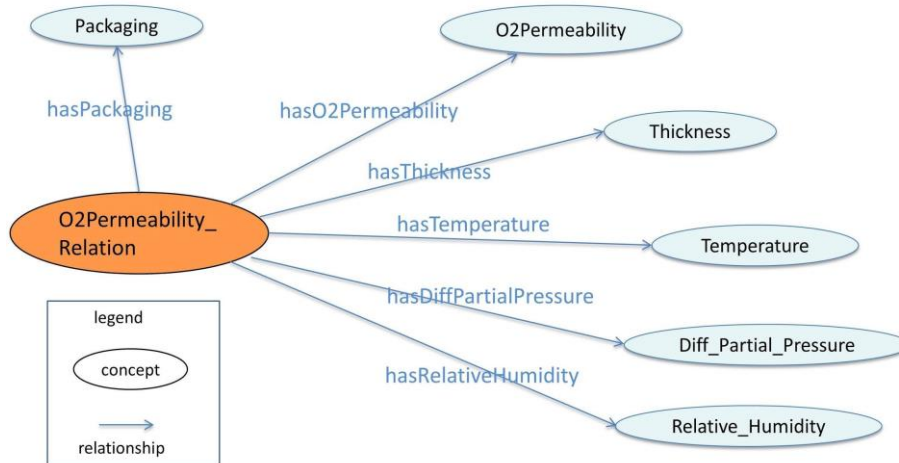


FIGURE 2 - An example of ontological conceptual modelling: the O_2 permeability n-ary relation concept.

@Web application is composed of two sub-systems. The first one is an annotation sub-system for the acquisition and annotation, with concepts of the OTR, of experimental data extracted from scientific documents; those annotated data are being stored into a database. This sub-system also allows the reliability of data sources to be assessed using the approach of (Destercke et al., 2013). The second sub-system is a flexible querying system based on the approach presented in (Destercke et al., 2011). @Web is implemented using the semantic web standards (RDF, OWL, SPARQL): the OTR is defined in OWL2-DL, annotated tables in XML/RDF and the querying in SPARQL. Section 4 will present a decision support system using the outcome (i.e. the database) of the @Web application.

3.2 PO²

PO² (Process and Observation Ontology) (Ibanescu et al., 2016) allows one to represent a food transformation process described by a set of experimental observations available at different scales and changing over time through the different unit operations of the production process. PO² ontology is the outcome of the data and knowledge representation tasks of projects involving INRA researchers. The goal of one of these projects, the CellExtraDry project, was to build a multicriteria decision support system allowing to improve the environmental impact for the production of stabilized micro-organisms (e.g. yeast). The goal of a second project, the NutriSensAI project, was to collect and integrate data from the cheese production system and from the domain concerning the sensorial perception of food; these NutriSensAI data will be used to implement a decision support system allowing to produce new foods with better nutritional qualities and acceptable by the consumers.

PO core ontology has been developed using the Scenario 6 of the NeON methodology (Suarez-Figueroa et al. 2012), i.e. reusing, merging and re-engineering ontological resources. PO core ontology has been built from SSN/SOSA, QUDT (<http://qudt.org/schema/qudt/>), OWL-TIME (<http://www.w3.org/2006/time#>) and @Web OTR described in Section 3.1. Moreover, and for the goal of increasing semantic interoperability, particularly in the life

sciences domain, PO core ontology was fully integrated with the Basic Formal Ontology (BFO), a small, and genuine upper level ontology.

PO core ontology concepts and relations are given in Figure 3. Three types of concepts are used:

- the concepts **Process**, **Itinerary** and **Step** concern the description of the production process. An itinerary is composed of a set of steps. A step describes an operation unit and it is characterized by its participants and its temporal duration.
- the concepts **Product**, **Mixture**, **Material** and **Method** concern the participants involved in a production step. A **Mixture** is the aggregation of several raw products. Material represents all the objects which are used during a step. These materials can be sensing devices performing measurements or transformation equipments. A method is the description of the way the observation has been performed.
- the concepts **Observation**, **Attribute** and **Scale**. An observation concerns an attribute, e.g. pH or temperature.

PO core ontology v2.0, implemented in OWL 2, is published on the AgroPortal ontology library (Jonquet et al. 2018) (<http://agroportal.lirmm.fr/ontologies/PO2>).

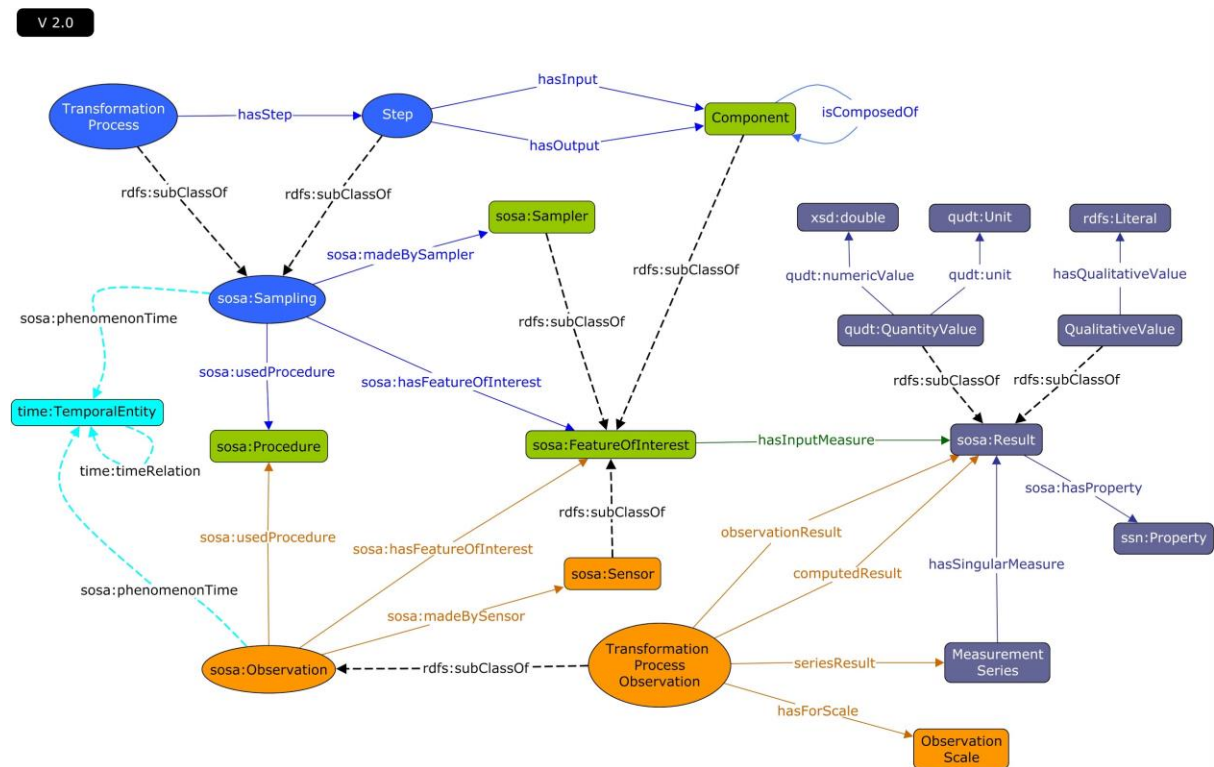


FIGURE 3 - The core ontology PO².

PO DG (DG for Dairy Gel) is a domain ontology built from PO core ontology in the field of dairy products (<http://agroportal.lirmm.fr/ontologies/PO2-DG/>). PO DG reuses concepts from Global Agricultural Concept Scheme (GACS) (<http://agroportal.lirmm.fr/ontologies/GACS>) which is aligned with AgroVoc (<http://agroportal.lirmm.fr/ontologies/AGROVOC>) and NALT (<http://agroportal.lirmm.fr/ontologies/NALT>). It allows to describe the production of cheese with attributes for sensory perception (e.g. texture or taste intensity) and rheological properties studied in the NutriSensAI project. Moreover PO DG ontology includes concepts needed to quantify the environmental impact using Life Cycle Assessment in the production of stabilized micro-organisms defined in the CellExtraDry Project.

PO DG contains 3475 concepts and 122 relations and it is available from the AgroPortal repository (http://agroportal.lirmm.fr/ontologies/PO2_DG) under the licence Creative Commons Attribution International 4.0 International (CC BY 4.0). Figure 4 gives an excerpt of the **Product** concept hierarchy in PO DG.

In (Penicaud et al. 2019), the presented results illustrate how the common vocabulary and the structure provided by the PO DG ontology allow the data combination from different domains and how this semantic approach can be useful to estimate missing data in NutriSensAI project. Moreover, an illustrative example is given to show how to transfer knowledge from the CellExtraDry project to the NutriSensAI project in order to evaluate the environmental impact using Life Cycle Assessment (LCA) by giving hints to the LCA expert about relevant parameters to be measured. A RDF repository **PO²DG_dataset** is under the process of data integration and will store the available data of the NutriSensAI project structured using the PO DG ontology. The decision support system build upon the integrated and semantically annotated data from the **PO²DG_dataset** is under construction.

The next section presents an example of a decision support system build upon the integrated and semantically annotated data presented in Section 3.1.

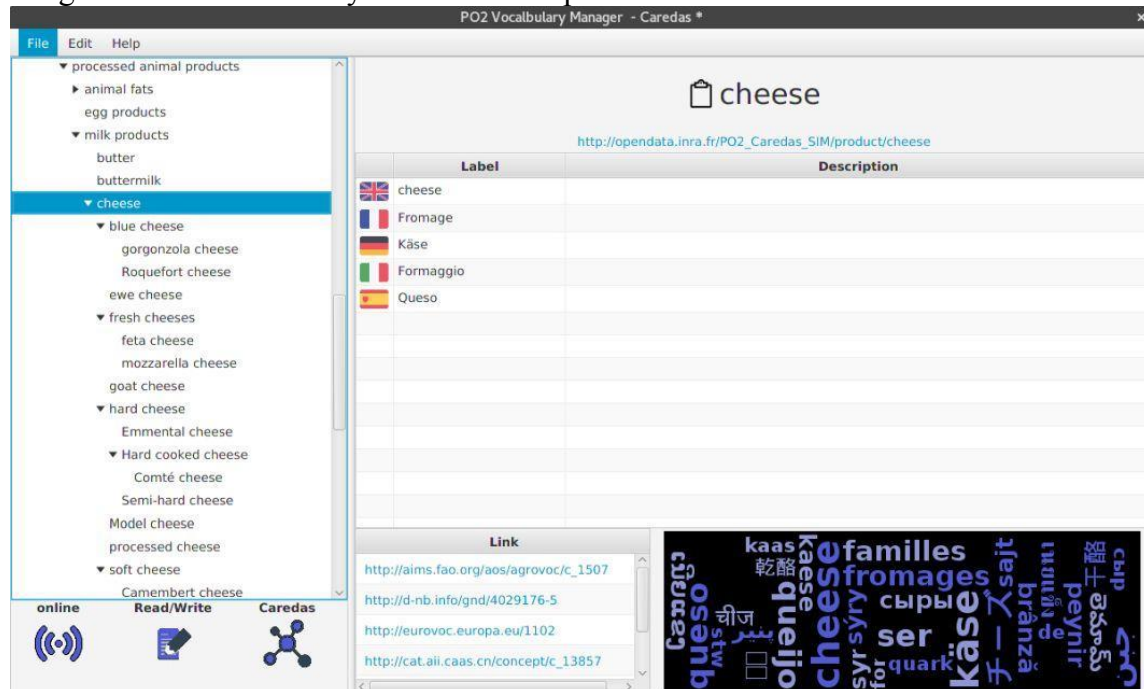


FIGURE 4 - An excerpt of the **Product** concept hierarchy in PO²DG.

4 Multicriteria decision support based on integrated data and food chain stakeholders' preferences and constraints

Knowledge engineering methods have been used to mimic the decision making process of the human brain and allow the development of decision support systems (DSS) like EcoBioCAP software that help users take the right decision in the field of food packaging (Guillard et al. 2015). EcoBioCAP is a powerful DSS tool able to answer a complex multicriteria query such as: "I want a packaging material that will maintain the quality of strawberries (i.e. with the permeability properties that match the respiration of strawberries), at a cost of less than €3 per kg, and if possible transparent and derived from renewable resources". Flexible querying methodologies employed in knowledge engineering were used to develop this tool (Destercke et al. 2011). Below we briefly present the EcoBioCAP DSS tool that have

been developed at the junction between different fields of expertise such as food engineering, computer science, knowledge engineering, argumentation and numerical simulation.

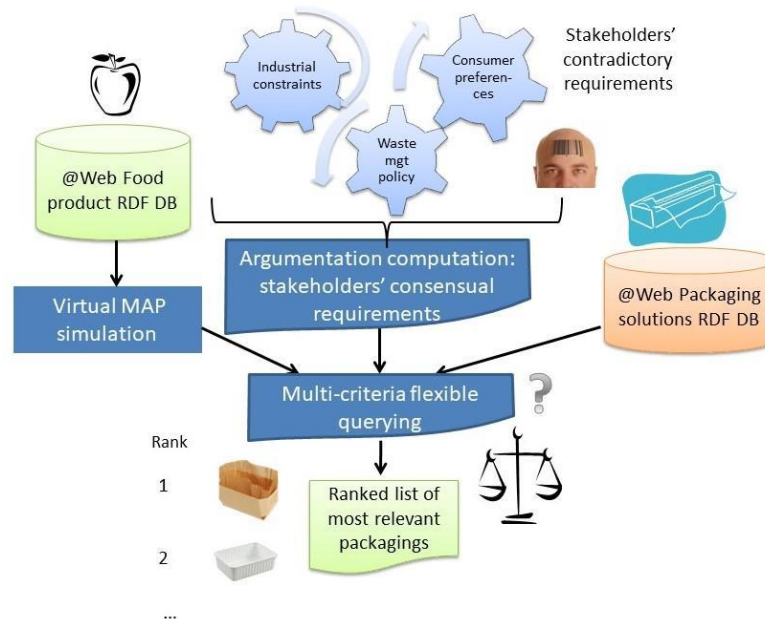


FIGURE 5 - Architecture of EcoBioCAP DSS.

EcoBioCAP DSS tool (see Figure 5) relies on data, the food product characteristics (O₂/CO₂ respiration) and the food packagings properties (O₂/CO₂ permeabilities), stored in the @Web RDF triplestore presented in Section 3.1. This tool retrieves respiration characteristics from the @Web food product database and uses this data plus other user-entered characteristics such as pack geometry to compute the optimal permeabilities for the food product. These permeabilities are automatically considered as mandatory preferences associated with selection criteria for the query, to which are added other mandatory or optional preferences that are determined by the user. The flexible querying module polls the @Web packaging database to retrieve the material that best satisfies the query preferences, and proposes as output a ranking of these materials. The DSS can manage both imprecise and missing data (Destercke et al. 2011). An answer is guaranteed even if no material satisfies the mandatory criteria. This type of tool marks a significant breakthrough, as it had never before been attempted in the field of food packaging. The first step in the process of building a DSS in the field of food packaging is to develop the numerical program that will serve to compute the evolution of food quality in relation to mass transfers in the food/packaging system. Several mathematical models have been developed that combine mass transfer models (based on Fick's laws) with food degradation models, such as the Mickaëlis-Menten equation for respiration or first-order reactions for oxidation (Penicaud et al. 2011). The EcoBioCap numerical model (Guillard et al. 2016) is used to adjust the packaging material to "the strict minimum", i.e. just those mass transfer properties necessary to maintain the protective atmosphere within a given range of values. Mathematical models for food engineering do allow some technical outputs to be computed but are not sufficient for decision making in an industrial world where choice of a packaging material is a multi-criteria decision. To take into account this aspect, the EcoBioCAP tool was developed to choose the most suitable packaging material for respiring produce from a dedicated database by answering bipolar multi criteria querying (currently four criteria considered in the first prototype). Bipolarity refers to the human reasoning that combines information on pros with information on cons to make decisions, choices or

judgments. Some preferences are modeled as constraints for which satisfaction is mandatory, while others are ‘nice-to-haves’ for which satisfaction is optional. Any packaging material that fails to satisfy the constraints is definitively discarded, while preference for a packaging increases the more it satisfies optional nice-to-haves. It is thus natural, in this context, for the querying process to make use of a bipolar approach, as it can handle compound preferences made of mandatory conditions and optional conditions. The web application and short demo videos are available at <https://umr-iate.cirad.fr/equipes/ico/resultats-marquants/ecobiocap-dss>.

We have seen that the EcoBioCap DSS tool is able to use preferences associated with multiple criteria in order to refine the decision. In the previous paragraph, we have mainly discuss technical criteria as optimal permeabilities values. Additional criteria have to be taken into account to obtain a sustainable decision. Those criteria are based on stakeholders’ preferences (by example consumers’ preferences about the packaging’s transparency or packaging’s end-of life) in the context of collective decision making, as different stakeholders might have different preferences. In such a situation, it is important to be able to input preferences that represent as accurately as possible the individual desiderata. To solve this, preferences aggregation techniques, i.e. voting rules, have been designed to merge a set of individual preferences into a unique, "global", preference. The Papow software (Karanikolas et al. 2018, Jedwabny et al. 2019) implements several of these voting rules, such as plurality, k-approval or Borda count, and allows for an in-depth exploration of individual preferences through filtering and clustering of the voters.

However, while aggregating preferences provides a direct answer to the collective decision-making problem, it only offers a partial solution with various shortcomings such as the inability to explain why agents’ preferences differ in the first place. In order to understand these disagreements, it is then necessary to study the justifications behind the preferences, i.e. the arguments an individual might put forward to support her preferences.

The DAMN software (<https://hamhec.github.io/damn/home>) (Hecham et al. 2018) has been implement to meet this need. Indeed, by using logical reasoning techniques, this software is able to compute justified preferences through automatic justification analysis. The process is the following: first, the participants are asked to provide a justification for each of their preferences; then agreements and contradictions between the reasoning steps provided by these agents are automatically detected; finally, the participants can discuss the diverging reasoning steps and potentially change their preferences or justifications. As a result, DAMN provide collectively assessed preferences that can then be aggregated thanks to Papow in order to obtain a unique, collectively justified, preference which can be considered as input of a query which is executed by the EcoBioCap DSS tool presented in Figure 5.

5 International initiatives to share agri-food ontologies and other semantic resources

The ontologies presented in this paper are both research materials and results. As such, it is important to, at a minimum, make them public, or at best make them reusable. The reasons to open our ontologies are many, including:

- data reusability: semantic resources used to structure or document a dataset should be made explicit, be findable, and accessible.
- data interoperability: ontologies should be made public and actionable for the construction of concept mappings which contribute to the semantic interoperability of datasets or systems using them
- research reproducibility and process transparency: (successive versions of) ontologies used in decision support systems or reasoner should be accessible and actionable by anyone willing to run the system

- avoidance of duplicate effort: the expertise and time spent on the formalisation of domain knowledge should be shared with the secondary effect of increasing the quality and modularity of semantic resources.

Recently, a repository and a catalogue have been created to respectively host and reference the semantic resources that are useful to research and industry in the agricultural and nutrition domain. Their use has shown that such domain specific portals contribute to the structuring of the community, and facilitate cross-discipline collaborations. Their development have been initiated or reinforced by initiatives such as Research Data Alliance (RDA) (<https://rd-alliance.org/>) or Global Open Data for Agriculture and Nutrition (GODAN) (<https://www.godan.info/>), presented and discussed at the end of this section.

5.1 AgroPortal and the Agrisemantics Map of Data Standards

Many vocabularies and ontologies are produced to represent and annotate agronomic data. For instances, the Plant Ontology, Crop Ontology, and more recently, the Agronomy Ontology, Food Ontology, or Process and Observation Ontology presented in Section 3.2.

Semantic interoperability is a key issue for agronomy, and the use of ontologies a way to address it (Lehmann et al. 2012), Ontologies have opened the space to various types of semantic applications, to data integration (Buche et al. 2013b), to process and transformation description (Lousteau-Cazalet et al. 2016, Muljarto et al. 2017) or decision support (Guillard et al. 2015).

However, those ontologies are spread out over the web (or even unshared), in many different formats and types, of different size, with different structures and from overlapping domains. Therefore, there is need for a common platform to receive and host them, align them, and enabling their use in agro-informatics applications. There exists a need of a one-stop-shop for ontologies in the agri-food domain enabling to identify and select an ontology for a specific task as well as offering generic services to exploit them in search, annotation or other scientific data management processes. The need is also for a community-oriented platform that will enable ontology developers and users to meet and discuss their respective opinions and wishes.

The AgroPortal project, is a community effort started in 2015-2016 by the Montpellier scientific community (LIRMM, IRD, CIRAD, INRA, Bioversity International) to build an ontology repository for agronomy and related domains. Our goal is to facilitate the adoption of metadata and semantics to facilitate open science and the production of FAIR data. By enabling straightforward use of existing ontologies, we expect data managers and researchers to focus on their tasks, without requiring them to deal with the complex engineering work needed for ontology management.

Mid-2015, by reusing the NCBO BioPortal technology (Noy et al. 2009), we have designed AgroPortal (<http://agroportal.lirmm.fr>), an ontology repository for the agronomy domain but also food, plant, and biodiversity sciences. AgroPortal (Jonquet et al. 2018) offers a robust and reliable advanced prototype that features ontology hosting, search, versioning, visualization, comment, and recommendation; enables semantic annotation, stores and exploits ontology alignments, and enables interoperation with the semantic web. By specifically addressing the requirements of the agronomy community, AgroPortal has kindled an important interest both at the national and international levels. The platform currently hosts 106 vocabularies with more than 2/3 of them not present in any similar ontology repository and 10 private ontologies. We have identified 80 other candidate ontologies that will be loaded in the future to complement this valuable resource. The platform already has more than 100 registered users and some vocabularies are visited more than 100 times per month.

In addition to its core repository of ontology mission, AgroPortal also offers many applicable tools, including a mapping repository, an annotator, an ontology recommender, and

community support features. Our vision was to adopt, as the NCBO did, an open and generic approach where users can easily participate to the platform, upload content, and comment on others' content (ontologies, concepts, mappings, and projects).

However, the current AgroPortal prototype only partially addresses the needs of the community: it is not multilingual, it is limited in terms of ontology alignment capabilities and does not provide semantic-search and retrieval of ag & biodiv data. Addressing some of these issues are among the objectives of a ANR-funded project (starting in June 2019) called D2KAB (Data to Knowledge in Agronomy and Biodiversity – www.d2kab.org).

In parallel, and based on previous work at Food and Agriculture Organization of the UN (UN-FAO), the Agrisemantics Map of Data Standards (<http://vest.agrisemantics.org>) (Pesce et al. 2018) is a catalog of semantic resources of interest for agri-food. This catalog lists any kind of data standards (not only ontologies & vocabularies) and we have developed semi-automatic synchronization mechanisms mainly to import ontologies and vocabularies metadata from AgroPortal to the Map.

The objectives of building the Agrisemantics Map of Data Standards task were: (i) To map currently available open and proprietary standards in use for the exchange of key data on agriculture and nutrition; (ii) To identify where a lack of standards is inhibiting the effective use of agricultural and nutritional data and the best methods for promoting open data standards. A gap analysis was also produced by GODAN Action.

In 2018, both AgroPortal and Agrisemantics Map of Data Standards content were incorporated in the FAIRsharing catalog (<https://fairsharing.org/>) for a larger dissemination.

5.2 Scientific and technical international fora

Exposing semantic resources in catalogues and repositories ensures their findability and accessibility by the community of semantic professionals of the agricultural domain and associated disciplines. But fostering their reuse in other contexts as well as reaching new users need to be supported by networking activities like scientific and technical meetings, and task force groups. The Research Data Alliance (RDA) and the Global Open Data for Agriculture and Nutrition (GODAN) initiative are two examples of international fora that have supported the social part of our work for more than five years. These are complementary to more classical scientific fora we are involved in like the IN-OVIVE International Workshop on sources and data integration in agriculture, food and environment using ontologies, or the AgroSem Track at the Metadata and Semantic Research (MTSR) conference.

For research organisations like ours, RDA represents the international arena where to check for latest advancements on the data side, gain inspiration for new ideas, and validate ongoing approaches. It is also the place to promote the approaches developed in our labs to varied stakeholders of the data chain (application developers, funders, data managers, etc.). Since the launch of RDA in 2013, the Interest Group on Agricultural Data (IGAD) has been a forum for sharing experience and providing visibility to research and work in agricultural data. With over 200 members, it represents stakeholders in managing data for agricultural research and innovation, including producing, aggregating and consuming data. Meetings every 6 months reinforce shared knowledge, technological transfer, and stir new collaborations while working groups created under its umbrella allow to develop a community approach to address data related issues.

In 2017, with a group of public and private organisations including INRA, Irstea, LIRMM, IRD, Embrapa, Wageningen University, UN-FAO, CABI, Agroknow and AgGateway, we launched the Agrisemantics Working Group to allow semantic experts and agriculture stakeholders meet and collaborate. There was the shared idea that results of scientific research and developed methodologies like those described above need to reach a

larger community of adopters. Indeed, while proved to efficiently address some complex data interoperability issues, semantic technologies globally suffer from 1) being perceived as too difficult, requiring advanced competencies, and 2) a lack of tools and methodologies allowing non-semantic experts to seamlessly edit, align, and use semantic resources on their data. A landscaping exercise (<https://www.rd-alliance.org/system/files/documents/Deliverable1%20-%20Landscaping.pdf>) and collection of requirements (Caracciolo et al. 2018, <https://www.rd-alliance.org/deliverable-2-use-cases-and-requirements>) by Agrisemantics substantiated this observation for the agricultural and nutrition community. Based on an open dialog with the community of researchers and practitioners of semantic technologies, we have identified a few basic points to address in order to make their use more widespread and effective. Then, we have translated those points in specific recommendations roles and activities in data stewardship (report under review by RDA secretariat for endorsement).

In the meantime, the GODAN initiative (<https://www.godan.info/>) announced at the Open Government Partnership Conference in October 2013, has contributed to the structuring of the community by supporting working groups (on soil data, nutrition data gap, capacity building, etc.), and specific actions like the Agrisemantics Map of Standards mentioned above. Through the promotion of success stories and the publication of reports and white papers, GODAN advocates for high level political and policy actions that enable action on the ground.

Since 2016, the FAIR principles (Wilkinson et al. 2016) have emerged as a supporting framework on the path of open science. While the FAIR principles are generic - aiming at making data (code, services, etc) more Findable, Accessible, Interoperable, and Reusable - their adoption and implementation require deep anchorage into the domain communities. To achieve this, on the initiative of INRA, Wageningen University and Research, and GODAN, has started a GO FAIR Implementation Network called "Food Systems" (<https://www.go-fair.org/implementation-networks/overview/food-systems/>). Our objective is to advance a global data ecosystem for agriculture and food by implementing FAIR data and services. In particular, building on the outputs of the RDA Agrisemantics Working Group and their knowledge and experience in the field, the French partners of the Food Systems IN will use the Food Systems IN to make their ontologies and other semantic resources more FAIR, as well as the data that use them. We will also collectively address the need for training on semantic technologies and methodologies with the aim of reaching data managers and application developers in particular.

6 Conclusion

We have shown in this article through different French computer science research works and initiatives some of the main locks to face and some proposed solutions using ontologies in data and knowledge integration for decision support in the Agri-food domain. Our aim was not to be exhaustive but to present illustrative examples in the Agri-food domain.

Several questions remain and we are only at the beginning of the data revolution in the Agri-food domain. First, networks of ontologies are promising answers (Muljarto et al. 2017) to link data from primary production to data from secondary production allowing therefore to be able to assess the impact of operations undertaken during primary production on the final quality of food product. Second, the data alignment problem remains a very challenging issues to face in particular in the Agri-food domain composed of many heterogeneous sub-domains that deal with very specific data treated at different scales by many distinct disciplines with their own objectives, scopes and methods. Third, the encounter of two worlds of data will have to be taken into account: the big data flow coming from more and more sophisticated equipment and the experimental data made by human in laboratories with many missing values. In this second kind of data, when trying to estimate missing data using available ones, important domain questions rise: i) what method can be used for the estimation? ii) what available data

should be used? and iii) what are “similar” data? Domain experts need to find answers to these questions and semantic techniques may help. The question *what are “similar” data?* is in particular very challenging in computer science and is investigated more and more in the context of linked data (Halpin et al. 2010; Beek et al. 2018).

References

BATTLE R. & KOLAS K. (2012). Enabling the geospatial Semantic Web with Parliament and GeoSPARQL. *Semantic Web* 3(4): 355-370.

BEEK W., RAAD J., WIELEMAKER J. & VAN HARMELEN F. (2018). sameAs.cc: The Closure of 500M owl: sameAs Statements. In *ESWC 2018*, p. 65-80.

BOWERS S. (2012). Scientific workflow, provenance, and data modeling challenges and approaches. In *J. Data Semantics* 1 (1), p. 19–30.

BUCHE P., DERVAUX S., DIBIE-BARTHELEMY J., SOLER L., IBANESCU L. & TOUHAMI T (2013a). Intégration de données hétérogènes et imprécises guidée par une ressource termino-ontologique. In *Revue d'Intelligence Artificielle* 27(4-5), p. 539-568.

BUCHE P., DIBIE-BARTHELEMY J., IBANESCU L. & SOLER L. (2013b). Fuzzy Web Data Tables Integration Guided by an Ontological and Terminological Resource. In *IEEE Trans. Knowl. Data Eng.* 25(4), p. 805-819.

CARACCIOLO C., AUBIN S., WHITEHEAD B. & ZERVAS P. (2018). Semantics for Data in Agriculture: A Community-Based Wish List. In *Contributions to Management Science* (pp. 340–345). Springer International Publishing. https://doi.org/10.1007/978-3-030-14401-2_32.

CIMIANO P., BUITELAAR P., MCCRAE J. & SINTEK, M. (2011). LexInfo: a declarative model for the lexicon-ontology interface. In *J. Web Sem.* 9 (1), 29–51.

COMPTON M., BARNAGHI P., BERMUDEZ L., GARCIA-CASTRO R., CORCHO O., COX S., GRAYBEAL J., HAUSWIRTH M., HENSON C. & HERZOG A. (2012). The ssn ontology of the w3c semantic sensor network incubator group. In *Web semantics : science, services and agents on the World Wide Web*, 17, 25–32.

COX S. (2011) Observations and measurements-xml implementation. <https://publications.csiro.au/rpr/pub?pid=csiro:EP115858>

DESTERCKE S., BUCHE P. & GUILLARD V. (2011) A flexible bipolar querying approach with imprecise data and guaranteed results. In *Fuzzy Sets Syst* 169:51–64. doi: 10.1016/j.fss.2010.12.014.

DESTERCKE S., BUCHE P. & CHARNOMORDIC B. (2013). Evaluating data reliability: an evidential answer with application to a web-enabled data warehouse. In *IEEE Trans. Knowl. Data Eng.* 25 (1), 92–105.

DOAN A., HALEVY A. & IVES Z. (2012). *Principles of Data Integration*. Morgan Kaufmann 2012, ISBN 978-0-12-416044-6, pp. I-XVIII, 1-497.

ETSI TS 103 410-2 - v1.1.1 (2017). SmartM2M ; Smart Appliances Extension to SAREF ; Part2 : Environment Domain. Rapport interne, ETSI.

GARDNER S.P. (2005). Ontologies and semantic data integration. In *Drug Discov. Today* 10 (14), 1001–1007.

GUARINO N., OBERLE D. & STAAB S. (2009). What Is an Ontology? In Handbook on Ontologies, p. 1-17.

GUILLARD V., BUCHE P. & DESTERCKE S (2015). A decision support system to design modified atmosphere packaging for fresh produce based on a bipolar flexible querying approach. In *Comput Electron Agric* 111:131–139. doi: 10.1016/j.compag.2014.12.010

GUILLARD V., COUVERT O., STAHL V. & BUCHE P. et al. (2016). Validation of a predictive model coupling gas transfer and microbial growth in fresh food packed under modified atmosphere. In *Food Microbiol* 58:43–55. doi: 10.1016/j.fm.2016.03.011.

HALLER et al. (2017). Semantic Sensor Network Ontology W3C Recommendation 19 October 2017 (Link errors corrected 08 December 2017) <https://www.w3.org/TR/vocab-ssn/>.

HALPIN H., HAYES P., MCCUSKER J., MCGUINNESS D. & THOMPSON H. (2010). When owl: sameAs Isn't the Same: An Analysis of Identity in Linked Data. In *International Semantic Web Conference (1) 2010*: 305-320.

HECHAM A., BISQUERT P. & CROITORU M. (2018). On a Flexible Representation for Defeasible Reasoning Variants. In *AAMAS 2018*: 1123-1131.

HOBBS J. & PAN F. (2004). An ontology of time for the semantic web. *ACM Trans. Asian Lang. Inf. Process.* 3(1): 66-85.

HODGSON R. (2016). Quantities, Units, Dimensions and Types (QUDT) Schema - Version 2.0 http://qudt.org/doc/2016/DOC_SCHEMA-QUDT-v2.0.html

IBANESCU L., DIBIE J., DERVAUX S., GUICHARD E. & RAAD J. (2016). PO² - A Process and Observation Ontology in Food Science. Application to Dairy Gels. In *MTSR 2016*: 155-165.

JEDWABNY M., BISQUERT P. & CROITORU M. (2019). Papow Aggregates Preferences and Orderings to select Winners. In *AAMAS 2019*, to appear.

JONQUET C., TOULET A., ARNAUD E., AUBIN S., DZALE YEUMO E., EMONET V., & GRAYBEAL J. (2018) AgroPortal: a vocabulary and ontology repository for agronomy. In *Computers and Electronics in Agriculture* 144 (2018): 126-143.

KARANIKOLAS N., BISQUERT P., BUCHE P., KAKLAMANIS C. & THOMOPOULOS R (2018). A Decision Support Tool for Agricultural Applications Based on Computational Social Choice and Argumentation. In *IJAEIS* 9(3): 54-73.

LEHMANN, R.J., REICHER, R., SCHIEFERA, G. (2012). Future internet and the agri-food sector: State-of-the-art in literature and research. In *Computers and Electronics in Agriculture*. 89, 158–174 (2012).

LEBO T., SAHOO S. & MCGUINNESS D. (2013). PROV-O: The PROV Ontology. W3C Recommendation. <http://www.w3.org/TR/2013/REC-prov-o-20130430/>

LOUSTEAU-CAZALET C., BARAKAT A., BELAUD J.P., BUCHE P., BUSSET G., CHARNOMORDIC B., DERVAUX S., DESTERCKE S., DIBIE J., SABLAYROLLES C. & VIALLE C. (2016). A decision support system for eco-efficient biorefinery process comparison using a semantic approach. In *Computers and Electronics in Agriculture* 127: 351-367.

MCCRAE J., SPOHR D. & CIMIANO P. (2011). Linking lexical resources and ontologies on the semantic web with lemon. In: *ESWC*, vol. 1, pp. 245–259.

- MADIN J., BOWERS S., SCHILDHAUER M., KRIVOV S., PENNINGTON D. & VILLA F. (2007). An ontology for describing and synthesizing ecological observation data. In *Ecological Informatics*, 2(3), 279–296.
- MULJARTO A., SALMON J-M., CHARNOMORDIC B., BUCHE P., TIREAU A. & NEVEU P. (2107). A generic ontological network for Agri-food experiment integration - Application to viticulture and winemaking. In *Computers and Electronics in Agriculture* 140: 433-442.
- NOY N., SHAH N., WHETZEL P., DAI B., DORF M., GRIFFITH N. & JONQUET C. (2009). BioPortal: ontologies and integrated data resources at the click of a mouse." In *Nucleic acids research* 37, no. suppl_2, W170-W173.
- PENICAUD C., BROYART B. & PEYRON S. (2011). Mechanistic model to couple oxygen transfer with ascorbic acid oxidation kinetics in model solid food. In *J Food Eng* 104:96–104. doi: 10.1016/j.jfoodeng.2010.11.033.
- PÉNICAUD C., Ibanescu L., ALLARD T., FONSECA F., DERVAUX S., PERRET B., GUILLEMIN H., BUCHIN S., SALLES C., DIBIE J. & GUICHARD E. (2019). Relating transformation process, eco-design, composition and sensory quality in cheeses using PO² ontology. In *International Dairy Journal*, vol 92, p. 1-10.
- PERROT N., DE VRIES H., LUTTON E., VAN MIL H.G.J., DONNER M., TONDA A., MARTIN S., ALVAREZ I., BOURGINE P., VAN DER LINDEN E. & AXELOS M. (2016). Some remarks on computational approaches towards sustainable complex agri-food systems. In *Trends Food Sci. Technol.* 48, 88–101. [http:// www.sciencedirect.com/science/article/pii/S0924224415002186](http://www.sciencedirect.com/science/article/pii/S0924224415002186).
- PESCE V., TENNISON J., MEY L., JONQUET C., TOULET A., AUBIN S., & ZERVAS P. (2018). A Map of Agri-food Data Standards. In *F1000 Research*, Technical Report 7-177, Global Open Data for Agriculture and Nutrition (GODAN), February 2018.
- POVEDA-VILLALON M., NGUEN Q.-D., ROUSSEY C., CHANET J.-P. & DE VAULX C. (2018). Ontological requirement specification for smart irrigation systems: a SOSA/SSN and SAREF comparison. In *Proceedings of the 9th International Semantic Sensor Networks Workshop SSN2018*, Monterey, USA,
- ROCHE C., CALBERG-CHALLOT M., DAMAS L. & ROUARD P. (2009). Ontoterminology – a new paradigm for terminology. In: *KEOD*, pp. 321–326.
- ROUSSEY C. & ABDERRAHMANI GHORFI T. (2018). Annotation sémantique pour une interrogation experte des Bulletins de Santé du Végétal. Dans les Actes des 29e Journées Francophones d'Ingénierie des Connaissances IC 2018, adossée à la 11e Plate-forme Francophone d'Intelligence Artificielle, 2018, Nancy, p 37-52. <https://hal.archives-ouvertes.fr/hal-01839545>
- SEEDAH D.P.K., SANKARAN B. & O'BRIEN W.J. (2015). Approach to classifying freight data elements across multiple data sources. In *Transp. Res. Rec.* 2529 (18), 56–65.
- SUÁREZ-FIGUEROA M.C., GÓMEZ-PÉREZ A., FERNÁNDEZ-LÓPEZ M. (2012). The NeOn Methodology for Ontology Engineering. In *Ontology Engineering in a Networked World 2012*, p. 9-34.
- WILKINSON M. D., DUMONTIER M., AALBERSBERG Ij. J., APPLETON G., AXTON M., BAAK A. & MONS B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. In *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.
- WOLFERT J., VERDOUW C., VERLOOP C. & BEULENS A. (2010). Organizing information integration in agri-food – a method based on a service-oriented architecture and living lab approach. In *Comput. Electron. Agric.* 70 (2), 389–405.