A Software Application with Ontology-Based Reasoning for Agroforestry

Raphaël Conde Salazar\(^1\), Fabien Liagre\(^2\), Isabelle Mougenot\(^3\), Jéôme Perez\(^1\), and Alexia Stokes\(^1\)

\(^1\) AMAP, University of Montpellier, CIRAD, CNRS, INRAE, IRD, Montpellier, France
\(raphael.condesalazar@free.fr\)

\(^2\) AGROOF, Anduze, France
\(liagre@agroof.net\)

\(^3\) UMR 228 ESPACE-DEV, Espace pour le Développement, University of Montpellier, Montpellier, France
\(espace-dev@ird.fr\)

Abstract. Agroforestry consists of combining trees with agriculture, both on farms and in the agricultural landscape. Within a context of sustainable development, agroforestry can improve soil conservation and reduce the use of toxic chemicals on crops, as well as improving biodiversity. Interdisciplinary by nature, the field of agroforestry mobilizes a large body of knowledge from environmental and life sciences using systemic approaches. In this framework, field observation data are acquired in partnership with several categories of stakeholders such as scientists, foresters, farmers, breeders, politicians and land managers. For data management efficiency, we propose the software application AOBRA (a software Application with Ontology-Based Reasoning for Agroforestry). The core of AOBRA is a domain ontology called “Agroforestry” which serves as a basis for capitalizing and sharing knowledge in agroforestry. By exploiting the capabilities of inference and linkages to other areas of expertise on the Web offered by the use of an ontology model, it aims to provide a broad view of agroforestry designs, and to allow the comparison between different spatial layouts of trees and crops.

Keywords: Agroforestry · Knowledge management system · Ontology · OWL

1 General Introduction

The development of intensive agriculture in Europe has led to a gradual depletion of soil over the last 50 years [2]. Combined with modifications in climate, leading
to more frequent extreme weather events, crop yields are decreasing [11]. Never-theless, a worldwide increase in the human population and the necessity to feed up to 10 billion people in 2050 has resulted in the creation of a new paradigm in agriculture, to enable the preservation of resources and use of sustainable techniques, whilst increasing crop production. Among the proposed alternatives to intensive agriculture, is agroforestry. Agroforestry consists in the reintroduction of trees in the agricultural sector. Its most common forms are i) agrisilviculture, where trees intended for wood production are intercropped with cereals such as maize (Zea mays) and barley (Hordeum vulgare) and ii) sylvopastoralism, when silvicultural and pastoral activities are integrated in the same area. Agroforestry is not a simple planting of trees, but the close association between trees and agriculture that provides a beneficial synergy [5]. This synergy allows the limitation of inputs such as water, fertilizers and pesticides with a view to sustainable development. To better manage resources within a modern and integrated agroforestry, we need diverse data, indicators and guidelines, to assist farmers in the transition from monospecific to plurispecific agriculture.

Field observations and data collected by different stakeholders such as biologists, foresters, farmers and breeders, in the framework of scientific agroforestry experiments have been accumulating [9]. The management and reuse of these data are made difficult by the multiplicity of media and formats used and by the diversity of stakeholders and their professional vocabularies. In addition, agroforestry studies require systemic approaches to understand in particular how to better manage a site in response to climate change, pests, or soil pollution. These responses must take into account the close links with other fields of knowledge such as climatology, zoology and pedology. To use and share these data and to report on the development and efficiency of the structural organisation of an agroforestry plot, we propose an application AOBRA, that used ontology-based reasoning. AOBRA will provide a global overview of the experiments already carried out and will allow the comparison of data in terms of tree and crop production. “Agroforestry” is the core ontology of the application, it will enable to describe the elements and its organization for the most common patterns available in agroforestry. In this paper we present a general view of the system, followed by the “Agroforestry” knowledge model. We will also discuss the implementation of AOBRA and the technological choices made. Finally, we illustrate the functioning of AOBRA using data from the monitoring of tree biomass in agroforestry plots at the Restinclières estate in the South of France.

2 Application Overview

Our approach consists of three parts (Fig. 1) described below:

**Modelling:** The data provided by farmers and researchers are in the form of an Excel sheet or database organized in a classical form of relational database. At this stage, we will transform these data into formal descriptions that can now be used as semantic data. These data are stored in an OWL serialization
format. The adoption of the semantic web standards will help to link easily our data with other knowledge database available of the web.

**Reasoning activities:** Inference mechanisms are applied to the semantic data in order to deduce new knowledge. During this phase, the data are aggregated in space and time, and several calculations are carried out, leading to the consolidation of quantitative data for the creation of qualitative indicators.

**Outcomes:** The data inferred from the reasoning activities are new semantic data that can be reused to participate in new deductions, thus amplifying the creation of new knowledge. Native and inferred data are then extracted using the SPARQL query language, to answer user queries via the application interface or to feed into other applications (e.g. a website via APIs).

During the modeling phase, we used the “Agroforestry” framework ontology that we have developed. The “Agroforestry” model use not only the physical description of the biotic and abiotic elements present on a agroforestery plot but also the interactions between them and the impact on the plot.

### 3 The “Agroforestry” Model

As presented in green in Fig. 2 the **StructuralElement** class is central. We distinguish the separate **SimpleElement** and **CompositeElement** daughter classes, which allow us to state that one element will admit other structural elements within it, or on the contrary, be a “leaf” element of the model. These modeling elements are disconnected from any temporal consideration. The **StructuralElement** class is associated via the `hasProperty` association to the **PropertyOfElement** class, that allows the enriching of the structural elements with various characteristics that evolve over time. The **PropertyOfElement** property is this time reified and will be specialized by exploiting different terminological references to the example of the total plant size in Trait Ontology (TO:1000012, named “whole plant size”). A superclass named **Resources** is used to generalize **StructuralElement** but also **PropertyOfElement**. The interest of this superclass is to be able to semantically enrich any resource of the model with terms
borrowed from ontoterminalogical resources made available on the Web. For this purpose, the Term class also specializes the Resource class and its instances complement and standardize the way of describing the other resources in the model. For example, the notes within the UML class diagram (Fig. 2) include terms from terminology ontologies that can be used to qualify the elements of an agroforestry site, as well as their properties and interrelationships. We created the “Agroforestry” knowledge base using existing ontologies and terminologies model, that have already been validated.

Integration of the SOSA ontology SOSA (Sensor, Observation, Sample, and Actuator)

In order agroforestry practitioners to capitalize on the various observations made in the field, the SOSA ontology [8] was integrated with the “Agroforestry” ontology. If some properties of our agroforestry elements have immutable values (such as site location), other properties (such as soil acidity through the measurement of soil pH), are quantified or qualified using a specific observation procedure. These data may vary over time and are dependent on the measurement method. It is essential to take into account this variability to obtain a rigorous outcome when comparing and analysing data from multiple observations. SOSA is the central (but self-contained) module of the SSN (Semantic Sensor Network) [6] ontology which allows the describing of “sensors” (including humans) and the observations acquired by these sensors. The SOSA core ontology places the observation (sosa:Observation) at the centre of its model: an observation then informs the value of a descriptive property for an element of interest, at a given instant or time interval. The observation is linked to an individual of the class sosa:FeatureOfInterest by the relation sosa:hasFeatureOfInterest. The observed property of the element (class sosa:ObservableProperty) is related to an observation by the relation sosa:observedProperty and the value of this property is related to observation by the relation sosa:hasResult. In the “Agroforestry” model, the StructuralElement class specialises in sosa:FeatureOfInterest, and will naturally benefit from all the modelling around the notion of observation. We can therefore, organize the information around the growth of a tree, considering it as an instance of the sosa:FeatureOfInterest class, and inform its size at regular intervals through several observations. The ssn:Property class (superclass of sosa:ObservableProperty) and extended by the PropertyOfElement class so that it can write in a complementary way the properties involved in the description of the key elements of agroforestry. The idea is to be able to have in the longer term, pairs of elements/collections of properties relevant to agroforestry and thus facilitate their reuse by the agroforestry community. We also extended the PropertyOfElement through RelationshipBetweenElements, to support relationships other than structural relationships between elements of agroforestry management. These relationships can be valued and dated in time, and will allow us to capture the interactions between biotic elements found in the same plots. Pilot sites in agroforestry allow us to highlight the effects of agroforestry practices through experiments conducted on test plots.
The classes `sosa:Sample` and `sosa:Sampling` will therefore also be useful; an experimental plot can be seen as a `sosa:Sample` individual. In purple inside Fig. 2, we resume the UML class diagram of the main SOSA classes, some of which are extended in the “Agroforestry” model.

### Extension of GeoSPARQL Ontology and OWL-Time Ontology

A double spatial and temporal dimension is necessary for any entity of the “Agroforestry” model if one wants to account for the evolution of a system where proximity relationships over time are of the utmost importance. Agroforestry systems and in particular, agrosilvicultural systems, are by their nature spatial, where the place of each element (tree, cereal plant and tree line) is of prime importance in the context of plant-plant interactions. The geographical location of the data of interest is able to bring together and link data from different fields such as hydrology or pedology. As an example, a river could border an agroforestry plot and so limit competition for water between trees and cereal plants. GeoSPARQL [1] is proposed by the OGC (Open Geospatial Consortium) international consortium, to provide the necessary elements for the representation and querying of spatialized data in the context of the semantic web. GeoSPARQL is organized in a modular way, and can offer services both to systems based on qualitative spatial reasoning and to those based on quantitative spatial calculations. Quantitative spatial calculations, such as distance calculations, require precise knowledge of the geometry of the elements studied (Euclidean geometry), whereas systems based on qualitative spatial reasoning rely on topological relationships between elements. These relationships (such as adjacency, intersection or inclusion) are, for example, described through the formalism “region connection calculus” (RCC8) [12]. GeoSPARQL also offers the possibility of associating one to several geometries, for example a point or a polygon, to any geographical entity through the `geo:Geometry` class, that represents the super class of all possible geometries. In the “Agroforestry” model, the `StructuralElement` class is subsumed by `geo:Feature` allowing us to take advantage of all the modeling proposed by GeoSPARQL around geometries. Time is an equally important notion in the context of agroforestry developments in which, for example, the annual seasonality of crops is confronted with the multi-annual life of trees. In order to better understand the interactions between trees and crops over time, we must precisely define the planting dates and the periods of the specific presence of each agroforestry element. Therefore, the `StructuralElement` class subsumed by `sosa:FeatureOfInterest` will be able to mobilize the modeling retained in SOSA, i.e., to maintain a `sosa:phenomenonTime` type relationship with a `time:TemporalEntity` temporal entity of the OWL-Time ontology [7]. The class `time:TemporalEntity` is then specialized in `time:Instant` and `time:Interval`, and so it is possible to introduce an agroforestry management element in a period of time, based on a precise time interval; or to define a temporal pattern for a specific observation (Fig. 2).

We have presented both the conceptual model corresponding to the minimum foundation for structuring knowledge specific to forest management, and portions of conceptual models of framework ontologies conveying three concepts that are...
Fig. 2. Integrated class diagram: “Agroforestry”, SOSA, GeoSPARQL and OWL-Time models and their relationships

of primary interest, i.e., observation, space and time. Figure 2 illustrates the articulations defined between these different models SOSA, GeoSPARQL, OWL-Time and “Agroforestry”, for the specific needs of agroforestry management. The color code allows to visualize the origin of each modeled element.
4 Implementation

For the implementation, we have chosen the Java Jena framework [10] for RDF graph editing and management and more specifically for:

- Permanent storage of RDF triplets in the TDB triplestore.
- The support of RDFS and OWL formalisms for the construction of ontologies and SPARQL for their consultation.
- The choice among different reasoners that take advantage not only of OWL and RDFS formalisms, but also of rule-based languages such as SWRL or the specific language “Jena Rule”.

For the reasoning activities, the general process is presented in Fig. 3. The inference mechanism takes place in several successive steps:

**First step: Inferences with the PELLET tool and OWL axioms**

The original model comprising a merger between the “Agroforestry” ontology (Tbox) and the data on amenities (Abox) is submitted to the PELLET inference engine. This procedure will have the effect of making explicit some implicit information by generating additional RDF triplets based on the properties (e.g., reflexivity and transitivity) of the predicates initially present in our model. For example, the transitivity and reflexivity of the `bfo:hasPart` relation and its inverse `bfo:partOf`, will be used to make explicit the content/container relations not expressed in the original model (Fig. 4). This mechanism of inference is made possible by the use of the OWL language which provides a formalization for the description of the properties of a predicate used in an RDF triplet. OWL is inspired by description logic and so has formal semantics defined in first order logic.
Second step: inferences based on the “JENA Rules” The possibilities of inference based on the OWL language are limited. In the interest of generating new data, we have developed a rules-based mechanism. Jena has its own rule-based reasoner (JENA Rules). JENA Rules are in line with Horn clauses (head and body) and have a syntax close to SWRL rules. Functions created by users or other programs (such as the statistics oriented R software) can be called up by the rules, allowing complex calculations to be performed, the results of which can be used when applying the rules. The algorithms in the form “IF . . . THEN” will subsequently be transposed into JENA Rules. The transition to JENA rules is facilitated by the fact that these characteristics are represented through properties that are usually described in a standardised manner in terminology repositories. Thus, instances of the class afy:StructuralElement are associated with a set of properties via the relation ssn:hasProperty. Similarly, the instances of the sosa:Observation class are linked to a property via the sosa:observedProperty relation. However, the instances of the class afy:StructuralElement are linked by the relation sosa:isFeatureOfInterestOf to the instances of the class sosa:Observation to translate the link between the observations and the observed entity. Figure 5 shows the example of an algorithm for the identification of the properties to be studied and the expression of the JENA rule derived therefrom (Fig. 6).

Third step: data extraction The SPARQL language allows us to query and manage graph data, whether native or inferred.

Fig. 4. “Inference with the PELLET engine”: Example of the exploitation of the transitivity and reflexivity of the relation hasPart and its inverse PartOf.
For all elements, properties and observations, IF there is an element who have a property AND that there is an observation whose observe this property AND that this observation has as feature of interest this element THEN this property is a property to study.

Fig. 5. Algorithm for retrieving the studied properties.

```
[RetrievingProperties: (?structuralElement ssn:hasProperty ?property)
 (?observation sosa:observedProperty ?property)
 (?observation sosa:hasFeatureOfInterest ?structuralElement)
 ->
 (?property :isPropertyToStudy 'true'^xsd:Boolean) ]
```

Fig. 6. JENA rule to identify the studied properties.

5 Case Study: The Restinclières Agroforestry System

Our application was tested in Restinclières domain [3], that has the particularity to be a research site that enables to acquire tree related data as well as data on the performance of agriculture practices in experimental and controls plot [4]. At this field station, we can compare the growth of trees without crops (forest controls) and crops without trees (agricultural controls). The data used were collected between 2015 and 2018 and collected from the agroforestry plot (PA3AF) and the associated control plot (PA4TF). The dendrometric data available for these two plots are total tree height, the height from the base of the trunk to the lower branches of the crown (height at which tree is pruned) and diameter at breast height (DBH, at 1.3 m). These measurements were taken annually at the same time over three consecutive years in 2015, 2016 and 2017. Due to their close proximity, it is considered that the climatic and soil conditions are identical for both plots. All trees are walnut hybrids (Juglans nigra x J.regia) and were planted in the same year (1994), for the production of ornamental wood. There are 140 trees on the PA3AF plot and 224 trees on the PA4TF plot. The data are made compliant with the “Agroforestry” model.

As a first analysis, we launched a statistical evaluation of the data. We focused on the values of the trees different characteristics that are recorded as observations.

- First, we retrieve all values of the same characteristic shared by all the trees of the same plot for one year.
- Second, we create of a vector with the retrieved values of the same characteristic shared by all the trees of the same plot for a year. The different values are grouped as vectors (c(33.1, 24.2, 42.3...)) by plot and year.
- Third, we calculate the quantiles distribution of a vector with the retrieved values of the same characteristic shared by all the trees of the same plot for one year. These vectors are then processed by functions of the R software used for statistical data processing. The later is carried out by different JENA rules which are able to call external functions during their execution.
Other JENA rules using other R-functions allowed us to also calculate the median, standard deviation and variance of tree property values by plot and year.

Our results showed and in particular when looking at the quantiles distribution of tree height and diameter for the agroforestry plot PA3AF and the forest control PA4TF (Table 1) that tree growth more in the agroforestry plot.

**Table 1.** Distribution of quantiles of tree diameters and heights for plots P3AF and P4TF for the year 2016.

| Statistics 2016 trunk diameters | Quantile 25 | Quantile 50 | Quantile 75 |
|---------------------------------|-------------|-------------|-------------|
| Plot <http://www.afy.fr/Restinclieres/PA3AF> | 25.6 | 29.7 | 32.5 |
| Plot <http://www.afy.fr/Restinclieres/PA4TF> | 14.9 | 19.9 | 23.9 |

| Statistics 2016 trunk heights | Quantile 25 | Quantile 50 | Quantile 75 |
|---------------------------------|-------------|-------------|-------------|
| Plot <http://www.afy.fr/Restinclieres/PA3AF> | 410.0 | 420.0 | 440.0 |
| Plot <http://www.afy.fr/Restinclieres/PA4TF> | 298.7 | 380.0 | 420.0 |

In light of these results, we continued our analysis for the **tree biomass**. We investigated whether we observed a stronger tree growth under agroforestry conditions. We studied the tree biomass by diameter category on the total tree biomass of a plot in order to confirm this phenomenon of increased production. Improved growth of agroforestry trees would imply a useful reduction in the operating time required to market the trees for the production of ornamental wood. There are several tree categories used in the ornamental timber trade, that are determined by the DBH. The categories add to the knowledge base and are defined using JENA rules. For a specified year and plot, we calculate the ratio between the tree biomass for a given category and the total tree biomass. For the calculation of this ratio, it is necessary to aggregate the calculation of biomass at the spatial and temporal level. Therefore, we obtain confirmation that agroforestry-grown trees grow faster than those in monospecific forestry plots, as shown by the ratio of average trees (Category II, Table 2) The presence of Category III trees for the PA3AF agroforestry plot and their absence for the PA4TF forest control reinforces this result over the three years studied.

**Table 2.** Ratio of the biomass of one tree category to the biomass of all trees per plot per year.

| Ratio of biomass of trees eligible for trading Category II/ total biomass per plot for the year 2015 |
|-----------------------------------------------|-------------|
| Plot <http://www.afy.fr/Restinclieres/PA3AF> | 0.78 |
| Plot <http://www.afy.fr/Restinclieres/PA4TF> | 0.46 |
To determine why there was a large difference in growth of trees between the agroforestry plot compared to the forest control plot, even though meteorological and soil conditions were the same, we examined whether the location of a river (River Lez) adjacent to plots, could influence tree growth. Table 3 show the mean value of the distances between the trees and the River Lez for each plot, tree category and year (Table 3). For the control forest plot, the average distance to the river was slightly lower for the tallest trees categories (38 m for category II versus 46 m for category I). On the contrary, for the agroforestry plot, the ratio was reversed, as the tallest trees were the furthest from the river (187 m for category III versus 173 m for category II and 169 m for category I). Taking into account the additional fact that the trees in the forest control were on average closer to the River Lez than their counterparts in the agroforestry plot (40 m versus 170 m on average), we can reject the hypothesis that the proximity of the river has an effect on tree growth. Results suggest therefore that other mechanisms occur that enhance agroforestry-grown trees, such as less competition for resources and access to fertilizer applied to the seasonal crops.

Table 3. Average distances between trees and the river Le Lez by plot, tree category and year.

| Plot                           | Category I | Category II | Category III |
|--------------------------------|------------|-------------|--------------|
| [http://www.afy.fr/Restinclieres/PA3AF>](http://www.afy.fr/Restinclieres/PA3AF> | 169.6      | 173.4       | 187.3        |
| [http://www.afy.fr/Restinclieres/PA4TF>](http://www.afy.fr/Restinclieres/PA4TF> | 46.0       | 38.46       |              |

Our knowledge base showed that agroforestry-grown trees had enhanced height and biomass production compared to forestry-grown trees. For agroforesters, this result means that the end product will be obtained more quickly, and so the economic value of individual trees will be higher. The agroforester will also benefit from the intercropped cereals. Using our knowledge base, it is also possible to estimate the relative yield of successive crops grown on the agroforestry plot (*Pisum sativum*, pea in 2016, *Triticum durum*, durum wheat in 2017 and *Hordeum vulgare*, barley in 2018), compared to their agricultural-grown controls. Analysing data of relative yields measured at harvest time, we found that the there was a decrease in yield for agroforestry-grown crops (peas - 18%, durum wheat - 34% and barley - 39%). The variability of these yields depend partially on tree age (the trees increasingly shade the crops as they grow) and motivated us to set up indicators combining agricultural yield and tree growth evolution. The constitution of simple rules on the principle of Horn rules and supported by reusable user functions, allows us to infer not only indicators, but also new observations such as those concerning the category of a tree that can be exploited again by other rules.
6 Conclusion

Our case study demonstrated that an application centered around a knowledge modele can provide evidences for agroforestery management. We were able to generate indicators for size and crop yields from an agroforestery plot, and to compare the data with control plots or other (agro)forestry plots with similar conditions (i.e., pedology, climatology, crop and identical planting period) using available data and ontologies. We chose to use ontologies already developed and selected those that were recognized and adopted by various communities of experts. In this case study, we used the SOSA framework ontology to express the observations and extracted the definitions of our elements and properties in terminological ontologies such as Agrovoc. Thus, the particularity of our model is to be above all a generic model, usable by all stakeholders and visible on the semantic web. Our application and model may one hand aid agroforesters, stake-holders and decision makers in managing ressources and taking decisions when planning their future tree-crop mixtures and on the other hand any individuals interested in the elements description on these plots.

References

1. Battle, R., Kolas, D.: Enabling the geospatial Semantic Web with Parliamentand GeoSPARQL. Semantic Web 3(4), 355–370 (2012). https://doi.org/10.3233/SW-2012-0065, https://content.iospress.com/articles/semantic-web/sw065
2. Calame, M.: Comprendre l’agroécologie: origines, principes et politiques. ECLM (2016)
3. Domaine de Restinclières: Domaine de Restinclières et agroforesterie (2020). https://restinclieres.herault.fr/549-l-agroforesterie.htm
4. Dufour, L.: Tree planting and management in agroforestery. In: Agroforestery for sustainable agriculture, p. 480 p. Burleigh Dodds Science Publishing, Cambridge (united kingdom) (2019). https://hal.archives-ouvertes.fr/hal-02439981
5. Dupraz, C., Liagre, F.: Agroforesterie: des arbres et des cultures. France Agricole Editions (2008)
6. Haller, A., et al.: The modular SSN ontology: a joint W3C and OGC standard specifying the semantics of sensors, observations, sampling, and actuation. Semant. Web 10(1), 9–32 (2018). https://doi.org/10.3233/SW-180320
7. Hobbs, J.R., Pan, F.: An ontology of time for the semantic web. ACM Trans. Asian Language Inf. Process. (TALIP) 3(1), 66–85 (2004). https://doi.org/10.1145/1017068.1017073
8. Janowicz, K., Haller, A., Cox, S.J., Le Phuoc, D., Lefrançois, M.: SOSA: a lightweight ontology for sensors, observations, samples, and actuators. J. Web Semant. 56, 1–10 (2019). https://doi.org/10.1016/j.jwebsem.2018.06.003
9. Labelle, R.: Ten years of work in agroforestry information and documentation. Agroforestry Syst. 5(3), 339–352 (1987)
10. McBride, B.: Jena: a semantic web toolkit. IEEE Internet Comput. 6(6), 55–59 (2002)
11. O’Neill, B.C., et al.: IPCC reasons for concern regarding climate change risks. Nat. Climate Change 7(1), 28–37 (2017). https://doi.org/10.1038/nclimate3179
12. Randell, D.A., Cui, Z., Cohn, A.G.: A spatial logic based on regions and connection. KR 92, 165–176 (1992)