Intelligent Environments 2017
C. Analide and P. Kim (Eds.)

doi:10.3233/978-1-61499-796-2-69

An Ontology-Based Approach for an Efficient Selection and Classification of Soils

Andrés MUÑOZ[a,1], José M. SORIANO-DISLA[b], Leslie J. JANIK[c]

[a]*Department of Polytechnic Sciences, Universidad Católica San Antonio de Murcia, Murcia, Spain. E-mail: amunoz@ucam.edu*

[b]*Sustainable Use, Management and Reclamation of Soil and Water Research Group (GARSA). Department of Agricultural Science and Technology, Universidad Politécnica de Cartagena. Paseo Alfonso XIII 48, 30203, Cartagena, Spain. E-mail: martin.soriano@upct.es*

[c]*CSIRO Environmental Contaminant Mitigation and Technologies Program, CSIRO Land and Water, Waite Campus, Waite Road, Urrbrae, 5064, South Australia, Australia. E-mail: les.janik@csiro.au*

Abstract. Soil libraries represent an invaluable resource in terms of research, management or planning. However, the access to such libraries and selection of soils is often tedious and time consuming thus limiting their usefulness. In this study we propose an ontology-based approach for an efficient and intuitive selection and classification of soils. For this test, a soil library of 458 soils from Australia was used. An ontology was then developed to model the fundamentals concepts and relationships found in the data. The basic capabilities of the ontology are shown to select samples with certain values for a number of attributes. In addition, an inference process known as realization is tested to automatically assign individuals to concepts, in our present case to a soil texture class or soil order (the latter known as soil classification). Results show the potential of ontology approaches to select samples from large libraries in an efficient and intuitive way. In addition, and through the use of reasoning processes, we were able to classify soils from different orders and textural classes with accuracy higher than 80% in most cases. This represents an additional application of ontology approaches to produce hidden data from the original data set.

Keywords. Knowledge representation, Ontology, Reasoning, Soil classification, Soil order, Soil selection, Soil texture.

1. Introduction

Soils represent one of the most important components of earth (FAO, 2015; FAO and ITPS, 2015). Adequate characterization is crucial for a good management and allocation of uses. Soil characterization frequently requires the analysis of a number of chemical, physical and biochemical properties. This is performed at a high sample density since soils are highly variable spatially, with depth and over time (Nocita et al., 2015). This has resulted in the development of soil libraries at different scales (i.e. local, regional, national and international) with associated values for a number of properties. The access to soil samples within soil libraries represents an invaluable resource for soil management, land use and planning, research and monitoring (Nocita et al., 2015).

[1] Corresponding author: amunoz@ucam.edu

In order to have access to these benefits there is the need of selecting samples with the required attributes in an efficient, intuitive, easy to use manner. However, such a method is yet to be developed with traditional tools being tedious and time-consuming. In this context, the use of Semantic Web technologies (Berners-Lee et al., 2001) such as ontologies may represent a compelling alternative to manage soil data in a more automatic and intelligent manner. In a nutshell, ontologies are a formal description of a domain by means of axioms (i.e., assertions representing facts on the domain). These axioms allow defining concepts, relationships among concepts and specific individuals of those concepts.

Due to their formal properties, ontologies offer two main advantages when dealing with any type of data. Firstly, data represented through ontologies can be easily exchanged among different communities since it is expressed in a common and shared vocabulary (Gruber, 1993). It means that different computer applications can access the data in the same manner. Secondly, different reasoning processes or *inferences* (Horrocks, 2008) can be performed on the data using ontology reasoners. In particular for soil data it is interesting to apply a kind of inference known as "realization", namely the automatic assignment of soil samples from a soil profile to a soil order (e.g., Calcarosols or Chromosols; known as soil classification) or to a soil texture (e.g., clay or loam) according to the characteristics of such samples. Soil classification and soil texture class are crucial attributes associated to soil samples which guide soil management and planning.

The modeling of soil libraries using ontologies enables selecting samples by means of queries posed in a specific language, The advantage in this case with respect to traditional database queries resides in that ontology queries can retrieve not only explicit asserted data, as in databases, but also inferred data from the reasoning processes, thus offering possible implicit or *hidden* data derived from the original dataset. Moreover, queries on an ontology model can be more expressive (and more complex, too) than in traditional databases because of the semantics associated to the ontology language (Horrocks, 2008). Thus, the main objective of this study is to test the usefulness of ontology approaches for an efficient selection of soils with characteristics as required by the application. It is also an objective to test such approach for soil classification and allocation of textural class purposes.

The manuscript is organized as follows. Section 2 reviews the state-of-art of this research topic. Section 3 describes our methodology for soil selection and modeling using ontology approaches. Section 4 shows some application examples of the proposed soil ontology for soil selection and classification. Section 5 highlights the conclusions of this study.

2. Related work

Most of the literature related to the application of ontologies in agriculture studies is devoted to achieving interoperability among heterogeneous agricultural libraries. In (Aparício et al., 2005) it is proposed to use ontologies to integrate different relational databases related to agriculture data. Ontology is devised as a software layer to achieve interoperability among the terms defined in each database. Some concepts in the ontology, especially related to soil, are only briefly explained; however the ontology itself is not shown. Moreover, authors only focus on the use of ontology as a common vocabulary and no attention is given to the reasoning capabilities as proposed in this manuscript. Athanasiadis et al. (2009) offer a more complete approach. The authors develop several interrelated small ontologies to define crop, farm and product concepts among others, as well as agricultural activities and policy assessment. Apart from achieving interoperability among databases, authors also highlight the benefit of cross-programming gained with ontologies for simulation processes in programming languages such as Java or C#, for example. Again, reasoning capabilities are not taken into account.

Similarly, but offering more advances features, Sánchez–Alonso & Sicila (2009) study how to convert AGROVOC, a well-known agricultural vocabulary, into an ontology. In particular, they focus on designing an ontology for fertilization purposes. They take into account reasoning processes enabled by the ontology, in particular for the classification process; i.e., computation of all the hierarchical relationships among the concepts in the ontology. Two case studies are given to show the application of the ontology for educational purposes. While the development of the ontology model is well explained, the domain and the purpose of their developed methodology are totally different to the one proposed here.

A second approach found in the literature consists in the design of ontology-based frameworks to integrate and manage different sources of data related to agriculture. In (Beck et al., 2010) authors use an ontology-based simulation environment named Lyra to build ontologies modeling complex soil–water and nutrient management systems. Ontologies are created using web-based visual tools with the aim of representing terms and equations needed to represent the management systems. These ontologies are then exported to mathematical formats to simulate the dynamics of the aforementioned systems in other programs. Although the methodology proposed in the manuscript offers a complete process to simulate agricultural environments, ontologies are used as a mere vehicle to represent system equations. Moreover, the reasoning process is performed by mathematical tools after exporting the ontology model to XML-based formats, thus losing the semantics in the ontology model. Shoaib & Basharat (2010) propose a different approach for obtaining a framework to integrate agricultural data into ontologies. While a complete framework architecture is presented, the ontology in which the system is based on is not shown. This ontology is claimed to model concepts such as soil, water, crop and production information. However, no case of study to neither illustrate its usage nor mention on reasoning processes are given in the paper.

3. Materials and methods

3.1. Soil samples

The samples used in this study were obtained from the Australian CSIRO National Soil Archive (CNSA), and can be found in http://www.asris.csiro.au/. The final selection comprised 80 soil profiles (n = 458 samples) from South Australia (66 %) and New South Wales (34 %). Soil samples were dried at 40°C and sieved < 2 mm. Samples were sourced from variable depths, from the surface down to 180 cm. Samples represented 9 soil orders (Australian soil classification) which are commonly used for cropping in Australia, mostly Calcarosols, Chromosols, Dermosols, Sodosols and Vertosols. Minor contributions of Dermosols, Kandosols, Kurosols, Ferrosols and Tenosols were observed.

3.2. Soil laboratory analysis

The following properties were determined by the methodology described in Rayment and Higginson (1992) and Rayment and Lyons (2011): Exchangeable bases calcium (Ca^{2+}), magnesium (Mg^{2+}), potassium (K^{+}) and (Na^{+}), alcoholic 1M ammonium chloride at pH 8.5, pre-treatment for soluble salts, and exchangeable bases by compulsive exchange, no pre-treatment for soluble salts; cation exchange capacity, automated determination of ammonium and chloride ions, and compulsive exchange, no pre-treatment for soluble salts; electrical conductivity (EC), 1:5 soil/water extract; pH, 1:5 soil/water suspension; organic carbon, Walkley and Black; bulk density (BD) and drained upper limit Moisture (DUL), volumetric APSRU (Agricultural Production System Research Unit) in situ methodology (Burk & Dalgliesh, 2013); saturated moisture (SAT), calculated from BD (Burk & Dalgliesh, 2013); particle size distribution (clay, silt and sand), hydrometer method (Gee & Bauder, 1986); total carbon and nitrogen (TC and TN), Leco analyser following the method proposed by Matejovic (1997).

3.3. Soil ontology

A basic ontology to model the fundamental concepts and relationships found in CNSA data is proposed (see Figure 1). The main concepts are Soil, SoilLayer, SoilOrder, SoilTexture and Location. SoilOrder is specialized following the soil orders stated in section 3.1. All these soil orders are modeled as disjoint among them (i.e. the same soil sample cannot be classified as two different soil orders). Regarding the texture class hierarchy, our proof-of-concept reduces the classes to clay, loam and sand.

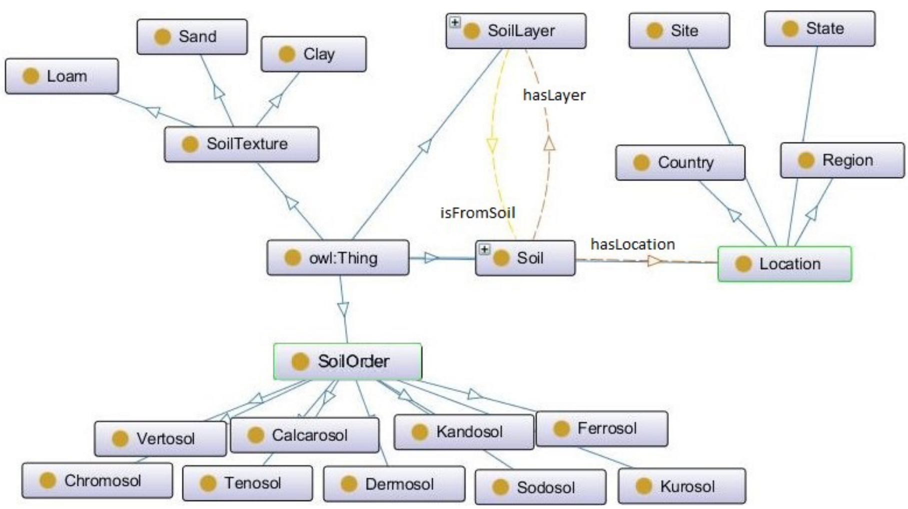

Figure 1. Soil ontology (hierarchical view).

The following relationships among soil concepts have been taken into account, as shown in Figure 1: A Soil has several SoilLayers (and inversely, a SoilLayer is from one and only one Soil) and a Soil has different Location that may range from a specific site to a country. In section 3.4 it will be described how any SoilLayer is related to a SoilTexture and how any Soil is related to a SoilOrder by means of the realization reasoning process.

Some important data properties have been defined for each concept. Thus, some Soil properties include ID, latitude and longitude. Regarding the SoilLayer concept, it includes properties such as layer depth (ranging from 1 to 4, being 1 the most superficial layer and 4 the deepest layer: typically 0-15 cm,15-30 cm, 30-60 cm and > 60 cm, respectively), as well as physical and chemical attributes described in section 3.2.

The ontology has been modeled in the OWL 2 language using Protégé 5.2[2]

3.4. Realization of soil samples

One of the most important features of ontologies is the ability of automatically assigning individuals to concepts based on the definition of such concepts. This inference process is known as realization. In the soil context presented here, it means that it is possible to assign soil samples to a soil texture class and to a soil order in an automatic manner. Thus, instead of manually classifying each soil sample one by one according to several parameters, it is possible to derive these classification rules in the ontology and perform the realization process to classify all samples in one click. This represents a practical example on how ontologies work. The application is twofold: one allows for the selection of samples of our interest using rules that we define beforehand; the other allows classifying using such rules, which can be very useful especially when a large number of samples are available.

A concept in an ontology can be defined as a logical expression relating concepts, relationships and properties by using operators such as "and", "or", "not" and restricting relationship cardinalities and data ranges. Then, any individual fulfilling the logical expression is inferred as a member of that concept.

In our ontology methodology we have defined several logical expressions for soil texture class and soil order. Let us see first the logical expressions related to soil textures. According to the USDA texture triangle, it is possible to classify the soil texture based on the percentages of clay, sand and silt in the sample. As an example, Figure 2 shows the logical expression for clay texture.

This logical expression can be read as "For any soil layer having a clay percentage ≥40% and a sand percentage ≥ 45% and a silt percentage ≤ 40%, then such a soil layer is classified as clay texture". Similar expressions are defined for loam and sand textures.

Figure 2. Logical expression (axiom) describing "Clay texture" concept.

[2] http://protege.stanford.edu/

For this first version, and regarding logical expressions related to soil classification in orders, we have defined axioms for Vertosol, Sodosol and Choromosol orders. Classification rules for these orders have been adapted from Isbell (2016) to allow us to classify soils using laboratory analyses without knowing horizon distribution. As an example, Figure 3 shows the logical expression for Vertosol. This axiom reads as "For any soil having both layers 1 and 4 with a percentage of clay particles ≥ 35%, then such a soil is classified as Vertosol".

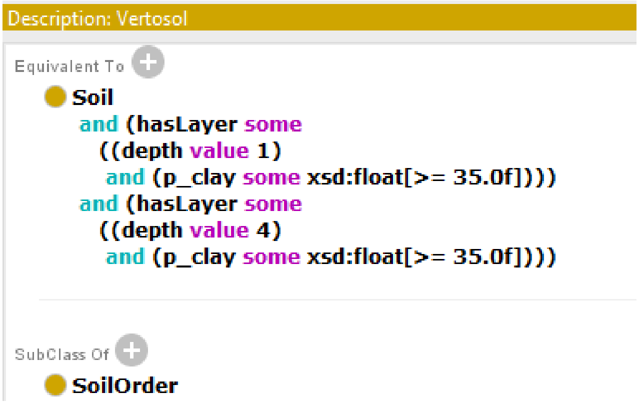

Figure 3. Logical expression (axiom) describing "Vertosol" order.

The logical expressions shown above need to be evaluated by an ontology reasoner in order to perform the realization process. For this study we have used Pellet[3], a well-known ontology reasoner for OWL 2 language included in Protègè.

4. Results

This section shows the application of our proposed ontology for the selection of soil samples. The methodology followed for the usage of the ontology can be summarized in three steps: 1) data loading (i.e. soil sample data) into the ontology; 2) performing reasoning processes on the data and ontology axioms (including the realization inference); and 3) writing the queries of interest. These steps are further explained below.

Before querying the ontology, we first need to load the soil samples data gathered from CNSA into our ontology. The data were originally exported into an Excel spreadsheet. We have used Cellfie[4], a plug-in for Protègè, in order to transform the soil samples in Excel format into ontology assertions. Soil order and soil texture class information for each sample have been excluded when loading the data as it will be inferred by the ontology axioms explained in section 3.4.

[3] https://github.com/stardog-union/pellet
[4] https://github.com/protegeproject/cellfie-plugin

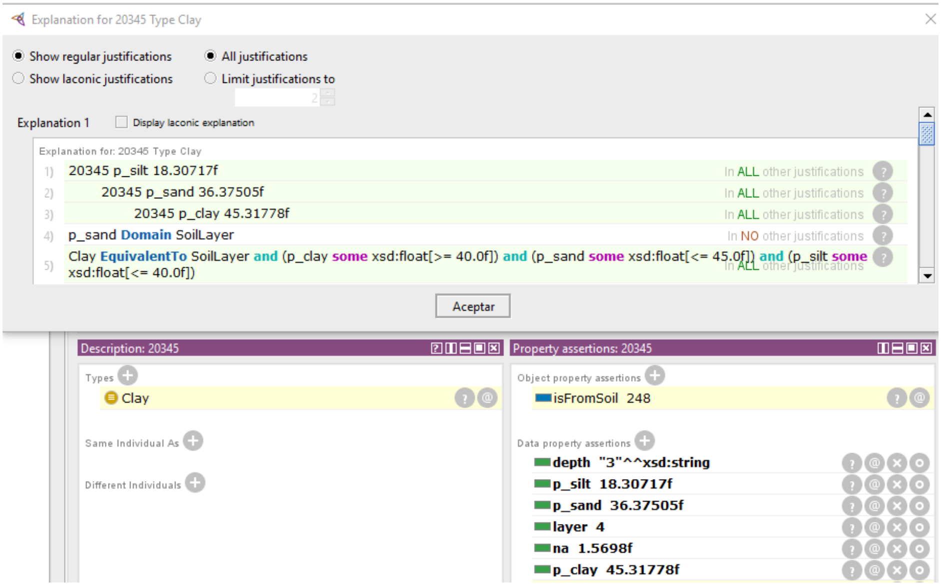

Figure 4. Realization process for soil texture. Sample 20345 has been inferred as clay texture. An explanation in logical language can be retrieved for any inference (in the top of the figure).

Then, we perform the realization process to automatically classify soil texture class and orders for each soil sample. Figure 4 shows an example for the classification of a soil sample (ID 20345) as a clay texture soil. Note that an explanation for any realization can be produced, as shown in the top of Figure 4.

After comparing the outcome of the realization process with the original soil texture and soil order classification given by the CNSA, promising results were found as shown in Table 1. Nevertheless, we will need to refine the logical expressions for sand texture and Chromosol order.

Table 1. Outcome of the realization process for soil texture and soil classification

Soil Texture	**CNSA data**	**Ontology realization**	**Accuracy (100%)**
Clay	206	208	99
Loam	8	8	100
Sand	6	10	60
Soil Order			
Chromosol	19	13	68
Sodosol	11	9	81
Vertosol	9	8	88

It is worth mentioning that soil ID 274 was classified as Vertosol and Sodosol at the same time. This generated an inconsistency in the ontology, as both concepts are disjoint, and the reasoner warns about this through a message in Protègè. As a result, we detected that the logical expression for Sodosol needed to be refined. This validation process is possible due to the formal properties provided by ontologies and enable developers to detect mistakes in the ontology development.

The ontology is then ready to receive queries for selecting soil samples. Similar to traditional databases, queries in ontologies are expressed using SPARQL, a query language similar to SQL but with a higher expressiveness. The first case is a typical search for a soil sample according to some features of interest. For example, query in Figure 5 searches for soils with any soil layer with Na ≥ 10.0 and clay percentage ≥ 50%. This query returns soils 268 and 274 with their respective values.

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX soil: <http://www.semanticweb.org/ucam/ontologies/2017/3/soil#>

SELECT ?soilID ?Na ?clay
        WHERE { ?soil rdf:type soil:Soil.
                ?soil soil:soilID ?soilID .
                ?soil soil:hasLayer ?layer .
                ?layer soil:na ?Na .
                ?layer soil:p_clay ?clay .
                FILTER(?Na >= 11.0) .
                FILTER(?clay >= 50.0) .
    }
```

Execute

?soilID	?Na	?clay
274	11.3	57.0
268	12.0	59.0

Figure 5. A simple SPARQL query for searching soil samples with Na ≥ 10.0 and clay percentage ≥ 50%

A more complex case is proposed. Let us suppose that given only the ID of a specific soil, we want to find other soils classified in the same soil order (we do not know the soil order for the original soil). The query in Figure 6 solves this issue and shows the results. In this case we asked for soil ID 247, which turned to be Vertosol, and the query returns all soils classified as Vertosol according to the realization process explained in Section 3.4. Note that this process will not be possible in traditional databases if soils are not manually classified in the first instance.

5. Conclusion and future work

This study has confirmed the potential of ontology approaches as an efficient, intuitive, and easy to use tool for the selection of soils from large soil libraries. As illustrative examples, we provide basic selection strategies along with more complex queries where samples are satisfactorily selected by their texture class or soil order. Ontologies

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX soil: <http://www.semanticweb.org/ucam/ontologies/2017/3/soil#>

SELECT  ?soilID ?soilOrder
          WHERE { soil:247 rdf:type ?soilOrder.
                  ?soilOrder rdfs:subClassOf soil:SoilOrder .
                  ?anotherSoil rdf:type ?soilOrder .
                  ?anotherSoil soil:soilID ?soilID
                  FILTER(?soilOrder !=soil:SoilOrder)
    }
```

?soilID	
260	soil:Vertosol
281	soil:Vertosol
280	soil:Vertosol
693	soil:Vertosol
264	soil:Vertosol
274	soil:Vertosol

Figure 6. A more complex query to retrieve similar soils to the one provided based on their orders

also warn about inconsistencies (i.e. a soil classified under 2 orders) which enable users to detect mistakes or special samples.

For the selection of samples that belong to given soil orders there is a need to use a reasoning process. This enables to quickly and automatically classify soils when this information is not available. Thus, we were able to classify soils from different orders, this showing an additional application of ontology approaches to produce hidden data from the original data set.

The proposed approach based on ontologies offers invaluable potential in agriculture and wider environmental applications. For example, ontology can be used for an easy and fast selection of the most suitable soils for specific crops. Likewise, best managerial practices can be easily selected in soil degradation, contamination and ecosystem services studies. In a generic environmental context, ontology can be used to activate alarm situations and to assist environmental managers in decision making processes. Irrespective of the area, tools like the presented in this study represent a necessary alternative for the management and access of increasing volumes of data being produced.

Acknowledgments

This work is supported by the Spanish MINECO under grant TIN2016-78799-P (AEI/FEDER, UE). José M. Soriano acknowledges the financial support from the Spanish Ministry of Economy, Industry and Competitiveness ("Juan de la Cierva-Incorporación Program"). The author also wishes to acknowledge the support of the Grains Research and Development Corporation (GRDC project CSO00045 "Soil infrared capability"), and the provision of soils and analytical data by Linda Karssies and Peter Wilson (CSIRO National Soil Archive).

References

Aparício, A. S., Farias, O. L., & dos Santos, N. (2005). Applying ontologies in the integration of heterogeneous relational databases. In *Proceedings of the 2005 Australasian Ontology Workshop-Volume 58* (pp. 11-16). Australian Computer Society, Inc.

Athanasiadis, I. N., Rizzoli, A. E., Janssen, S., Andersen, E., & Villa, F. (2009). Ontology for seamless integration of agricultural data and models. In *Research Conference on Metadata and Semantic Research* (pp. 282-293). Springer Berlin Heidelberg.

Beck, H., Morgan, K., Jung, Y., Grunwald, S., Kwon, H. Y., & Wu, J. (2010). Ontology-based simulation in agricultural systems modeling. *Agricultural Systems*, *103*(7), 463-477.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific american*,*284*(5), 28-37.

Burk L, & Dalgliesh N (2013). Estimating plant available water capacity. Grains Research and Development Corporation, Canberra (Australia). Available at https://grdc.com.au/Resources/Publications/2013/05/Estimating-Plant-Available-Water-Capacity, last accessed on 20/03/17.

FAO (2015). World Soil Charter. *Food and Agriculture Organization of the United Nations. Available at http://www.fao.org/3/a-mn442e.pdf.*

FAO and ITPS (2015). Status of the World's Soil Resources (SWSR). *Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils, Rome (Italy).* http://www.fao.org/documents/card/en/c/c6814873-efc3-41db-b7d3-2081a10ede50/

Gee G W, & Bauder J W (1986). Particle-size analysis. *Methods of Soil Analysis, (PART 1),* 383-411.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, *5*(2), 199-220.

Horrocks, I. (2008). Ontologies and the semantic web. *Communications of the ACM*, *51*(12), 58-67.

Isbell R, & National Committee on Soil and Terrain (2016). The Australian Soil Classification (Second edition). *CSIRO Publishing: Clayton* (Australia)

Matejovic I (1997). Determination of carbon and nitrogen in samples of various soils by the dry combustion. *Communications in Soil Science and Plant Analysis*, 28(17-18), 1499-1511.

Nocita M, Stevens A, van Wesemael B, Aitkenhead M, Bachmann M, Barthès B, Ben Dor E, Brown D J, Clairotte M, Csorba A, Dardenne P, Demattê J A M, Genot V, Guerrero C, Knadel M, Montanarella L, Noon C, Ramirez-Lopez L, Robertson J, Sakai H, Soriano-Disla J M, Shepherd K D, Stenberg B, Towett E K, Vargas R, & Wetterlind J (2015). Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. *Advances in Agronomy*, 132, 139-159.

Rayment G E, & Higginson F R (1992). Australian Laboratory Handbook of Soil and Water Chemical Methods. *Australian Soil and Land Survey Handbooks Series*, Vol 3. Melbourne: Inkata press.

Rayment G E, & Lyons D J (2011). Soil Chemical Methods-Australasia. Collingwood: CSIRO Publishing.

Sánchez-Alonso, S., & Sicilia, M. A. (2009). Using an AGROVOC-based ontology for the description of learning resources on organic agriculture. *Metadata and Semantics*, 481-492.

Shoaib, M., & Basharat, A. (2010). Semantic web based integrated agriculture information framework. In *Computer Research and Development, 2010 Second International Conference* (pp. 285-289). IEEE.