ONTOLOGY FOR DATA REPRESENTATION IN THE PRODUCTION OF COTTON FIBER IN BRAZIL

C. J. Santos, and A. R. Hirakawa

Department of Computer & Electrical Engineering
Escola Politécnica - Universidade de São Paulo
São Paulo – SP, Zipcode 05508-010, Brazil

ABSTRACT

An important feature in computer systems developed for the agricultural sector is to satisfy the heterogeneity of data generated in different processes. Most problems related with this heterogeneity arise from the lack of standard for different computing solutions proposed. An efficient solution for that is to create a single standard for data exchange. The study on the actual process involved in cotton production was based on a research developed by the Brazilian Agricultural Research Corporation (EMBRAPA) that reports all phases as a result of the compilation of several theoretical and practical researches related to cotton crop. The proposition of a standard starts with the identification of the most important classes of data involved in the process, and includes an ontology that is the systematization of concepts related to the production of cotton fiber and results in a set of classes, relations, functions and instances. The results are used as a reference for the development of computational tools, transforming implicit knowledge into applications that support the knowledge described. This research is based on data from the Midwest of Brazil. The choice of the cotton process as a study case comes from the fact that Brazil is one of the major players and there are several improvements required for system integration in this segment.

Keywords: ontology, fiber cotton production, agribusiness.

INTRODUCTION

In the 1950s the great expansion of agriculture started in Brazil. Before this period, the agricultural production was concentrated in the south and south-east of the country. The expansion of agriculture to the Midwest was concentrated in the state of Goias, and spread to other areas, especially to the state of Mato Grosso. The cotton production in the Midwest has been increasing since 1997, and nowadays plays an important role making changes in the domestic cotton market. Figure 1 shows the distribution of cotton production in Brazil.

Figure 1. Cotton Production in Brazil (source CONAB 2007).

The cultivation of cotton in Brazil is expected to increase in coming years. The planting area has grown especially in Mato Grosso, Mato Grosso do Sul and Bahia. The exportation of cotton plume increased by 2 times in just one season, from US$ 93 million in 2002 to US$ 188.5 million in 2003. In the 2003/04 season, Brazil produced 1.2 million tons of plume, and 847.5 million tons in the previous period, which results in a 46.3% growth. In the next years, the planted area should rise from 735.1 million hectares to 1 million hectares (MAPA, 2010).

As major challenges, the agricultural sector presents the heterogeneity of data that come from by different systems, regardless of the type of culture and different phases of the processes. To try to solve such challenges, data exchange standards are proposed, such as AgroXML and AGXML. With the AgroXML standard, software or system applications can exchange data among processes using a standardized format (SCHMITZ *et al*, 2009)(KUNISCH *et al*, 2009). AgXML is a standard developed to meet the formalization of the grains segment data, also favors the business information processing and related entities (AGXML GROUP, 2011). The working group is committed to developing standards for efficient and effective electronic communication of information in agribusiness.

A possible alternative to solve problems related to the heterogeneity of data is the creation of a common standard for all the players involved in this area. The creation of a new pattern must start from the identification of classes of data, followed by their main attributions and relationships.

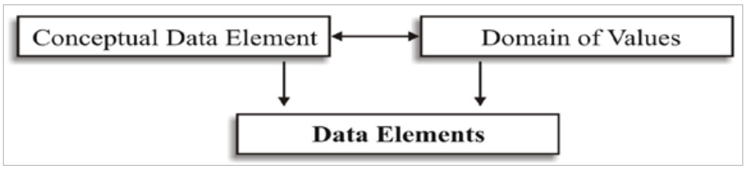
The proposal presented herein is focused on mapping the main data classes in cotton production, more specifically in cotton for fiber production. As results, a formal ontology for the data and its hierarchical structure are presented. The research will use the production scenario from the Midwest region of Brazil as a study case.

The main contribution of this work is the use of specific metadata for agribusiness in the cotton production chain as a standard for data transport in the different processes during the phases of planting, harvesting and processing.

Ontology provides a shared space and reusable knowledge about a specific domain and has been applied in several areas such as Semantic Web, e-commerce, medicine, tool manufacturing, automotive repair, financial management and information retrieval, among others (Li, 2005). An ontology defines the terms used to describe and represent an area of knowledge and can be shared by people, databases and applications that need to share domain information or a specific area. The ontology word has also been used to describe the devices with different degrees of structure (Cruz, 2004). From the definition of ontology to represent a hierarchical data structure, it is possible to

create a standard to provide metadata in the distribution of data among the elements involved in the processes.
Metadata is data about other data (FILETO *et al*, 2011). A metadata item can specify what it is given in a particular and intelligible piece of information by computer systems. Metadata facilitates the understanding of relationships, use and usefulness of the data (NISO, 2004). A conceptual diagram regarding metadata is shown in Figure 2.

Figure 2. Metadata Overview.



MATERIALS AND METHODS

The study conducted for this work was based on the field research of the processes involved in cotton production by the Brazilian Agricultural Research Corporation (EMBRAPA). After identifying the processes, the analysis of process continuity of cultivation and industrial processing was started (RICHETTI *et al*, 2003). From that, it was possible to compare the information collected at farms, processing companies and other related companies. The methods and procedures used for information collecting are presented in a sequential way. The starting point was the review of existing metadata standards and applications focusing on agriculture. Meetings were held with professionals from multidisciplinary fields to understand specific technical content and to obtain technical data, such as varieties of the species studied, pests, weather conditions and soil characteristics. Technical visits to companies and farms were performed for verifying and understanding existing processes in various phases of production of cotton lint. After that, technical discussions were finally held with multidisciplinary professionals to validate the ontology model proposed. Data for the formalization of the ontology were collected through technical references and visits to segments, such as academic institutions, certification of products, commercial enterprises, government institutions and technology companies, as shown in table 1. The main companies and institutions are located in Mato Grosso State.

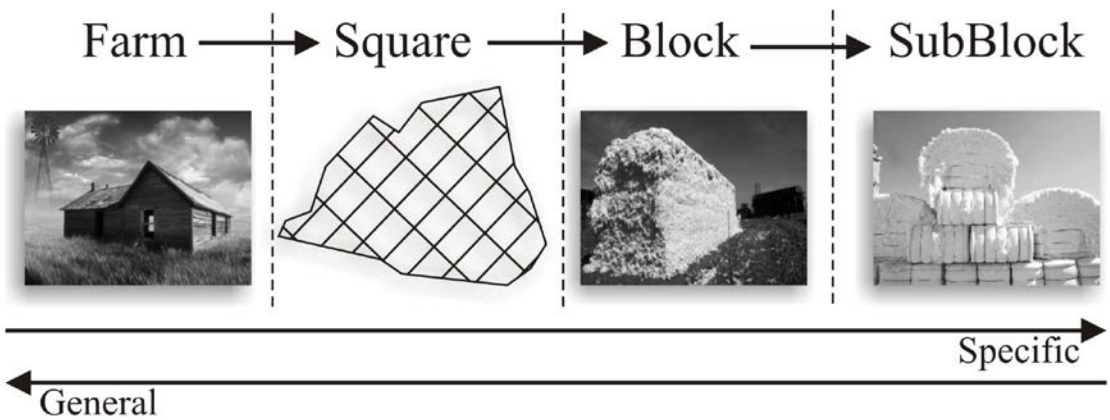
Table 1. Data Sources of Research.

Academic	
Name: Engenharia Agrícola Location: UFMT/Rondonópolis-MT Description: Master's degree in Agricultural Engineering	Name: Economics Location: UFMT/Rondonópolis-MT Description: M.A. in Economics
Certification	
Name: BMF (Bolsa Mercantil e de Futuros - commodities and futures exchange) Location: Rondonópolis-MT	Name: SGS Location: Rondonópolis-MT Description: Certifying quality rating of cotton fiber and HVI.

Description: Certifying quality rating of cotton fiber and HVI.	
Commercial	
Name: COPERBEN Location: Campo Verde-MT Description: Cotton processing cooperative.	Name: Grupo Bom Jesus Location: Rondonópolis-MT Description: Large-scale cotton producer.
Name: IBEL Location: Rondonólis-MT Description: Extraction of oil from cottonseed.	Name: TBM – Tecelagem Bezerra de Meneses Location: Rondonópolis-MT Description: Textile segment company.
Name: TecSolo Location: Rondonólis-MT Description: Analysis of soil quality.	Name: Yara Location: Rondonópolis-MT Description: Produces fertilizers for several types of plants.
Government	
Name: CODESA (Companhia Docas do Espírito Santo) Location: Vitória-ES Description: Commercial port in Vitoria, State of Espirito Santo	Name: INPEV (National Institute for Processing Empty Containers) Location: Rondonópolis-MT Description: Recycling of empty pesticide containers.
Name: Pro-Sementes Location: Rondonópolis-MT Description: Research on seed quality.	
Technology	
Name: Fundação Mato Grosso Location: Rondonópolis-MT Description: Agricultural research.	Name: Unisystem Location: Rondonópolis-MT Description: Development of agricultural software.

From the initial study of the large amount of data collected, four classes of data were identified: farm, square, block and sub block as shown in Figure 4. From those classes, other classes were derived.

Figure 4. Main Classes.



RESULTS

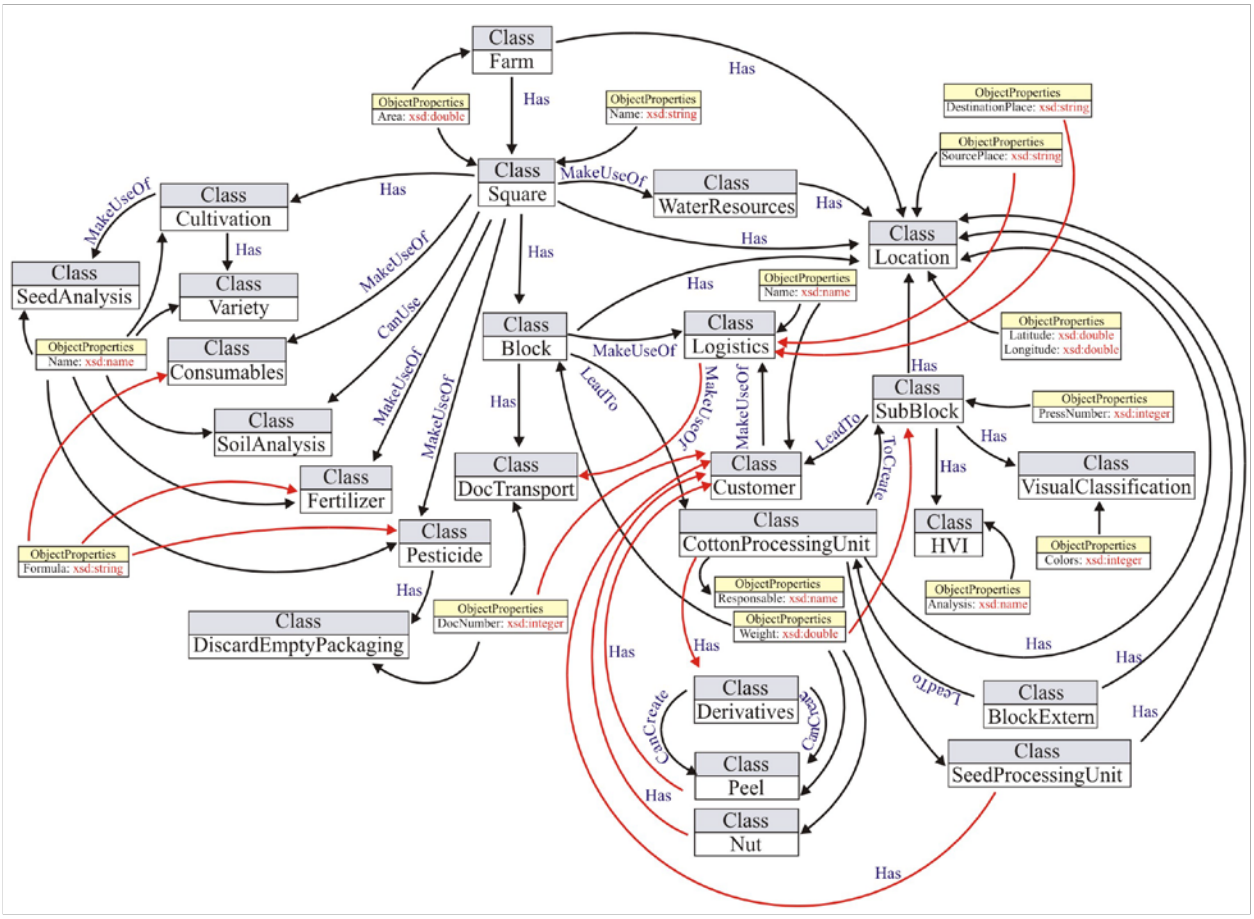
The results of the work are presented in three phases: the first refers to the proposal of the ontology, the second includes implementing a metadata standard in XML using the ontology; finally, a practical application is presented as a study case. Figure 5 presents an overview of the layered structure of results.

Figure 5. Layered Structure.

PROPOSED ONTOLOGY

The ontology proposed is focused on the key data generated in the cotton fiber production chain. The purpose is to define the structure of hierarchical data and their correlations. Figure 6 shows the main classes identified for the formalization of the ontology, and some relevant properties of each class are also represented.

Figure 6. Classes of the Proposed Ontology.



The mapping process purpose is to identify the relevant data required for the creation of ontology to represent knowledge and to provide better understanding of the overall cotton production process. In the upper lever, there is a farm that includes soil segments, and all other data are derived from this root.

One of the initial procedures to stress the ontology is the selection of the variety to be cultivated, and the choice depends on factors such as rainfall index (Class:Farm has Class:Square has Class:Cultivation makeuseof Class:SeedAnalysis), and, in some cases, as soil conditions: pH and fertility (Class:Square canuse Class:SoilAnalisys), which are obtained from laboratory analysis.

As a next step, it is necessary to consider the preparation of the soil consisting of any corrections made using consumables such as lime and plaster (Class:Square makeuseof Class:Consumables). After those activities, the actual planting is performed.

During the growth phase of the culture, technical monitoring and the use of pesticides are required according to a particular stage of maturity of the crop (Class:Square makeuseof Class:Pesticide). In more advanced stages of maturation, the use of pesticides has no effect due to the natural barrier created by the physical characteristic of the culture such as the presence of leaves.

The final activity in the field includes harvesting the cotton and its compression into bales of approximately twelve tons ("fardões") (Class:Square leadto Class:Bale).

The cotton bales are temporarily storage and moved to Cotton Processing Units (Class:Bale leadto Class:CottonProcessingUnit). At the Cotton Processing Units, the first activity is the gross weight of the "bale" still in the transport vehicle. After that, an analysis of moisture is performed to verify the cotton quality and, as these results are used as references to control the quality of Cotton after the ginning process. The fiber is separated into small bundles called subblocks ("fardinhos") (Class:CottonProcessingUnit leadto Class:SubBlock) and other products derived from cotton seed are separated for further processing in the Seed Processing Units (Class:CottonProcessingUnit has Class:Derivatives). Samples of subblocks("fardinhos") are separated for visual analysis (Class:SubBlock has Class:VisualClassification) of the presence of staining and fiber quality in classification rooms by specialized personnel. The classification groups into subblocks ("fardinhos") with the same characteristics and this is used as a key to product commercialization. The products for the overseas market have to undergo an analysis called *High Volume Instrument* or *HVI* (Class:SubBlock has Class:HVI) that complements the visual analysis performed in the previous step. The overall cotton production process finishes with the transportation logistics (Class:SubBlock has Class:Consumer has Class:Logistics). In addition to the key points described in the model presented, there are other issues such as water resource management (Class:Square makeuseof Class:WaterResources) and disposal of pesticide empty packaging (Class:Pesticide has Class:DiscardEmptyPackaging), which are related to the environment and supervised by government agencies.

SAMPLE APPLICATION OF ONTOLOGY: DATA HIERARCHY

A survey instrument of a file or a catalog of a library are descriptions of a collection of documents, organized in order to facilitate their recovery and access, and today is called metadata (Milsted,1999). The use of metadata in the Web context allows widespread dissemination of general information. In this scenario, the XML (Extensible Markup Language) arises, proposed and ratified by the W3C as a mechanism to encode metadata associated with an electronic document in a format that is readable by people and computer programs. The use of programs to process XML metadata is one of the pillars of the initiative called Semantic Web (Berners-Lee, 2001).
In this context, XML is used to implement the proposed ontology and can enable applications for a practical view of ontology and, as a study case, applied to webservice development.

Table 2 shows the description of the main classes in the proposed ontology, also presenting derived classes and attributes. In Table 2, XPath locator (and depth) is the path for each tag in the hierarchy and represents the relationship among tags to be followed; Rule(s) represents the consistency and validation rules for using tag; Annotation presents explanations of technical details concerning the tag and complementary information.

The XML hierarchy vision makes it usable in other applications that rely on an organized structure to store data in a database way or to create interfaces in Webservices.

Table 2. Definition of tags to represent the data hierarchy.

XPath locator (and depth)	Rule(s)	Annotation

1 Farm	required item	Root tag about data related to the farm and its elements
@Area	optional with "double" data type	Total farm area in hectares
2 Square	required item	Segments of farm
@Area	with data type of "double" required item	Total area of squares from the farm in hectares
@Name	required item with data type of "string"	Name of square
3 WaterResources	optional	Represents water resources used on square for cultivation
4 Location	required item repeatable item	Location of the source of water resources
@Latitude	with data type of "double" required item	Latitude value
@Longitude	required item with data type of "double"	Longitude Value
3 Cultivation	required item	Defines kind and variety of plants for cultivation on square
4 SeedAnalysis	required item	Analysis of seed for checking the quality
@Name	with data type of "string" required item	Name of the class of seed used in planting
4 Variety	required item	Variety chosen for cultivation
@Name	required item with data type of "string"	Name of riety chosen for cultivation
3 Consumables	optional	Data related to inputs used if there are any
@Formula	with data type of "string"	Formula of the consumables used
3 SoilAnalysis	optional	Soil analysis performed for the squares
@Name	required item	Name of the analyses
3 Fertilizer	required item	Fertilizers used in crop growth stages
@Formula	required item with data type of "string"	Formula composition of the fertilizer used
3 Pesticide	required item	Pesticides used for pest control
@Formula	with data type of "string"	Composition formula of defensive

	required item	
4 **DiscardEmpty Packaging**	required item	Data concerning the safe processing of empty pesticide containers
@DocNumber	required item with data type of "double"	Number of the document proving the safe processing of empty pesticide containers issued by the company in charge
3 **Block**	repeatable item required item	Cotton processed and ready for commercialization
4 **DocTransport**	required item	Document required to transport the processed cotton
@DocNumber	required item with data type of "integer"	Number of document required to transport the cotton
4 **Logistics**	required item	Data for the transport of the processed product
@Name	required item with data type of "string"	Name of company responsible for transportation
@DestinationPlace	with data type of "string" required item	Place of destination of the cargo to be transported
@SourcePlace	with data type of "string" required item	Place of source of the cargo to be transported
4 **Location**	required item	Geographical location of crop sites
@SourcePlace	required item with data type of "string"	Description of local crop
5 **DocTransport**	required item	Document required to transport the cotton
@DocNumber	required item with data type of "integer"	Number of document required to transport the cotton
4 **CottonProcessingUnit**	required item	Data related to the cotton processing unit
@Responsable	with data type of "string"	Responsible for the cotton processing unit
@Weight	with data type of "double" required item	Weights of crude product
5 **Derivatives**	optional	Other derivatives or cotton by-products
6 **Peel**	optional	Processing of the peel
7 **Costumer**	with data type of "string" optional	Data about costumers
6 **Nut**	optional	Nut Processing

7 Costumer	with data type of "string" optional	Data about costumers
5 SeedProcessingUnit	optional	Data concerning the processing of seeds for new planting
6 Costumer	required item with data type of "string"	Final destination for the processed seeds
6 Location	required item	Geographical location of Processing Unit
@Latitude	with data type of "double" required item	Latitude value
@Longitude	required item with data type of "double"	Longitude Value
5 SubBlock	required item	Cotton processed and packaged for shipment to final destination
@PressNumber	with data type of "integer" required item	Number of press used for creating sub blocks
6 VisualClassification	required item	Visual classification of the plume by staining
@Colors	with data type of "string" required item	Color identified for the classification of the plume
6 HVI	required item	High Volume Instrument
@Analysis	required item with data type of "string"	Types of analyses performed in HVI
6 Costumer	required item with data type of "string"	Data about costumers
6 Location	required item	Geographical location of subblock in processing unit
@Latitude	with data type of "double" required item	Latitude value
@Longitude	required item with data type of "double"	Longitude Value
3 Location	required item	Geographical location of Blocks
@Latitude	with data type of "double" required item	Latitude value
@Longitude	required item with data type of	Longitude Value

	"double"	
@SourcePlace	with data type of "string" required item	Description of local crop

STUDY CASE: WEBSERVICE DEVELOPMENT

Computational systems must not only be effective as a data management tool. They must have information to have clever ways to provide data distribution or decentralization. A very efficient way to accomplish this task is through Webservices. Webservices is a relatively new technology that has gained wide acceptance as an important implementation of service-oriented architecture. This is due to the fact that Webservices provide a distributed computing approach for integrating heterogeneous applications over the Internet. Webservices specifications are completely independent of the programming language (Li, 2005), operating system and hardware to promote the interaction between the service consumer (client) and provider (server) (NORDBOTTEN, 2009). The use of open standards provides broad interoperability between different vendor solutions. These principles mean that companies can implement Web services without any knowledge of consumers of services and vice versa. This integration facilitates just-in-time interaction and allows business to create new integrations easily and dynamically (ENDREI *et al*, 2004).
The case study illustrates the use of data model for a specific application to recover information on processes mapping of cotton fiber. Figure 7 shows an overview of the proposed case study implementation.

Figure 7. Structure and hierarchy of information for implementation of cotton webservice.

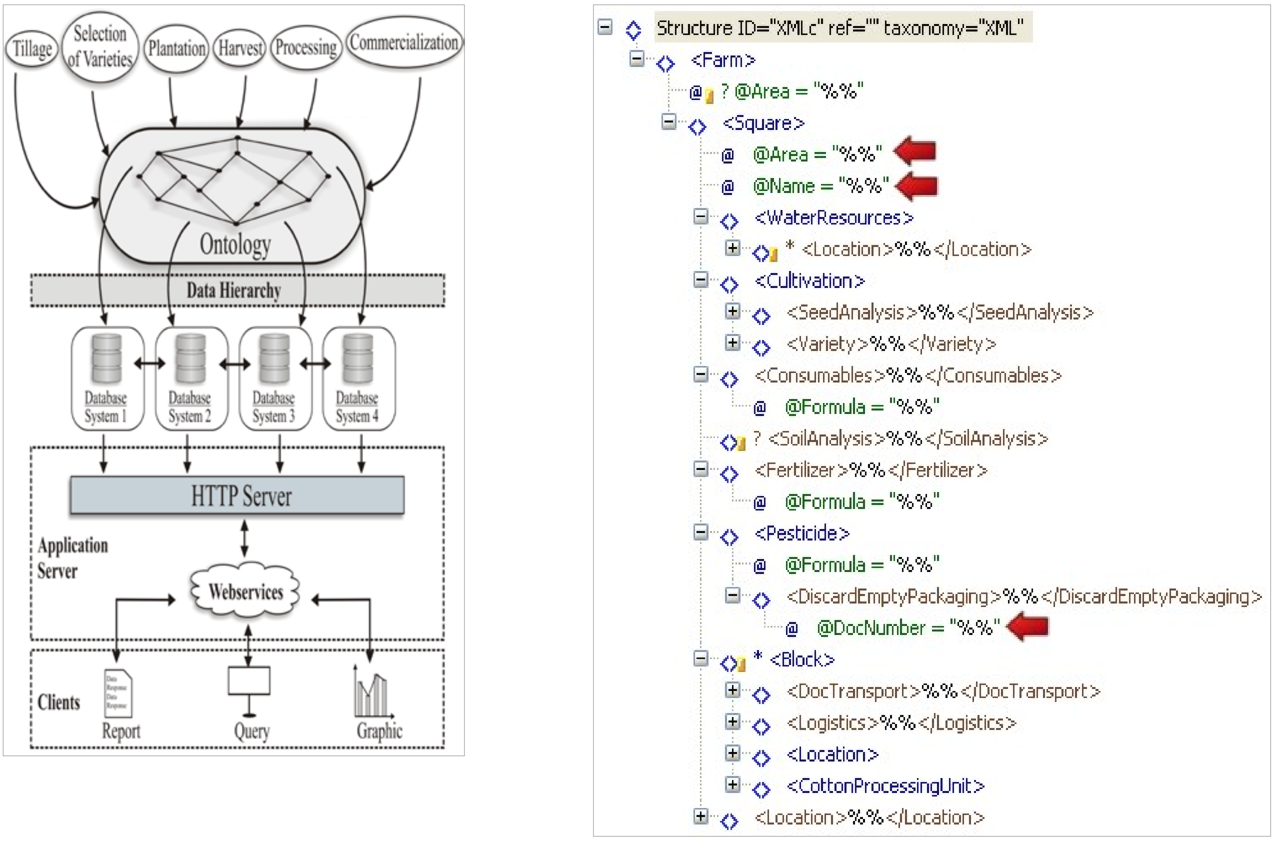


Figure 8 shows the relational database used in the Webservice implementation.

Figure 8. Entity relationship diagram for the implementation of the webservice.

.

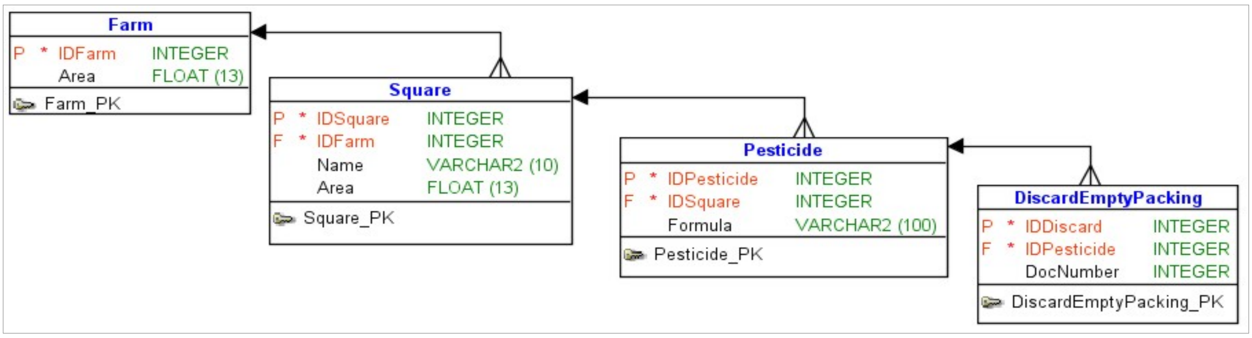


Table 3 lists the queries implemented for data recovery in the relational model of the proposed implementation of the Webservice. The development is based on PHP language with the Oracle database and Apache web server.

Table 3. Queries implemented in the Webservice.

Data (Tables)	Queries
Farm	**σ**{IDFarm=5}(Farm)
Farm, Square	**Π**{Farm.Area,Square.Name,Square.Area} (**σ**{Square.IDFarm=Farm.IDFarm ∧ Square.IDFarm=5}(Farm,Square))
Square Pesticide	**Π**{Square.Name,Square.Area,Pesticide.Formula}(**σ**{Square.IDSquare=Pest e.IDSquare ∧ Square.IDSquare=1}(Square,Pesticide))
Square Pesticide DiscardEmptyPacking	**Π**{Square.Name,Pesticide.Formula,DiscardEmptyPacking.DocNumber} (**σ**{DiscardEmptyPacking.IDPesticide=Pesticide.IDPesticide ∧ Square.IDSquare=Pesticide.IDSquare ∧ Square.IDSquare=1}(Pesticide,DiscartEmptyPacking,Square))

The queries are presented in hierarchical order according to the following data model: Farm□Square□Pesticide□DiscardEmptyPacking.

The Webservice implementation model includes simplified operations of a consultative role as a parameter receiving the unique identification of the farm. The final result is a structured list of data that abides by the ontology model using XML as shown in Table 3.

Table 4. Answer provided by Webservice in XML.

<?xml version="1.0" encoding="windows-1250"?> <Farm> <IDFarm>5</IDFarm> <Area>5000</Area> <Square> <Name>TH01</Name> <Area>700</Area> <Pesticide> <Formula>Randup</Formula> <DiscardEmptyPacking> <DocNumber>23434</DocNumber> </DiscardEmptyPacking> </Pesticide> </Square>	<Square> <Name>TH07</Name> <Area>500</Area> <Pesticide> <Formula>Tordon</Formula> <DiscardEmptyPacking> <DocNumber>23498</DocNumber> </DiscardEmptyPacking> </Pesticide> </Square> </Farm>

The results presented in Table 4 show a case study of a possible implementation using the proposed model.

CONCLUSIONS

The agribusiness sector is characterized by hardware and software solutions designed without concern about data sharing. The possibility of integration among systems is usually performed through Middleware applications. In more specific cases, in which it is not possible to create a Middleware solution for integration, the solution is the use of data exchange through exporting data in specific file formats, making it even more complex and less efficient. The use of a common standard for the production chain in segments, such as the agricultural, industrial, among other ones, provides an efficient way for sharing data using standards based on ontologies as references. From the model, standards such as metadata dictionaries or even a metadata standard can be derived. This works showed a specific ontology for cotton production, also presenting one case study that illustrates some application possibilities for the ontology proposed. It is worth noting that the model is still in the evolution process with the identification of new relevant situations and it requires a general proof of concept.

REFERENCES

AGXML GROUP. agxml Overview. Missouri – US. Avaliable in <http://www.agxml.org/Default.aspx?tabid=293> Access in 20/01/2011.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The semantic web. Scientific American, New York, n. 5, May 2001. Avaliable in < http://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf >. Access in 29/02/2012.

CRUZ, Isabel, Xiao Huiyon, Hsu, Feihong. An Ontology-based Framework for XML Semantic Integration. University of Illinois, pg 1-2, 2004

ENDREI, Mark, *et al*. Patterns: Service-Oriented Architecture and Web Services, IBM-REDBOOKS, 2004.

FILETO, Renato, *et al*. Using Domain Ontologies to Help Track Data Provenance. Institute of Computing, University of Campinas. College of Computing, Georgia Institute of Technology, Atlanta, GA, USA, pg 3-4, 2011

KUNISCH, Martin, *et al*. Entwicklung von agroXML. Kuratorium für Technik und Bauwesen in der Landwirtschaft e.V. (KTBL) 2009.

LI, Man, Du Xiao-Yong, Wang Shan. LEARNING ONTOLOGY FROM RELATIONAL DATABASE. Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, School of Information, Renmin University of China, Beijing 100872, China, pg2-5, 2005

NISO, National Information Standards Organization. Understanding Metadata. Published by: NISO Press.
Copyright © 2004.

NORDBOTTEN, N. XML and web services security standards. IEEE Communications Surveys and *Tutorials* [S.I.], v. 11, n. Compendex, p. 22-36, 2009.

MOLIN, J. P. Geração e Interpretação de Mapas de Produtividade para Agricultura de precisão. Avaliable in <http://www.esalq.usp.br/departamentos/ler/download/CLP%202000.01.PDF>. Access in: 17 Out. 2006. (in portuguese)

MAPA, Ministério da Agricultura, Pecuária a Abastecimento. Brazilian Agribusiness: An Investment Opportunity.http://www.agricultura.gov.br. Access in 28/03/2010 (in portuguese)

MILSTED, Jessica, FELDMAN, Susan. Metadata: cataloging by any other name. ONLINE, jan. 1999. (http://www.iicm.tugraz.at/thesis/cguetl_diss/literatur/Kapitel06/References/Milstead_et_al._1999/metadata.html, Access in 29/02/2012).

RICHETTI, Alceu, *et al*. Embrapa Algodão, Production Systems, 2, ISSN 1678-8710 Eletronic Version. url: http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Algodao/AlgodaoCerrado/index.htm
2003. (in portuguese)

SCHMITZ, M. *et al.* agroXML Enabling Standardized, Platform-Independent Internet Data Exchange in Farm Management Information Systems. *Metadata and Semantics* [S.I.], p. 463-468, 2009.