APPLICATION OF METADATA STANDARDS FOR GRAIN CLASSIFICATION

Emili Everz
IT Department – Ponta Grossa State University.
Av. Carlos Cavalcanti, 4748.
84030-900 – Ponta Grossa – PR – Brazil.
emilieverz043@gmail.com

Maria Salete Marcon Gomes Vaz
IT Department – Ponta Grossa State University.
Av. Carlos Cavalcanti, 4748.
84030-900 – Ponta Grossa – PR – Brazil.
salete@uepg.br

Abstract

Agricultural production in Brazil has great potential for expansion, either in the area produced or with the increase in productivity. The great national production of grains causes the number of commercializations involved to be high, for each commercialization of grains carried out in the national territory, the federal, state and municipal governments, together with MAPA, Ministry of Agriculture, Livestock and Supply, determine rules to be followed. One of them is that it determines the classification of grains destined for human consumption, either in natura or processed. Each grain must follow a different classification standard, according to regulations and ordinances. Metadata can be called data about data, that is, it is a way of describing data, thus facilitating the organization of information. The metadata standard is presented in a simple and understandable way, since it will be used in other systems managed, not necessarily, by different users and, since the purpose of this work is to facilitate the grain classification process, it would be impracticable to create a system with need for training to use it.

Keywords: Metadata Standards, Grain Classification, Technology, Agriculture.

1. Introduction

Grain classification is the act of determining the extrinsic and intrinsic qualities of a plant product, its by-products and residues of economic value, based on official MAPA standards. As Brazil is one of the largest grain producers in the world, According to Embrapa - Brazilian Agricultural Research and Agriculture Company (2020), its concern with the grain classification process is of great importance, and, given this premise, it causes astonishment to affirm that the process of classification of grains is still of the manual form and almost becomes a technology, excluding only scales for weighing grains and hygrometers (equipment used to measure the humidity present in the grain sample).

If the grains are not in according to the standard, the grain load may lose part of its value or even be removed from circulation. However, the data is largely tangled, causing difficulty in find what you want in a quick search. With the help of metadata, data can be summarized in order to make the search friendly to predict and demarcate the characteristics of the product end, facilitating the work of agricultural producers.

In order to improve the grain classification process, the proposal for metadata standards has become a means to facilitate technological understanding of the subject, given that grain classification is an area directly related to Agronomy, the metadata vision it is defined, basically, as data about other data or information about information,

being structured information that aims to identify, describe, explain, locate, facilitate the recovery and use of data (MANOUSELIS et al., 2010; NISO, 2017 ; STEINACKER; GHAVAM; STEINMETZ, 2001). And, from this point of view, Nascimento (2008) shows us that the metadata standards aim to:

a) Assist in the recovery of information and database;

b) Guarantee the registration of relevant information;

c) Facilitate the use and sharing of information and knowledge, for humans and machines;

d) Assist the exchange of information between different systems and databases.

Through this it is possible to present complex information about the classification of grains, in a simplified way, facilitating the creation of a structure to store, manage, sort and perform searches on classification data, as well as allowing the reuse of this structure by any systems. Knowing that the use of metadata facilitates the extraction of information about a data set, it helps in decision making within the system, since the use of metadata facilitates the way of retrieving and consulting specific information in a complex database within the system.

Despite their enormous value, metadata is generally scarce and incomplete, making services more difficult. Much digital information, from the existing mechanisms, for the automated creation of metadata relies mainly on content analysis, which can be expensive and inefficient. The automatic metadata generation system proposed by (RODRIGUEZ et al, 2009) takes advantage of relationships of resources generated from existing metadata as a means of propagating resources from rich metadata to poor metadata.

The use of standards has already been seen as a form of limitation among the developer community, today with the great growth of stored data, it is seen as a great ally. Investing in standards is a correct effort to solve known problems and bring several benefits to its users. It facilitates the analysis activity, since they are generally widely documented, bringing up solutions to problems that in some cases had not yet been predicted by analysts, facilitating communication between users and providing uniformity and integration between solutions.

Still, it is necessary to mention that the classification of vegetable products aims to guarantee the quality of the product to be imported, exported and sold internally. Law No. 9,972, of May 25, 2000, establishes the classification of all plant products intended directly for human consumption, and MAPA is responsible for determining classification standards by means of regulations or ordinances (BRASIL, 2000).

To determine classification levels for the most different types of plant products, MAPA determines ordinances or normative instructions that, in conjunction with Law No. 9,972, govern the classification of the main plant products. Through the classification of grains it is possible to determine the Group, Class and Type of each grain and final destination, human and animal consumption, production of non-food products, among others. It is also part of the classification process of returning the grains to the producer, in case he is interested in requalifying the grains for a new analysis and new framing in the characteristics mentioned above.

In addition to being manual, the classification process causes exorbitant expense of paper, since, for each grain, printing of classification tables is necessary, requiring

more than three printed sheets for each classifier, as each grain has a table according to its type, group, class and final destination.

2. Bibliographical Review

Vegetable classification is the act that determines the quality of a product through analyzes and by comparison between the analyzed sample and official standards approved by the federal government (MAPA), aiming to identify the extrinsic and intrinsic characteristics of plant products that meet the standards required by the consumer. Grain quality is a very relevant parameter for marketing and processing, and it can affect the final value of the product. Quality is ensured by making the classification of batches of these products.

However, in addition to the official classification, there is a commercial classification that is established in a contract between buyer and seller. In this case, the buyer determines the classification standard that must be met by the producer or holder of the grains, which cannot be less rigid than those established by the supervisory body.

The classification of grains was instituted by Law No. 9,972, of May 25, 2000 and regulated by Decree No. 3,664, of November 17, 2000. Art. 3 of the law defines that classification is the act of determining the intrinsic and extrinsic qualities of a plant product, based on official physical or descriptive standards. The law requires that, in certain situations, the beans are classified and stipulates who can perform this classification. Grains should be classified when:

a) They are intended directly for human consumption, that is, when the products meet all food safety standards to be offered to the final consumer.

b) In the purchase and sale of public authorities.

c) At ports, airports and border posts, upon import.

The standard is defined by the set of specifications or parameters of identity and quality of plant products. The qualitative parameters are represented by the defects and the quantitative ones, by the tolerance levels of these defects, which are expressed in percentages. The objectives of standardization are:

a) Assist in the commercialization of quality products, through standards set by the MAPA.

b) Determine the intrinsic and extrinsic qualities of products of plant origin.

SENAR - National Rural Learning Service (2017) presents a point of great importance in the process of creating the metadata standard for grain classification, aiming at its branching in typifications, states that the classification of grains can be:

a) Official - is one in which food safety standards are obeyed in order to offer safe products to final consumers.

b) Commercial - is one that can be adjusted according to the needs of each company.

For the commercialization of grains, there are quality standards established by the legislation and the non-compliance with them may result in financial discounts from the producer or seller of the products. The main differences found in the load in relation to the standards are the percentages of impurities and foreign matter, the humidity and

the percentage of grains with serious or slight defects above the tolerance of the law or the purchasing company.

Sampling is the procedure for obtaining the sample, which is part, fragment or representative unit of a lot. The sample will be analyzed to provide the qualitative information of the batch. It can be simple, composed, average and work. Obtaining the sample (or sampling) must take place in accordance with the provisions of art. 9 of Decree No. 3,664 / 00, which establishes the necessary terminology, criteria and procedures.

Each grain has specific defects, which define it, according to its quality and its tolerance limit, to be presented for purposes of use such as human food, animal, biodiesel production or to be classified as unusable due to the excessive amount of defects presented in the sample taken for evaluation. As a basis for the research, soybeans, corn and wheat were used to validate the metadata standardization process aimed at the grains. tablet 1 shows, as an example, the standards accepted by the law for soybeans classified in type I, in contrast, tablet 2 presents the standards accepted by the law for soybeans classified in type II, to show the difference in their percentage of defects.

	Damaged						
Type	Total Burned	Maximum of Burned	Moldy	Total	Greenish	Broken and Smashed	Total MEI
1	1,0	0,3	0,5	4,0	2,0	8,0	1,0
2	2,0	1,0	1,5	6,0	4,0	15,0	1,0

Tablet 1 – Maximum Tolerance Limits, Expressed as a Percentage for Type I Soybeans

	Damaged						
Type	Total Burned	Maximum of Burned	Moldy	Total	Greenish	Broken and Smashed	Total MEI
Basic Pattern	4,0	1,0	6,0	8,0	8,0	30,0	1,0

Tablet 2 – Maximum Tolerance Limits, Expressed as a Percentage for Type II Soybeans

As a contribution, this work proposes a division between the metadata storage architecture and the modeling of the patterns, that is, in one place the data of the metadata will be stored, and in another the patterns will be stored that will be used as templates (models) for the storage of data, according to the specification of each standard described. Another contribution of pattern modeling is the flexibility for constant extensions of the patterns used by the organization. It is not part of the scope of this work to determine the form or technology used for storage for metadata standards and data.

3. Methodology

A metadata standard must be complete and error-prone to zero, given this premise, it is necessary to understand how data can be generalized and organized within a database and how metadata is used. To do this, to base this research on standards such as: Dublin Core, MODS, EAD and PREMIS, ensures that several possible errors in the standards were removed, given that these had already been dealt with and presented in their respective directories.

The creation of the metadata standard for grain classification is based on the idea of the standards presented, with Dublin Core being a standard for support and discovery

of electronic resources on the web; MODS – Metadata Object Description Schema - standard of descriptive metadata derived from the MARC 21 bibliographic system, which focuses on electronic resources and library services; EAD - Encoded Archival Description - standard for coding archival access instruments, such as inventories and indexes; PREMIS - PREMIS Data Dictionary for Preservation Metadata - standard that defines a set of basic elements to encode, store, manage and exchange preservation metadata in the context of a repository system for environmental preservation.

The justification for its foundation is due to the fact that the metadata standard for classification makes room for the creation of web systems that use it as a research source replacing the current classification tables, as well as allowing the insertion of new grains based on the standard already imposed for corn, soybeans and wheat. And, as a consequence, create a data repository on the evaluated grains.

As shown, the grains have specific defects, and, thus, would result in the creation of a very polluted metadata standard, since many defects are the same among grains, only differing by their based nomenclature, mandatory by the criteria of Law No. 9,972. However, as the definition of metadata itself, it is possible to insert data into the standard and contain various information within it, for example, it is possible to create a generic defect that represents the similar defects presented within the grains and, from there, verify for this generic defect each of the specific defects presented by each grain. To abstract the idea, figure 1 presents the generic defect “chocho or broken” that appears in the soybean as “chocho”, in the corn grain as “chocho or broken” and in the wheat grain as “wheat”, however , represent the same situation, where the grain does not appear in its complete form, in other words, when it reached its maturity it did not present its common size.

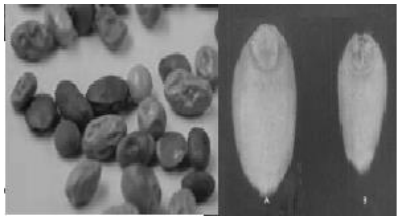

Figure 1 – To the left, soybean grains and to the right, a common wheat grain followed by a wheat grain.

By obtaining all the specific defects of each grain, and evaluating which defects are found to be similar, it is possible to create the generic defect (metadata) to be inserted within the metadata standard for grain classification to facilitate consultation and management of that data by the system that is using it.

For the creation of the metadata standard for the classification of grains, based on a metapattern model presented in figure 2 is of substantial importance, since when it is used in complex systems, it must obey the order and structure of tree-based data, to make the research models used in the sub-areas derived from the database more flexible.

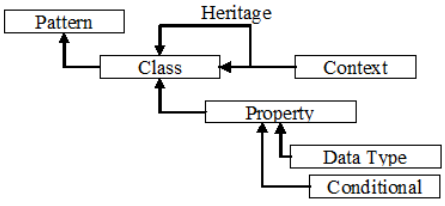

Figure 2 – Metapattern model

In the Standard entity, the standards that will be stored in the model will be registered, each attribute will be registered as a class, having properties that information will be registered in it, as specified in the properties by mandatory, optional or conditional. If it is necessary to use any rule to fill in the property attribute, the condition must be registered in the conditional entity.

It is mandatory to specify a context for each class. Since the classes can be used by any standard, provided that according to the standards specified for the standard it is necessary to use this class with this same context.

4. Results

With the use of the metapattern model idea and with a gathering of all data on defects of each grain, it was possible to create the final view of the description of the grain classification data using the metadata pattern shown in figure 3.

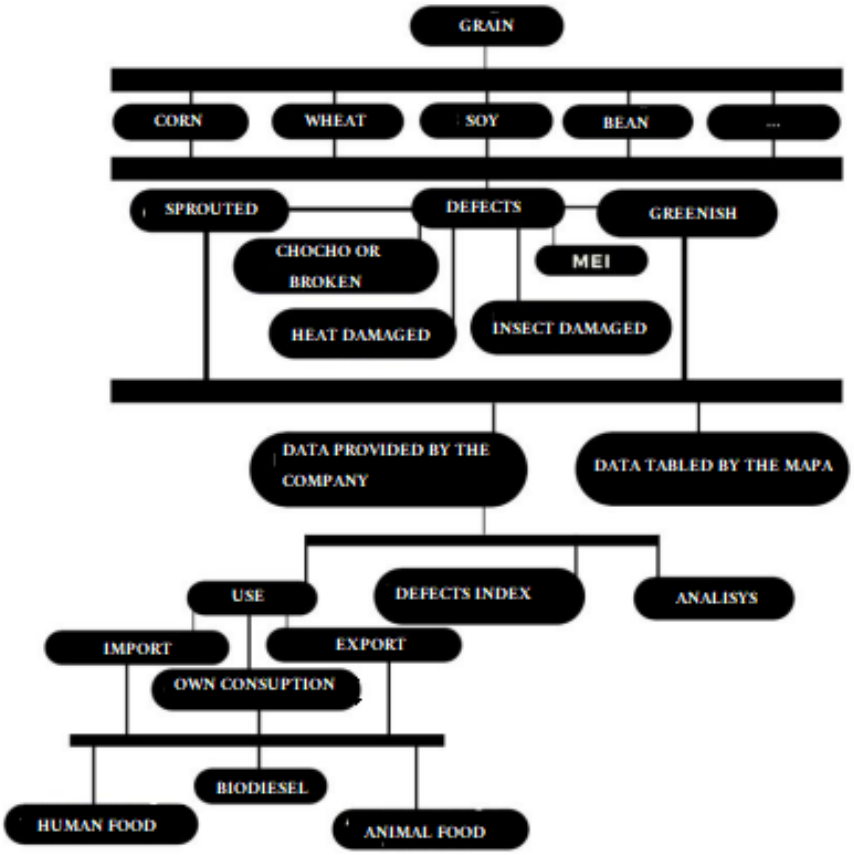

Figure 3 – Padrão de Metadados para Classificação de Grãos

Through the figure above it is possible to verify that the structure of the pattern is interconnected by means of a “tree”, used in order to become more abstract, tree structures serve to demonstrate data paths, where each branch, that is, each “branch of the tree” connects it to the “root”. All information is interconnected with the base, that is, this pattern was created specifically to be a classifier, where the root, that is, the beginning of the tree will have the representation of what will be classified, in this case, the grains and data on what are grains, after that we will be able to choose more specifications about each one of them descending more and more in the branches of the tree.

The standard used knowledge in data structures so that it could become a widely accessible standard, even though it is only focused on the grain area in this research, it is presented in a way to give greater opening of classification possibilities because it is a standard generic.

5. Conclusions

Through the use of metadata and metadata standards, it is possible to promote the integration, interpretation, localization and reuse of data, ensuring its existence over time. The proposed Metapatterns model provides a repository where the functionalities for the management and standardization of data within an organization can be integrated, providing the Administrator with a complete metadata management environment, allowing registered metadata to be accessed and used by people and applications. It can be considered that the main advantage of the proposed model is the reusability of the application itself, which makes the environment very simple to be used by its users, as the system can be used to record data on all standards and there may be the integration of information between distinct patterns.

Acknowledgments

The authors would like to acknowledge the support of National Council of Scientific and Technological Development – CNPq – Brazil.

References

BRASIL. Câmara dos Deputados. **LEI Nº 9.972, DE 25 DE MAIO DE 2000**, Márcio Fortes de Almeida.

EMBRAPA (2020) **Grandes Contribuições para a Agricultura Brasileira: GRÃOS.** Ciência que transforma. In: www.embrapa.br/grandes-contribuicoes-para-a-agricultura-brasileira/graos.

MANOUSELIS, N. et al. (2010) **Metadata interoperability in agricultural learning repositories: An analysis**. Computers and Electronics in Agriculture. v. 70, n. 2, p. 302-320, ISSN 0168-1699.

NISO, National Information Standards Organization. **UNDERSTANDING METADATA WHAT IS METADATA, AND WHAT IS IT FOR?** In: <http://www.niso.org>.

SENAR (2020) **Grãos: Classificação de Soja e Milho**. Coleção SENAR. v.178, p. 1-156, CDU 633.34:633.15.

STEINACKER, A.; GHAVAM, A.; STEINMETZ, R. **Metadata Standards for Web-Based Resources**. IEEE MultiMedia, v. 8, n.1, p. 70-76. 2001.

NASCIMENTO, L. U. **Um Padrão de Metadados para Indexação e Recuperação de Objetos Multimídia.** 2008. Dissertação (Mestrado em Informática). Universidade Federal do Paraná, Curitiba, 2008.