

SPECIFICATION OF DATA DIMENSION TO MEASURE THE DATA QUALITY ON COTTON PRODUCTION

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

The research proposes dimensions of data quality for qualifying masses of data with the specific farm. Despite of the evidence of the need for information technology resource to improve management in agriculture. In general there are relatively few initiatives developed or under development for this purpose. In general the requirements used in this process are generic and do not include the necessary privacy in agriculture. The aim of this research is to map the main requirements of data quality in agriculture and in particular the fiber cotton production in Brazil, providing subsidies to establish dimensions of the need data to measure data quality in this segment. The research focus is about the dimensions specification for analysis of specific data quality for cotton production; the proposal is to establish qualitative parameters to determine data enrichment actions when it is necessary. The specification of criteria for used dimensions qualification depends entirely on the context in which it is bounded on the research, metrics may take different amounts in different context. The absence of requirements for defining data structures can be seen in other dimensions like fill and duplication columns. There is no specific dimensions for measuring data quality in agriculture , dimensions may take different amounts depending on its context , the research presents a solution with the purpose of contributing to the qualification of data masses and hence with possible enrichments. The contribution presented will create criteria to provide a methodology for data qualification given to the particularities in this segment in Brazil, contributing to the complementary expert systems to generate information .

Key Words: Data Quality, Ontology, Metadata

INTRODUCTION

The cotton industry activities consist of processes providers of data at each stage of the chain production, as shown in research held by (Sofiatti, 2009). The management of cotton cultivation and agriculture in general, depend on data quality enabling the retrieving of useful information as an aid in decision making related to management techniques and farm management. Part of this task depends on intelligible data generated through the processes that make up this segment. Creating means for enabling the classification data is the starting point for making decisions regarding any corrections or adjustments in the mass data. The heterogeneity of data structures in farming systems creates gaps in the sharing and integration of management systems. Generally, the problem is the vertical integration of these solutions, currently the integration of databases in applications such as business intelligence, data warehousing and enterprise resource planning or ERP illustrate the problem. Although there

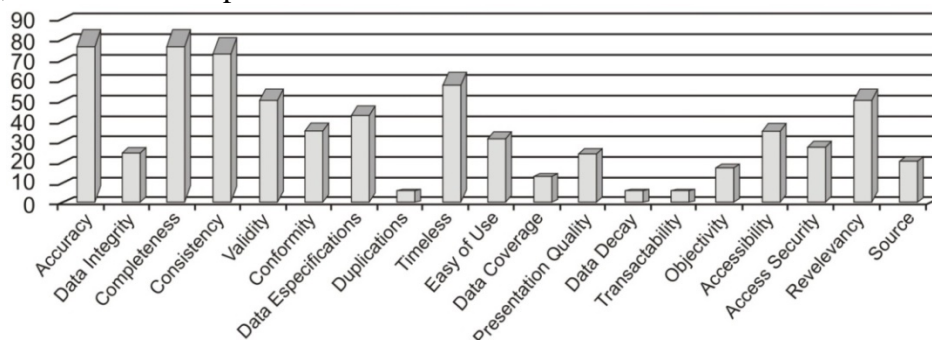
are computational solutions guided consolidated the concepts above, there are still difficulties in terms of efficiency due to the insignificant or incomplete data.

MATERIALS AND METHODS

An important point for research is delimitation of the classes of data with greater importance in this context. It was used as initial reference research developed by the Brazilian Enterprise for Agricultural Research, EMBRAPA and compiled by (Richetti et al, 2003). The study was also used to establish the degree of granularity required for the representation consistently data for this segment. The identification of most important classes provides relevant information about the data to be qualified reference, allowing identification to the primary and secondary data.

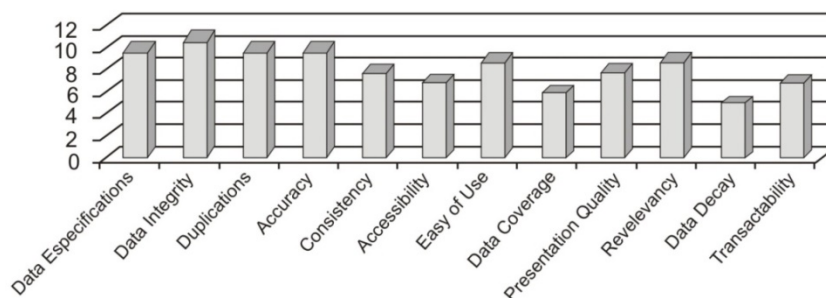
RESULTS

The research focus is on specification of dimensions for analysis of the specific data quality for cotton farming, specifying criteria for qualification of dimensions used depends on the context in which it is bounded to the survey, metrics can take different amounts in different context. The criteria may also take subjective aspects. Quantitative evaluation was performed by collecting data at specific groups of skilled professionals as academic and technical users with knowledge in this segment. The qualification requirement has been constructed using the 19 dimensions with a range of values between zero and four, as shown in Graph 1.



Graph 1. Analysis of the dimensions of data quality.

After that it was verified the proportionality for each dimension thus defining the degree of importance of the research context, as shown in Graph 2, although research in 19 dimensions, only 12 were classified as helpful to agriculture in this context.



Graph 2. Definition of dimensions for the agricultural context.

CONCLUSIONS

Concerning dimensions addressed, stands up data integrity due to the structural characteristics of the data used in agricultural systems, the absence of requirements for defining data structures with efficiency commits other areas such duplicity and accuracy. There are not specific dimensions for measuring data quality in agriculture, dimensions may take different amounts depending on its context and agriculture in particular is bounded in the research setting. The contribution presented proposes a methodology using data dimensions for classification of data with the characteristics found in this agriculture segment, also contributing to the adoption of criteria for qualifying data providing identification of flawed aspects to be improved.

REFERENCES

Richetti, A., et al., 2003. Embrapa Algodão, Production Systems, 2, Eletronic Version. url:

<http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Algodao/AlgodaoCerrado/index.htm>. (*in Portuguese*).

Santana, A., 2008. Software Management, Market Study - SEBRAE/ESPM - 2008. Brazilian Service of

Support for Micro and Small Enterprises – Sebrae 2008. (*in Portuguese*).

Sofiatti, V, et al., 2009. Brazilian Enterprise for Agricultural Research. Embrapa Algodão.

Ministério da Agricultura, Pecuária e Abastecimento. Algodão em Pluma. Embrapa

Informação Tecnológica. Brasília, DF. 2009. (*in Portuguese*).