African Journal of Business Management Vol. 7(10), pp. 769-779, 14 March, 2013
Available online at http://www.academicjournals.org/AJBM
DOI: 10.5897/AJBM12.1117
ISSN 1993-8233 ©2013 Academic Journals

Full Length Research Paper

Metrics for data accuracy improvement in a production scheduling software: An application in the Brazilian Meat Industry

Eduardo Scherer Rücker[1], Luis Henrique Rodrigues[1], Daniel Pacheco Lacerda[1]* and Ricardo Augusto Cassel[2]

[1]Production and System Engineering Program – UNISINOS, Research Group on Modeling for Learning – GMAP, Av. Unisinos, 950 – São Leopoldo – Rio Grande do Sul – Brasil.
[2]Production Engineering Program – UFRGS, Av. Osvaldo Aranha, 99 – 5º. Andar – Porto Alegre – Rio Grande do Sul – Brasil.

Accepted 4 March, 2013

This study aims to develop a method for improving the data accuracy of a specific production scheduling software for the meat industry. The proposed method was based on the results obtained in the implementation phase of the scheduling tool in Empresa Alfa, that produces food based on chickens, turkeys and porks. The research method used was a case study, by means of which the influence of data accuracy on the information generated by the system during the project is reported and analysed. The development of the proposed method was based on the theoretical framework on production scheduling, data quality and information quality, the author's perceptions about the project which was applied in the case study and the contributions of experts in the subject of the work. The method was structured in processes and subprocesses, hierarchy that enabled the implementation of the objectives of each process regarding the data accuracy in stages (subprocesses).

Key words: Data accuracy, meat industry, production scheduling, data quality.

INTRODUCTION

The competitive context of the business environment demands quality on decision-making. Given this situation, the data and consequently the information represent a significant input in the decision-making process. Investments in information systems are an attempt to make the choice of the alternative actions (decision-making) fast and qualified. Thus, in order to meet the increasingly demanding consumer markets, manufacturing industries depend on qualified data. These data will be used in the decisions related to the production scheduling process. Therefore, the quality of the production scheduling will be a result of the quality of the database and consequently of the information used. In organizations, the operationalization of information systems for this purpose depends on the availability of data whose accuracy faithfully reflects the reality of the company.

The accuracy of the data collected, particularly in manufacturing environments, is important due to the impacts that its quality can have on the decision-making process and on production planning and scheduling. This perspective is analogous to the input-transformation-output model of processes shown in Figure 1. In a production system, the raw materials go through a manufacturing process in which they ultimately acquire form and economic value. In the same way, the data are processed for generating information. Thus, if the data (raw materials) are not properly collected and fed into the system (by inaccuracy or by differing of the nature from the expected data), the results generated (end product) will differ from reality. It is noteworthy that the nature of the expected data refers to the pattern established for the input of data into the system.

These divergences have a significant set of implications for the organization as a whole, such as: i) the information on costs can be questioned; ii) the stock levels may not be adequate; iii) delivery deadlines may not be met; and; iv) the best use of the company's capacity may be compromised. Goldratt (1991, p. 3)

*Corresponding author. E-mail: dlacerda@unisinos.br. Tel: +55 51 35908273.

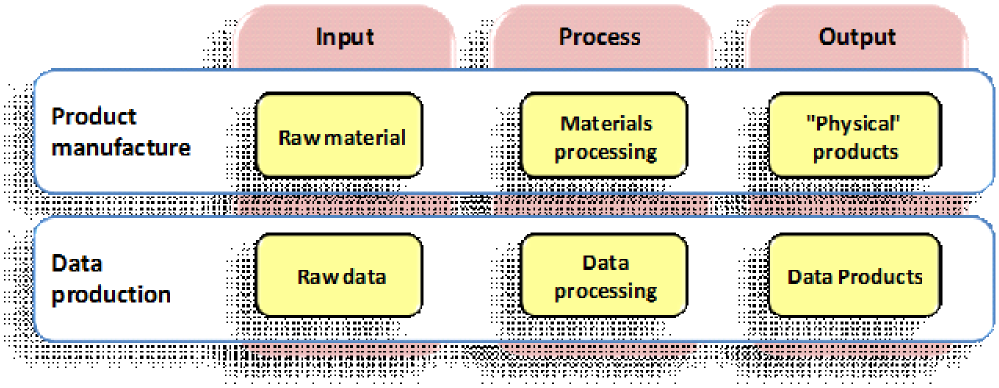

Figure 1. Analogy between product manufacturing and data generation.
Source: Constructed from Wang et al. (1995).

states that "we are drowning in oceans of data; however, it seems that we rarely have enough information." In turn, Olson (2003, p. 3) says that "data is the fuel we use to make decisions." These authors confirm the evidences that data influence decisions, being accuracy an important factor to be considered. Yet, Wang et al. (1995, p. 637) discuss what they consider to be essential for the improvement of data accuracy: "it is necessary to study the link between the poor quality of data and the procedures used to detect and eliminate problems." The research in the data quality area motivated the creation of the Total Data Quality Management (TDQM) at the Massachusetts Institute of Technology (MIT), a project that had its origin in the need of the industry for better indexes of data quality (Tdqm, 2008).

Seeking to contribute in the construction of "procedures to detect and eliminate problems in data quality" (Wang et al., 1995 p. 637), this paper proposes an approach for the improvement of data quality. This approach consists of processes of identification and correction of data, and is supported by indicators and control indexes. To support the research, a case study on the implementation of a production scheduling system in a meat company was used, through the measurement of the input and output data (information or results of the system).

The Brazilian meat-based food companies have led the country to the position of world leader in the export of this kind of product in the 2004 to 2007 period (Abef, 2008). The implementation of production scheduling systems in this type of companies allows the analyses of impact of the peculiarities inherent to the production environment of the meat-based food companies on the decisions on orders scheduling.

The developed approach and the results of its application were derived from a single case study. The study occurred in one of the largest Brazilian meat-based food companies, whose project for implementing the production scheduling and control system presented pitfalls, in part due to the lack of data quality.

The next section presents the theoretical framework necessary for the development of this work. Later, the methodological approach that led the study and supported its findings will be shown. Then, the case study and the proposed approach are discussed. Finally, conclusions, limitations and suggestions for further studies are presented.

THEORETICAL FRAMEWORK

Data quality has a close connection with the information quality. The concept of data quality is widespread in the literature according to its suitability for use, that is, the quality of the data is bound to answer a need (Tayi and Ballou, 1998; Helfert, 2001; Lee et al., 2006). Redman (2004) argues that data quality relates to "obtaining the right and proper data, in the right place and at the right time to accomplish a task" (Redman, 2004, p. 2).

Different methodologies were developed due to the need to treat data with the proper focus. Wang et al. (1995) highlight the importance of having methods that assist in the qualification of data as a way to prevent them from damaging the information generated for the decision-making process. In the scope of these methods, there are identification, analysis and data correction procedures. In addition, there are also the data measurement procedures.

Each one of the studies about data quality developed by Wang et al. (1995) proposes different dimensions (or categorizations) to analyse the quality of the data. Among the terms used, the *accuracy*, the *completeness*, the *consistency* and the degree of data updating (*timeliness*) stood out, and are detailed in Table 1.

The accuracy dimension is intertwined with data quality in terms of representing the level of conformity of the data in relation to its value in the real world. If a data is correct, it will be reliable for decision-making; therefore, its quality, which could be evaluated under other dimensions

Table 1. Dimensions of data quality.

Dimension	Description
Accuracy	It is the level of conformity of a data in relation to its value in the real world. An example of inaccuracy would be the ability of a machine in a database registered as 1,000 parts per hour, but which in reality can process up to 1,500 pieces per hour.
Completeness	It refers to the ratio between the existing and the necessary data. For example, the table of stocks in a database. If storage units lack in the database registration, the data will be incomplete; therefore, the completeness is low.
Consistency	It occurs when the representation of the data is the same in different sources. An example of consistent data is the ZIP (Postal Code) of a correspondence referring to the ZIP code of the destination of the same correspondence.
Updating degree	It refers to how much the data reflects its current value at the source where it was collected in relation to the moment when it should have been updated. An outdated data, for example, is a production batch of a product that is listed in the inventory system, but that in fact was dispatched to the final customer.

Source: Authors (2010).

sions, is evaluated according to its accuracy.

Olson (2002) adds the shape and the content of a data to the concept of accuracy, also pointing out some of the causes of inaccuracy: i) wrong values inserted into a database or on a report; ii) errors made by people who do not pay attention to the handling of the data; iii) confusing or contradictory screens or forms; iv) procedures that allow data not to be inserted or not registered within the deadline; v) procedures that promote the inclusion of erroneous data; and; vi) databases structured mistakenly.

According to Barchard and Pace (2011), to err is human. And human data entry can result in errors that ruin statistical results and conclusions. When humans do data entry, errors are therefore expected. Unfortunately, data entry errors can have devastating effects on research results. Simple data entry errors – such as typing an incorrect number, typing a number twice, or skipping a line – can ruin the results. Data entry errors can have serious effects on the results of a statistical analysis.

The incorrect input of data in a system has effects extremely harmful to the corporation. One such effect is the difference between the values of physical (reality) and system (virtual) stocks of a corporation. This ends up generating missing values and balances of finished products, raw materials, information, among others. These values have direct implications in the moment of decision-making in companies, as well as in corporate balance sheets.

An approach used in the attempt to improve the quality of the data is the TDQM (Total Data Quality Management).

The TDQM was developed with the aim of constructing theoretical foundations and methods that would contribute to the improvement of data quality (Tdqm, 2008). This is a mixed approach, which combines the application of statistical (control charts, standard deviation, among others), technological (automatic validations of data errors identification, for example) and managerial methods (analysis techniques and data problem solution). The scope of the TDQM researches is divided into three components, which are detailed in Table 2.

In Table 2, it is possible to understand how the TDQM addresses the data quality issue, because the components proposed lead to a systematic view of the problem. The TDQM methodology is inspired in the Deming's Cycle of the Total Quality Management. Similar to the PDCA, instead of the known stages of *Plan*, *Do*, *Check* and *Act*, the TDQM cycle analogously contemplates the Set, Measure, Analyse, and Improve activities, respectively. TDQM's proposal, based on Wang (1998), is shown in Figure 2.

The "Set" stage of the TDQM cycle refers to the identification of data quality dimensions and the respective requirements that will be considered in a project. Later, in the "Measure" phase, metrics for the evaluation of data quality are developed. Based on the results of the measurements, the causes of errors in data are identified in the "Analyse" component. Finally, the "Improve" stage is characterized by data correction. In Figure 2, one can verify that the process must be continuous. This means that after the last improvement stage, it proceeds again to the set stage, resuming the cycle. Thus, a process of continuous data improvement is pursued. The methodological approach that guided the work is highlighted next.

METHODOLOGICAL APPROACH

In terms of its nature (Silva and Menezes, 2001), the present study is characterized as an applied research. It is so characterized

Table 2. Data quality research components.

Component	Characteristics
Definition of data quality	The theme research sustains itself in the study of the concept dimensions. Aspects such as the measurement of data quality and the support of algorithms for the identification of errors are studied in this component.
Analysis of the impact of data quality in organizations	The activities involve the comparison between the data quality and some key parameters identified in a company, such as customer satisfaction.
Improvement of data quality	The third component comprises methods for the improvement of data, which are divided into 4 categories: – Processes redesign: simplification and processes rationalization, in order to minimize the chances of data errors occurrence; – Motivation for the data quality: deals with rewards and benefits to employees for the purpose of sensitizing them on the handling of data; – Use of new technologies: automated techniques for data capturing and communication among systems; – Data interpretation technologies: assist employees in data capturing and imply the redesign of processes.

Source: Constructed from TDQM (2008).

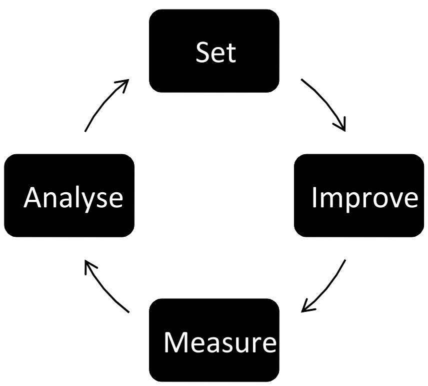

Figure 2. TDQM cycle. Source: constructed from Wang (1998).

considering the objective of generating knowledge for practical application aimed at solving specific problems. In terms of goals, several authors (Gil, 2007; Yin, 2004; Silva and Menezes, 2001; Vargas and Maldonado, 2001) classify researches as: i) descriptive; ii) exploratory; and; iii) explanatory. This research is characterized by an exploratory objective, since it aims to empirically construct and verify an approach to the improvement of data quality. Describing the nature of the existing problems in data quality, as well as the causal relations that imply the quality of the data, does not make part of the objective of this paper.

In this sense, according to the objective to which it was set, this study is characterized by the need of a qualitative approach. The objective is related to the specific environment where the construction and the implementation of the proposed approach occur. In situations with this configuration, several authors (Srivastava and Teo, 2006, Sobh and Perry, 2006; Amaratunga et al., 2002; Mangan et al., 2004; Neves, 1996; Godoy, 1995a, 1995b, 1995c) recommend the use of a qualitative approach.

Within the qualitative approach, there are different research methods, among which are the case studies (Godoy, 1995b). Case Study is a research method that is characterized by a thorough analysis of a given reality. According to Dubé and Paré (2003), Case Studies are appropriate when the object of research is complex, when it requires a vision of the whole and when the phenomenon in question cannot be analysed outside the context where it occurs. This research method is also appropriate in cases where a deeper study is necessary (Dubé and Paré, 2003; Einsenhardt, 1989). In addition, according to Ellram (1996), the Case Study can be used to: (1) explore a particular issue or problem, understanding it deeply; (2) explain a phenomenon; (3) describe a phenomenon; and; (4) predict its characteristics. Some of the purposes for using the Case Study described by Ellram (1996) justify the use of this method for this research.

Some procedures were established for conducting this work. These procedures are summarized in Figure 3. However, it is noteworthy to report that the present study covered the project design stage (pre-implementation) until the early months of the effective use of the developed production scheduling tool.

Initially, a wide review about production planning and scheduling concepts was carried out, since the system, which is the object of this research, consisted of the features of this area of knowledge. At the same time, the different existing methods for data qualification were analysed. After this review, the main methods, their characteristics and possible weaknesses were summarized. After that, it was possible to empirically verify how the data quail-fication issue was being dealt with in the system being implemented. In addition, the implications of data quality on the information generated by the system under analysis was sought to be verified.

The first data collection instrument was made viable through the personal observation of the project team members of the production scheduling system. These observations had an open characteristic, since the researcher had permission to carry out the research inside the Alfa Company. On the Alfa Company's side, among the project team members, there were managers, professionals of the industrial, commercial and Information Technology areas, besides the key users of the production sequencing tool. In the implantation team, there were business consultants and systems analysts of the company that developed the production scheduling system.

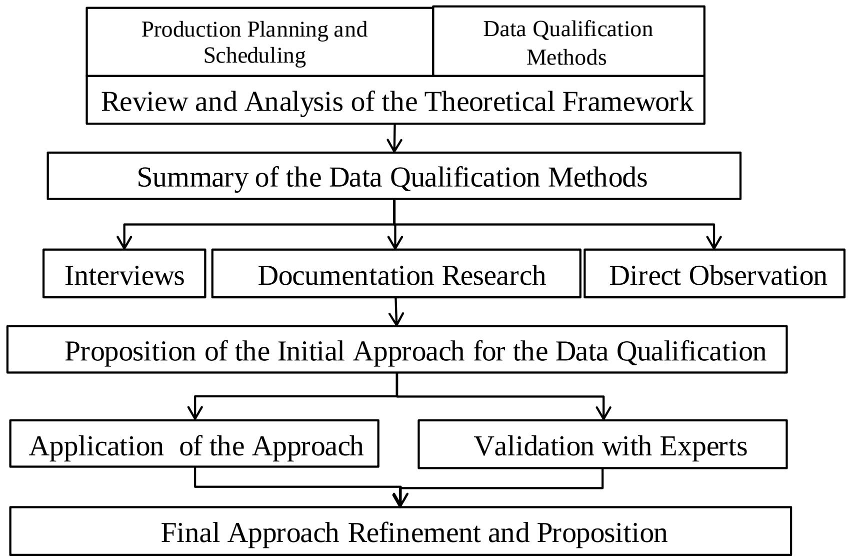

Figure 3. Work Method. Source: Authors (2010).

The second source of evidence (documents) was used to retrieve the information that was stored in printed or electronic media. Examples of this source are the electronic messages (emails), records of meetings, reports of the company that developed the system, tool guides (interface and operationalization) and the guides, prepared by the implementation team, of the procedures for identifying and solving data problems. Another instrument used to collect data during the application of the case study was the retrieval of diaries of the project. With them, it was possible to analyse the history of the system implementation, which enabled to identify, for example, the events that influenced the behavior of the indicators created for the information generated by the tool.

After collecting the information, it was possible to establish an initial approach to the data qualification. Two parallel procedures were adopted in order to allow a greater robustness of the approach. The first one was the implementation of the approach in a real situation during the tests of the system under development. The second procedure consisted of the explanation and validation of the proposed approach with different experts. It is noteworthy that, prior to the validation meetings, the project, strategies, critical success factors and flowcharts of the approach stages were sent to the experts. Table 3 briefly describes the experts who validated the research.

Finally, the initial approach was revised in light of the observation of the behavior that was presented in the application and of the contributions provided by the experts. Next, the case study and its results are presented.

CASE STUDY

In this section, the necessary information for understanding the case study is presented. Initially, the production system of the company is detailed; then, the structure of the data in the scheduling system researched is presented. Finally, an indicator used in the case study, as well as an analysis of the results will be shown.

The study was conducted in the Production Scheduling and Control (PSC) area of one of the largest organizations of the meat industry in Brazil. This organization slaughters poultry, pigs and turkeys, and exports approximately 80% of its production. The company owns nearly a dozen slaughterhouses and factories of Industrialized meat-based products, in addition to other units of a verticalized structure (feed mills, hatcheries, farms, among others). Next, its production system is detailed.

Production scheduling systems

The company where the study was conducted has two types of productive systems in its factories. To facilitate the understanding, the V-A-T analysis of the Theory of Constraints (Cox III and Spencer, 2002) will be used. One of the productive systems is characterized by the divergence of the production processes (plants of the "V" type). This kind of organization has a small number of raw materials if compared to the number of end products generated (Cox III and Spencer, 2002). A slaughterhouse has a divergent manufacturing system, where poultry and pigs go through cutting processes until they are transformed into end products.

The other productive system is represented by the convergence of the manufacturing operations (plants of the "A" type), where a small number of end products is produced from a larger number of raw materials (Cox III and Spencer, 2002). The meat-based products (sausage, bologna, among others) are produced based on the formulation defined by the Research and Development (R&D) area, and are composed by raw meat and condiments.

The production scheduling system implemented was

Table 3. Experts and their qualifications.

S/N	Qualification
1	Bachelor’s Degree in Marketing by Universidade Anhembi Morumbi and Graduate Degree in IT Governance by Instituto Mauá de Tecnologia. Experience as a consultant and in management and directorship positions in IT companies. Currently, is the Director of Data Quality Products in a Brazilian company that develops systems for this purpose.
2	Bachelor’s Degree in Business Administration by Universidade de Caxias do Sul and 9-year experience in the PSC (Production Scheduling and Control) areas of two companies of the meat industry, having held the position of corporate management in both. Coordinated the implementation of an aggregated and optimized system for production and sales planning, and of a production scheduling tool, both for the meat industry.
3	Bachelor's Degree in Production Engineering by Universidade do Vale do Rio dos Sinos (UNISINOS) and Master’s Degree student of the Graduate Program in Production Engineering and Systems at the same institution. Since the year 2003, works in the area of production and sales scheduling in the meat area, having worked at three companies of the industry in Brazil. During this period, participated in the implementation of an aggregated and optimized system for production and sales planning, and currently coordinates the demand management team of a Brazilian company in the area of meat-based foods. In that position, uses a system of production and sales planning.
4	Bachelor's Degree in Production Engineering by Universidade do Vale do Rio dos Sinos (UNISINOS) and Master’s Degree in Production Engineering and Systems by the same institution. Professor in the Undergraduate Programs of Business Administration, Accounting, Production Management and in Graduate Programs at Feevale. Professor in the Logistics Technological Program at UNISINOS. Business and consultancy experience in the Management of Productive and Quality Systems, implantation of Production Planning and Scheduling systems and in projects of Modeling, Optimization and Systemic Thinking in companies of the poultry, metal-mechanic, oil and automotive industries.
5	Bachelor’s Degree in Mechanical-Aeronautical Engineering by ITA and Master’s Degree in Electrical Engineering by UNICAMP. Has 15 years of experience in the development of decision support systems for the optimization of decision-making processes of Supply Chains (Supply Chain Management). Is Partner-Stockholder of a company in Brazil that is reference in the development and implantation of production scheduling systems for the meat industry.
6	Bachelor’s Degree in Business Administration by Instituição Educacional São Judas Tadeu, Master’s Degree in Business Administration by University of Vale do Rio dos Sinos (UNISINOS) and Ph.D. in Production Engineering by COPPE/UFRJ. Has experience in the Management area, with emphasis in Production Management, working especially on the following topics: Business Process Engineering, Theory of Constraints, Costs, Strategy.

Source: Authors (2010).

endowed of its own heuristic sequencing; therefore, not optimizing. This heuristic was defined together with the representatives of the Production Planning area of the company. The structuring of the system, from the information need point of view, included the development of three modules: planning, order tracking and scheduling. The planning module encompasses the production plans for the factories, the raw material requirements, the purchase and consumption of material, the transfers between units and the occupation of productive resources.

9The order tracking module consists on the display of: i) the stage at which the orders are (approved or not by the corporative PSC area); ii) the backlog; and; iii) the adherences (comparison between the delivery date requested by the customer and the dispatch date of the order).

Finally, the scheduling module is characterized by allowing the quotation of the delivery date for the customers, as a “Promise to Order” system. This process is performed prior to the confirmation (or not) of the order, since the date suggested by the system is submitted to the evaluation of the corporative Production Planning area and the customer.

In order for the production scheduling system to work properly, there is a dependency of the data input. In this sense, a global and integrated view of the data used for the operation of the production scheduling system will be presented next.

The structuring of data in the system

Table 4 presents the groups of data that were collected to be updated in the database of the production scheduling

Table 4. Description of the data groups in the production scheduling system.

	Description
Registration	It aggregates the data used for the structuring or modeling of other data groups; therefore, it is basic for the system to function. Example: an order may be composed by the registration data of a country, customer and product.
Slaughters	It comprises the data related to the animals that generate the end products or raw materials through the product structures (the bill of materials, also called cutting trees, which represent the "dismantling" of animals). Among these data, the slaughter scales inform the type of animal to be slaughtered (chicken, turkey or pig), its weight band, the factory where the slaughter will be done, the date, the shift and the number of animals.
Industrial	It represents the data that are collected in the productive area of the company and which are related to the products and raw materials. Examples: cutting trees, productive resources (bottlenecks and other constraints) and work schedules.
Commercial	It comprises the data related to the volumes of confirmed orders and to the demand forecast by market and product. The function of these data is to direct the decisions of the PSC system for the programming of the end products in the factories, as well as of the raw materials, as they are demanded according to the need of the formulated production (sausages, for example).
Logistic	It is composed of the data related to product transportation and storage of supplies and products. In the transportation case, times and permissions of product transfers between factories are considered, as well as storage facilities and boarding points (dry port, seaport and airport). Meanwhile, the data related to the storage include product and supply stocks in the factories, as well as storage capabilities.

Source: Authors (2010).

tool. The registration data, in addition to being inserted into the system, have a relationship with the other groups of data. The insertion of the data from the other groups depends on the availability of the registration data. Once all these data are populated in the database, the system uses them in the processing, which generates information for decision-making (production plans).

The approach applied in the case

After researching the population of the production scheduling system database in a project environment, the tests of the tool began. The project team concluded that it was necessary to adopt analysis, identification and data error correction procedures. This need was due to the existing inconsistencies between the information generated by the system and the real data of a particular selected period. The adoption of indicators and a checklist were proposed to identify data errors in the case the results did not achieve the goals set, which were determined by the high management of the company.

One of the indicators used was the planned slaughter (Figure 4), which represented the percentage of the total volume of poultry and pig (kg) planned by the system in relation to the volume (kg) actually produced by the company. The indicator covered the planning horizon (three months) of all factories. The main idea was to visualize the metric behavior in a global dimension, instead of using it separated for each production plant or in a short term (days or weeks, for example). This indicator was established for representing the beginning of the process of generating the end products; therefore, basic for further analyses (such as orders planning, for example).

In Figure 4, it can be verified that in most part of the period in which the indicator was measured, the results did not reach the goal and remained unstable. Throughout the whole period, a checklist (Table 5) was proposed for the identification and correction of data errors if the results did not reach the goal. With the support of this checklist, errors were identified according to the results of the total global slaughter indicator shown in Figure 4.

The checklist was created based on the problems encountered in the data migrated from the legacy systems of the company, as well as on the data registered by the team that worked on the project. In the case of data, there were cases where the wrong data was collected at source (e. g the measurement capacities of machines in the factory) and situations in which, after collecting the data correctly, wrong data were registered in the system during the typing process. Therefore, the team concluded that an approach to improve the accuracy of data should begin with these issues. Four focus points were chosen

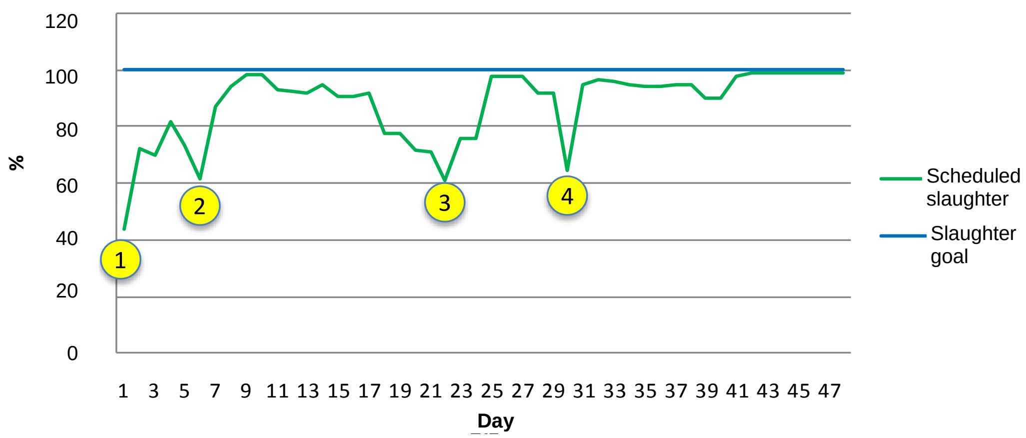

Figure 4. Indicator of the total global slaughter. Source: Authors (2010).

Table 5. Checklist for the identification and correction of data errors.

S/N	What to analyse
1	- Check if there are resources (machines, for example) without a calendar, which makes it unviable for the system to generate the production plans. The calendar sets the net available time of the resource in each day and shift of the week.
2	- Check if the orders (clients' needs) were loaded without any error in the part numbers, quantities, delivery dates, etc.
3	- Check if the unit weights registered for the animals are correct.
4	- Check if there are resources with an occupancy greater than or equal to 99%. In this case, the resource should be identified, as well as the day and the shift in which the problem occurs. - Check if the available times in the shifts are correct (calendar registration). - Analyse the products that are consuming the resource. - Check the bill of material of the product.

Source: Authors (2010).

to analyse the metric.

At point 1 in Figure 4, it was identified that there were productive resources saturated or close to this level (occupancy of 95 to 100%). Data errors were found in the registration of shifts of some resources calendars, which were corrected. At points 2 and 3, it was verified that there were errors in the bill of materials. Finally, at point 4, a malfunction in the production scheduling system was identified. Unlike the initial perception, in fact, there were data errors that took the software to a bug. It was identified an error in the bill of materials where the same product was registered in the system with a convergent and a divergent structures, which made the system go into loop and thus the output data differed from reality.

Based on the knowledge acquired during the measurement, it was diagnosed that the analysis should be improved by detailing the steps, focusing on the input data that most influenced the output data (slaughter planning); also, the analysis should be done in a sequence that could optimize the analysis and correction of the data.

In other words, after applying the checklist shown in Table 5, it was realized that a deepen detailing of the items to be checked was necessary (analyzing the planning of the production plants, for example). Furthermore, it was concluded that the analysis should be done in a sequence, starting from the analysis of the output data and going through the accuracy of the groups of data (those on Table 4). This expanded list is shown in Table 6.

Next, the proposed approach for data accuracy is presented.

Proposal of an approach for the improvement of data accuracy

Given the learning acquired in the case and the

Table 6. Expanded checklist for the identification and correction of data errors.

S/N	What to analyse
1	Generate a listing of the planned volumes and of the slaughter goals, by type of animal and factory.
2	Sort the list in a descending order of the absolute differences between the projected and the planned volumes, prioritizing the analysis of the major differences.
3	For each selected difference, identify in which factory, animal and cutting tree the difference occurred.
4	Check if the cutting tree in analysis is habilitated for the factory.
5	Check if the unit weight of the animal in analysis is correct.
6	Check if there are resources without a calendar, which makes it unviable for the system to generate the production plans.
7	Analyse the product "dismantling" structure (cutting trees).
8	Check if the orders have been loaded from the legacy system which stores that information.
9	Generate a list of all orders, in which it is possible to visualize the amount requested in the order and the order volume programmed by the system.
10	Sort the list in a descending order of the absolute differences between the allocated and the requested amount in orders, prioritizing the analysis of the major differences.
11	For every selected order, analyse the following situations: • Check if the amounts requested for the products of the order are consistent with the reality. • Check if the products of the order are habilitated in the factories that can produce them in reality.
12	Analyse the resources occupancy: • Check if there are resources with an occupancy greater than or equal to 99%. In this case, the resource should be identified, as well as the day and the shift in which the problem occurs. • Check if the available hours in the shift are correct (calendar registration). • Analyse the products that are consuming the resource.

Source: Authors (2010).

knowledge gained through the theoretical framework, the conditions for proposing an approach for the improvement of the data accuracy in the system studied were established. A preview of this model was then submitted to the evaluation of the experts, whose skills were displayed in Table 3. The final version of the approach, presented in Figure 5, was prepared based on the experts´ contributions and their validation of the proposal.

In addition to the use of checklists and a metric to measure the quality of the slaughter data, other points must be listed as learnings during the case study. First of all, there must be a premise that data must be entered in the tool with accuracy – the software should not run until it is guaranteed a minimum data quality, with the previous identification of outliers, for example. The model proposes the application of metrics in the input data, and not only for the output data, as done in the case study. On the other hand, the results after the run of the software (output data) should also be submitted to the measurement to make sure that the information will be in accordance with reality.

In Figure 5, it is verified the use of the control indexes. These correspond to the metrics that are applied to the data before the system's processing, unlike the indicators, which are verified based on the information (output data) generated by the tool's processing. It is noteworthy that a premise of the approach is that there will be no errors when importing data into the system.

Another learning, which is not shown in Figure 5, but facilitates the analysis of the results of the software, is the load of the input data gradually. These data were presented as mentioned earlier in the previous section

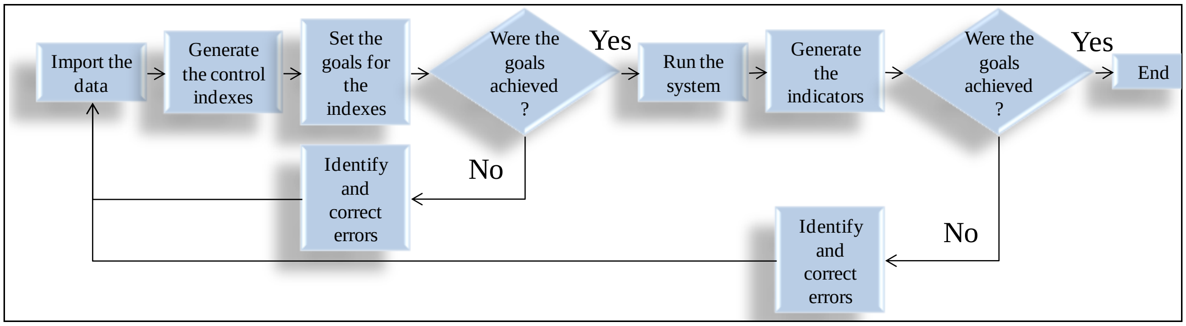

Figure 5. Approach for the improvement of the data accuracy of a production scheduling system for the meat industry.
Source: Authors (2010).

Table 7. Detailing of the identification and correction of errors based on the results of the stock average occupancy indicator.

S/N	What to analyse
1	Create a list in which it is possible to visualize the products and the respective volumes planned for the stock over the horizon.
2	Select the planned stocks in the last day of the planning horizon.
3	Sort the list in a descending order of the differences between the stock planned volumes and the maximum level considered consistent with the reality of the company.
4	For each product selected for analysis, the following situations should be verified: - Verify if the initial stock of the product is correct. - Verify if the data of the firm orders and of the demand forecast are correct. - Verify if the orders and the demand forecasts are being properly planned by the system. If not, analyse the product structures and the manufacturing routings. - Verify the permissions and the transportation times for the boarding points.

Source: Authors (2010).

and should be loaded in a predetermined sequence and in different cycles, cumulatively. The sequence was defined based on the dependency relationship among the data, which is linked to the architecture of the software database.

The stages in Figure 5 represent the generic model of the proposal, since all the flow should be applied to each data grouping in the following order: i) registration data; ii) logistical data; iii) commercial data; iv) product structures; and; v) manufacturing routes. It is important to highlight that the data should not be tested individually, but cumulatively – in other words, the commercial data will be validated while the registration and logistics data will remain populated in the database system after having been submitted to their specific control indexes

In the scope of the proposal, there are also a number of sub-processes for each data group (shown in Table 4), which in turn contain the details of the necessary steps to implement each stage of the flow. An example of a sub-process is shown in Table 7.

Next, the conclusions and recommendations for future studies are presented.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE STUDIES

Based on the research carried out, this article pointed out the data quality problem and the advantages obtained by adopting a methodological approach to improve the quality of the data used in a production scheduling system of a meat product company. The results achieved in the case presented, despite of being a direct consequence of the use of a checklist for the identification of data errors, were also supported by the indicators generated on the information of the system. Therefore, the contribution of the indicators is not only in regards to the improvement of the data accuracy, but also on the understanding of the status of a systems implementation project. They point out the possible discrepancies or anomalies – and if there were no indicators, all data should be checked at all times, activity which would be expensive in any situation.

The application of indicators and the use of the control indexes refer the proposed approach to the flow of the input-transformation-output model. Thus, both ends of the process are contemplated. Furthermore, the logic of the database population in an orderly manner, differently from all data groups at the same time, leads to the use of the Parsimony Principle, contributing for certain factors that impact the results of the system to be better identified and corrected.

The measurement of data complements the subjective analysis that is always needed in any situation. It also helps to show how close or far we are from the desired state (that is consistent in terms of scheduling).

The checklist provides a guide for the analysis, but it should be flexible in face of the variables of planning and control. It is important to know when a control index verifies the resources capacities; one should consider all the issues that surround them, in a broad sense, being open to the possibility of finding different results in the reality. Sometimes, it may be possible that they are consistent and the reference metrics used by the company in fact have some divergence of measurement and analysis.

The challenge is to identify the cause of an error and take preventive measures so that the problem does not recur, as well as opportunities for improvement arising from the analysis to avoid the appearance of other errors not occurred so far. In other words, the search for the cause of data errors can enable to identify opportunities to improve the accuracy of data.

The operationalization of this challenge is to adopt techniques of analysis, identification and correction of data errors, which in this study pointed to the metrics as the central point, and beside the checklist for error chec-king and sequential data loads, among other techniques presented in this study.

In terms of future studies on this article's theme, the implementation of the proposal could be made at another company in the same area in which the research was conducted, so that the robustness of the approach could be evaluated. Another study possibility would be the development of an array of metrics for the data and information considered important in the meat industry, initiative that would not be limited to localized issues, but would provide a global view of the data accuracy.

REFERENCES

Abef (2008). Exportação mundial de carne de frango – principais países. União Brasileira de Avicultura. Available at: <http://www.abef.com.br/noticias_portal/exibenoticia.php?notcodigo=74>. Accessed: 23 feb.

Amaratunga D, Baldry D, Sarshar M, Newton R (2002). Quatitative and Qualitative Research in the built environment: application of "mixed" research approach, Work Study, Emerald 51(1):17-31.

Barchard KA, Pace AL (2011). Preventing human error: The impact of data entry methodson data accuracy and statistical results. Comput. Hum. Behav. pp.1834-1839.

Cox III JF, Spencer MS (2002). Manual da Teoria das Restrições. Porto Alegre: Bookman.

Dubé L, Paré G (2003). Rigor in Information Systems Positivist Case Research: Current Practices, Trends and Recommendations, MIS Quart. 27(4):597-635.

Einsenhardt KM (1989). Building theories from case study research. Academy of Management Review. Stanford 14:532-550.

Ellram LM (1996). The use of the case study method in logistics research. Journal of Business Logistics. Arizona 17(2):93-138.

Gil AC (2007). Métodos e técnicas de pesquisa social. São Paulo: Atlas.

Godoy AS (1995a). Introdução à Pesquisa Qualitativa e suas possibilidades, RAE–Revista de Administração de Empresas 35(2):65-71.

Godoy AS (1995b). Pesquisa Qualitativa – Tipos Fundamentais, RAE – Revista de Administração de Empresas 35(3):20-29.

Godoy AS (1995c). A Pesquisa Qualitativa e seu uso em Administração de Empresas, RAE – Revista de Administração de Empresas 35(4):65-71.

Goldratt EM (1991). A síndrome do palheiro: garimpando informações num oceano de dados. São Paulo: IMAM.

Helfert M (2001). Managing and measuring data quality in data warehousing. Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics.

Lee YW, Pipino LI, Funk JD, Wang RY (2006). Journey to Data Quality. Boston: MIT Press.

Mangan J, Chandra L, Bernanrd G (2004). Combining quantitative and qualitative methodologies in logistics research, Int. J. Phys. Distrib. Logist. Manage., Emerald 34(7):565-578.

Neves JL (1996). Características, usos e possibilidades. Caderno de Pesquisas em Administração, Pesquisa Qualitativa 1(3):1-5.

Olson J (2003). Data quality: the accuracy dimension. San Francisco: Morgan Kaufmann.

Olson J (2002). Data accuracy: the challenge, DM Review Magazine. Available at: <http://www.dmreview.com/dmdirect/20021108/6019-1.html> Accessed: 20 may 2008.

Redman TC (2004). Data: an unfolding quality disaster. DM Review Magazine. Available at: <http://www.dmreview.com/issues/20040801/1007211-1.html> Accessed: 20 may 2008.

Silva E, Menezes EM (2001) Metodologia da Pesquisa e Elaboração de Dissertação. 3. ed., Florianópolis: UFSC Distance Learning Laboratory.

Sobh R, Perry C (2006). Research design and data analysis in realism research, Eur. J. Mark. Emerald 40(11/12):1194-1209.

Srivastava SC, Teo TSH (2006). Understanding, Assessing and Conducing Interpretative Management Research. IIMB Management Review, Bangalore, June.

Tayi GK, Ballou DP (1998) Examining Data Quality. Commun. ACM 41(2):54-57, February.

Tdqm (2008). What is Tdqm. Total Data Quality Management. Available at: <http://web.mit.edu/tdqm/www/about.shtml>. Accessed: 12 apr 2008.

Vargas L, Maldonado G (2001). Guia para apresentação de trabalhos científicos. 3. ed., Porto Alegre: Business Administration Graduate Program – PPGA/UFRGS.

Wang RY (1998). A product perspective on Total Data Quality Management. Commun. ACM 41(2):58-65.

Wang RY, Storey VC, Firth CP (1995). A framework for analysis of data quality research. IEEE Trans. Knowl. Data Eng. 7(4):623-640.