A Business Intelligence Methodology for e-Government Reverse Auctions

Rafael Souza, Adriano Pereira
Department of Computer Science – Federal University of Minas Gerais
Av. Antônio Carlos 6627 – ICEx – Pampulha – CEP 31270-010
Belo Horizonte – Minas Gerais – Brazil
{rbarra, adrianoc}@dcc.ufmg.br

Abstract

In the first quarter of year 2008, electronic reverse auctions (e-RAs) allowed the Brazilian Government to save up to US$ 270 million, which account for 87% of its acquisitions in the period, against 1% in the same period 6 years earlier – according to the Ministry of Planning, Budget and Management. A tool of such importance is subjected to fraud or even anomalous behaviors, which are difficult to detect with simple analysis over the current system. In this paper we look forward into identifying suspicious behaviors in e-government procurement systems, through the use of business intelligence techniques. The results confirm that our methodology can help discovering interesting aspects that can be used to help market players in decision support and auctioneer’s management.

1. Introduction

The Brazilian Federal Government signed in May 2005 a decree that regulates the use of electronic procurements (e-procurement) for the acquisition of common goods and services. The objectives of such publication action are to increase the transparency of the governmental procurements, to speed the acquisition process of goods and services and, mainly, to reduce the acquisition costs by public agencies.

The results (in terms of transactions, suppliers, and economy) associated with the governmental electronic acquisitions are remarkable. At ComprasNet portal [1], Brazilian Government’s public sales portal, one million transactions are handled daily by the second semester of each year, when the majority of acquisitions occur. The amount of registered users has been growing by an average tax of 3,200 new suppliers per month, reaching more than 306,000 different suppliers by October 2008. The amount of e-procurements over the last 5 years increased by 42% on average, representing an increase of 179% in monetary values. By the first trimester of 2008, 4,537 of a total of 5,228 acquisition requests (87%) were handled by electronic reverse auctions (e-RAs). In the first trimester of 2002, these percentages barely reached 1 percent [2].

There is a great concern with the functioning of the whole auctioning system, which goes far beyond the monitoring of services activity and infrastructure. Several factors, not explicitly found in these numbers, either in the results of procurements, can represent threats to the public administration or even spoil suppliers’ attendance to e-RAs. These factors could derive directly from the procurement process, since governmental agencies have autonomy to handle their own contracts, misuse of the system, and, finally, corruption practices among suppliers.

Considering the potentiality of such problems and that they are not of simple detection, the use of business intelligence, such as characterization and data mining techniques, can be an important and powerful tool to identify anomalous or suspicious behavior in e-procurements. Such tools can assist in the e-procurement management process, providing not only the recognition of fraudulent activities, but serving in the design of tighter business rules for this process.

In this work, the main objective is to evaluate hypotheses that could help answering useful questions about such important e-governmental application. It is also important to emphasize that this paper presents a novel evaluation of e-RAs using one of the most representative governmental sales portal in Latin America and the methodology and results can be used to improve the current e-procurement systems and the acquisition process in public administration.

The remainder of this paper is organized as follows: Section 2 explains some basic concepts about e-reverse auctions. Section 3 describes some relevant related works. Sections 4 and 5 present our research methodology and ComprasNet case study, respectively. This case study is explained in Section 6, and finally Section 7 outlines the conclusion and future work.
2. Electronic Reverse Auctions (e-RAs): Definition and Overview

The concept of electronic reverse auctions (e-RAs) is defined as a buying procedure that is used by companies to electronically trade goods and services [3], thus representing a business-to-government (B2G) acquisition model. Further definitions and concepts about it in the economics literature can be found at [4]. E-RAs have shown to be an efficient electronic negotiation process to drive down prices of the acquisitions maintained by government agencies. As stated by the Brazilian Ministry of Planning, Budget and Management, the positive results of e-procurements show that e-RA is a successful modality in the Federal Government acquisitions.

2.1 e-RA Life Cycle

The e-RA life cycle is summarized in Figure 1. The first step is represented by publishing an announcement of the e-RA, followed by its registration in the system (Figure 1a and 1b) – in the Brazilian scenario, the system corresponds to the sales portal. When this first step is completed, interested suppliers must register in the system, if not yet registered (Figure 1c), then post their initial proposals for specific lots of the auction they want to participate (Figure 1d). Lots are merely groups of items, often following some criteria, which can receive bids independently. As scheduled in the announcement, an auctioneer, usually a member of the purchasing agency, starts the e-RA session (Figure 1e) and the proposed starting prices are published and then classified (Figure 1f). Then, individual bidding sessions for each e-RA lot can initiate (Figure 1g). The auctioneer then invites the suppliers to submit their bids, which must necessarily improve the initial proposal (Figure 1h). The bids not necessarily need to beat each other, but the acquisition will be made considering the best offered price.

The bidding sessions provide an exclusive real-time chat system along with time stamped bidding values. The supplier identifications are kept in secret, being only visible from a certain stage of the e-RA, so that suppliers do not recognize each others, fact that could interfere in the bidding results.

The auctioneer has the autonomy of finishing the bidding sessions of each lot as far as a fair price is obtained. A fair price is usually a price as near as possible from the last acquisitions of the same goods – which are kept in track by the sales portal. Once he/she decides to finish the bidding session, a random counter is started in order to stimulate more bids. At this time, the prices can present huge dropdowns: as the suppliers do not know the exact time the random time will run off, they often bid their best offers.

By the end of the bidding sessions of each lot, the auctioneer proceeds with the next phase in the auction, known as “Acceptance phase” (Figure 1i). At such moment the final price for the lot can be negotiated between the auctioneer and the current winning supplier. Once a proposal is accepted, the auctioneer should proceed to the “Enabling phase” (Figure 1j), when the documentation of the supplier is checked for inconsistencies. Once a supplier is enabled, the period of appeals is started (Figure 1k).

During the period of appeals, any supplier who feels harmed can express an intention to lodge an appeal. The auctioneer can accept the intention or not, and if accepted, the supplier has a deadline to submit its founded reasons. Everyone who is directly affected by the appeal, like the enabled supplier, for example, has the right to submit a response to it.

Once the resources are judged by a higher authority (Figure 1l), the e-RA session can take different courses: lots or even the whole e-RA session can be cancelled or revoked, another supplier of a lot can be chosen, and even the bidding session can be repeated.

As each lot reaches a final state (Figure 1m), having an accepted proposal and an enabled supplier, the winner can be finally declared and the lot reaches an approval state (Figure 1n). If not, the lot can be considered desert. During the approval state, an evaluation technician determines whether the item meets the requirements of the announcement, and therefore the supplier is authorized to perform the sale to the government. All the collected data concerning the e-RA session is stored, along with the ranking of suppliers. If any problem arises between the approval state and the final delivery of the product, any other classified supplier can be called to accomplish the sale.

Figure 1. E-RA life cycle

Now that the main concepts about e-RA and its life cycle have been explained, the next section describes some related work.

3. Related Work

Many articles address the buyer-supplier relation in e-RAs. Results of a study based on a theory concerning buyer-supplier perspective are presented in [5]. The work investigates parameters that affect the possible savings during auctions, including lotting strategies
and the degrees of supplier visibility. The same buyer-supplier relationship is the focus of a recent work [6], which shows how the management of e-RAs has a positive impact into the suppliers.

When it comes to suppliers, [7] shows how increased levels of opportunism can harm supplier performance during the e-RA. With more variables, such as number of bidders, economic stakes and price visibility, [8] examines how the auction design can affect suppliers during an auction.

In [9] the authors review the use of e-RA compared to traditional physical auctions and describe some conditions and methods for using such type of electronic procurement. The same authors exploit the introduction of online reverse auctions in [10], highlighting its positive effects and deficiencies. Those aspects are also discussed in [11] and [12]. Despite some deficiencies, the adoption of e-RAs is shown to very be positive because of its highly competitive process [13]. Allowing companies to make considerable savings, as described in [14] and [15], e-RAs help reducing administrative costs [16] and increasing transparency of the purchasing process [17].

When it comes to public administration, the potential impact of the implementation of e-RAs can have a demonstrable economic benefit, according to [18], which investigates the impact of reverse auctions in five Southeast Asian nations. Under the trade-off fashion analysis, e-procurement is widely discussed in [19] with its applicability in public administration. Still regarding e-procurement in particular, an empirical investigation on price reduction and governance structure is presented on [20]. The adoption of e-RAs for sourcing is the discussed theme in [21], where different perceived barriers to its implementation are discussed. Such barriers are also discussed in [5].

To the best of our knowledge, this is the first work to propose a business intelligence methodology that can be applied to real scenarios to provide support decision to e-business in public administration.

4. Research Methodology

The methodological strategy adopted in this work is based on the one previously presented in [22]. Figure 2 represents such strategy in a workflow diagram.

According to the diagram, each step of the workflow is described, as follows:

i) **Initial Idea**: Defines what the work addresses, in this case, characterization and investigation of anomalous behaviors in e-RAs.

ii) **Related Work**: Identifies related work in literature that fit as an instrument to improve the knowledge and contributions in the boarded scenario.

iii) **First Observations**: Some peculiarities or even novel ideas can arise as the first observations of the environment are documented. Some questions can be made at this stage, contributing with the formulation of hypotheses that will guide the research work.

iv) **Hypothesis**: As the first observations are made, questions are formulated and it is important to try to answer those questions, leading to guesses or formulation of hypotheses. A hypothesis must be stated in a way that can be tested by the scientific method.

v) **Model**: In order to test formulated hypotheses a model is needed, so that if none is available, one should be created and data must be collected.

vi) **Experiments**: The experiments guide the validation – or neglecting – of the hypotheses.

vii) **Result Analysis**: The evaluation of the results come as a validation tool for the hypotheses and for the whole experiment. If it brings the expected results, if new questions should be answered, this activity will help proving or disproving some hypothesis.

viii) **Conclusions**: Conclusions obtained with the experimental work has to be announced and explained, providing also new important questions and topics that would be evaluated as future work.

Following the diagram, steps i, ii and iii were accomplished in sections 1 and 3 of this work. Performing steps iv, v, vi and vii for the first time, we obtain a large amount of results (rules) from the data, but only a few evidences. Following the methodology, we performed a new execution of steps iv, v, vi and vii, achieving satisfactory results. These new results are presented in section 6, containing a basic characterization and a quantitative analysis.

5. Case Study: ComprasNet

This section describes our case study using actual data from ComprasNet, which is the sales portal of the Brazilian Federal Government. It is a site maintained by the Ministry of Planning, Budget and Management, to provide information concerning the auctions and contracts promoted by the Federal Government, and allow the execution of e-procurement processes. The Web site shows the calls for proposals, the undertaken contracts, purchasing procedures for bidding and other information related to negotiations conducted by the Federal Government directly or by its agencies.

The Brazilian government sales portal is a case of success, as pointed out by some values. It had, until the

![Figure 2. Research methodology](image-url)
end of October 2008, registered an average of 3,200 new suppliers per month. By the same month, 45,925 auctions were disclosed in the year until then, surpassing the entire amount of auctions in year 2007 by a number of 6,466 auctions. It is expected that this number will be fairly surpassed, as the last trimester of the year always register the highest acquisition volume in the year. Considering year 2007, up to 1.3 billion dollars were saved using e-RAs. The first quarter of 2008 allowed the saving of up to U$S 270 million.

The dataset used in our case study consists of all government acquisitions made through e-RAs during a certain period of year 2006. This dataset is available through TAMANDUA [23], a platform of data mining services that aim at supporting public administrations, especially in what concerns to purchasing and contracting activities. In 2006 the Brazilian Federal Government bought nearly U$S 480 million in common goods and services through electronic bidding only in the first quarter of the year. With such performance, the electronic bidding accounted for 46% of the value of acquisitions performed in that period that reached nearly U$S 1 billion. The main attributes of this dataset are:

- **E-RA Key**: Key that identifies an e-RA uniquely in a certain year, used to identify the e-RA of an item.
- **Agency Key**: Key that identifies the government agency responsible for the item acquisition.
- **Official auctioneer**: Auctioneer responsible for the e-RA.
- **Description**: Short description of the item to be purchased.
- **Reference value**: The value the agency expects to pay for the item, obtained as a reference from the last item acquisition or some market survey.
- **Situation**: The state of the e-RA in its life cycle.
- **Accepted to**: Identification of the supplier who won the bidding for the item.
- **Smallest bid**: Smallest bid for the item during the bidding session.
- **Negotiated value**: Final value of the item, after the negotiation between the auctioneer and the winning supplier.
- **Negotiation gain**: 100 * (Smallest bid – Negotiated value) / Smallest bid.
- **Auction gain**: 100 * (Reference value – Smallest bid) / Reference value.

Among the items available at the 2006 dataset from ComprasNet, there are some attributes that identify important entities, which deserve a detailed characterization as a startup for the first observations and as a guide for the formulation of the main hypothesis we want to evaluate in this work. Those entities and some quantitative information are shown in Table 1.

In the next section we explain our methodology for characterizing and mining ComprasNet transactions, through the use of real examples.

### Table 1: Entities and amounts

<table>
<thead>
<tr>
<th>Entity</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>e-RAs</td>
<td>1,057</td>
</tr>
<tr>
<td>Government Agencies</td>
<td>1,568</td>
</tr>
<tr>
<td>Auctioneers</td>
<td>2,772</td>
</tr>
<tr>
<td>Items</td>
<td>471,106</td>
</tr>
<tr>
<td>Item types</td>
<td>10,204</td>
</tr>
<tr>
<td>Suppliers</td>
<td>10,926</td>
</tr>
</tbody>
</table>

### 6. Characterizing and Mining ComprasNet Transactions

The universe of possible hypotheses and questions to be formulated concerning the entities and attributes involved in the analyzed dataset is vast. Therefore, in order to present the whole characterization methodology workflow, a few specific entities were chosen to be presented, as follows.

As a first observation, among various others that could arise by looking at the entities, we chose to examine the relationship between auctioneers, government agencies and negotiation gain. The reason that explains why we chose this specific relation to investigate is due to its importance and relevance among the other investigations we have performed – which we cannot present here due to the lack of space.

The remainder of this section describes the complete evaluation and analysis of this relationship. At section 6.1, we perform a basic characterization, considering the main entities involved as an initial approach to strengthen the hypothesis. The basic characterization is followed by a quantitative analysis at section 6.2, where useful elements for the formulation are identified. The data mining procedure is described in section 6.3. Finally, at section 6.4, some results and remarks are presented.

#### 6.1. Basic Characterization

An interesting starting point to examine the available entities is through the visualization of their probability distribution through histograms. This can be done by plotting graphs showing the cumulative distribution function (CDF) of the entities. For a random variable X, its CDF is given by \( F_X(x) = P(X \leq x) \), where the right hand side represents the probability that a variable X takes a value less than or equal to x (i.e., how often a variable is below a certain volume of occurrence).

Starting with the analysis of existing government agencies willing to acquire goods and services, we can analyze the CDFs with respect to the potential of acquisition of the agencies. The potential of acquisition is understood as the amount of money an agency demands for the acquisition of an item – defined by its reference value. The higher the intention to purchase, the greater the quantity or variety of items will be.
Figure 3 shows the list of items (that account for 471,106) demanded by the agencies (1,568). The standard deviation of it is relatively high, explained by the long distance of the points and the average. The standard deviation in this case is 817, with an average equal to 300. Therefore, this indicates a covariance of 2.72, confirming a significant variability in the demand from the agencies. Analyzing this CDF, we can see that 90% of the items are concentrated in 30.55% of total agencies with intention to buy items.

Figure 4 shows the CDF of the reference value of the items requested by each agency. As can be observed, it shows that 6.44% of the agencies hold 90% of the buying potential of the whole universe of items involved in auctions.

As can be seen, 90% of total funds are held by 6.85% of the auctioneers. Obviously, this analysis is not as simple because we have the same financial volume that was considered for the agencies, however the amount of auctioneers is not the same. 190 auctioneers hold 90% of the capital, amount held by 101 agencies. Regarding the previous graphics, it is suggested that most items purchased in an auction may not demand more auctioneers. Despite this, we can conclude that a greater volume of acquisition intent may indicate the need of more auctioneers.

Figure 5. Items per auctioneer

Looking up the relation between auctioneers and reference value of the items, showed in Figure 6, we can analyze the distribution of the acquisition intention among the auctioneers. As can be seen, 90% of total funds are held by 6.85% of the auctioneers. Obviously, this analysis is not as simple because we have the same financial volume that was considered for the agencies, however the amount of auctioneers is not the same. 190 auctioneers hold 90% of the capital, amount held by 101 agencies. Regarding the previous graphics, it is suggested that most items purchased in an auction may not demand more auctioneers. Despite this, we can conclude that a greater volume of acquisition intent may indicate the need of more auctioneers.

Figure 6. Reference value per auctioneer

Considering the set of possible rules in a smaller and denser universe should be more effective, so it is possible to focus the mining task on two groups. The first one is formed by the agencies that hold the majority of acquisition intent and the second group is the one defined by the auctioneers responsible for bidding. As our first intention is to investigate auctioneers, agencies and negotiation gain, we have a
small group of agencies and auctioneers, however they hold the majority of acquisition intent – making it a very dense space for further mining, as who buys more will surely negotiate in a higher number of occasions.

6.2 Quantitative Analysis

The first part of the characterization activity served as an initial approach to open remarks for a later work of data mining, guided by the methodology used in this research project.

From the relationship between auctioneers and agencies, followed up by Figures 7 and 8, it is easy to see that the predominant case is one auctioneer working for one unique agency. In the other hand, it is somehow common to have more than one auctioneer working for the same agency, so that it distributes the auctions between them, although the majority has a single auctioneer (51.66%), as shown in Figure 8.

![Figure 7. Amount of auctioneers per agency](image7.png)

**Auctioneers per agency**

- 51.66% of auctioneers work for one unique agency.
- 28.57% work for two agencies.
- 11.42% work for three agencies.
- 3.57% work for four agencies.
- 1.08% work for more than four agencies.

![Figure 8. Amount of agencies per auctioneer](image8.png)

**Agencies per auctioneer**

- 96.68% of auctioneers work for one unique agency.
- 2.49% work for two agencies.
- 0.40% work for three agencies.
- 0.14% work for four agencies.
- 0.11% work for more than four agencies.

As a preliminary observation, the role of the auctioneer in different agencies seems to be a determining factor in their performance. This fact motivates a deep analysis of it using data mining, as we are going to do in the next session.

6.3 Data Mining

Following the investigation of the relation between auctioneers, agencies and auction gains, in this section we present an example to exploit the previous formulated hypothesis. In order to start the mining procedure, two aspects are then defined:

- **Fact:** Auctioneers have a variety of gains in bidding for items of the sessions from which they are responsible for.
- **Hypothesis:** Investigate whether an auctioneer has different gains for different agencies.

TAMANDUA, the data mining system we use in this research, uses the ECLAT algorithm [24], which tries to generate a set of rules considering cause and consequence from a certain group of data. The algorithm has two configuration parameters, confidence and support, essential for the generation of association rules. The confidence parameter is the degree of certainty one can have that a set of entities establishes a relation of cause and consequence. Support is the minimum frequency a group of items should have to be considered frequent.

Therefore, to investigate the given hypothesis, only the auctioneers who work for more than one agency were used for mining.

In order to analyze the results generated by TAMANDUA, it is important to define two measures (both described in [23]):

![Figure 9. Amount of items in each gain range](image9.png)

**Auction gain**

- Over 50% of items lie in the range of over 10% - 50%.
- The auction gain ranges for different auctioneers can be observed in Figure 9. It is also shown the percentage of items whose auction gain lies in the respective margin.
By analyzing those results together with rules generated by other auctioneers who work for more than one agency, it is noticed that more auctioneers working for an agency can inflict on higher auction gain rates. That does make sense, as more than one auctioneer working for the same agency should encourage a healthy capability competition among them.

This research also allows the identification of interesting remarks, which can be used to improve the sales portal system and provide insights for public administration. Some of them are:

- Other auctioneers were analyzed and the majority of auctions that they were responsible for had greater gain if they were the only auctioneer of the purchasing agency.
- The auctioneers responsible for the majority of the acquisitions belong to a reduced group of suppliers.
- Agencies with more than one auctioneer present higher auction gains.
- The failure of auctions is not related to the agency or auctioneer, being itself a stochastic occurrence.
- Auctioneers and agencies that hold more auctions present nearly the same failure percentage as the ones who hold fewer auctions.

7. Conclusion

This work presents a methodology of business intelligence that is used to characterize an e-procurement application. We apply this methodology in an actual e-government application from one of the most representative sales portal of Latin America.

Through the characterization of the e-RA dataset, we note that all CDF graphs follow a Pareto-like distribution. Such distribution is widely used in the observation of social, scientific, geophysical and actuarial phenomena, and its analysis can play an important role for the evaluation of business strategies. Because of that, this work could expand into new interests. A deeper analysis of the dataset alongside with observations collected during the research process suggested new hypotheses to be verified, but complementary data would be needed to enrich the data mining task. Such additional compilation would include more information from the suppliers, explanations for the failed auctions, geographic location of auctioneers and agencies, among others.

It is important to emphasize that the activity of data mining is much richer and more complete than just the single task of mining data. Preliminary activities such as characterization provide a universe of observations and knowledge concerning the dataset which are impossible to detect exclusively with the use of mining. The characterization allows consistent initial observations, improves hypotheses formulation and helps addressing existing assumptions either to validate or refuse presumptions.
In the data mining field, this work contributes with a Business Intelligence methodology with focus on data, highlighting the role of characterization. In e-Government, it can help suggesting guidelines for the detection of behaviors and formulation of questions that could raise new requirements on the current e-RA systems.

Future work includes more complete and different e-RA datasets, so that new aspects can be observed. As an example, the Brazilian Government recently launched new laws concerning the support of small suppliers and their attendance in e-RAs as a mean of promoting economic and social development. The characterization and data mining methodology could be used to investigate the effectiveness of such rules and its effects in government acquisitions.

8. References