

# Improving Social Accountability of Government Purchases Through Financial Statements Data Mining

[Demo]

Rommel N. Carvalho, Ricardo S. Carvalho, Henrique A. da Rocha, and Gilson L. Mendes

Department of Research and Strategic Information  
Brazil's Office of the Comptroller General  
SAS, Quadra 01, Bloco A, Edifício Darcy Ribeiro  
Brasília, DF, Brazil

{rommel.carvalho,ricardo.carvalho,henrique.rocha,liborio}@cgu.gov.br

## ABSTRACT

Since 2004, Brazil's Office of the Comptroller General (CGU) has been publishing several data related to government expenditures as open government data in the Transparency Portal (<http://transparencia.gov.br/>), which was considered "one of the five best practice in strategies for transparency and fight on corruption in the world" by the United Nations in 2008. It has more than 1.5 billion transactions summing up to about 6 trillion US Dollars. In 2010, CGU started publishing daily every financial statement produced by the Federal Government. Nevertheless, inconsistencies which hinder accountability have been found in this data set. Among these inconsistencies, there is diesel fuel being bought for less than a cent per liter as well as for far beyond the price paid by a regular citizen. Moreover, the structured data for classifying a product is not well organized and used, thus we rely on unstructured data to identify the product we want to compute the reference price of. These difficulties and the sheer volume of data we have to analyze make this project a non-trivial endeavor. The price reference tool we developed at CGU uses statistics, data mining, and text mining techniques to retrieve essential information for a better accountability, which includes among others: what was bought, the price paid per item, a price reference per product, which agencies are buying below this price reference (in order to identify best practices), which agencies are buying above this price reference (in order to find where the Government can improve), and which agencies are deviating too much from this price reference (in order to identify possible causes of inconsistencies). Just by analyzing the last three years of purchases from a sample of 52 different products varying from fuel, food, and office supply, CGU was able to identify a potential in savings around 30 million US Dollars, which represents about 20% of the total

spent. This analysis has allowed CGU to learn guidelines on purchasing best practices in different product categories, forwarding them to the agencies in need, as well as to identify and solve problems related to the use of a centralized financial system, improving accountability. Finally, this information will be incorporated in the Transparency Portal, allowing every citizen to understand how much the Government is really paying, in general, for products. In fact, this project is one of the commitments Brazil has made in its second action plan within the Open Government Partnership (OGP): *"to develop a database containing reference prices for the most purchased items by the Federal Government, from data published on the Transparency Portal. The interface shall provide for the identification of items average prices, thus constituting an efficient strategy for formulating budgets and procurements, disseminating best practices in public purchases, as well as for supporting actions aimed at fighting corruption, especially in circumstances where over-price purchases are identified."* Therefore, with the publication of the results obtained from this endeavor, it will be possible to improve social control and provide a solid accountability not only to CGU, as an internal control agency, but also to Brazil's citizens who, in the end, are the ones paying the bill.

## Categories and Subject Descriptors

I.3 [Industrial and Government Track]: Emerging category

## General Terms

Application

## Keywords

price reference, accountability, data mining, text mining, transparency, open government data

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

KDD '14 New York, New York USA

Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

## 1. SCREENSHOTS



BRASIL

Acesso à informação

Participe

Serviços

Legislação

Canais

Controladoria-Geral da União

Portal da Transparência

GOVERNO FEDERAL

Perguntas frequentes

Contato

Glossário

Links

Manual de navegação

Acesso rápido

Selecione...

OK

Você está em:

Início » Detalhamento Diário das Despesas » Detalhamento do Documento

Detalhamento Diário das Despesas

Detalhamento do documento: 2013NE800205

DADOS BÁSICOS

Fase:	Empenho		
Documento:	2013NE800205	Tipo de Documento:	Nota de Empenho (NE)
Data:	09/10/2013		
Tipo de Empenho:	ORDINARIO	Espécie de Empenho:	Original
Órgão Superior:	52000 - MINISTERIO DA DEFESA		
Órgão / Entidade Vinculada:	52921 - FUNDO DO EXERCITO		
Unidade Gestora Emitente:	167245 - POLICLINICA MILITAR DE NITEROI		
Gestão:	00001 - TESOURO NACIONAL		
Favorecido:	04.960.002/0001-83 - DIBOA COMERCIAL LTDA.		
Valor:	R\$ 1,387.50		

DADOS DETALHADOS

Observação do Documento:	AQUISIÇÃO DE MATERIAL DE CONSUMO DE EXPEDIENTE. MAPA 145. 2013NC418217 - FEX, DE 12SET13. SUBITEM 17, PDR 00123/10/13. PROC ORIGEM: 2013PR00002		
Esfera:	2 - ORÇAMENTO DE SEGURIDADE SOCIAL	Tipo de Crédito:	A - INICIAL (LOA)
Fonte de Recursos:	50 - RECURSOS NAO-FINANCEIROS DIRETAM. ARRECADADOS		
Unidade Orçamentária:	52921 - FUNDO DO EXERCITO		

Detalhamento do Gasto

Subitem da Despesa	Quantidade	Valor Unitário (R\$)	Valor Total (R\$)	Descrição
17 - MATERIAL DE PROCESSAMENTO DE DADOS	150	9.25	1,387.50	150,00000 resma CONJUNTO MULTIUSO ESCRITÓRIO Papel Sulfite - material papel alcalino, branco, comprimento 297 mm, largura 210mm, formato A4, aplicação em impressora a jato de tinta, laser e copiadoras, gramatura 75 g/m2, resma com 500 fls - OBS: ao ser utilizado, o papel não poderá manchar ou se deformar após sua passagem pelos referidos equipamentos. Marcas de referência: Copimax, chamex ou de melhor qualidade. MARCA: Chamex ITEM DO PROCESSO: 00145 ITEM DE MATERIAL: 000150460

Figure 2: An example of a financial statement available at the Transparency Portal (<http://transparencia.gov.br/>). It has information about the agency that bought the product, the supplier, date of purchase, quantity, value, among others. The description in the lower right corner of the figure (*Descrição*) is the one used to help identify a specific product. Once the system computes the reference price, it allows the user to come back to the financial statements used to compute that price.



## Papel A4 75 g-m2 Branco

de R\$ 9,96 a R\$ 10,13 (resma)

### Observações:

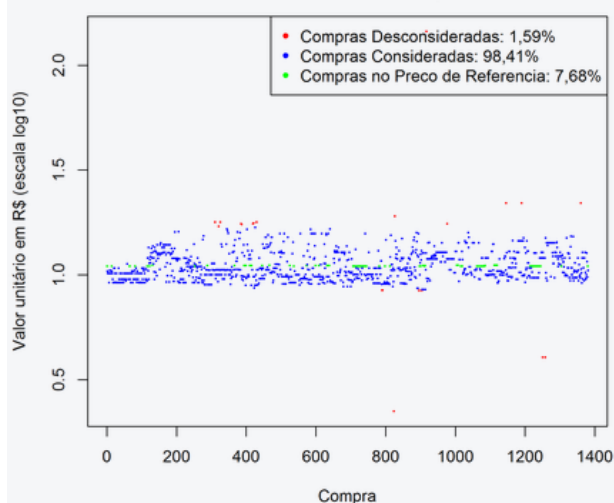
Essa ficha foi criada a partir da metodologia desenvolvida pela DIE/CGU, baseada em técnicas de mineração de dados, para extrair as informações necessárias dos empenhos disponibilizados no Portal da Transparência de forma a possibilitar o cálculo de faixas de preço para produtos comprados pelo Governo Federal. Além disso, são apresentadas também algumas análises feitas com base nessas faixas de forma a enfatizar sua importância para a melhoria da gestão dos recursos públicos tanto na identificação de gestores com boas políticas de compras como na identificação de oportunidades de economia.

### RESUMO

Quantidade total de Compras	1.381
Valor total das compras	R\$ 13.263.367,22
Faixa de preço geral	R\$ 1,23 a R\$ 144,00
Faixa de preço comum	R\$ 7,50 a R\$ 15,80
Faixa de preço ideal	R\$ 9,96 a R\$ 10,13

## Definição das Faixas de preço

Gráfico de Dispersão  
Valor Unitário vs Compra



Valor Unitário vs Compra

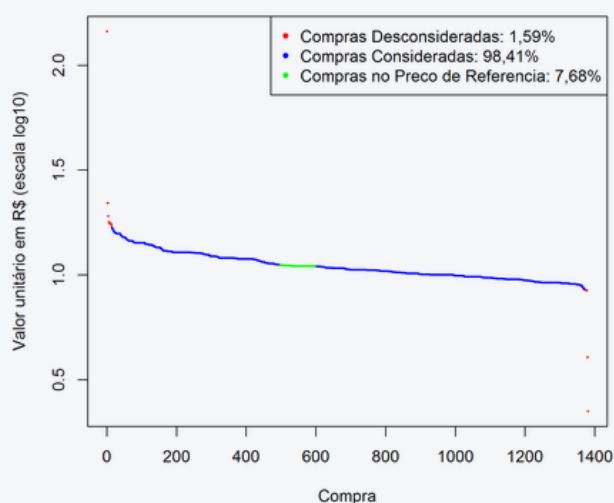


Figure 4: A summary of the results, including the price reference, in 2013 for national purchases of 75 g/m<sup>2</sup> A4 paper. The plots show the identified anomalies in red, the most frequent paid prices in blue, and the purchases in the ideal price (95% confidence interval for the mean within the non-anomalies) in green.

## Análise dos Resultados

**Economia em Potencial de R\$ 331.556,57 ou 2,50%**

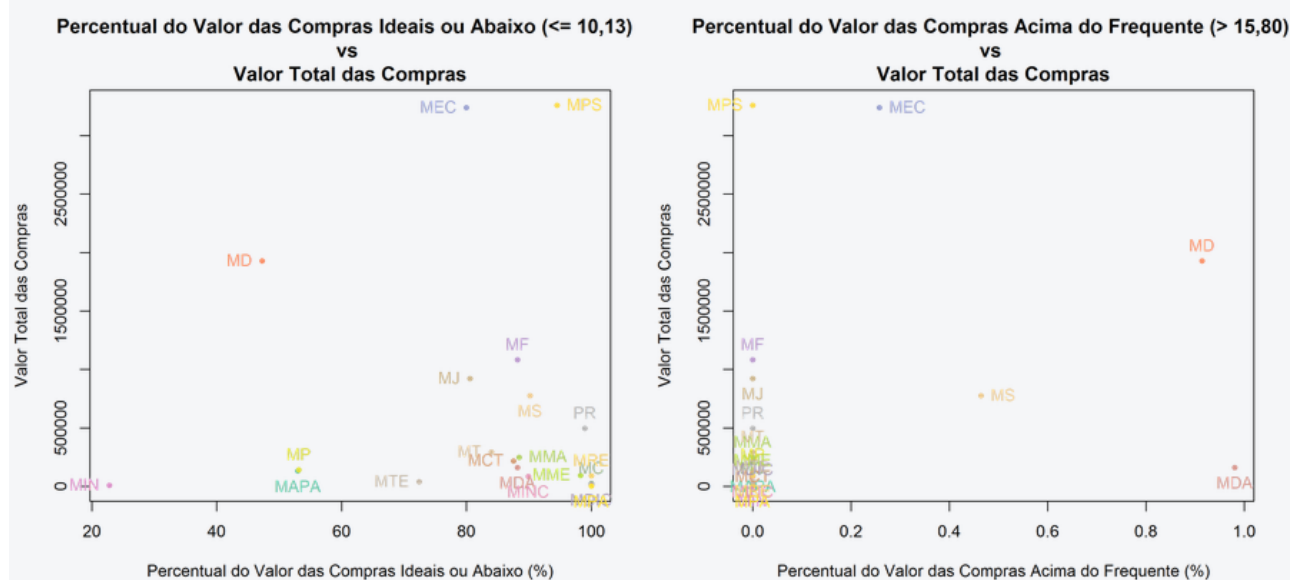


Figure 5: Part of the analysis of the results obtained from a reference price, which is generated automatically for every product. It presents a summary of how much the government could potentially save with this product. Furthermore, it allows the identification of best practices by identifying the agency that buys in the price reference or cheaper.