

Extração, Tratamento e Análise de Dados Relativos às Cotas Parlamentares de Deputados Federais no Brasil

Thiago Capaverde, M. Carmo Nicoletti, Eduardo J. Huerta Yero
UNIFACCAMP

Campo Limpo Paulista - SP, Brasil

thiagocapaverde@yahoo.com.br,
carmo@cc.faccamp.br,
edu.huerta@gmail.com

Abstract. This article presents and describes a work in progress that aims at the extraction, treatment and analysis of public data, related to the financial quotas granted to Brazilian federal deputies, for the exercise of their functions. It is intended, with the statistical analysis of such data, to show trends in the use of quotas, with the objective of establishing a profile of the investments made, as well as detecting any irregularities and problems. The data will also be analyzed, considering various perspectives, particularly, the year in which they were collected, the state of the federation represented by the federal deputy who made use of the quota and the different categories that were contemplated with the use of the quota.

Keywords: data extraction, data processing, data analysis, parliamentary quotas, data mining.

Resumo. Este artigo apresenta e descreve um trabalho em andamento que tem por finalidade a extração, tratamento e análise de dados públicos, relativos às cotas financeiras concedidas a deputados federais brasileiros, para o exercício de suas funções. Pretende-se, com a análise estatística de tais dados, evidenciar tendências no emprego das cotas, com o objetivo de estabelecer um perfil dos investimentos feitos, bem como detectar eventuais irregularidades e problemas. Os dados serão analisados, também, considerando várias perspectivas, particularmente, ano em que foram coletados, estado da federação representado pelo deputado federal que fez uso da cota e as diferentes categorias que foram contempladas com o uso da cota.

Palavras-chave: extração de dados, tratamento de dados, análise de dados, cotas parlamentares, mineração de dados.

1. Contextualização

A Constituição Brasileira (CB) [Brasil 2020] informa que a República Federativa do Brasil é formada pela união indissolúvel dos Estados e Municípios e do Distrito Federal. Atualmente o Brasil (União) é dividido política e administrativamente em 26 estados e o Distrito Federal, que somam um total de 27 unidades federativas. A CB estabelece os princípios da soberania popular e da representação, em que o poder político pertence ao povo e é exercido pelo povo por meio de representantes eleitos ou diretamente, como rege seu art. 1º. Em seu art. 2º, a CB constitui três Poderes independentes, nomeados Legislativo, Executivo e Judiciário. Como determina a CB, o Poder Legislativo do Brasil no âmbito federal é exercido pelo Congresso Nacional (CN), constituído pela Câmara dos Deputados (CD), também conhecida como Câmara Baixa e pelo Senado Federal (SF). A CD é composta por deputados federais (DF) e o SF, por senadores. Particularmente, a CD é composta pela Mesa da Câmara dos Deputados do Brasil, pelo Colégio de Líderes e pelas Comissões, que podem ter caráter permanente, temporal, especial ou de inquérito.

Dentre as categorias apresentadas anteriormente que compõem o CN, o foco desse trabalho considera apenas aquela dos representantes eleitos para comporem a CD *i.e.*, a categoria dos DFs. A representatividade dos estados na CD é feita levando em consideração o número de habitantes por estado. Essa proporcionalidade, entretanto, é limitada a um mínimo de oito deputados e a um máximo de setenta deputados por estado e 513 deputados no total [Wikipedia 2020a], lembrando que deputados federais são eleitos por voto direto, com mandato de quatro anos, sendo que um indivíduo pode concorrer a sucessivas reeleições. Eleitores podem votar apenas em candidatos de seu estado. Dentre as várias atribuições a um DF destaca-se o ato de legislar, com possibilidade para propor, alterar, revogar, etc. leis, leis complementares e emendas. Como informado no Portal da Câmara dos Deputados [Camara-Deputados 2020], deputados federais são eleitos pelo sistema proporcional, em que a votação final obtida por um candidato é influenciada por dois fatores: (1) pela soma dos votos obtidos por todos os candidatos do mesmo partido ou coligação e (2) pelo número de votos de legenda (sigla do partido). Dentre os auxílios, cotas, reembolsos e verbas a que DFs têm direito, estão (1) ajuda de custo, (2) assistência à saúde, (3) auxílio-moradia e apartamento funcional, (4) cota para o exercício da atividade parlamentar, (5) cotas gráficas, (6) serviços de transporte rodoviário, (7) verba de gabinete, (8) viagens em missões oficiais, a serviço ou para treinamento (ver detalhes em [Camara-Deputados 2020b]).

Em [Controladoria-Geral 2020] o Portal da Transparência (PT) é apresentado como um *site* do governo federal administrado pela Controladoria-Geral da União (CGU), que atua na defesa do patrimônio-público, da transparência e no combate à corrupção. O PT disponibiliza informações sobre a gestão pública e, principalmente, sobre a aplicação do dinheiro público no Brasil, possibilitando que cidadãos possam acompanhar o uso de recursos federais arrecadados, particularmente aqueles advindos de impostos.

No contexto das análises específicas pretendidas, a aquisição e familiarização com o conceito de *transparência* é fundamental. O Portal da Transparência explica o conceito de transparência adotado pelos órgãos públicos, com o comentário: “... uma pessoa também pode ser transparente: não porque se enxerga através dela, mas sim porque ela é sincera e não tem nada a esconder. O mesmo vale para órgãos públicos.” Outro conceito relevante para os objetivos do projeto de pesquisa é o da *prestação de contas ampliada*, também referida na literatura como *responsabilização*; ambos os termos são acepções da tradução do termo *accountability*. Em [Wikipedia 2020b] as várias acepções do termo são apresentadas por meio de uma descrição geral, que envolve responsabilidade com ética, implicando a obrigação e transparência, de participantes de órgãos administrativos ou representativos, de prestar contas a instâncias controladoras ou a seus representados. De uma maneira simplificada, a prestação de conta ampliada implica órgãos/pessoas que desempenham funções sociais relevantes, explicarem regularmente à sociedade o que estão fazendo, porque estão fazendo e como estão fazendo, particularmente no que tange ao uso/aplicação de bens e do erário.

Este artigo descreve os principais objetivos bem como as atividades planejadas para atingi-los, de um Projeto de Pesquisa (PP) em nível de mestrado que está em desenvolvimento junto ao PMCC da UNIFACCAMP. O principal objetivo do projeto como um todo é o da extração, tratamento e análise de dados disponibilizados publicamente, relativos às cotas financeiras concedidas a deputados federais, para o exercício de suas funções. Pretende-se, com a análise de tais dados, evidenciar áreas prioritárias em que cotas foram investidas, com o objetivo de estabelecer um perfil dos

investimentos feitos. Os dados serão analisados, também, considerando várias perspectivas, particularmente, ano em que foram coletados, estado da federação representado pelo deputado federal que fez uso da cota, categoria discriminada na CB, em que a cota foi usada. O artigo está organizado nas seguintes seções. A Seção 2 apresenta a metodologia utilizada durante o levantamento bibliográfico inicial conduzido, que abordou documentos, *sites*, fatos e trabalhos acadêmicos que têm intersecções com a pesquisa sendo conduzida. A Seção 3 focaliza brevemente como os dados estão publicamente disponíveis para acesso, e os vários formatos em que tais dados estão disponibilizados. A Seção 4 descreve as principais atividades previstas para a extração, tratamento e análise dos dados relativos às Cotas Parlamentares de deputados federais, bem como as técnicas e métodos que são pretendidos para a condução de tais atividades.

2. Metodologia Adotada na Condução da Revisão Bibliográfica

Esta seção aborda a metodologia do trabalho inicial realizado com o objetivo de conduzir um levantamento bibliográfico de pesquisas que utilizam dados governamentais, publicamente disponibilizados com objetivos variados e, particularmente, com o objetivo de analisar o emprego de cotas financeiras com vistas ao exercício do cargo parlamentar, no caso, o de deputado federal.

A ideia inicial da proposta de pesquisa foi induzida pela possibilidade de acesso às informações governamentais, disponibilizadas publicamente. No caso, informações referentes às possíveis utilizações das cotas financeiras atribuídas a deputados federais, com vista ao escrutínio e análise das várias possibilidades de seu emprego. A metodologia que foi sendo definida ao longo do processo de revisão bibliográfica teve quatro passos sequenciais distintos, como segue:

(1) Levantamento de documentos e de vários *sites* governamentais que abordam verbas públicas, acompanhado de leituras/estudos dos respectivos conteúdos, com vista à aquisição de conhecimento na área de políticas públicas e, também, dos principais conceitos envolvidos nela. Foram lidos e examinados os seguintes documentos e *sites*: [Constituição 1988] [Camara-Deputados 2020a] [Camara-Deputados 2020b] [Wikipedia 2020a] [Controladoria-Geral 2020].

(2) Devido ao foco da pesquisa envolver o conceito de *transparência*, o conteúdo do *site* nomeado Portal da Transparência (PT) [Controladoria-Geral 2020] foi cuidadosamente estudado. Em [Kopits & Craig 1994] são abordadas e discutidas várias das muitas dimensões que esse conceito tem. O trabalho apresentado em [Zuccolotto & Teixeira 2014] investiga o conceito de transparência fiscal no contexto brasileiro.

(3) O conceito de *prestação de conta ampliada (accountability)*, também relevante para a pesquisa sendo realizada, foi adquirido por meio da leitura de [Wikipedia 2020b], [Telhado 2019], [Vivyan 2012].

(4) Durante o processo de busca realizado foram também identificadas algumas publicações tangencialmente relacionadas ao tema da pesquisa, que permitiram aprender sobre uso dos conceitos de transparência e prestação de contas em situações diversas das que direcionam esse trabalho. Merecem destaque:

(4.a) O artigo [Vivyan 2012] examina a prestação de contas associada a processo eleitoral, após o escândalo associado às despesas na Inglaterra.

(4.b) Em [Fomina & Vynnychenko 2017] é feita uma avaliação do impacto da transparência fiscal no desempenho de indicadores orçamentários. O

principal objetivo do estudo foi determinar a relação entre índices de transparência fiscal e indicadores de desenvolvimento dos países envolvidos no estudo.

- (4.c) No estudo apresentado em [Brito *et al.* 2014] são abordados, entre outros, os gastos dos parlamentares da câmara de deputados e a construção do *site* intitulado *Meu Congresso Nacional*, sediado no endereço *meucongressonacional.com*, desde novembro de 2013, que teve por objetivo difundir dados disponibilizados por órgãos públicos, em uma formatação mais acessível ao cidadão comum.
- (4.d) Os artigos [Eggers & Fisher 2011] [Besley & Larcinese 2011] analisam o envolvimento de parlamentares em escândalos relacionados ao uso do erário inglês.
- (4.e) Um estudo de caso de jornalismo computacional, com foco no escândalo financeiro envolvendo membros do parlamento inglês é apresentado em [Daniel & Flew 2010].

3. Dados Relativos às Cotas Parlamentares

No que segue é brevemente abordado como os dados relacionados às cotas federais atribuídas a deputados federais, para exercício da atividade parlamentar, estão disponibilizados para acesso público, na página [Câmara-Deputados 2020c]. Na página do *site* os dados estão em arquivos compactados do tipo: XML, JSON, CSV e XLSX, como listados na Tabela 3.1. A compactação é necessária devido ao volume de dados que tais arquivos contem.

Como anunciado na página inicial *do site*, a justificativa para a disponibilização ao público dos dados em tais tipos de arquivo é a de permitir que possam ser utilizados diretamente por sistemas de gerenciamento de banco de dados (SGBD) existentes, para serem consultados bem como tratados, com o objetivo de análise e extração de informações que possam ser consideradas relevantes para acompanhamento das mais variadas metas. É importante lembrar que os arquivos sofrem atualizações diárias, uma vez que tais dados são atualizados diariamente.

Além dos dados relativos ao ano corrente, o site também disponibiliza os dados relativos ao ano anterior e aos demais anos anteriores (2017–2009). Cada instância de dado é um vetor que descreve uma instância de utilização da verba federal recebida por um particular deputado federal. Instâncias são descritas por 28 atributos, que têm diferentes tipos (numérico real, numérico inteiro, textual, binário, etc.).

Tabela 1 Tipos de arquivos de dados disponibilizados na página [Câmara-Deputados 2020c].

| Extensão do arquivo com os dados | Descrição |
|--------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| XML (<i>Extensible Markup Language</i>) | especialização da SGML (<i>Standard Generalized Markup</i>), usada para descrever dados. |
| JSON (<i>JavaScript Object Notation</i>) | linguagem com vista à troca de dados entre diferentes sistemas. |
| CSV (<i>Comma Separated Values</i>) | formatação usada em arquivos de texto, em que unidades de texto são separadas por um delimitado; a vírgula é usualmente utilizada como delimitador. |
| XLSX | arquivos cujos conteúdos são planilhas de dados gerados sob o Microsoft Excel 2010 e versões seguintes. |

4. Atividades em Progresso e Considerações Finais

Algumas das atividades elencadas a seguir estão sendo realizadas, enquanto as demais estão previstas para serem realizadas em continuidade, em um período estimado de cinco meses.

Em relação à extração de dados, foi decisão inicial considerar todos os dados disponibilizados no Portal da Câmara dos Deputados [Câmara-Deputados 2020c], ou seja, um arquivo de dados referente a cada ano de 2009 até 2018, o que totaliza 10 arquivos de dados.

O tratamento dos dados em princípio envolverá:

- (1) escolha do tipo de arquivo de dados, dentre os disponibilizados no Portal (ver Tabela 1);
- (2) serão necessários, inicialmente, o estudo, escrutínio e identificação de descritores efetivamente relevantes que possam viabilizar análises e análises comparativas, sem a interferência de descritores irrelevantes;
- (3) assume-se que em todos os 10 arquivos de dados, os dados estejam descritos pelo mesmo conjunto de descritores (atributos). Na eventualidade desse fato não se verificar, o foco dessa atividade do trabalho será apenas em descritores que compõem nas descrições dos dados nos 10 arquivos; será necessário, portanto, determinar esse conjunto reduzido de atributos;
- (4) decisões deverão ser tomadas com relação a qual procedimento adotar, quando da ausência de valor de atributo(s) em instâncias de dados;
- (5) como os dados cobrem um período de 10 anos, serão necessários alguns ajustes, particularmente no que tange à inflação anual, uma vez que vários dos valores de atributos são valores monetários; os dados precisam ser corrigidos levando em consideração as taxas anuais de inflação;
- (6) cogita-se, para o armazenamento dos dados, o uso do banco de dados relacional PostgreSQL [PostgreSQL 2020], que é considerado um dos mais avançados gerenciadores de bancos de dados de código aberto.

Como adiantado na Seção 1, a análise dos dados tratados tem por objetivo detectar irregularidades e tendências no emprego de cotas, evidenciar áreas prioritárias em que cotas foram investidas, com o objetivo de estabelecer um perfil dos investimentos feitos. Os dados serão analisados, também, considerando várias perspectivas, particularmente, ano em que foram coletados, estado da federação representado pelo deputado federal que fez uso da cota, categoria discriminada na CB, em que a cota foi usada.

Agradecimentos O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES)-Código de Financiamento 001.

Referências

- [Besley & Larcinese 2011] Besley, T.; Larcinese, V. (2011) Working or shirking? Expenses and attendance in the UK Parliament, *Public Choice*, v. 146, pp. 291–317.
- [Brito et al. 2014] Brito, K. S., Santos, M.; Garcia, V. Meira, S. R. L. (2014) Usingn parliamentary Brazilian open data to improve transparency and public participation in Brazil, In: Proc, of the 15th Annual International Conference on Digital Government Research, pp. 171–177, <https://doi.org/10.1145/2612733.2612769>
- [Câmara-Deputados 2020a] Portal da Câmara dos Deputados. Ac: 17/7/2020. <https://www.camara.leg.br/noticias/544742-deputados-sao-eleit-os-pelo-sistema-proporcional-veja-como-funciona/>

- [Câmara-Deputados 2020b] Portal da Câmara dos Deputados. Ac: 17/7/2020.
https://www2.camara.leg.br/transparencia/aceso-a-informacao/copy_of_perguntas-frequentes
- [Câmara-Deputados 2020c] Portal da Câmara dos Deputados–Cota Parlamentar, Ac: 17/7/2020.
<https://www2.camara.leg.br/transparencia/cota-para-exercicio-da-atividade-parlamentar/dados-abertos-cota-parlamentar>
- [Constituição 1988], Constituição da República Federativa do Brasil de 1988, Brasília, DF: Presidência da República. Ac: 17/7/2020
<https://www2.camara.leg.br/atividade-legislativa/legislacao/constituicao1988.html>.
- [Controladoria-Geral 2020] Portal da Transparência, CGU, Ac: 17/7/2020.
<http://www.portaltransparencia.gov.br/download-de-dados>
- [Daniel & Flew 2010] Daniel, A.; Flew, T. (2010) The Guardian reportage of the UK MP expenses scandal: a case study of computational journalism, In: *Communications, Policy and Research Forum*, Sydney, Australia.
- [Eggers & Fisher 2011] Eggers, A. C.; Fisher, A. C. (2011) Electoral Accountability and the UK Parliamentary Expenses Scandal: Did Voters Punish Corrupt MPs? *Social Science Research Network* (SSRN), <https://ssrn.com/abstract=1931868> or <http://dx.doi.org/10.2139/ssrn.1931868>
- [Fomina & Vynnychenko 2017] Fomina, J.; Vynnychenko, N. (2017) Fiscal transparency: cross-country comparisons, *Business Ethics and Leadership*, v. 1, no. 2, pp. 39-46.
- [Kopits & Craig 1994] Kopits, G.; Craig, J. (1994) Transparency in government operations, Occasional Paper no. 158, *International Monetary Fund*, Washington DC, 42 pgs.
- [PostgreSQL 2020] PostgreSQL Ac: 17/07/2020.
<https://www.postgresql.org/>
- [Telhado 2019] Telhado, L. A. S. (2019) Governos locais e *accountability*: um estudo sobre a difusão do controle institucional da União, Dissertação de Mestrado, Programa de Pós-Graduação em Ciência Política, Instituto de Ciência Política, UnB, 106 pgs.
- [Vivyan 2012] Vivyan, N. (2012) Representative misconduct, voter perceptions and accountability: Evidence from the 2009 House of Commons expenses scandal, *Electoral Studies*, pp. 750–763.
- [Wikipedia 2020a] Wikipedia. Câmara dos Deputados do Brasil. Ac: 17/7/2020.
https://pt.wikipedia.org/wiki/C%C3%A2mara_dos_Deputados_do_Brasil
- [Wikipedia 2020b] Wikipedia. Accountability. Ac: 17/7/2020.
<https://pt.wikipedia.org/wiki/Accountability>
- [Zuccolotto & Teixeira 2014] Zuccolotto, R.; Teixeira, M. A. C. (2014) As causas da transparência fiscal: evidências nos estados brasileiros, DOI: 10.1590/1808-057x201410820