

Método e Aplicação de Automatização Para Download de Dados Abertos em Portais de Transparência

Cledisson Patrick Pereira Martins¹, Giliad de Souza Silva²

¹Faculdade de Sistemas de Informação – Universidade Federal do Sul e Sudeste do Pará (UNIFESSPA)
Marabá, PA – Brasil

²Instituto de Estudos em Desenvolvimento Agrário e Regional – Universidade Federal do Sul e Sudeste do Pará (UNIFESSPA)
Marabá, PA – Brasil

{p4tmars,giliad.souza}@unifesspa.edu.br

Abstract. *This paper presents partial results of methods and applications for automating data downloads on public transparency portals, relating access to open data and the difficulty in accessing them by researchers and the Brazilian public. A method and application was developed to facilitate the use of open data by a research group on the use of mineral compensation in affected towns.*

Resumo. *Este artigo apresenta os resultados parciais de métodos e aplicação de automatizar o download de dados em portais de transparência pública, relacionando o acesso aos dados abertos e a dificuldade de acesso aos mesmos por pesquisadores e o público brasileiro. Foi desenvolvido um método e aplicação para facilitar o uso de dados abertos por um grupo de pesquisas sobre a utilização de compensação mineral nos municípios afetados.*

1. Introdução

Os dados abertos podem ser definidos com dados que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa [Open Data Handbook 2024]. Ainda segundo o Open Data Handbook, para que esses dados sejam realmente abertos, eles devem ter um dos aspectos mais importantes: a disponibilidade e acesso [Open Data Handbook 2024]. Isto quer dizer que, os dados devem estar disponíveis como um todo, preferencialmente possíveis de serem baixados pela internet, e esses dados devem também estar disponíveis de uma forma conveniente e modificável [Open Data Handbook 2024].

No Brasil, a Lei de Acesso à Informação (LAI) - Lei nº 12.527/2011, define que cabe aos órgãos e entidades do poder público assegurar a gestão transparente da informação, propiciando amplo acesso a ela e sua divulgação [BRASIL 2011]. Sendo assim, os municípios brasileiros contratam sistemas de informações para executar a legislação vigente. No entanto, alguns dos portais contratados não atendem o aspecto de disponibilidade e acesso como descrito, dificultando o trabalho de pesquisadores que utilizam os dados.

Uma maneira de acessar e disponibilizar dados desses portais é utilizando um software de *web scraping*, que seria uma técnica para capturar e organizar informações de um website [Cambridge Dictionary, 2024]. Esta técnica pode ser usada em diversas

áreas de negócio, como no mercado imobiliário [Williams e Wheaton 2021] e em biomedicina através da bioinformática [Glez-Peña et al. 2013].

O grupo de pesquisa “De Olho na CFEM” que, utilizando o acesso aos portais da transparência, buscam responder duas perguntas: 1. Os recursos que a mineração deixa nos cofres das prefeituras na forma de CFEM – Compensação Financeira pela Exploração Mineral, estão sendo usados para melhorar as condições de vida da população? 2. Eles estão servindo para construir um desenvolvimento não dependente da exploração mineral, que mais dia menos dia se esgotará? [De Olho na CFEM 2024]. O grupo então, analisa dados abertos disponibilizados pelos municípios brasileiros, e tem como tema e preocupação chave, dentre outros, a transparência, isto é, a forma como a prefeitura apresenta as informações sobre os valores recebidos da CFEM e sobre o seu uso.

O grupo de pesquisa “De Olho na CFEM”, procurou auxílio para automatizar o acesso aos dados abertos no portal da transparência “Governo Transparente”, que é utilizado por municípios abordados na pesquisa.

Os dados abertos requisitados pelos grupos são os de documentos de empenhos, liquidações e pagamentos de despesas dos municípios. Como exemplo, para os documentos de pagamentos, são necessários os dados: 1. Número do pagamento; 2. Data do pagamento; 3. Valor do pagamento; 4. Número do empenho; 5. Data do empenho; 6. Tipo; 7. Credor; 8. Modalidade da licitação; 9. Unidade orçamentária; 10. Função; 11. Subfunção; 12. Programa do governo; 13. Programa/Atividade; 14. Natureza da despesa; 15. Fonte do recurso; 16. Histórico. Sendo assim, para determinado documento de pagamento, o grupo precisa dos dados do pagamento e os dados do documento de empenho de despesa ao qual o pagamento é referente.

Este artigo apresenta os resultados parciais obtidos através do apoio em desenvolvimento de software de web scraping para o grupo “De Olho na CFEM” no acesso de dados abertos. Mostrando como estudo de caso a necessidade de acesso aos dados abertos pelo grupo de pesquisa, utilizando o portal da transparência “Governo Transparente” [Aspec Informática 2024], que é contratado por alguns municípios abordados pelo grupo de pesquisa.

2. Metodologia

2.1. Portal da Transparência Governo Transparente

O portal da transparência “Governo Transparente” não disponibiliza os dados da maneira em que o grupo de pesquisa necessita, sendo possível apenas, no exemplo para documentos de pagamentos de despesas, visualizar e baixar a lista de documentos de pagamentos (ver Figura 1). Os documentos de dados abertos disponíveis para download possuem os dados: Data do pagamento, Número do documento de pagamento, Número do empenho, Unidade gestora, Credor, CPF/CNPJ, Natureza da despesa, Modalidade da licitação, Histórico, Registro e Valor do pagamento. Ou seja, os dados disponíveis para download estão incompletos para os fins do grupo de pesquisa.

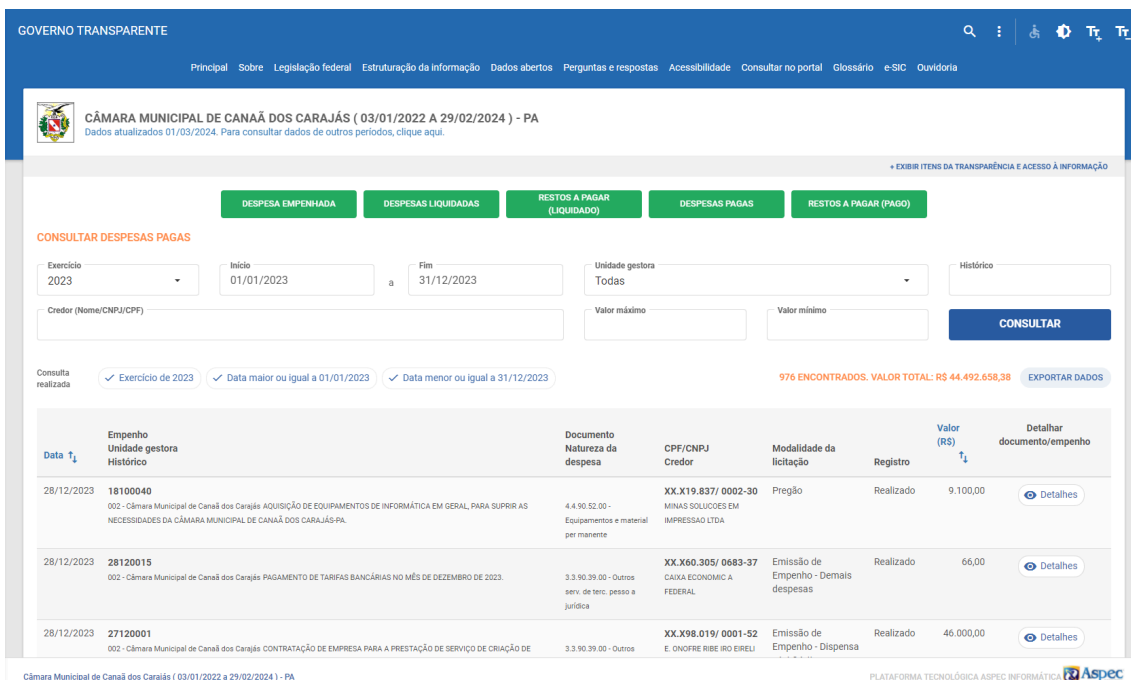


Figura 1. Resultado da consulta de documentos de despesas pagas no portal “Governo Transparente”.

Para consultar os demais dados, o usuário do portal deve acessar os detalhes do documento de pagamento clicando em um botão de acesso aos detalhes do documento, onde então finalmente, é possível acessar os dados solicitados pelo grupo de pesquisa (ver Figura 2).

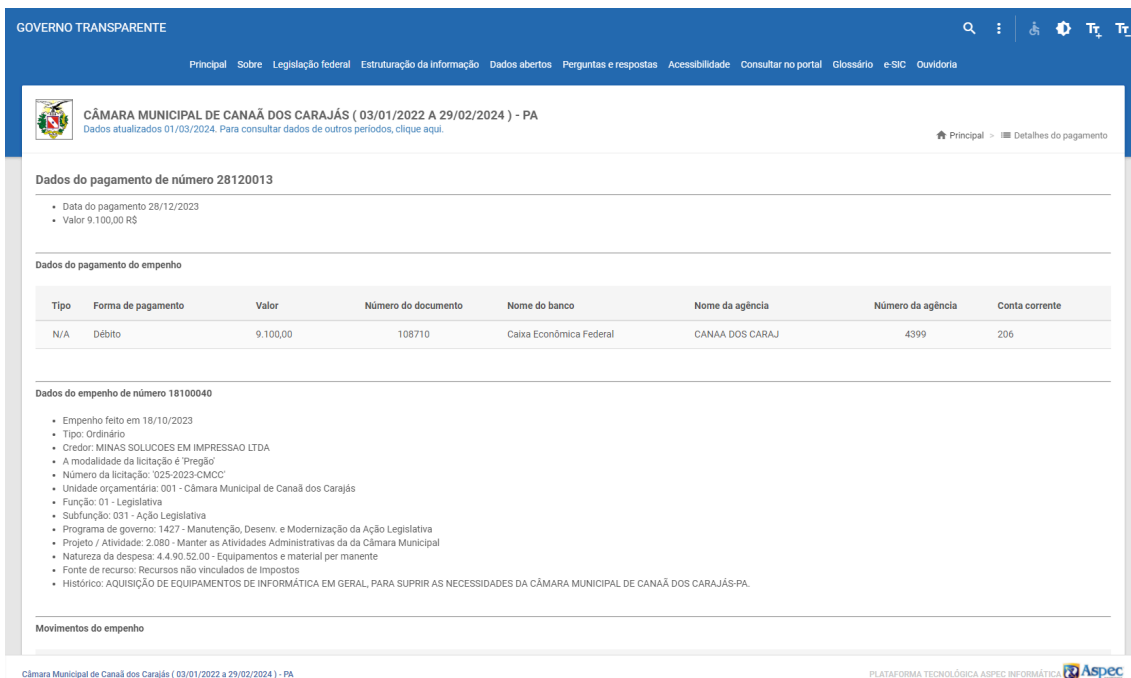


Figura 2. Resultado do detalhamento de um documento de despesa paga no portal “Governo Transparente”.

O portal atende em parte à Lei de Acesso à Informação (LAI) - Lei nº 12.527/2012, implementando parcialmente alguns dos requisitos listados na lei, sendo os mais importantes que foram implementados [BRASIL 2011]: I - conter ferramenta de pesquisa de conteúdo que permita o acesso à informação[...]; II - possibilitar a gravação de relatórios em diversos formatos eletrônicos[...]; IV - divulgar em detalhes os formatos utilizados para estruturação da informação; V - garantir a autenticidade e a integridade das informações disponíveis para acesso; VI - manter atualizadas as informações disponíveis para acesso.

Apesar dos requisitos implementados, o requisito III - possibilitar o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina [BRASIL 2012], não está disponível. Os dados disponíveis para download são incompletos para o grupo de pesquisa, necessitando do requisito mencionado para obter os dados abertos de forma completa.

2.1. Proposta Para Exploração e Download dos Dados

Já que os dados no portal da transparência estão disponibilizados de forma incompleta e necessitam de processo repetitivo humano para a obtenção, faz-se necessário um programa de computador para o download automatizado dos dados abertos.

O método utilizado para explorar e automatizar o download de dados abertos em um portal da transparência é mostrado através de um fluxograma na Figura 3.

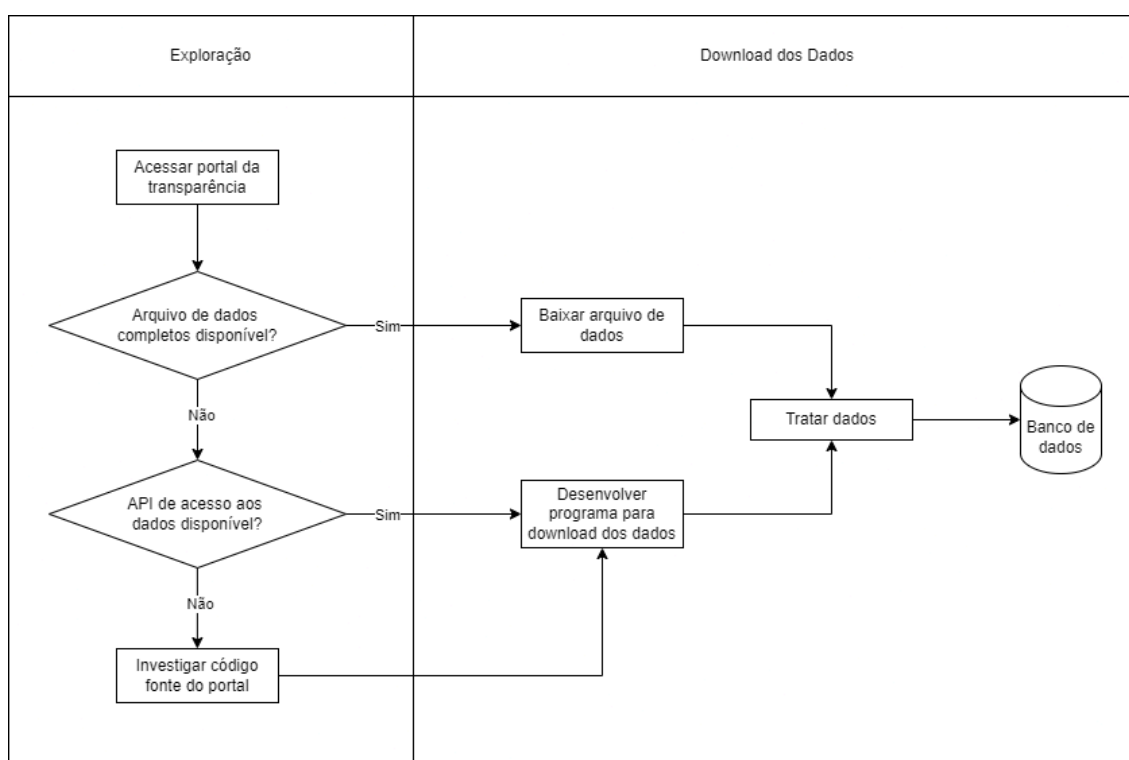


Figura 3. Fluxograma do método para exploração e download de dados abertos

O método proposto para acesso aos dados é dividido em duas partes: exploração e download dos dados. Na primeira parte, o portal da transparência é acessado e, se os dados abertos necessários já estão disponíveis para download em formato aberto e

acessível, partimos para a segunda parte, onde os dados são baixados e tratados, depois armazenados em um banco de dados para uso dos pesquisadores. Se ainda na primeira parte não temos arquivos de dados disponíveis, partimos para a exploração no portal de alguma API (*Application Programming Interface*) disponível abertamente pelos fornecedores do portal. Caso esteja disponível, partimos para o download desses dados através de um programa que baixe, trate e armazene esses dados no banco de dados. Por último, se nenhuma das situações anteriores ocorrerem, partimos para a investigação do código fonte do portal.

Utilizando as ferramentas de desenvolvedor de um navegador *web*, podemos investigar as requisições feitas durante o uso do portal, além do código de marcação HTML das páginas, possibilitando encontrar acessos à APIs que os portais usam para apresentar as páginas de visualização. No caso do portal “Governo Transparente”, encontramos a URL de API para acesso de todos os documentos que quisermos. Como exemplo, temos a URL para consulta de pagamentos: [https://www.governotransparente.com.br/portal/api/v1/json/pagamento\[...\]](https://www.governotransparente.com.br/portal/api/v1/json/pagamento[...]). Essa URL completa (ver Tabela 1) nos permite desenvolver uma aplicação de download dos dados abertos de forma automatizada.

Tabela 1. Esquema da URL de API encontrada

URL da API	
https://www.governotransparente.com.br/portal/api/v1/json/pagamento/{entidade}?{exer}&{tipo}&{inicio}&{fim}	
Parâmetro	Explicação
entidade	O número de identificação da entidade do município no portal "Governo Tansparente"
exer	O exercício da consulta dos dados
tipo	O número de identificação do tipo de despesa no portal "Governo Transparente"
inicio	A data de início da consulta dos dados
fim	A data final da consulta dos dados

Assim, após investigados e encontrados dos demais parâmetros através da ferramenta de desenvolvedor, o download de todos os dados de documentos abertos pela Prefeitura do município de Marabá, PA, durante o ano de 2020, é representada pela URL:

<https://www.governotransparente.com.br/portal/api/v1/json/pagamento/44669490?exer=3&tipo=3&inicio=01/01/2020&fim=31/12/2020>

Figura 4. URL de API completa

Desenvolvemos então um programa que baixa os dados através da API encontrada, tratamos e transformamos os dados baixados e os armazenamos em um banco de dados.

3. Resultado parcial

No momento temos uma aplicação que é executada em ambiente que permite execuções do programa em dados intervalo de tempo, por exemplo, executando aplicação diariamente para baixar, tratar e armazenar dados.

Atualmente os dados estão disponíveis para o grupo de pesquisa “De Olho na CFEM” por meio de arquivos CSV, que podem facilmente serem acessados e lidos pelos membros do grupo. Além do programa de *scraping* para o portal apresentado, os portais de transparência dos outros municípios pesquisados também passarão pelo método de exploração, análise e download dos dados.

O grupo de pesquisa “De Olho na CFEM” já publicou nove notas técnicas (De Olho na CFEM, 2024), abordando a transparência de como a CFEM é utilizada nos municípios pesquisados, produzidas com auxílio dos dados baixados pelo método descrito. Atualmente, o grupo está produzindo a décima nota técnica, que aborda a limitação dos portais de transparência dos municípios. Ainda em andamento, a página de download de dados tratados estará disponível no site do grupo de pesquisa.

Como resultado futuro, esperamos disponibilizar uma aplicação de portal da transparência com dados abertos, através do portal do grupo “De Olho na CFEM”, exemplificando como os portais de transparência poderiam disponibilizar dados abertos ao público.

Essa aplicação define um marco, em que os sistemas de informação dão apoio ao desenvolvimento de políticas públicas, através da facilitação da análise de dados da mineração, que se apresentam de forma dispersa. Por exemplo, ele viabiliza avaliar a maneira como os *royalties* da mineração se associam com questões sociais dos municípios mineradores, como questões relacionadas a gênero, meio ambiente, bioeconomia, comércio, disputas por terra, etc. Embora a metodologia proposta tenha sido aplicada a organização de dados da CFEM, ela pode ser ajustada a quaisquer tipos de informações que se encontrem disponíveis de forma aberta, mas sem acesso massivo automatizado, como por exemplo, portais de transparência de instituições públicas.

Agradecemos o apoio financeiro do grupo “De Olho na CFEM”.

Referências

- Open Data Handbook (2024) “O Que São Dados Abertos?” em Open Data Handbook, https://opendatahandbook.org/guide/pt_BR/what-is-open-data/.
- BRASIL (2011) “LEI No 12.527, DE 18 DE NOVEMBRO DE 2011.” em L12527, https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm.
- Cambridge Dictionary (2024) “Web scraping definition”. <https://dictionary.cambridge.org/us/dictionary/english/web-scraping>.
- Williams, Oscar, and William C. Wheaton (2021) “Identifying Real Estate Development Opportunities: Web-Scraping, Regex Patterns & String-Searching Algorithms.” Identifying Real Estate Development Opportunities: Web-Scraping, Regex Patterns & String-Searching Algorithms, n.d.
- Glez-Peña, Daniel, Anália Lourenço, Hugo López-Fernández, Miguel Reboiro-Jato, and Florentino Fdez-Riverola (2013) “Web Scraping Technologies in an API World.” Briefings in Bioinformatics 15, no. 5: 788–97. <https://doi.org/10.1093/bib/bbt026>.
- De Olho na CFEM (2024) “O Que É: De Olho Na Cfem.” De Olho na CFEM, <https://www.deolhonacfem.org/de-olho-na-cfem/o-que-e>.
- Aspec Informática (2024), Governo Transparente, <https://www.governotransparente.com.br/>
- De Olho na CFEM (2024) “Notas Técnicas | De Olho na CFEM” em De Olho na CFEM, <https://www.deolhonacfem.org/publicacoes/notas-tecnicas>.