

# Um Estudo dos Dados Governamentais Abertos do Estado de Alagoas

Marcos Vinicius Batista Costa<sup>1</sup>, Taígo Ítalo de Moraes Pedrosa<sup>1</sup>, Ênio Andrade Pimenta<sup>2</sup>  
Leonardo Melo de Medeiros (Orientador)<sup>1</sup>

<sup>1</sup>Curso técnico integrado de Informática para Internet - Coordenação de Informática  
Instituto Federal de Alagoas (IFAL) – CEP: 57020-600 – Maceió – Alagoas – Brasil.

<sup>2</sup>4ª Procuradoria de Contas - Ministério Público de Contas do Estado de Alagoas  
CEP: 57055-000 - Maceió – Alagoas – Brasil

{marcosvbcosta10, ifal.taigoitalo}@gmail.com, eniopimenta@hotmail.com  
leonardo.medeiros@ifal.edu.br

**Abstract.** *Access to public information is essential for the sake of transparency on governmental actions. Budgetary transparency is a key factor to better monitor public expenditure and this kind of information improves the knowledge and participation of the Brazilian population [Beluzo and Craveiro 2014]. In Brazil, laws have been defined that require the publication of public expenditure in transparency portals, allowing the emergence of civil applications that facilitate the access to government's budget data. This study addresses the difficulties identified by using a transparency portal to build a data warehouse, such as: the manipulation of open data, the lack of metadata and the semantic loss of the Entity-Relationship model in the data schema.*

**Resumo.** *O acesso à informação pública é essencial para a transparência das ações de governo. A transparência orçamentária é fator chave para um melhor acompanhamento dos gastos públicos e participação efetiva da população brasileira nos mesmos [Beluzo and Craveiro 2014]. No Brasil, já estão definidas leis que obrigam a publicação de dados orçamentários nos portais de transparência pública, permitindo o surgimento de aplicações civis que facilitam o acesso a informação dos dados abertos governamentais. Neste estudo, estão sendo tratadas dificuldades já identificadas: a manipulação dos dados abertos, a falta de metadados e a perda semântica do modelo Entidade-Relacionamento sobre o esquema de dados.*

## 1. Introdução

É essencial a transparência nas despesas públicas e o acompanhamento delas pela população brasileira. As leis que regulamentarizam o acesso aos dados governamentais são: a Lei da Transparência [República 2009], a Lei de Acesso a Informação [República 2011] e a Política de Dados Abertos [República 2016].

No estado de Alagoas, o Portal Alagoas em Dados e Informações [Seplag/AL 2015] é uma iniciativa do Governo de Alagoas de produzir um sistema voltado para a pesquisa e disseminação de dados e informações sobre o estado.

Considerando a falta de ferramentas que integrem os dados governamentais, este estudo está desenvolvendo um ambiente de integração (Data Warehouse) dos dados referentes ao estado de Alagoas e que será disponibilizado ao público para consultas.

Um Data Warehouse é um banco de dados com a função de centralizar dados provenientes de fontes distintas e relacioná-los, fornecendo ao usuário estatísticas que o auxiliarão na tomada de uma decisão.

## 2. Justificativa

Este estudo é o primeiro passo para a criação de um projeto maior de monitoramento dos gastos públicos para o Estado de Alagoas. Com a integração dos dados, será possível, num futuro próximo, utilizá-los para identificar pontos de ineficiência na administração desses recursos e, por fim, atender a urgente demanda do Estado de Alagoas em prover de ferramentas que facilitem o acesso e a eficiência na busca por informações de despesas do governo.

## 3. Metodologia

A análise exploratória de dados (AED) é uma abordagem de análise de um conjunto de dados para identificar características e padrões presentes nesses dados de uma forma visual [Cook et al. 2007]. Neste projeto de pesquisa, está em criação um ambiente para AED que proverá, em um futuro próximo, pesquisas de modelos estatísticos possibilitando formular e testar hipóteses quanto ao uso dos recursos públicos.

Visando atingir o objetivo da pesquisa, estão sendo executadas as seguintes etapas: limpeza e proveniência dos dados, elaboração do banco de dados e acesso ao ambiente de Data Warehouse a público.

## 4. Desenvolvimento

Nesta etapa, foi construído um algoritmo, escrito em Python, que realiza a seleção e o download dos arquivos, a limpeza dos dados e, por fim, o armazenamento deles em um banco de dados MySQL. Além disso, será desenvolvido relatórios para os dados serem visualizados (Figura 1).

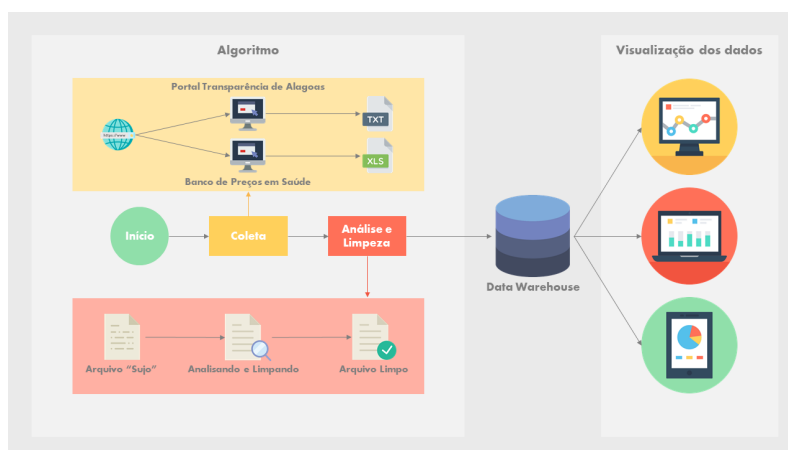


Figura 1. Diagrama de funcionamento do estudo.

#### **4.1. Coleta dos dados**

A fonte de dados referente as despesas governamentais é o Portal da Transparência de Alagoas [SEFAZ/AL 2007], por onde é disponibilizado arquivos anuais ".txt" com dados sobre as despesas do Governo de Alagoas. O Banco de Preços em Saúde (BPS) [CGES 1998] é uma base de preços no mercado de medicamentos e produtos para a saúde. A partir do BPS, serão obtidos arquivos anuais ".xlsx" com dados sobre esses produtos.

Para realizar a coleta desses dados, o algoritmo inicia o browser Chrome, em modo "headless" (sem interface gráfica), navega por uma lista predefinida de páginas e faz a raspagem de links de download, previamente requisitados, durante a análise dos diretórios e arquivos. Depois da coleta dos links, eles são baixados. O algoritmo verifica se todos os arquivos foram baixados completamente e logo após, se tudo tiver ocorrido como o esperado, irá descompactá-los, para que sejam obtidos todos os arquivos ".txt". Os arquivos ".xlsx" não passam pelo processo de descompactação.

#### **4.2. Análise dos dados e modelagem do banco de dados**

Para construir o algoritmo, existiam dois problemas que impediam isso: a falta de clareza nos relacionamentos entre os dados e quais informações são importantes para o estudo. Para resolvê-los, criou-se um algoritmo, também escrito em Python, que analisou todas as despesas e separou seus dados em arquivos distintos, de acordo com o modo que vinham dispostos. Esse algoritmo foi criado exclusivamente para essa etapa do estudo, não tendo utilidade nas seguintes.

Com os dados devidamente separados, analisou-se como cada tipo de dado interagiria com os outros, observando a quantidade de vezes que apareciam juntos (tentativa de identificar relações de acordo com a cardinalidade), características semelhantes como número de dígitos das chaves primárias (os "ids"), entre outras formas. Isso possibilitou a identificação dos relacionamentos entre os dados, permitindo a modelagem do banco de dados.

#### **4.3. Limpeza e armazenamento dos dados**

Durante a análise, pôde-se ter a noção de que nem todos os dados referentes as despesas seriam utilizados, pois não acrescentariam relevância ao serem utilizados nos relatórios.

Sabendo o que é útil ao estudo, a parte do algoritmo responsável pela limpeza dos dados categoriza os arquivos, de acordo com o seu conteúdo. Essa categorização ocorre pois, para cada tipo de arquivos, existe uma quantidade de dados diferentes em posições diferentes a ser extraída. Após essa classificação, o arquivo é aberto, recolhe-se apenas os dados necessários e os insere no banco de dados. Nenhuma despesa é descartada previamente desse processo.

#### **4.4. Visualização dos dados**

Para elaboração dos relatórios de visualização dos dados, está sendo utilizada a ferramenta de análise de negócios Power BI [Microsoft Corporation 2019].

Sendo o objetivo do estudo melhorar o gasto do dinheiro público, os relatórios girarão em torno da comparação entre os preços dos itens comprados pelo Governo de Alagoas e os preços dos itens correspondentes do Banco de Preços em Saúde [CGES 1998] e do fornecimento de informações sobre a nota de empenho que item em análise está associado.

## 5. Resultados

O banco de dados relacional foi estruturado e criado.

O algoritmo que realiza o tratamento os dados já se encontra em fase de testes e aprimoramento. Os resultados obtidos, até o momento, estão dentro da expectativa. Estima-se que até abril de 2019, o algoritmo esteja em total funcionamento.

Os relatórios de visualização dos dados se encontram em desenvolvimento e têm a estimativa de estarem prontos até maio de 2019.

Todos os resultados apresentados foram obtidos durante 7 meses de pesquisa que ainda segue em andamento, com tempo limite para conclusão até agosto de 2019.

## 6. Conclusão

Espera-se que, com a finalização do estudo e a disponibilização da ferramenta de visualização dos dados na web, esta iniciativa contribua para investigações relacionadas a fraudes e superfaturamentos em compras do Governo de Alagoas para a saúde e também com a melhora na transparência do dados governamentais.

## Referências

- Beluzo, J. R. and Craveiro, G. d. S. (2014). Criação de um repositório integrado de dados da execução orçamentária brasileira proveniente de diferentes fontes de dados em um modelo dimensional Data Warehouse. *Anais*.
- CGES (1998). Banco de Preços em Saúde. <http://bps.saude.gov.br>. Accessed: 2019-03-17.
- Cook, D., Swayne, D. F., and Buja, A. (2007). *Interactive and dynamic graphics for data analysis: with R and GGobi*. Springer Science & Business Media.
- Microsoft Corporation (2019). Power BI Desktop (version 2.67.5404.581). <https://powerbi.microsoft.com>.
- República, P. (2009). Lei Complementar Nº 131, de 27 de maio de 2009 que estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências. *Diário Oficial da União*. Seção 1, p.2.
- República, P. (2011). Lei Nº 12.527, de 18 de novembro de 2011 que regula o acesso a informações e dá outras providências. *Diário Oficial da União*. Seção 1, p.1.
- República, P. (2016). Decreto Nº 8.777, de 11 de maio de 2016 que institui a Política de Dados Abertos do Poder Executivo federal. *Diário Oficial da União*. Seção 1, p.21.
- SEFAZ/AL (2007). Portal da Transparência de Alagoas. <http://transparencia.al.gov.br>. Accessed: 2019-03-17.
- Seplag/AL (2015). Alagoas em Dados e Informações. <http://dados.al.gov.br>. Accessed: 2019-03-17.