IGOV: UM SISTEMA DE INTEGRAÇÃO DE DADOS GOVERNAMENTAIS

RESUMO

Nos últimos anos, as instituições públicas de boa parte dos governos democráticos têm se empenhado em oferecer mecanismos que favoreçam a política de dados governamentais abertos. No caso do Brasil, instrumentos como a Lei de Acesso à Informação e os portais da Transparência e de Dados Abertos representam algumas iniciativas que trilham esse caminho. Todavia, o fato de apenas liberar o acesso a dados governamentais não é suficiente para que os mesmos possam ser explorados com todo potencial sob o ponto de vista da pesquisa. Precisam-se desenvolver soluções que promovam a integração de dados vindos de diferentes fontes em níveis de heterogeneidade distintos (formato, acesso, etc.). Este trabalho discute uma solução para o processo de integração de dados oriundos da saúde pública brasileira e do programa de transferência de renda Bolsa Família.

PALAVRAS-CHAVES: Integração de Dados; Análise de Dados; Dados Abertos.

IGOV: AN INTEGRATION SYSTEM OF GOVERNMENT DATA

ABSTRACT

In recent years, most of democratic governments have engaged in providing mechanisms to develop open government data policies. Brazilian initiatives such as the Access to Information Federal Law and the creation of web portals in order to publish government data represent some strategies towards Government 2.0. However, the access to government data is not enough to explore the full research potential in the open government area. Current solutions seek to promote data integration coming from distinct sources in different levels of heterogeneity (format, access, etc.). This paper discusses one solution to the data integration process of the Brazilian public health data and the cash transfer program Bolsa Família.

KEYWORDS: Data Integration; Data Analysis; Open Data.

*Revista Brasileira de
Administração Científica,
Aquidabã, v.5, n.2, Out 2014.*

ISSN 2179-684X

SECTION: Articles

TOPIC: *Sistemas e Tecnologia da
Informação*

*Anais do Simpósio Brasileiro de
Tecnologia da Informação (SBTI 2014)*

DOI: 10.6008/SPC2179-684X.2014.002.0001

João Paulo Clarindo dos Santos

Universidade Federal de Alagoas, Brasil
<http://lattes.cnpq.br/6416924936563691>
jpaulocpm@gmail.com

Fábio José Coutinho da Silva

Universidade Federal de Alagoas, Brasil
<http://lattes.cnpq.br/6352083877939941>
fabiojcsilva@gmail.com

Received: 11/08/2014

Approved: 15/10/2014

Reviewed anonymously in the process of blind peer.

Referencing this:

SANTOS, J. P. C.; SILVA, F. J. C.. IGOV: um sistema de integração de dados governamentais. *Revista Brasileira de Administração Científica*, Aquidabã, v.5, n.2, p.8-16, 2014. DOI: <http://dx.doi.org/10.6008/SPC2179-684X.2014.002.0001>

INTRODUÇÃO

A desnutrição ainda é um grande problema mundial, segundo dados da Unicef cerca de 3 milhões de crianças morrem todo ano devido a uma alimentação desprovida dos nutrientes necessários (UNICEF, 2013). Esse número corresponde à causa de aproximadamente metade das mortes de crianças abaixo de 5 anos. No Brasil, a realidade não é diferente, apesar da taxa de mortalidade infantil estar caindo desde 1990, atualmente, ainda tem um número superior ao de países com semelhantes índices de desenvolvimento econômico. Dados da ONU mostram o Brasil atrás de México, Peru, Venezuela e Argentina (ECLAC, 2013). Vale ressaltar ainda que segundo a Lista Brasileira de Evitabilidade (LBE) ampliada, cerca de 70% dos óbitos infantis em 2010 poderiam ser evitados por ações do setor Saúde e de outros (MALTA et al., 2010).

Dentre as iniciativas do Governo Federal que visam tratar a questão da desnutrição, o programa social denominado Bolsa Família (BF) é apontado como o principal mecanismo de transformação da realidade atual. O programa tem como objetivo apoiar às famílias mais pobres e garantir a elas o direito à alimentação e o acesso à educação e à saúde, sendo citado por alguns analistas como um dos responsáveis pela redução do índice de miséria no Brasil.

Todavia, apesar de decorridos dez anos desde sua criação, o programa BF ainda apresenta diversos problemas que, afora a discussão política, põem em dúvida sua capacidade de fiscalização e controle. Neste aspecto, o programa tem buscado um modelo de gestão descentralizada, envolvendo os entes da federação (união, estados e municípios). Entretanto, dados recentes da CGU mostram que os beneficiários do BF ainda carecem de um melhor acompanhamento. Um exemplo disso é a informação do relatório da CGU (CGU, 2012) que mostrou que apenas 38,9% (amostra obtida a partir de levantamento feito em 401 municípios visitados) das Instâncias de Controle Social (ICS) realizavam visitas às escolas e aos postos de saúde para acompanhamento das condicionalidades.

As condicionalidades representam os compromissos assumidos tanto pelas famílias beneficiárias do BF quanto pelo poder público para ampliar o acesso dessas famílias a seus direitos sociais básicos. Por um lado, as famílias devem assumir e cumprir esses compromissos para continuar recebendo o benefício. Por outro, as condicionalidades responsabilizam o poder público pela oferta dos serviços públicos de saúde, educação e assistência social (MDS, 2010). Os compromissos das condicionalidades em saúde são o acompanhamento do calendário vacinal, do crescimento e desenvolvimento das crianças menores de sete anos, e o acompanhamento de gestantes e nutrízes.

O sistema IGOV propõe uma alternativa para a integração de dados governamentais. Neste contexto, foram consideradas duas fontes de dados que mantêm informações acerca dos óbitos ocorridos em municípios de Alagoas e dos beneficiários do programa Bolsa Família. O objetivo da proposta de integração é promover a análise de dados sobre a mortalidade de crianças, buscando casos em que a causa da morte esteja entre as causas evitáveis por

intervenções do governo (MALTA et al., 2010) a fim de verificar se eram pessoas cobertas por programas do governo. Neste trabalho, programamos a análise de casos de crianças que morreram devido a deficiências nutricionais para descobrir se faziam ou não parte de famílias cobertas pelo programa Bolsa Família.

Os detalhes da proposta de integração encontram-se organizados neste artigo da seguinte maneira. A Seção 2 detalha os dados que vão ser utilizados na integração: Dados do Sistema de Informações sobre Mortalidade (SIM) e dados abertos de beneficiários da Bolsa Família. Já a Seção 3 detalha o funcionamento do sistema IGOV, assim como sua implementação. A Seção 4 discute os resultados obtidos através do sistema. Finalmente, a Seção 5 traz as considerações finais e perspectivas futuras do trabalho.

METODOLOGIA

Esta seção descreve os conjuntos de dados utilizados neste trabalho. Inicialmente, foi feito um levantamento sobre os dados governamentais abertos relacionados aos temas discutidos na seção anterior. Em seguida, os dados obtidos foram manipulados a fim de fornecer uma visão integrada através do sistema IGOV.

O ativista de dados abertos David Eaves estabelece 3 leis para a classificação de *dados governamentais abertos* (EAVES, 2009): (i) se o dado não pode ser encontrado e indexado na Web, ele não existe; (ii) se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado; e (iii) se algum dispositivo legal não permitir sua replicação, ele não é útil.

Apesar das recentes iniciativas favoráveis ao modelo de dados abertos no Brasil tal como a Lei de Acesso a Informação Pública (Lei 12.527/2011), a obtenção de dados dos serviços públicos ainda é um processo difícil. A propriedade descrita em (ii), por exemplo, não é atendida pelo Portal da Transparência no que se refere à disponibilização dos dados de beneficiários da Bolsa Família em seu período de vigência, conforme será discutido posteriormente. As seções seguintes discutem o processo de obtenção dos dados referentes ao SIM e aos dados do BF.

DISCUSSÃO TEÓRICA

Dados SIM

O Sistema de Informações sobre Mortalidade (SIM) é uma das mais importantes fontes de informação para o planejamento e avaliação das ações de saúde (LAURENTI et al., 2006). A partir de seu uso é possível construir o perfil da mortalidade de um município, região ou estado e construir indicadores que permitam acompanhar melhor a situação de saúde da população, comparando a evolução dos mesmos ao longo do tempo e identificando áreas prioritárias.

Os dados que abastecem a base SIM são obtidos através do preenchimento da ficha de Declaração de Óbito, a qual é padronizada pelo Ministério da Saúde. A Declaração de Óbito contém informações básicas acerca do paciente tais como nome, idade, nome dos pais, endereço, assim como a causa do óbito e dados do médico responsável.

Neste trabalho, foram objeto de análise os dados do SIM que abrangem a área da grande Maceió, incluindo algumas outras cidades do Estado de Alagoas. Tais dados foram obtidos através da cooperação entre a Secretaria da Saúde do Estado de Alagoas, o grupo GSD (Grupo de Pesquisa em Sistemas Distribuídos) e o grupo LaTIM (Laboratório de Telemedicina e Informática Médica). Ambos os grupos são formados por bolsistas PIBITI, PIBIC e alunos dos programas de mestrado do IC/UFAL.

Os dados obtidos, sobre a mortalidade dos alagoanos, compreendem o período desde o ano de 2007 até 2012 e nos foram disponibilizados em formato JSON, o que permite sua manipulação e processamento através de ferramentas como o mongoDB e Hadoop. É importante destacar que a base de óbitos é manipulada respeitando as questões éticas e legais que garantem o sigilo e a privacidade dos pacientes. A seguir, são apresentados os dados SIM analisados pelo sistema IGOV.

O Quadro 1 descreve o conjunto de dados do SIM utilizados para o casamento com as informações sobre os beneficiários do Bolsa Família, conforme será discutido na Seção 3. Os dados SIM estão de acordo com o Manual de Instruções para o Preenchimento da Declaração de Óbito oferecido pela Fundação Nacional de Saúde (FUNASA).

Quadro 1: Campos utilizados para o casamento de informações.

Campo	Descrição do Campo
Campo 12	Nome do Pai do Falecido
Campo 13	Nome da Mãe do Falecido
Campo 15	Idade do Falecido
Campo 24	Município de Residência
Campo 25	Unidade Federativa
Campo 49/II	Causa da Morte – Parte II

O campo 49 (Causa da Morte) corresponde à causa principal do óbito (FUNASA, 2001), ressaltando que a manipulação desses dados no sistema IGOV é para fins meramente estatísticos, identificando apenas a cidade na qual ocorreram os óbitos, bem como o número de óbitos naquela cidade.

Dados dos Beneficiários da Bolsa Família

Os dados de beneficiários do programa Bolsa Família (BF) são disponibilizados publicamente através do Portal da Transparência do Governo Federal. Esses dados são referentes aos pagamentos realizados mensalmente aos mais de 14 milhões de famílias beneficiárias do programa social.

Tais dados estão disponíveis no portal para consulta online a partir do ano de lançamento do programa (2004), sendo exibidos numa interface gráfica que limita em quinze, o número de registros na tela. Outra forma de disponibilização dos dados do BF presente no portal é através da área de downloads que mantém arquivos em formato. CSV. Entretanto, tais dados estão somente disponíveis a partir do ano de 2011 (CGU, 2014). O IGOV não possui um extrator que permita recuperar os dados disponíveis online no Portal da Transparência, logo, foi utilizado apenas os arquivos. CSV disponibilizados para download. No Quadro 2, são listados os dados que serão usados para o casamento de informações pelo sistema IGOV.

Quadro 2: Campos utilizados para o casamento de informações

Campo	Descrição do Campo
UF	Unidade Federativa
Código_SIAFI_Municipio	Código do Município de Residência do Beneficiado
Nome_Favorecido	Nome do Representante da Família Beneficiada pelo programa

A Plataforma IGOV

O sistema IGOV (Integração de Dados **G**overnamentais) foi desenvolvido utilizando o software Pentaho Data Integration, também conhecido como Kettle (PENTAHO, 2011). O Pentaho Kettle disponibiliza um ambiente ETL (*Extract, Transform and Load*) que faz parte do pacote de programas de código aberto denominado Pentaho Business Analytics.

Figura 1: Visão geral do Sistema IGOV.

Através do uso dessas ferramentas é possível fornecer uma visão integrada de diferentes fontes de dados governamentais. A Figura 1 representa uma visão geral do Sistema IGOV e a seção seguinte descreve o processo de integração em detalhes, considerando as fontes mencionadas anteriormente.

Processo de Integração

A Figura 2 descreve os passos envolvidos no processo de integração desenvolvido a partir da plataforma Pentaho Kettle. Os quadros coloridos ilustram as etapas desse processo, as quais são detalhadas a seguir.

Figura 2: Passo-a-passo do processo de integração.

Passo 1a (em azul): O arquivo .CSV dos beneficiários do Bolsa Família é carregado, reduzindo para apenas beneficiários do estado de Alagoas, com apenas os dados do nome do beneficiário e a cidade na qual reside (que está descrita como código SIAFI). Em seguida, inicia-se a Passo 2.

Quadro 3: Doenças relacionadas à desnutrição.

CID-10	Descrição da Doença
E40	Kwashiorkor – Uma forma de desnutrição que ocorre quando há pouca proteína na alimentação.
E41	Marasmo Nutricional – Uma forma de desnutrição que ocorre quando há uma baixa distribuição proteico-calórica equilibrada.
E42	Kwashiorkor Marasmático – Uma forma mista dos casos E40 e E41, com uma baixa distribuição proteico-calórica, porém desequilibrada.
E43	Desnutrição Proteico-calórica Grave Não Especificada.
E44	Desnutrição Proteico-calórica de Graus Moderado e Leve.
E45	Atraso do Desenvolvimento Devido à Desnutrição Proteico-calórica.
E46	Desnutrição Proteico-calórica Não Especificada.

Passo 1b (em turquesa): A base de óbitos SIM é carregada diretamente do mongoDB. Em seguida, é feita a redução para apenas pessoas menores de 18 anos que tenham falecido devido a doenças relacionadas com a desnutrição. As doenças ocasionadas pela não ingestão dos nutrientes básicos consideradas neste trabalho são listadas juntamente com seu respectivo código CID-10 no Quadro 3.

Além destes códigos, também foram analisados os códigos de doenças descritas como anemias nutricionais, as quais possuem CID-10 que variam de D50* a D53*. Sendo assim, após a seleção na base de óbitos para as causas de óbito citadas anteriormente, os dados obtidos são levados ao Passo 3. Note-se que os passos 1a e 1b são executados paralelamente, devido à configuração implementada no Pentaho Kettle.

Passo 2 (em vermelho): Nos dados manipulados neste projeto, cada fonte de dados possui um padrão de identificação distinto para a representação da cidade. Na fonte SIM, as cidades são representadas através do padrão de 6 dígitos desenvolvido pelo IBGE (FUNASA, 2001). Atualmente, os códigos costumam ser disponibilizados em 7 dígitos, sendo o último como dígito verificador. Tabelas de conversão desses códigos são disponibilizadas abertamente em diversos setores. Entretanto, na fonte dos beneficiários da Bolsa Família, as cidades são representadas através do padrão SIAFI (Sistema Integrado de Administração Financeira) (CGU, 2014), o qual utiliza apenas 4 dígitos em sua representação. Para ser realizada esta conversão, foi necessária a criação de um sistema de mapeamento que gera uma nova base de dados a partir de dados que relacionem os códigos IBGE e SIAFI. A Figura 3 representa a transformação dos padrões de códigos dos municípios brasileiros.

Tal sistema foi desenvolvido em Python, e analisa cidade por cidade, e criando uma nova base de dados assimilando os códigos IBGE de seis e sete dígitos e SIAFI. Tal base de dados é exportada em um novo arquivo CSV, que por fim é utilizado no Pentaho Kettle para converter os códigos de cidades dos dados obtidos no Passo 1a. Por fim, os dados convertidos são levados ao Passo 4.

Figura 3: Detalhes dos procedimentos realizados no Passo 2.

Passo 3 (em verde): É feito o casamento dos dados dos passos 1b e 2. Os dados são extraídos apenas caso o nome da mãe ou do pai, assim como a cidade na qual residem casarem com os dados dos beneficiários. Essas informações são levadas ao Passo 4.

Passo 4 (em laranja): Os dados são refinados, e exportados um novo arquivo .csv, com a cidade, data e causa do óbito (em código CID-10). Os nomes dos pais do falecido são omitidos. Os dados finais gerados são apenas para fins estatísticos, e podem ser manipulados em programas de edição de planilhas, como o Microsoft Excel e o LibreOffice Calc.

RESULTADOS

Esta seção descreve os resultados obtidos neste trabalho. A pesquisa verificou que ocorreram 16 óbitos de crianças devido à desnutrição no período de 2011 a 2012, sendo que 13 desses óbitos foram de crianças cujos pais não eram beneficiários do programa Bolsa Família durante o período em questão. Tal situação foi verificada nos seguintes municípios: São Luiz do Quitunde, Maceió, Maragogi, Coruripe, Pindoba, Cacimbinhas, Delmiro Gouveia, Cajueiro, Novo Lino e Santana do Mundaú.

Na Figura 4 é exibido um mapa do estado de Alagoas cujos municípios destacados em vermelho ocorreram os óbitos. O mapa evidencia a incidência de casos tanto em municípios da Zona da Mata Alagoana, assim como Maceió, que segundo dados oferecidos pelo Programa das Nações Unidas para o Desenvolvimento (PNUD), Maceió possui o maior índice Maceió possui o maior Índice de Desenvolvimento Humano (IDH) relacionado a renda per capita, com valor de 0,739, índice considerado acima da média (PNUD, 2013).

Figura 4: Óbitos por desnutrição de crianças não beneficiadas pela Bolsa Família.

CONCLUSÕES

Os resultados obtidos a partir do sistema IGOV demonstram um número de óbitos elevado na capital, revelando que mesmo em centros economicamente mais desenvolvidos, ainda existem problemas na distribuição dos benefícios da Bolsa Família. Devido à ampla cobertura do programa sua gestão tem se apresentado como um desafio complexo (JESUS, 2011).

Embora o programa seja promovido pelo Ministério do Desenvolvimento Social (MDS), ficam a cargo dos municípios as ações de 'identificação, cadastramento e atualização de dados das famílias'. (MDS, 2012). Diversos motivos apresentados pelo MDS podem acarretar no cancelamento do programa para determinada família tais como: recusa em prestar informações; comprovação de omissão de dados; e declaração de afirmações falsas. Além disso, algumas famílias necessitadas podem não ser incluídas no programa devido à possível inoperância dos agentes municipais, seja por motivos logísticos ou de ordem geográfica nos casos de comunidades em áreas longínquas.

O IGOV utiliza informações obtidas através do Portal da Transparência, as quais são limitadas e restritas aos pagamentos mensais, fornecendo assim uma visão momentânea da situação dos beneficiários. Tal fato pode, por exemplo, desconsiderar situações aonde uma criança veio a óbito por desnutrição cuja família teve seu pagamento cancelado pelos motivos citados anteriormente pouco antes do óbito.

Para contornar este problema, foi feita uma solicitação de dados do Cadastro Único junto ao MDS, a qual se encontra em fase de análise. A obtenção desses dados irá permitir uma análise mais detalhada das famílias, promovendo maior precisão no processo de *matching* a partir do uso de informações tais como os nomes de todos os componentes da família e o detalhamento do endereço. A disponibilização dos dados do Cadastro Único permitirá em trabalhos futuros indicar não apenas os casos de óbitos não cobertos pelo programa, mas também será possível averiguar óbitos de crianças pertencentes a famílias beneficiárias do programa.

Nossos agradecimentos ao Núcleo de Inovação Tecnológica (NIT) da Universidade Federal de Alagoas (UFAL) por patrocinar este trabalho a partir do Programa Institucional de Bolsas de Iniciação em Desenvolvimento Tecnológico e Inovação (PIBITI); aos professores André Lage e Marcelo Oliveira e ao aluno Jorge Matias pelo apoio e colaboração para a obtenção de dados; finalmente, a Dra. Andréa Coutinho pelas informações a respeito das questões relacionadas às enfermidades nutricionais.

REFERÊNCIAS

- CGU. **Relatório de Avaliação da Execução de Programas de Governo Nº 7**. Brasília: 2012.
- CGU. **Acesso a dados a partir do Portal da Transparência**. Brasília: 2014.
- EAVES D.. **The Three Laws of Open Government Data**. Ottawa: 2009.
- ECLAC. **Statistical Yearbook for Latin America and the Caribbean**. Santiago: United Nations Publication, 2013.
- FUNASA. **Manual de Instruções para o Preenchimento da Declaração de Óbito**. 3 ed. Brasília: 2001.
- JESUS, A. C. S.. Desafios e diversidades na gestão do Programa Bolsa Família em pequenos e médios municípios. In: JORNADA INTERNACIONAL DE POLÍTICAS PÚBLICAS - ESTADO, DESENVOLVIMENTO E CRISE DO CAPITAL, 5. **Anais**. São Luís: Universidade Federal do Maranhão, 2011.
- LAURENTI R.; MELLO JORGE M. H. P.; GOTLIEB S. L. D.. **O Sistema da informação sobre mortalidade: passado, presente e futuro**. São Paulo: CBCD; 2006.
- MALTA D. C.; SARDINHA L.; MOURA L.; LANSKY S.; LEAL M. C.; SZWARTWALD C. L.; FRANÇA E.; ALMEIDA M. F.; DUARTE E. C.. **Atualização da lista de causas evitáveis por intervenções do Sistema Único de Saúde**. Brasília, v.19, n.2, p.173-176, 2010.
- MDS. **Guia para Acompanhamento das Condicionalidades do Programa Bolsa Família**. Brasília: Secretaria Nacional de Renda de Cidadania (Senarc), 2010.
- MDS. **Manual de Gestão do Cadastro Único para Programas Sociais do Governo Federal**. Brasília: Secretaria Nacional de Renda de Cidadania (Senarc), 2012.
- PENTAHO. **Pentaho Data Integration (aka Kettle) Documentation**. Orlando: 2011.