

1 Introdução

Os avanços na tecnologia da informação, bem como o crescimento da sociedade da informação através do uso da Internet, obrigaram os governos de inúmeros países, em seus mais variados níveis, a reformularem suas formas de atuação. O uso de novas tecnologias, em especial a Internet, oferece novas formas de interação entre governos, cidadãos e empresas que se manifestam em serviço e acesso aos dados governamentais antes inimagináveis. É neste contexto que surge o conceito “Governo Eletrônico”¹ ou e-Governo.

Um dos projetos do e-Governo é partilhar os dados governamentais a fim de oferecer maior abertura e transparência aos cidadãos, bem como permitir a melhoria das operações internas e entre governos (W3C-Brasil, 2009). Os dados governamentais são produzidos no decorrer das atividades do governo e através de seus órgãos de pesquisa, resultando em um imenso volume de dados.

Estatísticas públicas são dados sociais, demográficos e econômicos coletados, compilados e disponibilizados regularmente pelas agências ligadas ao planejamento governamental e outras instituições públicas para a sociedade civil, governo e empresas (Jannuzi, 2006).

Além de subsidiar o planejamento público, nas três últimas décadas, as estatísticas se tornaram objeto de estudo, atraindo a atenção de pesquisadores das mais diversas formações: demografia, economia, história, filosofia, saúde coletiva, sociologia, antropologia, ciência política, ciência da informação, entre outras porque a sua produção envolve valores científicos e tecnológicos, além de

¹ Governo eletrônico é um termo abrangente, adotado por diferentes autores para tratar dos diversos aspectos ligados à adoção de tecnologia no setor público, com o objetivo de aumentar o desempenho e eficiência da administração e dos serviços públicos. Existem também outros termos com o mesmo significado, como: electronic government, e-gov, e-government, wired-government, governo eletrônico, governo digital, governo virtual, estado virtual, e governança eletrônica.

perspectivas e abordagens típicas de campos de investigação (Senra & Camargo, 2010).

Com o intuito de compartilhar as estatísticas públicas, muitos órgãos de pesquisas dos governos aderiram ao movimento “Open Data” disponibilizando as estatísticas públicas na web em padrões abertos e estruturados -“Open Government data” (W3C-Brasil, 2009). Esta adoção permitiu que as estatísticas públicas fossem exploradas por outros pesquisadores ou entidades de pesquisas que detinham a tecnologia para transformar o conjunto de dados estatísticos em informação estatística.

Um avanço dos governos na tentativa de compartilhar os dados públicos, inclusive as estatísticas públicas, foi o desenvolvimento de aplicações Web conhecidas como *mashup* (Merrill, 2006). Estas aplicações permitem que, em um único ponto, o usuário final possa acessar outros aplicativos ou fontes de dados diferentes, agregando os dados coletados e tendo como resultado um novo conjunto de dados para análise. Embora, estas aplicações ofereçam ao usuário final um excelente caminho para encontrar e misturar os dados das diversas fontes, o resultado delas ainda é um novo conjunto de dados em formato bruto.

Pode-se citar como um exemplo marcante, lançado pelo Governo dos Estados Unidos, o site *data.org*, em 2009. Este site utiliza técnicas de *mashup* que combinam conjuntos de dados gerados e disponibilizados por diversas instituições de pesquisa do governo em formatos abertos e estruturados (UN-Survey, 2010). O Reino Unido, a Austrália e a Nova Zelândia, em 2010, também lançaram sites governamentais para servirem como o único ponto de acesso aos dados públicos. O Estado de São Paulo no Brasil também teve a iniciativa de disponibilizar cópias das bases de dados públicas, em caráter aberto e gratuito pela internet (Decreto, 2010).

Diante de todos estes esforços, o que se observa, porém, é que quem está oferecendo acesso às informações governamentais esta sendo o terceiro setor (ONGs², empresas, institutos de pesquisa), pois este possui a capacidade técnica para transformar os dados governamentais em informação e divulgá-las em seus sites. Conforme mostra, na Figura 1.1, a recente pesquisa realizada

² Organizações Não Governamentais

pela Organização das Nações Unidas (UN-Survey, 2010), somente 23,4% dos sites dos governos oferecem interatividade no acesso às informações governamentais e 57,1 % dos sites não governamentais apresentavam a mesma interatividade.

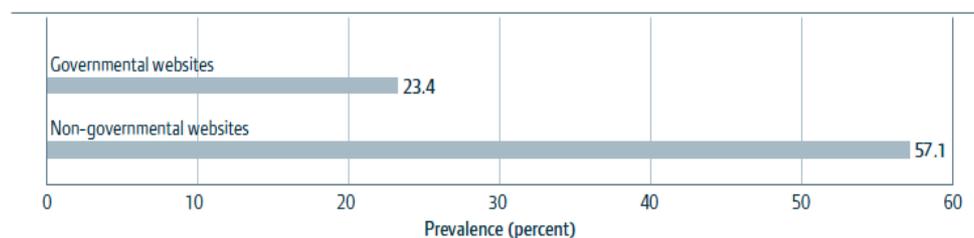


Figura 1 - Websites com interatividade para acesso às informações governamentais, outubro 2009 (UN-Survey, 2010)

A falta de um serviço público que disponibilize as informações estatísticas na Web cria uma dependência do terceiro setor para a obtenção das informações governamentais, demandando, ao usuário final, mais tempo e custo para se obter a informação desejada. Além disso, a disponibilização dos dados governamentais brutos é uma preocupação dos governos. Isso, porque, além de permitir a transferência dos dados para fora do órgão governamental, os dados acabam sendo modificados e manipulados para qualquer finalidade por terceiros. Dessa forma, a confiança do usuário final pode ser abalada pela incerteza do caráter original das informações visualizadas (W3C-Brasil, 2009).

Diante do exposto, percebe-se que o verdadeiro potencial do governo eletrônico, quanto ao compartilhamento de dados públicos, está em oferecer facilidades de acesso às informações governamentais para a sociedade (UN-Survey, 2010). Isso é viável de ser implementado na prática, porque, atualmente existe tecnologias avançadas e o conceito conhecido como Inteligência de Negócio (ou BI - Business Intelligence). Trata-se de um conjunto de processos que permitem, a partir dos dados da empresa, adquirir e apresentar informações úteis, orientadas a negócios e a resultados(Biere, 2003).

As ferramentas de BI têm a função de disponibilizar aos usuários uma interface simples, com uma variedade de relatórios e aplicações analíticas. Dentre estas, destaca-se a tecnologia OLAP -On-line Analytical Processing-(Codd, 1993) que é utilizada para investigar, analisar e apresentar dados agregados por diversas combinações em uma interface unificada. Para que as

ferramentas de BI tenham todo o apoio computacional possível, uma complexa infra-estrutura tecnológica é necessária, cujo principal componente é o Data Warehouse.

Um Data Warehouse (DW) é um banco de dados centralizado, orientado por assunto, integrado, não volátil e histórico, criado com o objetivo de dar apoio ao processo de tomada de decisão e que estrutura os dados em uma arquitetura analítica bastante distinta da arquitetura relacional utilizada nos bancos de dados transacionais (Inmon, 2000).

Construir um DW oferece a oportunidade de integrar diversas fontes de dados em uma única plataforma, de forma sincronizada e padronizada, permitindo assim uma maior qualidade dos dados. O processo de construção, acesso e manutenção de um DW é denominado de Data Warehousing (DWing).

Utilizar um DW como fonte para as ferramentas de BI é uma boa estratégia, pois este banco tem uma estrutura eficiente no armazenamento de grandes volumes de dados, oferece uma visão única do negócio e é fisicamente separado do ambiente transacional, evitando prejuízo no desempenho dos servidores transacionais. Além disso, o DW tem características estruturais que buscam maximizar o desempenho analítico e utiliza um modelo de dados adequado a este fim (Kimball, 1996).

Historicamente, o uso de ferramentas de BI estava restrito a grandes organizações, devido ao alto custo do desenvolvimento dos modelos e das ferramentas de software necessárias. Atualmente, com o desenvolvimento das tecnologias Web, que permitem maior dinâmica e interação com o usuário, a maioria das empresas de software que desenvolviam ferramentas de BI integraram em suas suites uma plataforma Web para que seus clientes acessem seus dados através de um Web browser.

Muitas são as ferramentas de BI disponíveis na Web, em especial, destacam-se os softwares que são desenvolvidos para oferecerem suas funcionalidades em forma de serviços, permitindo que o usuário final use e acesse estas aplicações sob a demanda de suas necessidades. O modelo de comercialização destas aplicações é denominado Software como um Serviço (*SaaS-Software as a Service*) e permite que os usuários deixem de comprar

licenças e passem a ser assinantes de softwares, que são acessados pela Internet, não sendo necessário instalar nenhum sistema no computador do usuário. Um dos benefícios deste modelo é o custo reduzido.

Segundo Rudin (1996), em um alto nível, o objetivo do data warehousing e o objetivo da Web são o mesmo: facilitar o acesso aos dados. Desta forma todos esses conceitos e tecnologias existentes podem ser utilizados para a elaboração de uma sistemática apoiada em Data Warehousing para apoio ao governo eletrônico.

1.1. Objetivos

O objetivo principal da pesquisa é propor uma sistemática baseada em Data Warehousing para gerar informações analíticas de apoio ao e-Governo, possibilitando a disseminação de informações estatísticas na Web. Com essa sistemática busca-se organizar as etapas necessárias à transformação de dados públicos, visando oferecer um serviço de acesso às informações analíticas para a sociedade na Web e sem custo.

Do objetivo principal é possível desdobrar objetivos específicos, a saber:

- Identificação dos principais conceitos utilizados na produção das estatísticas públicas e das possíveis formas de análise;
- Identificação das diretrizes e padrões estabelecidos pela W3C no Brasil, bem como seus principais interesses nas aplicações do governo eletrônico;
- Estudo e definição das principais etapas do processo de Data Warehousing a ser concebido;
- Estudo e identificação das principais tecnologias e componentes necessários para construir o protótipo no estudo de caso.

1.2. Delimitações da Pesquisa

Alguns pontos que constituem delimitações do trabalho, a saber:

- Não se pretende estudar as metodologias de produção das estatísticas, nem tão pouco as instituições que detêm a sua produção. Isso por que as estatísticas públicas são geradas a partir

de metodologias próprias de pesquisas, com suas limitações e potencialidades (Senra & Camargo, 2010).

- Não faz parte do escopo desse trabalho discutir as questões relacionadas à política de acesso aos dados públicos. Atualmente, no Brasil, inexistente um protocolo legal, como o *Freedom of Information Act* (FOIA) americano, que estabeleça as regras atinentes aos mecanismos para a garantia do acesso à informação pública (Lopes, 2007).
- Não se pretende realizar um estudo comparativo entre as diversas ferramentas e e tecnologias passíveis de aplicação no processo de desenvolvimento de um Data Warehouse. Isso por que a sistemática busca organizar as etapas necessárias à transformação das estatísticas públicas.

1.3. Motivação

Percebe-se que o uso das estatísticas públicas vai muito além dos interesses de quem as produz e que, apesar dos inúmeros esforços, ainda há uma dificuldade quanto ao seu uso. Assim, torna-se motivador empregar esforços para facilitar o uso para análise exploratória das informações estatísticas, visando responder às questões dos usuários finais, sem custo na Web.

O projeto desenvolvido pelo Departamento de Informática da PUC-Rio para o Centro Internacional de Estudos e Pesquisa sobre a Infância (CIESPI) mostrou-se uma oportunidade importante para aprimorar as soluções na área de governo eletrônico. Cabe explicar que o CIESPI é um centro de estudos e de referência dedicado ao desenvolvimento de pesquisas e projetos sociais voltados a crianças, adolescentes, jovens e seus elos familiares e comunitários (CIESPI).

1.4. Contribuições Esperadas

As contribuições esperadas desse trabalho são as seguintes:

- A aplicação de novas tecnologias disponíveis na web visando atender a necessidade de facilitar o acesso analítico às estatísticas públicas.
- A redução de tempo e custos para a exploração das informações contidas nas estatísticas públicas;
- A concepção da sistemática para transformação dos dados estatísticos em informação facilmente acessível na Web;
- Disponibilização de uma versão “demo” na Web contendo às informações da PNAD (ano 2007 e 2008);
- Estímulo ao uso das informações estatísticas pelo público interessado, incluindo cidadãos, organizações governamentais e não governamentais e acadêmicos devido à facilidade de acesso.

1.5. Organização da Dissertação

Esta dissertação está organizada em sete capítulos, a saber:

Capítulo 1 – Introdução: Apresenta o problema específico da pesquisa dentro de sua área e tema, expõe os objetivos almejados com o trabalho e a lógica de estruturação geral do documento.

Capítulo 2 – Definições Conceituais: são apresentados os conceitos, fundamentos, tecnologias e componentes necessários ao entendimento dos diversos assuntos abordados nos demais capítulos.

Capítulo 3 – A Sistemática Proposta: exhibe a sistemática proposta baseada em Data Warehousing para apoio ao e-Governo.

Capítulo 4 – Aplicação da Sistemática: descreve a aplicação da sistemática proposta utilizando dados reais de uma estatística pública (PNAD) como prova de conceito.

Capítulo 5 – A Usabilidade das Estatísticas Públicas: apresenta o processo de análise exploratória dos dados da PNAD no protótipo construído neste trabalho.