

Legibilidade em Dados Abertos: uma Experiência com os Dados da Câmara Municipal de São Paulo

Fernanda C. Bico, Leandro N. Trindade
Roberto A. R. Caracciolo Jr., Ronaldo J. S. Paiva Jr., Sarajane M. Peres¹

¹Escola de Artes, Ciências e Humanidades - Universidade de São Paulo (USP)
Av. Arlindo Béttio – 03828-000 – São Paulo – SP – Brasil

{fernanda.bico,leandro.trindade,
roberto.caracciolo,ronaldo.p,sarajane}@usp.br

Abstract. *São Paulo City Council's open data are published on its Website using XML (eXtensible Markup Language). XML format provides an organizing semistructured data, facilitating the data reading and reusing by computer applications, however supplies poor legibility for direct analysis about the information comprised therein. The work presented here aimed to study such organization to making possible building a Web application able to generate graphical views, making more democratic the opening data initiative. Both Web application development process and implemented filter options and graphical visualizations are presented in this paper.*

Resumo. *Dados abertos da Câmara Municipal de São Paulo são publicados no seu portal Web usando o formato XML (eXtensible Markup Language). O formato XML confere uma organização semiestruturada aos dados, facilitando a leitura e reúso dos mesmos por meio de aplicações computacionais, mas fornece pouca legibilidade para análise direta da informação nele contida. O trabalho aqui apresentado teve o objetivo de estudar tal organização para que fosse possível construir uma aplicação Web capaz de gerar visualizações gráficas, tornando mais democrática a iniciativa de abertura de dados. O processo de desenvolvimento da aplicação Web bem como suas opções de filtros de informação e de visualização são apresentadas neste artigo.*

1. Introdução

Em um país de tamanha densidade populacional e diversidade educacional, cultural e social como o Brasil, é frequente o estabelecimento de lacunas de comunicação entre o governo e a população. De certo modo, a informação referente ao trabalho do governo é acessível de forma resumida e pré-formatada e, quando disponibilizada, é apresentada em uma organização difícil de ser interpretada pelo cidadão comum. A atenção para iniciativas de “abertura de dados” se apresenta como um movimento recente no país, que direciona esforços para democratização da informação.

No âmbito governamental, dados abertos se referem à publicação de dados em formato natural (*raw*), porém que os tornem acessíveis, prontamente disponíveis para todos e passíveis de reúso [Machado e Parente de Oliveira 2011]. Assim, a divulgação de dados governamentais pode ser identificada como um grande passo rumo a um maior envolvimento dos cidadãos na gestão e desenvolvimento da sociedade, uma vez que informação

de qualidade sobre as decisões das esferas governamentais poderiam se tornar de amplo conhecimento da população.

Ainda que esse movimento seja recente no Brasil, e muitos cidadãos desconheçam o quão eficiente este recurso pode ser, algumas iniciativas para a regulamentação do movimento de dados abertos no Brasil vem motivando diversos órgãos públicos a disponibilizarem seus dados em meio eletrônico. Entre eles está a Câmara Municipal de São Paulo (CMSP), que vem divulgando na Web os dados sobre suas ações, sob o Programa de Dados Abertos do Parlamento¹, criado pelo *Ato da Mesa nº 1156/2011*, o qual “dispõe sobre a implementação do Programa de Dados Abertos do Parlamento no âmbito da Câmara Municipal de São Paulo”².

Entretanto, apesar das vantagens que potencialmente se vislumbra com estas iniciativas, ao acessarem os dados disponibilizados pela CMSP, os cidadãos paulistanos deparam-se com dados em formato semiestruturado, com baixa granularidade de informação. Para que estes dados possam ser efetivamente compreendidos, e para que informação de qualidade possa ser extraída da análise deles, é necessário possuir conhecimentos técnicos sobre o formato usado na disponibilização e sobre como usar a tecnologia para transformar os dados em informação.

O projeto aqui apresentado teve como objetivo estudar a estrutura de publicação dos dados da CMSP e implementar uma aplicação Web para extrair e disponibilizar informação contida nestes dados para os cidadãos. Por meio desta aplicação o usuário pode realizar consultas referentes ao assunto que mais lhe interessar e visualizar as respostas em uma interface gráfica. Este trabalho traz ainda um relato referente ao processo de entendimento dos dados e de desenvolvimento da aplicação que pode contribuir com iniciativas similares, já que na disponibilização de dados abertos, é comum a inexistência arquivos de meta-dados, o que insere um dificultador a mais para a compreensão dos dados e consequente projeto de aplicações sobre eles.

A fim de melhor organizar o conteúdo deste artigo, as seções que o divide são apresentadas como segue: na Seção 2 algumas características de iniciativas relacionadas são brevemente comentadas; a descrição dos dados abertos da CMSP é apresentada na Seção 3; a Seção 4 representa a primeira contribuição deste artigo - a descrição das tomadas de decisão que permitiram a construção da aplicação em questão; a segunda contribuição, a aplicação Web em si, é apresentada na Seção 5; finalmente, as considerações finais são apresentadas seguidas das referências bibliográficas.

2. Iniciativas relacionadas

Com o movimento em prol da divulgação de dados públicos e com a efetiva disponibilização de dados, começaram a surgir iniciativas que tem como objetivo ana-

¹<http://www.camara.sp.gov.br>

²Legislações correlatas a este tema são: a **Lei nº 301/2011**, de 22 de junho de 2011. Institui a política municipal de dados abertos e acesso à informação na administração pública direta e indireta, no tribunal de contas, e na Câmara Municipal de São Paulo, e dá outras providências (Requerimento de coautoria RDS 848/2011 - Aprovado.) Publicação Oficial da Cidade de São Paulo, 22 de junho de 2011, p.101.; a **Lei nº 41/2010**, de 30 de abril de 2010. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. Portal Atividade Legislativa.

lisar os dados, processá-los e disponibilizar aplicações para visualização de informação ou para prestação de serviços a partir das informações geradas. Esse movimento é mais evidente em países da América do Norte e da Europa. Um exemplo comum a países destes dois continentes é o *The Govwild site*³ [Böhm et al. 2010], onde estão disponíveis dados provenientes de diferentes agências governamentais americanas e européias, sob o formato de formulários de consultas que além de disponibilizar funções de filtragem da informação, também trazem consultas pré-projetadas para destaque dos dados mais recentemente publicados e das consultas mais acessadas pelos usuários.

No Brasil, existem alguns exemplos de projetos de destaque. O sistema *LegisDado*⁴, que se constitui como uma aplicação Web que consome dados diretamente dos sites da Câmara dos Deputados e do Senado Federal do Brasil, permitindo que a tramitação dos Projetos de Lei sejam acompanhados nas duas casas e criando um ambiente de mapeamento de todo o processo legislativo federal brasileiro. Em uma linha similar está site “Para onde foi o meu dinheiro”⁵, desenvolvido pelo *OpenData-BR*⁶, movimento criado pela comunidade de software livre *São Paulo Pearl Mongers*, utiliza o “Portal da Transparência”⁷ como fonte de dados. O site acessa dados do governo federal desde 2005 e do Estado de São Paulo desde 2010, e exibe de forma clara e objetiva, os valores e percentuais de como foram gastos os recursos públicos. Na área de “estatísticas” tem destaque o portal do IBGE⁸ que organiza dados de todas as cidades brasileiras, referentes a trabalhos de censo nos mais diversos domínios, incluindo desde indicadores como Produto Interno Bruto das cidades, até estatísticas sobre número de casamentos e divórcios realizados por ano em cada uma delas.

Para além de iniciativas com dados governamentais, a publicação de dados abertos também suporta aplicações que facilitam a vida dos cidadãos em diferentes sentidos. O site *After School Special*⁹ combina dados sobre escolas, livrarias e restaurantes da cidade de São Francisco nos Estados Unidos, de forma que os cidadãos possam planejar, por exemplo, atividades extra-escolares para crianças, levando em consideração informações referentes a nutrição, localização geográfica e tipo de vizinhança. Ainda em relação a informações sobre a cidade de São Francisco, outra aplicação interessante é a *SFScores*¹⁰ que oferece visualização gráfica sobre as inspeções de qualidade de todos os restaurantes e caso um local de interesse, selecionado pelo usuário, possua uma violação de padrões de segurança alimentar, a aplicação mostra os motivos e esclarecimentos sobre o fato. Na área de aplicativos específicos para *Smart Phones*, a cidade de Nova York se destaca. Repositórios de dados abertos sobre esta cidade permitiram a criação de aplicativos como o *Trees Near You*¹¹, para *iPhones*, que identifica todas as árvores da cidade; ou ainda o *Routesy*¹², também para *iPhones*, que mostra os horários de ônibus e a localização de estações através de um mapa. Uma reportagem destacando a importância de ferramentas

³govwild.hpi-Web.de

⁴<http://data.votaio.com/>

⁵<http://www.paraondefoiomeudinheiro.com.br>

⁶<http://opendatabr.org>

⁷<http://www.portaldatransparencia.gov.br/>

⁸<http://www.ibge.gov.br/cidadesat/topwindow.htm?1>

⁹<http://www.afterschoolsf.org/>

¹⁰<http://sfcores.com/>

¹¹<http://www.treesnearyou.com/>

¹²<http://www.routesy.com/>

como essas, e algumas outras mais, pode ser encontrada em [Miller 2006].

Assim como há projetos de disponibilização de dados e aplicativos para uso público, há também iniciativas de desenvolvimento de ferramentas para suportar a organização e disponibilização dos dados e para proporcionar que os próprios usuários transformem os dados em informação [Böhm et al. 2010]. Exemplos interessantes são o *dcat GridWorks*¹³ e as ferramentas encontradas no site do *Linked Data*¹⁴.

Em uma linha mais abrangente do que a criação de aplicações, está o trabalho desenvolvido pelo Serviço Federal de Processamentos de Dados (SERPRO) do Brasil, que propõe uma arquitetura para permitir o acesso a dados, por ele chamados “primários” [Machado e Parente de Oliveira 2011]. Este acesso se dá via desenvolvimento de software tal que seja possível, por exemplo, combiná-los em aplicações do tipo *mashup*, gerando novas informações. Esta arquitetura tem o objetivo de oferecer concordância semântica entre fontes de dados heterogêneos.

3. Dados Abertos da Câmara Municipal de São Paulo

A CMSP disponibiliza dados referentes às sessões plenárias ordinárias e extraordinárias, a informações sobre funcionários e à área contábil. A área contábil é a que possui mais dados disponibilizados e eles estão organizados nas seguintes subáreas: pagamentos gerais e contrato; demonstrativo de despesas realizadas; balancete da despesa; demonstrativo diário da execução da despesa (estorno); demonstrativo diário da execução da despesa (empenho); despesa de pessoal; dotações orçamentárias da Câmara Municipal de São Paulo; e dotações orçamentárias da Secretaria Municipal de Educação.

Esses dados estão divulgados seguindo o formato XML - padrão definido pelo W3C (*World Wide Web Consortium*) e disponibilizados para *download* via protocolo HTTP (*Hypertext Transfer Protocol*). A adoção da linguagem XML sob o padrão W3C permite a criação de uma comunicação universal para aplicações computacionais, visto que a sintaxe da linguagem XML permite que elas consumam e distribuam dados com facilidade. O conjunto de dados estudado no presente projeto está organizado em 8 arquivos XML conforme descrito na Tabela 1.

A Figura 1 apresenta um pequeno trecho do arquivo XML 8. Note que os dados são apresentados na forma semiestruturada, discriminados por *tags* que, em algum nível, passam informação semântica sobre o contexto que o dado representa. Observando o conteúdo do XML é possível verificar que se tratam de dados da área de educação porque a *tag* <orgão> discrimina a proveniência dos dados. Outras *tags* podem ser analisadas para inferência de mais informações, por exemplo: a *tag* <programa> especifica o contexto de uso da dotação; a *tag* <grupoDespesa> indica que o gasto é considerado como “investimento” e a *tag* <valor> especifica o montante referente ao investimento.

4. Decisões de Projeto e Modelagem de Dados

A complexidade do contexto contábil e a dificuldade em encontrar metadados que pudessem ajudar na compreensão dos relacionamentos existentes entre os dados dos arquivos XML representa um complicador tanto para a compreensão dos dados disponi-

¹³<http://lab.linkeddata.deri.ie/2010/dcat/#what>

¹⁴<http://linkeddata.org/tools>

Tabela 1. Resumo do contexto do conjunto de dados estudado no presente trabalho. A coluna referente a “tamanho” diz respeito ao número de registros XML contidos no arquivo. Tamanho variável (var.) foi adotado aqui para indicar arquivos que possuem atualizações frequentes.

	Tamanho	Sobre os dados
1	135	Pagamentos gerais e contratos de julho/2011
2	22	Demonstrativo de despesas referente a julho/2011 - mensal e acumulado
3	22	Balancete da despesa referente a julho/2011 - mensal e acumulado
4	26	Demonstrativo diário da execução de despesa referente a julho/2011 - estornos
5	169	Demonstrativo diário da execução de despesa referente a julho/2011 - empenhos
6	1	Despesa de pessoal - maio/2010 a abril/2011
7	873	Dotação orçamentária da Secretaria Municipal de Educação - exercício de 2011
8	29	Dotação orçamentária da Câmara Municipal de São Paulo - exercício de 2011
9	var.	Presença de Vereadores em Sessões Plenárias Ordinárias e Extraordinárias
10	var.	Votações em Sessões Plenárias Ordinárias e Extraordinárias

bilizados quanto para a construção de uma aplicação para produção de informação. Esta dificuldade é exemplificada pela situação ilustrada na Figura 1.

Com base nos dados visualizados nesse arquivo é possível inferir que “9 milhões de reais foram investidos na forma de obras e instalações para a área de educação da cidade de São Paulo, durante o ano de 2011”. Porém, o dado representado na tag **<modalidadeDespesa>**, por exemplo, possui um sentido semântico que não faz parte do vocabulário do cidadão comum e seu significado não pode ser facilmente determinado. Evidentemente que o acompanhamento de um especialista da área de Ciências Contábeis, com experiência em Contabilidade Pública, poderia suportar análises mais críveis e verossímeis, entretanto, num contexto de democratização da informação, a exigência deste conhecimento específico corrobora contra a plena legibilidade da informação.

Face a essas circunstâncias, os autores deste estudo identificaram duas formas de explorar os dados da CMSP para construção de aplicações para geração e disponibilização de informação:

- a primeira, mais restrita a um grupo de pessoas que tenham conhecimento especialista do contexto contábil dos dados ou acesso aos sistemas de informação legados da CMSP produtores desses dados, e respectiva documentação, possibilitaria a construção de aplicações mais complexas com potencial para geração de informação de alto nível e, inclusive, descoberta de conhecimento;
- a segunda, mais democrática, gira em torno do trabalho de extração de informação a partir de inferências, que apesar se constituírem como um mapeamento direto dos arquivos XML para uma modelagem de dados estruturadas, tem o potencial de permitir a construção de filtros de dados e geração de visualização gráfica da informação, atendendo a demanda de democratização da informação, ainda que parcialmente.

Adotando a segunda possibilidade como método de extração de informação, para a realização do presente projeto foram selecionados os arquivos 7 e 8 (Tabela 1). A partir da seleção desses arquivos, foram desenvolvidos duas modelagens de dados: (a) a primeira,

```

<CMSP>
- <linha>
  <idOrgao>16</idOrgao>
  <orgao>Secretaria Municipal de Educação</orgao>
  <idUnidade>10</idUnidade>
  <unidade>Gabinete do Secretário</unidade>
  <idFuncao>12</idFuncao>
  <funcao>Educação</funcao>
  <idSubFuncao>122</idSubFuncao>
  <subFuncao>Administração Geral</subFuncao>
  <idPrograma>1124</idPrograma>
  <programa>Acesso à educação e qualidade do ensino - Educação integrada</programa>
  <idProjetoAtividade>1848</idProjetoAtividade>
  <projetoAtividade>Reforma e Ampliação de Unidades Educacionais Integradas</projetoAtividade>
  <idDespesa>44905100</idDespesa>
  <despesa>Obras e Instalações</despesa>
  <idCategoriaDespesa>4</idCategoriaDespesa>
  <categoriaDespesa>Despesas de Capital</categoriaDespesa>
  <idGrupoDespesa>4</idGrupoDespesa>
  <grupoDespesa>INVESTIMENTOS</grupoDespesa>
  <idModalidadeDespesa>90</idModalidadeDespesa>
  <modalidadeDespesa>Aplicações Diretas</modalidadeDespesa>
  <idElementoDespesa>51</idElementoDespesa>
  <elementoDespesa>Obras e Instalações</elementoDespesa>
  <idFonteDespesa>00</idFonteDespesa>
  <fonteDespesa>Tesouro Municipal</fonteDespesa>
  <dotacao>16.10.12.122.1124.1848.4.4.90.51.00</dotacao>
  <valor>9000000</valor>
</linha>

```

Figura 1. Dotações Orçamentárias da Secretaria Municipal de Educação de 2011, em formato XML.

uma modelagem de dados relacional, para receber os dados transferidos dos arquivos XML e criar um ambiente adequado para realização de consultas que viabilizasse um estudo sobre o contexto dos dados; (b) a segunda, uma modelagem multidimensional e hierárquica, para organizar e persistir os dados para a aplicação Web.

4.1. Estudando os Dados com a Modelagem Relacional

A fim de criar um ambiente para facilitar o estudo sobre os dados, foi concebido um modelo relacional que representa a organização de dados usada nos arquivos XML. Esse modelo não apresenta relacionamentos entre os dados pois não foi construído para representar a estrutura relacional que os dados devem possuir no contexto real; é portanto um modelo não-normalizado. No entanto, a motivação para tal modelagem foi persistir os dados em um Sistema de Gerenciamento de Banco de Dados para criar cenários, através de consultas SQL (*Structured Query Language*), que pudessem revelar requisitos para construção da aplicação Web. Para construção deste modelo de dados, as seguintes regras de mapeamento foram estabelecidas:

- cada arquivo XML é representado, na base de dados, como uma relação (tabela);
- cada elemento-filho contido no XML é representado na base de dados como uma tupla (linha);
- cada atributo de um elemento-filho contido no XML é representado, na base de dados, como um atributo (coluna) de uma relação (tabela).

Portanto, o modelo relacional construído é um “espelho” do conteúdo do arquivo XML, ou seja, a organização dos dados contida nos arquivos XML permanece neste modelo. Um aplicativo console foi desenvolvido exclusivamente como uma ferramenta cujo objetivo é promover a interface de dados entre arquivos XML e banco de dados, estando apto a ler os arquivos XML e a popular este novo modelo, este último assumiu um caráter temporário no processo de desenvolvimento da aplicação Web aqui discutida.

Com a migração para o ambiente estruturado de um modelo de dados relacional foi possível melhorar a compreensão sobre os dados. Usando a linguagem SQL, foram projetadas e executadas diferentes consultas de agrupamento, contabilização e ordenamento de valores, o que possibilitou o alcance de graus de liberdade na manipulação dos dados - antes dificultada pelo fato dos dados estarem contidos em arquivos semiestruturados; e foi possível constatar alguns fatos.

Um exemplo dessas constatações é ilustrado pela situação presente na Figura 2. Note que vários elementos-filho referentes a contexto de “despesas”, nos dois arquivos selecionados, além de apresentarem a mesma estrutura (com atributos idênticos), apresentam valores de atributos em comum, tanto em elementos-filho do mesmo arquivo quanto de arquivos diferentes. Esta constatação motivou a criação de um modelo de dados único, adequado para receber dados de mais de um arquivo XML, para a aplicação Web.

<pre> <linha> <idOrgao>09</idOrgao> <orgao>Câmara Municipal</orgao> <idUnidade>10</idUnidade> <unidade>Câmara Municipal</unidade> <idFuncao>01</idFuncao> <funcao>Legislativa</funcao> <idSubFuncao>031</idSubFuncao> <subFuncao>Ação Legislativa</subFuncao> <idPrograma>2610</idPrograma> <programa>Suporte administrativo</programa> <idProjetoAtividade>1000</idProjetoAtividade> <projetoAtividade>Construção de Edificações da Câmara Municipal</projetoAtividade> <idDespesa>44903900</idDespesa> <despesa>Outros Serviços de Terceiros - Pessoa Jurídica</despesa> <idCategoriaDespesa>4</idCategoriaDespesa> <categoriaDespesa>Despesas de Capital</categoriaDespesa> <idGrupoDespesa>4</idGrupoDespesa> <grupoDespesa>INVESTIMENTOS</grupoDespesa> <idModalidadeDespesa>90</idModalidadeDespesa> <modalidadeDespesa>Aplicações Diretas</modalidadeDespesa> <idElementoDespesa>39</idElementoDespesa> <elementoDespesa>Outros Serviços de Terceiros - Pessoa Jurídica</elementoDespesa> <idFonteDespesa>00</idFonteDespesa> <fonteDespesa>Tesouro Municipal</fonteDespesa> <dotacao>09.10.01.031.2610.1000.4.4.90.39.00</dotacao> <valor>11200000</valor> </linha> </pre>	<pre> <linha> <idOrgao>16</idOrgao> <orgao>Secretaria Municipal de Educação</orgao> <idUnidade>10</idUnidade> <unidade>Gabinete do Secretário</unidade> <idFuncao>12</idFuncao> <funcao>Educação</funcao> <idSubFuncao>122</idSubFuncao> <subFuncao>Administração Geral</subFuncao> <idPrograma>1124</idPrograma> <programa>Acesso à educação e qualidade do ensino - Educação</programa> <idProjetoAtividade>1848</idProjetoAtividade> <projetoAtividade>Reforma e Ampliação de Unidades Educacionais</projetoAtividade> <idDespesa>44905100</idDespesa> <despesa>Obras e Instalações</despesa> <idCategoriaDespesa>4</idCategoriaDespesa> <categoriaDespesa>Despesas de Capital</categoriaDespesa> <idGrupoDespesa>4</idGrupoDespesa> <grupoDespesa>INVESTIMENTOS</grupoDespesa> <idModalidadeDespesa>90</idModalidadeDespesa> <modalidadeDespesa>Aplicações Diretas</modalidadeDespesa> <idElementoDespesa>51</idElementoDespesa> <elementoDespesa>Obras e Instalações</elementoDespesa> <idFonteDespesa>00</idFonteDespesa> <fonteDespesa>Tesouro Municipal</fonteDespesa> <dotacao>16.10.12.122.1124.1848.4.4.90.51.00</dotacao> <valor>9000000</valor> </linha> </pre>
--	--

Figura 2. Identificação de pontos comuns nos arquivos de dados XML, facilitada pela organização dos dados em um modelo relacional e pelo uso da linguagem SQL. Note que os valores dos atributos *grupoDespesa*, *modalidadeDespesa* e *fonteDespesa* são iguais nos dois arquivos.

Idealmente, um processo de normalização no modelo relacional seria necessário para que a modelagem alcançasse maturidade para suportar o desenvolvimento de uma aplicação. Contudo, devido à motivação, então já existente, para criação de um segundo modelo capaz de representar as constatações elicitadas, uma organização de dados mais suscetível para o processamento de dados e geração da informação mostrou-se interessante e um modelo de dados multidimensional e hierárquico foi projetado.

4.2. Organizando e Persistindo Informação com a Modelagem Multidimensional e Hierárquica

Conceitos de *OnLine Analytical Process* (modelagem, linguagem e produtos) envolvem, segundo [Thomsen 2002], noções ou ideias de dimensões múltiplas e hierárquicas, usadas para melhorar o entendimento sobre o mundo (domínio de aplicação). Assim, com o intuito de promover a geração de informação de forma eficiente, foi concebida uma modelagem de dados multidimensional e hierárquica para persistir os dados da aplicação aqui discutida.

pendências funcionais em [Elmasri e Navathe 2005]) referentes a esse contexto não foram encontradas e, inclusive, constatou-se a violação de dependências funcionais similares àquelas encontradas no contexto de “Função”. Assim, hierarquias no contexto “Despesas” não foram estabelecidas.

Note que essas decisões de projeto estão baseadas em um estudo sobre instâncias de dados, e dependências funcionais que são válidas nas instâncias podem não ser requeridas pelo modelo (ou esquema) [Elmasri e Navathe 2005]. Para definir as dependências funcionais no modelo, fazendo com que fosse exigidas nas instâncias, seria necessário usar conhecimento especializado do domínio contábil, e como já discutido antes neste artigo, não foi considerada a semântica contábil inerente ao domínio da aplicação.

A Figura 4 representa o processo de criação do modelo, o qual considerou os seguintes aspectos:

- os atributos de um elemento-filho do arquivo XML, cujas nomenclaturas se iniciam por “id”, tornaram-se chaves-primárias da relações que armazenam dados das “Dimensões” do modelo;
- cada tabela “Dimensão” possui atributos de descrição que acompanham os atributos cujas nomenclaturas se iniciam por “id”;
- os elemento-filhos dos arquivos XML são representados na tabela “Fato” do modelo, que possui chave primária numérica e incremental;
- os relacionamentos existentes entre a tabela “Fato” e as dimensões são do tipo “um-para-muitos” significando que cada tupla da tabela “Fato” se relaciona a uma tupla de cada “Dimensão”;
- todos os demais atributos que não se tornaram “Dimensões”, tornaram-se atributos da relação “Fato” constituindo os atributos de “Métricas” do modelo;
- dependências funcionais verificadas no estudo dos dados geraram normalização no modelo.

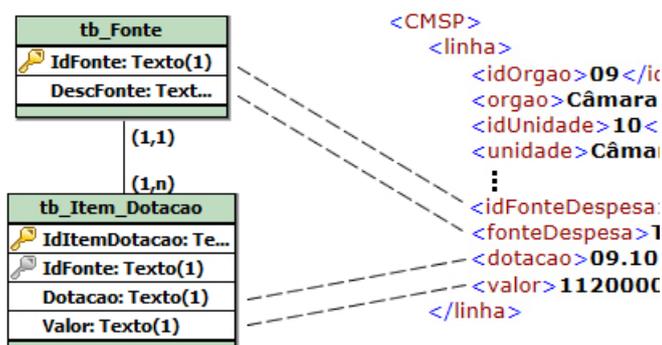


Figura 4. Processo de concepção do modelo multidimensional e hierárquico: a relação “Fato” *tb_Item_Dotacao* representa o elemento-filho *linha* do arquivo XML. Ela possui os atributos *Valor* e *Dotacao*, além da chave-primária *IdItemDotacao* e da chave-estrangeira *idFonte* proveniente da relação de “Dimensão” *tb_Fonte*. Esta relação, por sua vez, possui chave-primária *idFonte* e atributo *DescFonte*, originados, respectivamente, dos atributos *idFonteDespesa* e *fonteDespesa* do arquivo XML.

5. Câmara Aberta: um aplicativo Web para dados abertos

Sobre o modelo de dados multidimensional e hierárquico gerado foi desenvolvida uma aplicação Web intitulada “Câmara Aberta”. A aplicação disponibiliza relatórios e

representações gráficas, com informações geradas a partir dos dados abertos da CMSP. O objetivo desta aplicação é mostrar a informação de maneira simplificada. Com esse tipo de iniciativa, dados brutos de difícil compreensão podem ser visualizados na forma de informação.

Uma das principais características desta aplicação Web é oferecer ao usuário a liberdade de exploração da informação por meio de uma interface gráfica de fácil operação. Uma descrição mais detalhada das possibilidades de filtros que a aplicação disponibiliza é apresentada na Tabela 2¹⁵. As Figuras 5, 6 e 7 exibem, respectivamente, a página onde o usuário pode escolher os filtros a partir dos quais os relatórios serão gerados; um exemplo da organização da informação em um relatório; um exemplo de visualização gráfica.

Tabela 2. Relação de filtros disponíveis na aplicação.

	Órgão	Unidade	Categoria	Despesa	Elemento	Fonte	Função	Sub Função	Grupo	Modalidade	Programa	Atividade
Relatórios	X	X	X	X	X	X	X	X	X	X	X	X
Gráficos									X		X	X

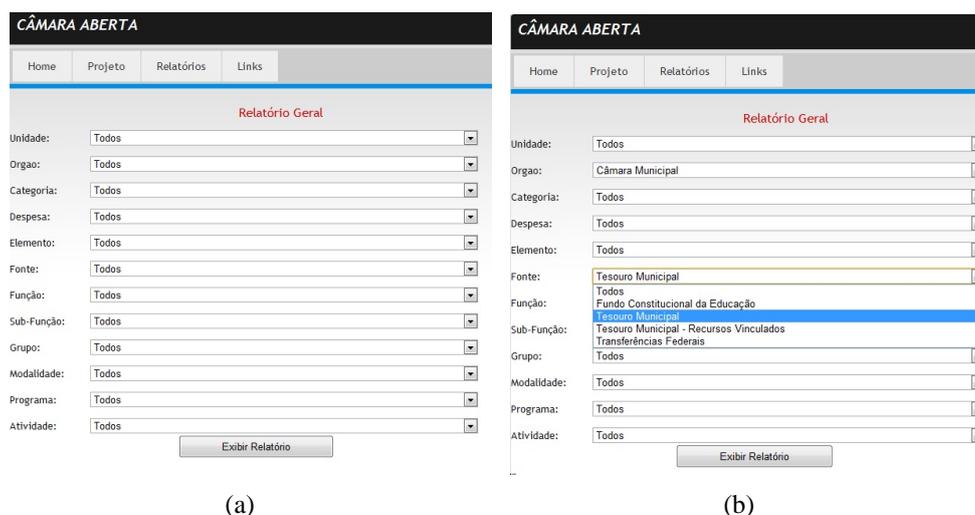


Figura 5. (a) Escolha de filtros para geração de relatórios (filtros escolhidos: Órgão e Fontes); (b) Escolha de opções dentro de filtros (opções escolhida: Câmara Municipal e Tesouro Municipal). A possibilidade de escolher vários filtros e várias opções dentro deles gera o cruzamento de dados.

6. Considerações Finais

A falha na comunicação entre os cidadãos e seus governos é um problema eminente, prejudicando a disseminação das informações públicas. O incentivo à divulgação de dados públicos no Brasil é um grande passo rumo à redução desta separação, embora esta ainda seja uma atividade muito recente no país e muitos cidadãos desconheçam o quão

¹⁵As células não marcadas na tabela referem-se a filtros possíveis para os quais ainda não foram implementadas as interfaces gráficas.

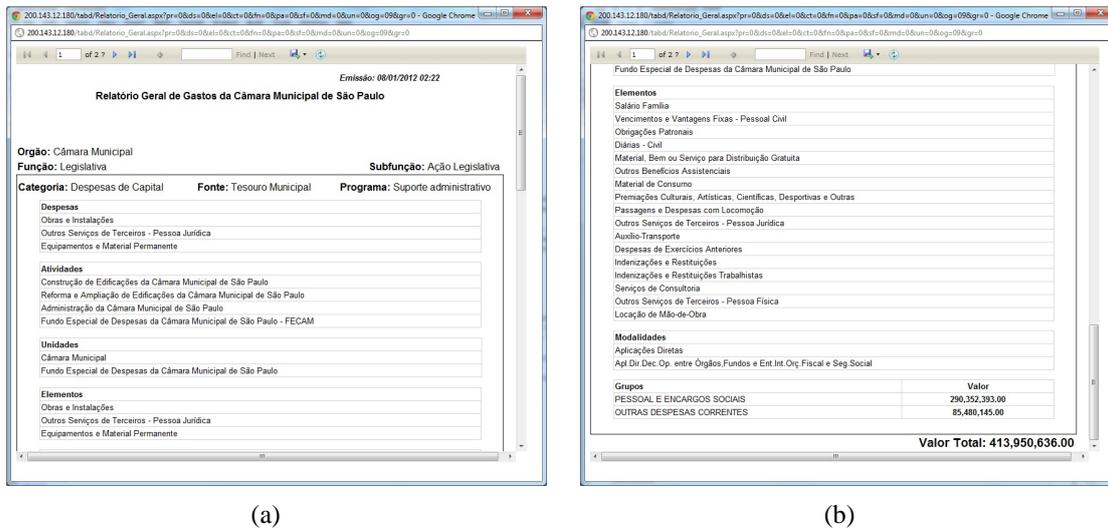


Figura 6. Relatório contendo todas as informações, organizadas, referente a gastos da Câmara Municipal sob a fonte Tesouro Nacional:(a) cabeçalho e informações iniciais; (b) informações finais do relatório.

eficiente pode ser este recurso. Neste contexto, um dos grandes obstáculos encontrados é a forma como estes dados são disponibilizados. Em suma, a maioria dos dados abertos é disponibilizado em um formato de difícil entendimento, ou porque exigem conhecimentos específicos para interpretação do contexto em que se inserem, muitas vezes desconhecidos pela maioria dos cidadãos; ou porque estão formatados em planilhas ou arquivos semiestruturados cuja legibilidade é totalmente inadequada para análise manual.

Este contexto evidencia a importância da utilização de ferramentas automatizadas que agregam valor a esses dados (e iniciativas) gerando informação útil e disponibilizando-as em formato gráfico e organizado. Dentre as diversas soluções existentes, a experiência relatada neste artigo tratou especificamente de uma solução Web, que conferiu legibilidade à parte dos dados disponibilizados pela CMSP, principalmente porque a modelagem de dados realizada abriu uma série de filtros que possibilitam a criação de visões com cruzamento de informação.

A aplicação aqui descrita explora dois arquivos XML que são disponibilizados pela CMSP e, portanto, ainda precisa de mais esforço de estudo, modelagem e implementações para que, de fato, consolide-se como uma ferramenta que fornece total visibilidade aos dados abertos da CMSP. Algumas necessidades prementes de extensão desta aplicação são:

- estudar a possibilidade de modelar a contexto de “Despesas” como uma nova tabela “Fato”, transformando o modelo atual (*Snowflake*) em um modelo *Constelação*;
- acrescentar a dimensão **Tempo** ao modelo, para comportar futuras atualizações de dados, conforme eles forem disponibilizados pela CMSP.
- explorar um número maior de arquivos XML;
- melhorar detalhes técnicos da aplicação para que ela possa ser disponibilizada para o público em geral. Entre essas melhorias estão: implementar a interface gráfica para mais filtros e cruzamentos de dados; otimizar o processo de con-

sumo dos dados do portal da CMSP; desenvolver funcionalidades de cadastro do usuários, envio de e-mail com avisos de atualização da base de dados; e integração do aplicativo com redes sociais.

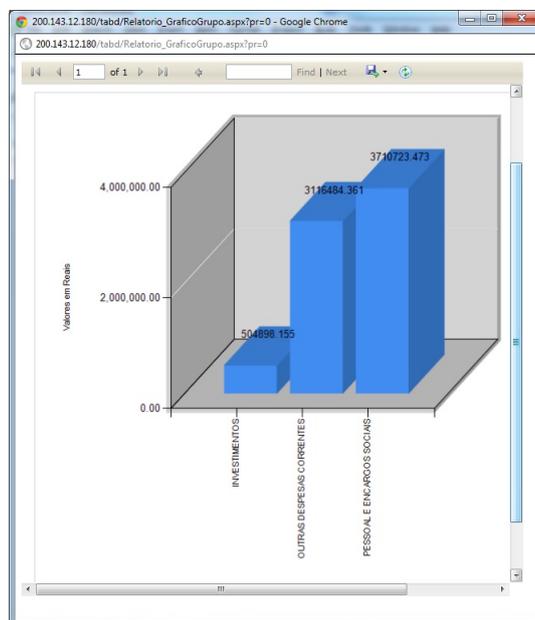


Figura 7. Exemplo de visualização gráfica: valores em reais referentes a três diferentes grupos de gastos da CMSP.

Mais do que relatar uma experiência, este artigo teve o intuito de mostrar o processo executado pelos autores para conduzir o estudo, esperando assim motivar outros desenvolvedores a explorar bases de dados abertos para disseminar a cultura de exploração destes dados e de criação de ferramentas correlatas.

Referências

- Böhm, C., Naumann, F., Freitag, M., George, S., Höfler, N., Köppelmann, M., Lehmann, C., Mascher, A., e Schmidt, T. (2010). Linking open government data: what journalists wish they had known. In *Proceedings of the 6th International Conference on Semantic Systems, I-SEMANTICS '10*, pages 34:1–34:4, New York, NY, USA. ACM.
- Elmasri, R. E. e Navathe, S. (2005). *Sistemas de Banco de Dados - Fundamentos e Aplicações*. Addison Wesley, 4a edition.
- Han, J. e Kamber, M. (2006). *Data Mining: Concepts and Techniques*. Diane Cerra, San Francisco, CA, 2nd edition.
- Machado, A. e Parente de Oliveira, J. (2011). Digo: An open data architecture for e-government. In *IEEE International Enterprise Distributed Object Computing Conference Workshops (EDOCW)*, pages 448–456.
- Miller, C. C. (2006). Local governments offer data to software tinkerers. Newspaper: The New York Times.
- Thomsen, E. (2002). *OLAP Solutions*. Robert Ipsen, Canada, 2nd edition.