

Aplicação OLAP para Segurança Pública: um Estudo de Caso a Partir de Dados Governamentais Abertos do Estado do Rio de Janeiro

Adriano Barboza Leite¹, Ellen Polliana Ramos Souza¹, Jorge da Silva Correa Neto¹,
Marcelo Iury de Sousa Oliveira¹

¹Unidade Acadêmica de Serra Talhada - Universidade Federal Rural de Pernambuco (UFRPE)
Caixa Postal 063 – 56.903-000 – Serra Talhada – PE – Brazil

adrianobarboza06@gmail.com, {eprs, jcorreia, iury}@uast.ufrpe.br

***Abstract.** Give transparency to public security data has been a practice in some Brazilian states, however, the tools available for manipulation and analysis have failed to meet the expectations of the population. The use of data marts and online analytical processing (OLAP), used by this project, permits a richer and effective interaction with the civil society. Therefore, using dashboards, it is possible to analyze the criminality indexes provided by the Public Security Institute of the Rio de Janeiro State and, for sure, to contribute with the citizenship and the civil society empowerment.*

***Resumo.** Dar transparência aos dados sobre segurança pública já tem sido uma prática em alguns Estados brasileiros, contudo, as ferramentas de manipulação e análise não tem conseguido atender às expectativas da população. O uso de data marts e de ferramentas de processamento analítico online (OLAP), empregados no desenvolvimento deste projeto, permite uma interação mais rica e efetiva com a sociedade. Também, por meio de painéis de bordo é possível analisar os índices de criminalidade disponibilizados pelo Instituto de Segurança Pública do Estado do Rio de Janeiro e, dessa forma, contribuir com o exercício da cidadania e o empoderamento da sociedade civil.*

1. Introdução

Na chamada era da informação, cada vez mais dados são publicados na *Web* permitindo livre acesso a qualquer interessado, que pode criar suas próprias consultas agregando, unindo, filtrando ou transformando informações como bem lhe servir. Este novo conceito é chamado de Dados Abertos e consiste na publicação e disseminação de informações na Internet, compartilhadas em formatos abertos, legíveis por máquinas, e que possam ser livremente reutilizadas de forma automatizada pela sociedade (BIZER *et al.*, 2008).

Neste contexto, é cada vez maior o número de organizações governamentais que têm disponibilizado seus dados, com o intuito de promover a democracia por meio do incremento da interação entre sociedade e governo. E, mais recentemente, no atendimento da demanda social por transparência – esta regida pela Lei de Acesso à Informação, Lei Federal nº 12.527/11 (BRASIL, 2011).

Dentre as áreas que integram a atuação do setor público, existe a de segurança, que, segundo Soares (2010), é um dos grandes problemas que os gestores públicos enfrentam e

representa uma das principais preocupações dos brasileiros. Muitas instituições estão ligadas direta ou indiretamente com a segurança pública. Os dados produzidos e armazenados, como por exemplo, as ocorrências policiais, normalmente resultam em volumes expressivos, dispersos em vários tipos de sistemas, informatizados ou não. Tais dados quase sempre estão em formato não estruturado, com problemas de inconsistência e redundância, de forma que a manipulação e análise tornam-se complexas. Conforme Bico *et al.* (2012), “dados abertos” geralmente são apresentados de maneira que desfavorece o entendimento: seja pelo conteúdo exigir capacidade de interpretação específica do usuário (cidadão); ou devido à formatação semiestruturada em arquivos ou planilhas que inviabilizam qualquer análise manual.

Diante do cenário exposto, o presente trabalho trata do desenvolvimento de uma solução baseada em dados, mais especificamente, uma aplicação OLAP que facilita a análise dos índices de criminalidade do Instituto de Segurança Pública do Estado do Rio de Janeiro (ISP-RJ). A proposta é relatar o projeto que visou agregar uma forma de visualização mais adequada aos dados, possibilitando que a sociedade possa acessar, analisar e compreender as informações disponibilizadas.

2. Dados Abertos e Segurança Pública no Brasil

Instituições governamentais têm adotado “dados abertos” com vários objetivos: promoção da transparência; enfrentamento da corrupção; melhoria da participação social; inovação; otimização na eficiência e efetividade dos serviços governamentais; verificação do impacto das políticas e geração de novos conhecimentos pela combinação de fontes de dados (W3C BRASIL, 2011). No Brasil, uma parcela significativa dos dados publicados e de aplicações de consumo de dados abertos no Brasil está voltada para as áreas contábeis.

No que tange a Segurança Pública, ainda não há muitos projetos concretos em andamento. Entre as iniciativas existentes destacam-se o Portal do Instituto Jones dos Santos Neves e o Retrato da Violência Contra a Mulher no RS. A primeira, embora apresente dados de segurança, possui um foco mais amplo não permitindo o detalhamento especializado das ocorrências e a avaliação de indicadores específicos de segurança. A segunda iniciativa está voltada para uma área de segurança pública e possui uma interface amigável para usuários comuns, contudo a mesma não oferece muitas customizações para que o usuário possa elaborar suas próprias consultas.

O aumento da criminalidade tem se refletido como uma das principais preocupações da população em geral e por isso deve ter atenção adequada. Os índices de criminalidade demonstram que as políticas públicas de segurança adotadas têm apresentado pouca eficiência em reverter a tendência de crescimento dos índices, apesar dos investimentos feitos na área de segurança (WAISELFISZ, 2013). Crescimento desordenado das cidades, desemprego, leis com pouco rigor, impunidade e pouco investimento em educação, estão entre os fatores que promovem a violência como um dos principais desafios a ser enfrentado pelas cidades brasileiras.

3. SegRioDeJaneiro: Aplicação OLAP para Análise e Consulta de Índices de Criminalidade

A necessidade de qualificar as informações obtidas, a partir dos dados de segurança pública, desperta o interesse no desenvolvimento de alternativas que contribuam no acesso aos dados resultantes da gestão pública. Adicionalmente, os conceitos de OLAP, *data mart* e dados

abertos utilizados neste trabalho, compõem um conjunto de instrumentos capaz de disponibilizar informações com melhor potencial de contribuição nos processos de gestão e, de forma mais amigável, à consulta pelos cidadãos.

Os dados utilizados neste projeto são provenientes da Secretaria de Segurança do Estado do Rio de Janeiro, que através de seu Instituto de Segurança Pública, disponibiliza dados oficiais sobre incidências criminais. Os dados estão acessíveis para *download* por meio de arquivos, no formato proprietário Microsoft Excel®, contendo as principais ocorrências de crime, publicadas mensalmente no Diário Oficial. Categorizados em 39 títulos, os dados criminais estão distribuídos conforme a Região, Área Integrada de Segurança Pública (AISP) e área de abrangência da Delegacia de Polícia (DP).

Mediante análise das planilhas e dos dados relativos aos registros dos índices de criminalidade, obteve-se como maior nível de granularidade possível, quanto à relação Criminalidade x Local x Tempo, que foi: DESCRIÇÃO (do registro) x DP x MÊS.

O armazenamento dos dados no *data mart* seguiu a estrutura Relacional OLAP, pois a mesma possui a vantagem de usar sistemas gerenciadores de banco de dados *open source* que fazem uso da linguagem aberta e padronizada SQL. E para orientar a construção do modelo dimensional optou-se por utilizar o esquema estrela proposto por Kimball (1998), por propiciar um arranjo que favorece a realização de consultas de forma simples e com resultados mais rápidos.

Durante o processo de desenvolvimento do modelo dimensional foram identificadas três tabelas dimensão (*dim_tempo*, *dim_local* e *dim_criminalidade*) e uma tabela fato (*fato_ocorrencias*), que estão associadas conforme o modelo constante na Figura 1.

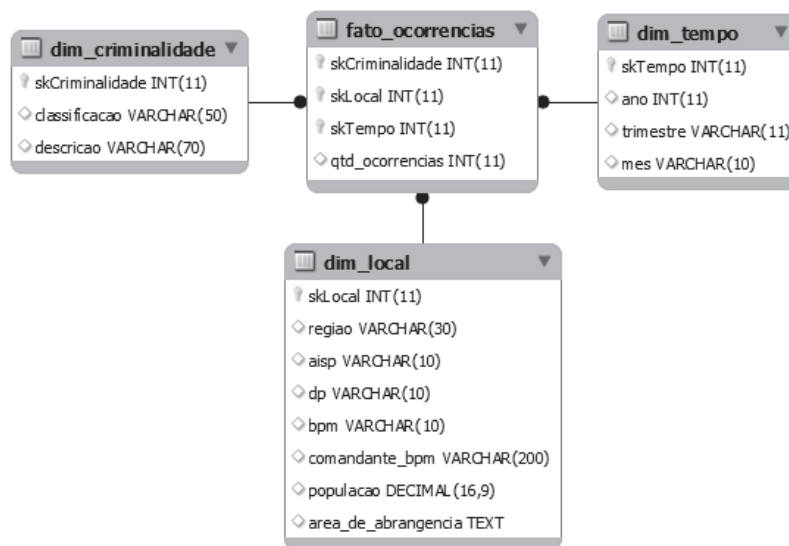


Figura 1 – Modelo dimensional

A dimensão tempo possui como chave primária o campo *skTempo*, do tipo autoincremento, destinado a manter a integridade referencial com a tabela fato – função similar ocorre com os campos chave das outras dimensões. Os valores de *skTempo* têm sua semântica por meio dos campos Ano, Trimestre e Mês. Grande parte das consultas realizadas, como por exemplo, para observação histórica de tendências e de comparação de valores, envolve a dimensão temporal.

Outro canal de entrada para o modelo é a dimensão local, cuja responsabilidade é a de possibilitar a filtragem dos resultados por meio de áreas de abrangência geográfica, conforme os campos Região, AISP e DP. Os demais campos da tabela têm função descritiva.

Questões relacionadas aos tipos de registros de criminalidade e/ou à classificação das ocorrências registradas podem ser formuladas através do uso da dimensão criminalidade. Os campos desta tabela são Classificação e Descrição.

3.1 Extração, Transformação e Carga

O processo de Extração, Transformação e Carga (ETC) foi realizado por meio da ferramenta Pentaho Data Integration, integrante da plataforma Pentaho BI. Como mencionado, a origem dos dados foi um conjunto de 48 arquivos em formato Excel, cada qual contendo os registros criminais relativos a um determinado mês, distribuídos entre 40 planilhas. Os registros foram extraídos, filtrados os que estavam vazios e gravados em um arquivo texto de acordo com a sequência de *steps* apresentados na Figura 2.

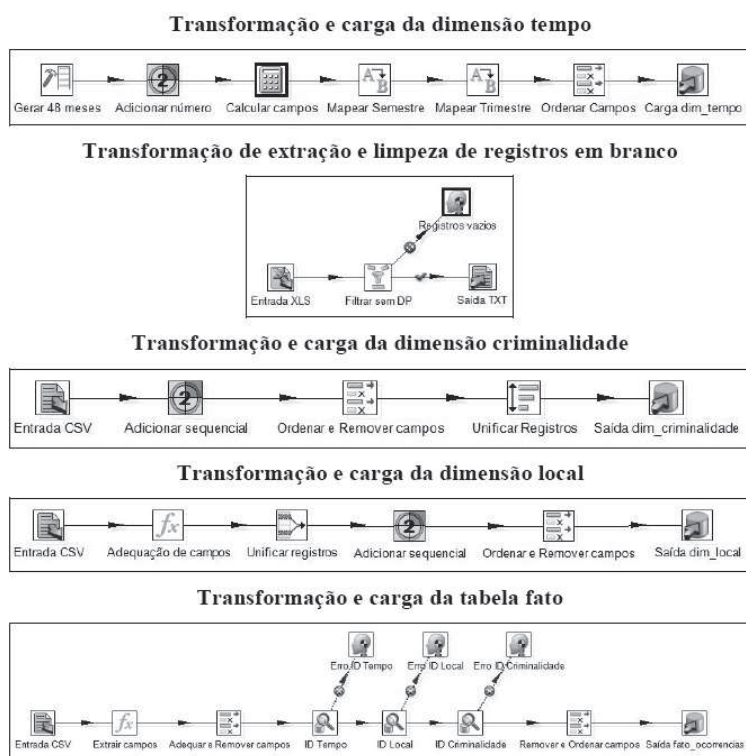


Figura 2 – Planos de ETC usados para montagem do DM

3.2 Painéis de Bordo

Como pré-requisito para construção dos painéis de bordo, fez-se necessária a elaboração de um Cubo OLAP. Tal estrutura é responsável pelo acesso e manipulação dos dados do *data mart*, armazenados mediante modelagem dimensional. A Figura 3 mostra a disposição das tabelas dimensão, seus campos associados hierarquicamente, a tabela fato e a métrica, na composição do Cubo OLAP montado com o uso da ferramenta Schema Workbench.



Figura 3 – Estrutura do cubo OLAP

Foram criados três painéis de bordo. Um deles para consulta das ocorrências criminais foi implementado para visualização dos dados sobre um mapa do Estado do Rio de Janeiro, subdividido por região. Nele é possível definir parâmetros para ano, trimestre e mês, além da classificação e descrição relativas aos registros para filtrar os resultados. Com este painel de bordo o cidadão pode verificar, mês a mês, os índices de criminalidade disponíveis no *data mart*. A Figura 4.a mostra uma consulta relativa ao mês de agosto de 2010 do número de ocorrências de prisões na região do Interior.



Figura 4 – (a) Mapa para consulta de índices; (b) Ranking de ocorrências por região

Outro painel de bordo foi o de análise do *Ranking* por Localidade, que permite consultas mediante a escolha da classificação, da descrição dos registros e do ano de interesse para a pesquisa. Na composição do painel de bordo foram utilizados gráficos de barra horizontal: o primeiro gráfico mostra as regiões do Estado; o segundo é relativo às AISPs; e o terceiro às DPs. O número de apreensões de armas pode ser verificado, ano a ano, conforme a Figura 3.b, que apresenta uma consulta com os dados relativos ao ano de 2010.

O terceiro painel de bordo permite a análise da evolução dos registros criminais referentes a um período de quatro anos. Campos para seleção da região, AISP, DP e classificação dos registros permitem a parametrização dos resultados exibidos no *dashboard*. Comparando-se áreas de abrangência diferentes (Estado, região, AISP ou DP) é possível observar se a evolução nos registros apresenta tendência similar entre as áreas, por exemplo.

4. Considerações Finais

Observou-se que o projeto SegRioDeJaneiro desenvolvido neste trabalho facilita uma melhor compreensão dos dados de índices de criminalidade do ISP-RJ – usados no projeto como estudo de caso –, pois originalmente estes dados têm sua apresentação disponibilizada através de planilhas segmentadas por mês. O SegRioDeJaneiro, por outro lado, permitiu organizar, agregar e visualizar as informações de várias formas, com agilidade e utilizando uma interface mais amigável, com possibilidade de ser acessada e manipulada por usuários em geral. Acredita-se que o modelo pode ser usado para qualquer Estado, com as adaptações necessárias.

Como proposta de trabalho futuro apresenta-se como fundamental proceder análise de testes de aceitação. Questões como o quanto foi facilitada a compreensão dos índices criminais são cruciais para a consolidação do SegRioDeJaneiro enquanto ferramenta.

É importante ressaltar que algumas dificuldades foram enfrentadas para o desenvolvimento deste trabalho, a saber: a necessidade de adequação na formatação da estrutura dos dados originais; e a pouca disponibilidade de materiais de apoio ao entendimento funcional dos *plugins* usados para criação dos painéis de bordo.

Adicionalmente, iniciativas como esta podem contribuir para a construção de “cidades inteligentes”, pois a mesma disponibiliza um instrumento de empoderamento do cidadão para acompanhamento e cobrança por resultados das políticas de segurança pública, e que também pode ser usada por gestores no apoio às suas decisões.

Referências

- BIZER, C.; HEATH, T.; IDEHEN, K.; BERNERS--LEE, T. Linked data on the web (LDOW2008). In: **Proceedings of the 17th international conference on World Wide Web**. ACM, 2008. p. 1265-1266.
- BRASIL. Casa Civil. **Lei 12.527/11**. 2011. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm>. Acesso em: 03 jul. 2013.
- BICO, F. C.; TRINDADE, L. N.; CARACCILO, R. A. R.; PAIVA JR, R. J. S.; PERES, S. M. Legibilidade em Dados Abertos: uma Experiência com os Dados da Câmara Municipal de São Paulo. In: VIII Simpósio Brasileiro de Sistemas de Informação, 2012, São Paulo. **Anais do VIII Simpósio Brasileiro de Sistemas de Informação**: Trilhas Técnicas, 2012. v. 1. p. 678-689.
- KIMBALL, R. *et al.* **The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses**. John Wiley & Sons, New York, 1998.
- SOARES, L. E. **Segurança pública no Brasil contemporâneo**. Disponível em: <http://www.nuso.org/upload/articulos/p6-6_1.pdf>. Acesso em: 01 jul. 2013.
- WASELFISZ, J. J. **Mapa da violência 2013 - Mortes Matadas por Armas de Fogo**. Centro Brasileiro de Estudos Latinos-Americanos, 2013.
- W3C BRASIL. **Manual dos dados abertos: governo**. Traduzido e adaptado de opendatamanual.org. Disponível em: <<http://www.w3c.br/divulgacao/pdf/gov-web.pdf>>. Acesso em: 19 jun. 2013.