

**SECRETARIA ESPECIAL DE POLÍTICAS PARA AS MULHERES**

**ONU MULHERES**

**Sistema de Informação da SPM**

**Produto 5 - Incorporação dos indicadores internos**

Jaqueline Juvencio de Sá

# Produto 5 – Incorporação dos indicadores internos ao painel e das bases de dados relacionadas aos indicadores internos do Sistema de Informações (DATASPM)

---

**Contrato n. 028/2015**

**Objeto da contratação:** Aperfeiçoamento de aplicativos eletrônicos de gestão da informação relacionados à implementação e acompanhamento de políticas, apoiando a implementação de procedimentos e mecanismos que aumentem o potencial de uso de ferramentas de gestão de informações no processo de coordenação e articulação relacionados à SPM.

Valor do produto: R\$ 5.200,00 (cinco mil e duzentos reais)

Data de entrega: 16/05/2016

Nome do consultor: Jaqueline Juvencio de Sá

Nome do supervisor: Filipe Hagen E. da Silva

De Sá, Jaqueline Juvencio

Título do produto: Incorporação dos indicadores internos / 2016.

Total de folhas: 15

Supervisor: Filipe Hagen E. da Silva

Secretaria Especial de Políticas para as Mulheres

Palavras-chave: *Business intelligence, OLAP, dados.*

## SUMÁRIO

RESUMO.....	5
1. INTRODUÇÃO.....	6
1.1 Contexto e importância da consultoria.....	6
1.2 Contexto e importância do Produto.....	6
2. DESENVOLVIMENTO.....	7
2.1 Ferramenta DATASPM.....	7
2.2 Instalação.....	11
3. CONCLUSÃO.....	13
Abreviações.....	14
ANEXOS.....	15

## RESUMO

Após o desenvolvimento do banco de dados com bases importantes sobre políticas públicas para as mulheres, outro passo importante foi a customização de uma ferramenta de *Business intelligence* para análise desses dados. A tecnologia usada é conhecida como Processamento Analítico On-Line (OLAP), onde os dados são estruturados em cubos. Ela trabalha em conjunto com Data Warehouse implementado no produto anterior dessa consultoria. Assim, a ferramenta permite que dados brutos sejam transformados em informações úteis a fim de analisar o negócio, ainda que o usuário não tenha conhecimento em banco de dados.

Palavras-Chave: *Business intelligence*, *OLAP*, dados.

## 1. INTRODUÇÃO

- 1.1 Esta consultoria terá como objetivo geral apoiar a SPM no aperfeiçoamento de aplicativos eletrônicos de gestão da informação relacionados à implementação e acompanhamento de políticas, apoiando a implementação de procedimentos e mecanismos que aumentem o potencial de uso de ferramentas de gestão de informações no processo de coordenação e articulação relacionados à SPM. Nesse sentido, este Produto específico tem como objetivo a customização de uma ferramenta que permita às/aos gestoras/es analisar grandes volumes de dados, realizando consultas em um alto nível de abstração.
- 1.2 Contexto e importância do Produto: No produto anterior dessa consultoria (Produto 4), foi implementado um servidor de banco de dados capaz de armazenar as informações sobre políticas públicas para as mulheres. Dessa forma, é de grande importância a criação de uma ferramenta de análise e visualização dos dados armazenados naquele banco. Uma ferramenta que permite ao gestor buscar, agregar, filtrar e visualizar graficamente as informações presentes no banco de dados da SPM. Essas funções são normalmente denominadas de “*business intelligence*” (BI) na área de tecnologia da Informação.

## 2. DESENVOLVIMENTO

Uma vez existindo o banco de dados centralizado na SPM, o próximo passo é desenvolver a capacidade de se extrair informações úteis dessa base. As aplicações em “*Business Intelligence*” oferecem métodos de apresentação das informações em visualizações interativas. Assim, o usuário pode gerar gráficos e painéis (*dashboards*) que apresentam os indicadores de desempenho e de resultados de maneira intuitiva, como por exemplo, aplicando filtros, agregando e desagregando informações.

No Sistema de Informação da SPM, que está sendo implementado por meio desse projeto, foram utilizadas tecnologias modernas de BI que permitem uma grande capacidade de análise de dados sem necessidade de intermediação de um setor de tecnologia informação. A tecnologia OLAP (*On-Line Analytical Processing*) é o estado da arte em análise de dados organizacionais (*business analysis*), que tem como funções básicas a visualização multidimensional, exploração, rotação e vários modos de visualização dos dados (ex: tabelas, arquivos .csv, gráficos de barras, colunas, pizza, entre outros).

O OLAP e o Data Warehouse trabalham juntos. Enquanto o DW armazena as informações de forma eficiente, o OLAP deve recuperá-las com a mesma eficiência, porém com muita rapidez. O OLAP não é uma forma de armazenamento de dados, mas se utiliza do armazenamento para apresentar as informações e permitir uma interface com o usuário. No modelo OLAP, os dados são estruturados em cubos. Um cubo é uma estrutura de dados que agrega as medidas pelos níveis e hierarquias de cada uma das dimensões. As dimensões de um cubo podem ser organizadas de forma hierárquica para facilitar a agregação e desagregação de informações.

Por exemplo, a dimensão “local” pode ter três níveis de hierarquia: 1- Região; 2- Estado; e 3 – Município. O usuário pode consultar dados a uma região, em seguida escolher informações de um Estado específico e, por fim, selecionar somente valores de um município. O sistema automaticamente entende como agregar e desagregar os dados dentro da hierarquia. Também podem ser aplicados filtros, permitindo selecionar, por exemplo, todas as informações de uma determinada região onde o número de habitantes é maior que 100 mil.

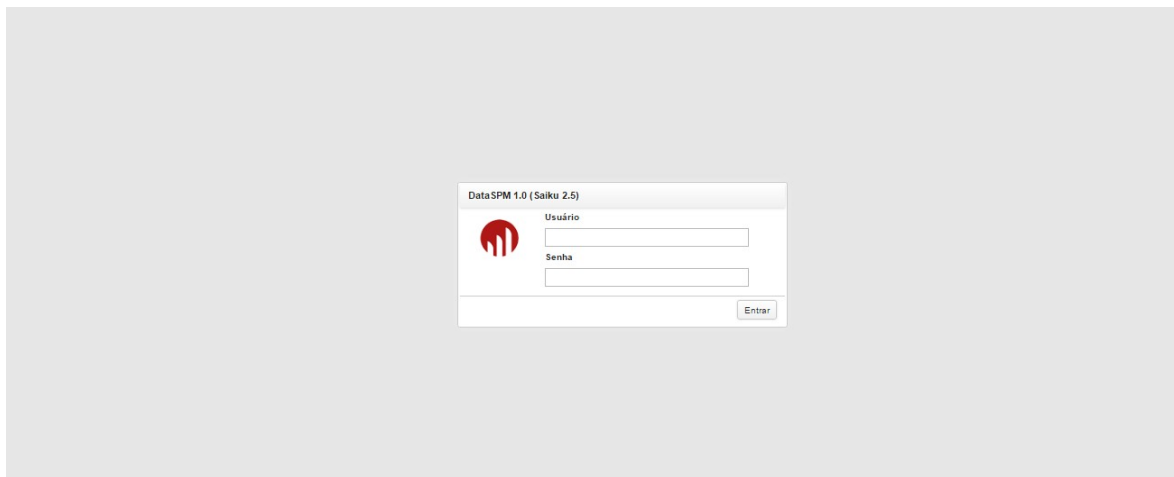
### 2.1 Ferramenta DATASPM

Neste projeto, a ferramenta utilizada para suprir a necessidade da visualização das informações do banco de dados é a *Saiku Analytics*. O código dessa ferramenta foi produzido pela *Meteorite Consulting* e é um aplicativo de BI que utiliza a tecnologia OLAP. O sistema OLAP adotado foi produzido pela Pentaho, chamado de Mondrian. Ele utiliza um banco de dados relacional como *backend* para armazenamento de dados. Ele é compatível com o banco PostgreSQL, um banco de dados gratuito que é utilizado pelo DATASPM. Uma das vantagens do Mondrian é que ele é uma ferramenta de código aberto e que é facilmente inserida em outras aplicações. Ao adotarmos o Mondrian no projeto do DATASPM,

garantimos que todos os cubos montados poderão ser acessados não só neste sistema, mas também por ferramentas de terceiros que sejam compatíveis com o mesmo padrão. Isso porque o Mondrian utiliza um padrão conhecido e aberto para armazenamento das configurações de cubos.

Uma vez que o DATASPM é uma aplicação web, ela deve ser acessada por meio do navegador. Ao acessar o sistema, a tela inicial será exibida, conforme a figura 1.

*Figura 1: Tela inicial do DATASPM*



As figuras abaixo ilustram a interface gráfica do DATASPM já com a customização realizada com base no Saiku.

*Figura 2: Interface gráfica do DATASPM*

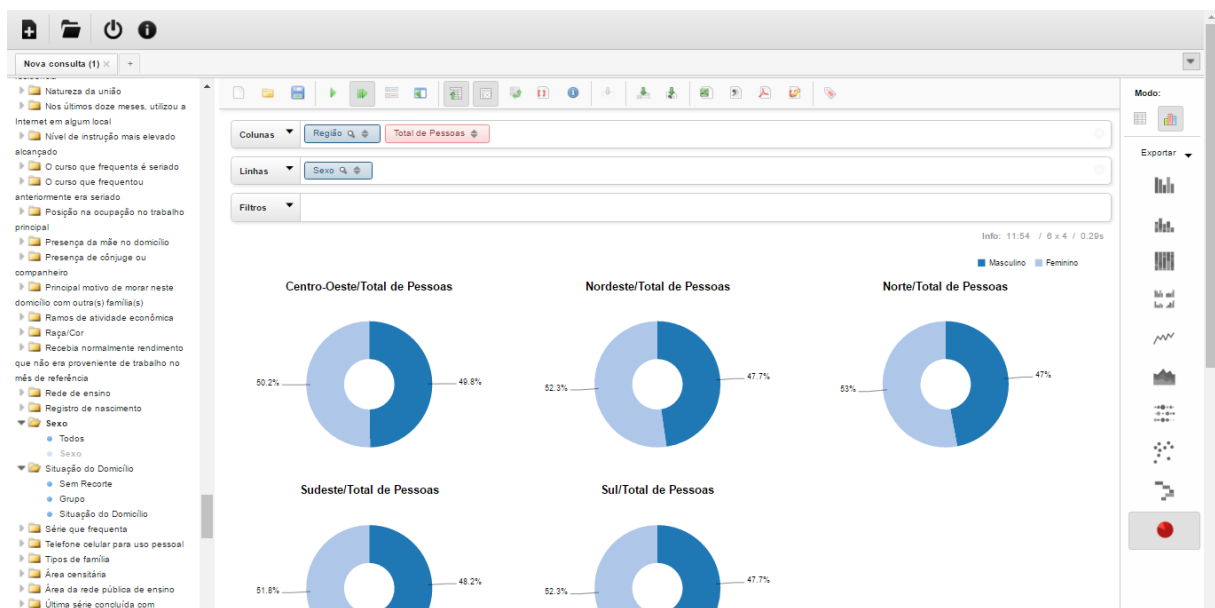




Figura 3: Exemplo de consulta montada

The screenshot shows the DATASPM software interface. On the left is a tree view of available variables. The main area displays a query configuration with columns and lines, and a resulting data table.

**Query Configuration:**

- Colunas:** Total de Domicílio, Região
- Linhas:** Grupo, Situação do Domicílio
- Filtros:** (Empty)

**Data Table:**

		Total de Domicílio				
Grupo	Situação do Domicílio	Centro-Oeste	Nordeste	Norte	Sudeste	Sul
Rural	Aglomerado rural de extensão urbana	4.457	87.836	25.306	144.138	35.224
	Aglomerado rural, isolado, povoado	59.288	1.064.141	191.017	90.745	12.275
	Aglomerado rural, isolado, núcleo	13.770	24.888	5.302		
	Aglomerado rural, isolado, outros aglomerados		72.418	7.597	19.972	
	Zona rural exclusive aglomerado rural	450.218	3.146.633	951.394	1.635.453	1.401.630
Urbana	Cidade ou vila, área urbanizada	4.549.574	12.986.564	3.779.579	26.848.435	8.459.662
	Cidade ou vila, área não urbanizada	79.926	166.243	43.051	394.710	362.820
	Área urbana isolada	7.926	117.994		233.170	57.358

Para que a ferramenta DATASPM seja criada, é necessária a implementação de um “esquema”, que consiste em um arquivo contendo a descrição de cada base em uma notação que é compreendida pela ferramenta. Esse esquema define, por exemplo, que uma coluna raça/cor em uma base do IBGE a cor “Branca” seja representada pelo código “2” e a cor “Preta” pelo código “4”. Dessa maneira, quando o usuário criar uma tabulação a partir desta base, será exibido na ferramenta os dados rotulados com as palavras “Branca” ou “Preta” ao invés dos códigos numéricos presentes na base. No esquema, também pode ser definido que a população de raça “Negra” consiste na combinação das populações de cor “Preta e Parda”, permitindo ao DATASPM calcular automaticamente dados agregando-se essas duas categorias quando desejado.

A elaboração desse esquema é complexa, devido ao grande número de variáveis existente em cada base. A base da PNAD (IBGE), por exemplo, contém mais de 700 variáveis, sendo que todas são representadas por codificações numéricas que precisam ser mapeadas na ferramenta para os seus significados em termos textuais, conforme o dicionário de variáveis disponível em cada base. Além disso, é preciso incluir no esquema as equações que definem cada métrica que pode ser calculada no cubo, tais como “Taxa de Alfabetização” e “Razão de Sexo”.

O esquema construído neste projeto para alimentar o DATASPM apresenta os cubos agregados em duas categorias: “Históricos”, que correspondem às bases de períodos anteriores até o atual, conforme indicadas no Produto 4; e “Bases Atuais”, que correspondem

à versão mais atual de cada base e assim tornando a consulta mais rápida. Essa divisão acontece devido ao grande volume de dados que são processados. Os cubos de histórico apresentam a dimensão de tempo, permitindo a criação de gráficos de séries temporais.

Os seguintes cubos estão disponíveis no DATASPM:

- **IBGE Censo Demográfico Domicílio;**
- **IBGE Censo Demográfico Pessoas;**
- **IBGE PNAD Domicílio;**
- **IBGE PNAD Pessoas;**
- **INEP Censo Superior Alunos;**
- **INEP Censo Superior Cursos;**
- **INEP Censo Superior Docente;**
- **INEP Censo Superior IES;**
- **MS SIM;**
- **MS VIVA SINAN;**
- **MTE RAIS;**
- **TSE Candidatos;**
- **TSE Eleitorados;**
- **TSE Resultados.**

O DATASPM foi customizado para ser compatível com a ferramenta utilizada pela SEPIR que, devido ao grande volume de dados, os cubos da categoria “Históricos” são reduzidos, nem sempre contendo todas as variáveis disponíveis na versão atual da base. Porém, no DATASPM tanto o cubo “Histórico” quanto o cubo “Bases atuais” possuem as mesmas dimensões e métricas. Da mesma maneira aconteceu com o banco de dados criado no Produto 4, as tabelas de Histórico e Bases atuais estão com todas as variáveis disponíveis pelo órgão produtor. Por isso, é necessária uma infraestrutura com capacidade de armazenar o banco de dados da forma que foi criado. Mais detalhes sobre o banco de dados estão descritos no referido produto.

Cabe ressaltar que as variáveis das bases que foram utilizadas neste produto e que são em comum com as variáveis das bases utilizadas pela SEPIR, foram sinalizadas nos arquivos de “esquemas” para que as ferramentas sejam integradas quando necessário.

A documentação para o desenvolvedor pode ser encontrada no material disponível na web sobre as ferramentas Saiku e Mondrian, as quais são as bases para o DATASPM. Os

endereços a seguir oferecem documentação sobre a construção e manutenção de cubos e a manutenção do Saiku.

- Definição de Cubos:

<http://mondrian.pentaho.com/documentation/schema.php>

- Saiku: <http://wiki.meteorite.bi/display/SAIK/Saiku>

## 2.2 Instalação

O DATASPM é uma aplicação Java Web, o que significa dizer que ela é executada em um servidor de aplicações Java e a interação com o usuário é feita por meio do navegador de Internet. Ela foi desenvolvida para funcionar no servidor Apache Tomcat. Caso a máquina em que será instalada a aplicação não tenha uma versão do Java 7 instalada, é preciso que o Java seja instalado antes que o servidor seja executado. O Tomcat requer uma variável de ambiente "JRE\_HOME" que aponte para o diretório em que o Java está instalado. No Linux, normalmente pode-se configurar simplesmente a variável "JAVA\_HOME=/usr" no profile de inicialização do ambiente.

O DATASPM está configurado para utilizar o banco de dados a partir da versão 8.4 (porém, recomendado utilizar a versão mais recente). Essa configuração aponta para um banco na mesma máquina que executa o Tomcat e na porta padrão. A configuração padrão da instalação utiliza o super-usuário:

- **Usuário:** postgres
- **Senha:** admin

Essas configurações podem ser alteradas posteriormente. Após a instalação do PostgreSQL, é preciso criar o banco de dados chamado "dbspm" utilizando o padrão de codificação UTF-8. Em seguida, restaurar o backup disponível como anexo do Produto 4. O PostgreSQL oferece uma ferramenta gráfica por meio da qual todas as configurações podem ser realizadas, chamada pgAdmin III.

Uma vez que as configurações acima tenham sido realizadas, o sistema está pronto para ser executado. Basta que o servidor Tomcat seja inicializado. Isso é feito executando-se o script "startup" no diretório "bin" que se encontra na pasta do Tomcat. As configurações do banco podem ser alteradas nos arquivos de configuração da aplicação "saiku", que se encontra no Tomcat (no diretório webapps). Os seguintes arquivos precisam ser alterados se as configurações do banco forem modificadas:

- webapps/saiku/WEB-INF/applicationContextspring-security-jdbc.xml
- webapps/saiku/WEB-INF/classes/saikudatasources/agregado
- webapps/saiku/WEB-INF/classes/saikudatasources/microdados

As configurações do Apache Tomcat, tais como porta de acesso e memória, podem ser realizadas seguindo-se a especificação da documentação do Tomcat disponibilizada pela Apache Foundation (<http://tomcat.apache.org/tomcat-9.0-doc/index.html>). Para se melhorar o desempenho da aplicação, é importante configurar os parâmetros -xms e -mx do Java no Tomcat, de acordo com a disponibilidade de memória RAM. A configuração deve ser feita no arquivo "bin/setenv.bat".

A configuração de senhas para acesso ao sistema é realizada diretamente na tabela **tbl\_users** dentro do PostgreSQL. É necessário ainda inserir na **tbl\_user\_authorities** o tipo de usuário.

### 3. CONCLUSÃO

O DATASPM, por ser uma ferramenta de BI, permite que pessoas sem conhecimento específico sobre banco de dados possam analisar grandes volumes de dados e transformá-los em informações significativas e úteis. É importante salientar que o DATASPM é apenas uma ferramenta. A sua eficácia na gestão de políticas públicas dependerá do seu uso sistematizado, do incentivo e orientação ao uso de seus recursos. O usuário deve conhecer as bases de dados a serem trabalhadas por intermédio do sistema, suas limitações e problemas de consistência e precisa compreender os princípios básicos da estatística descritiva.

## ABREVIATÖES

BI	<i>Business intelligence</i>
DATASPM	Customização da ferramenta Saiku Analytics
DATSEPPIR	Customização da ferramenta Saiku Analytics
DW	<i>Data Warehouse</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
IES	Instituição de Ensino Superior
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
MS	Ministério da Saúde
OLAP	Processamento Analítico On-Line
PNAD	Pesquisa Nacional por Amostra de Domicílios
RAIS	Relação Anual de Informações Sociais
SIM	Sistema de Informações de Mortalidade
SINAN	Sistema de Informações de Agravos de Notificações
SPM	Secretaria Especial de Políticas para as Mulheres
TSE	Tribuna Superior Eleitoral
VIVA	Vigilância de violências e acidentes

## ANEXOS

1. Arquivo contendo pacote *saiku.war* – aplicação que implementa a infraestrutura de análise de dados em tempo real (on-line analytical processing – OLAP);
2. Arquivo contendo pacote *analise.war* – sistema DATASPM, que oferece uma interface de tabulação e análise OLAP por meio da navegação em cubos, tendo o “saiku” como infraestrutura base.