

**SECRETARIA ESPECIAL DE POLÍTICAS PARA AS MULHERES  
SPM**

**ONU MULHERES**

**Sistema de Informação da SPM**

**Produto 3 - Avaliação das bases a serem incorporadas  
no banco de dados da SPM**

Jaqueline Juvencio de Sá

# Produto 3 – Avaliação das bases a serem incorporadas no banco de dados a ser criado para a SPM, a partir do DATASEPPIR.

---

**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: 25/01/2016

Nome da consultora: Jaqueline Juvencio de Sá

Nome do supervisor: Filipe Hagen E. da Silva

De Sá, Jaqueline Juvencio

Título do produto: Avaliação das bases a serem incorporadas no banco de dados da SPM / 2016.

Total de folhas: 20

Supervisor: Filipe Hagen E. da Silva

Secretaria Especial de Políticas para as Mulheres

Palavras-chave: Data Warehouse, Sistema de Informação, base de 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 Organização das Fontes de Dados.....	7
2.2 Nível de Estruturação das Fontes de dados.....	9
2.3 Mapeamento das Fontes de Dados.....	11
3. CONCLUSÃO.....	18
4. BIBLIOGRAFIA.....	19
Tabelas, Gráficos, Mapas e Figuras.....	20
Abreviações.....	21

## RESUMO

Para tornar possível a construção dos painéis de indicadores e das outras ferramentas do Sistema de Informação, que ainda não foram implementadas, serão descritas nesse relatório as etapas de extração, transformação e armazenamento das bases de dados necessárias ao funcionamento do sistema, a partir do modelo de estruturação das fontes de dados elaborado pela SEPPIR.

Palavras-Chave: *Data Warehouse*, Sistema de Informação, base de 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 suprir as demandas de informações das bases de dados necessárias ao projeto.

1.2 Contexto e importância do Produto: O objetivo desse produto é detalhar o processo de aquisição, tratamento e organização dos dados que vão servir de insumo na construção e funcionamento do sistema de informação de políticas públicas da SPM.

## 2. DESENVOLVIMENTO

### 2.1 Organização das Fontes de Dados

No Sistema de Informação da SPM serão utilizadas diferentes fontes de dados, que são, principalmente, pesquisas socioeconômicas, registros administrativos e informações sobre serviços públicos oferecidos à sociedade. Isso possibilita que sejam construídas diferentes visões de um mesmo aspecto, possibilitando uma maior riqueza analítica. Por outro lado, pode criar diversas dificuldades de organização e uso dessas, como problemas de operacionalização do processo de extração, transformação e carga de dados. Isso acontece devido a diferentes graus de estruturação das organizações responsáveis pela produção e divulgação das fontes de dados (Detalhes sobre esse assunto estão na seção “Nível de Estruturação das Fontes de dados”).

As bases de dados usadas nesse projeto, na grande maioria, estão descentralizadas, organizadas e estruturadas de formas diferentes, portanto, precisam ser tratadas para que ocorra uma boa integração entre elas. A solução de armazenamento e processamento de dados, para contornar essas dificuldades, é conhecida como um *Data Warehouse (DW)* que possibilita estocar um grande volume de informações.

A primeira etapa da solução de DW é a aquisição dos dados, que consiste em selecionar as fontes de dados e realizar o processo de ETL (Extração, Transformação e Carga). O ETL (do acrônimo inglês *Extract, Transform, Load*) é o processo de extração dos dados de fontes externas, o tratamento por meio de algoritmo de limpeza e armazenamento no *Data Warehouse*. Nessa fase também é importante armazenar os metadados, que são informações que possibilitam o entendimento dos dados que serão utilizados.

Na segunda fase, acontece o armazenamento das informações processadas na fase de aquisição. Para que o Sistema de Informação da SPM esteja em consonância com o sistema implantado pela SEPPIR, utilizaremos como referência o modelo apresentado por esta Secretaria, que de acordo com Carvano (2013), foi dividido em três partes distintas:

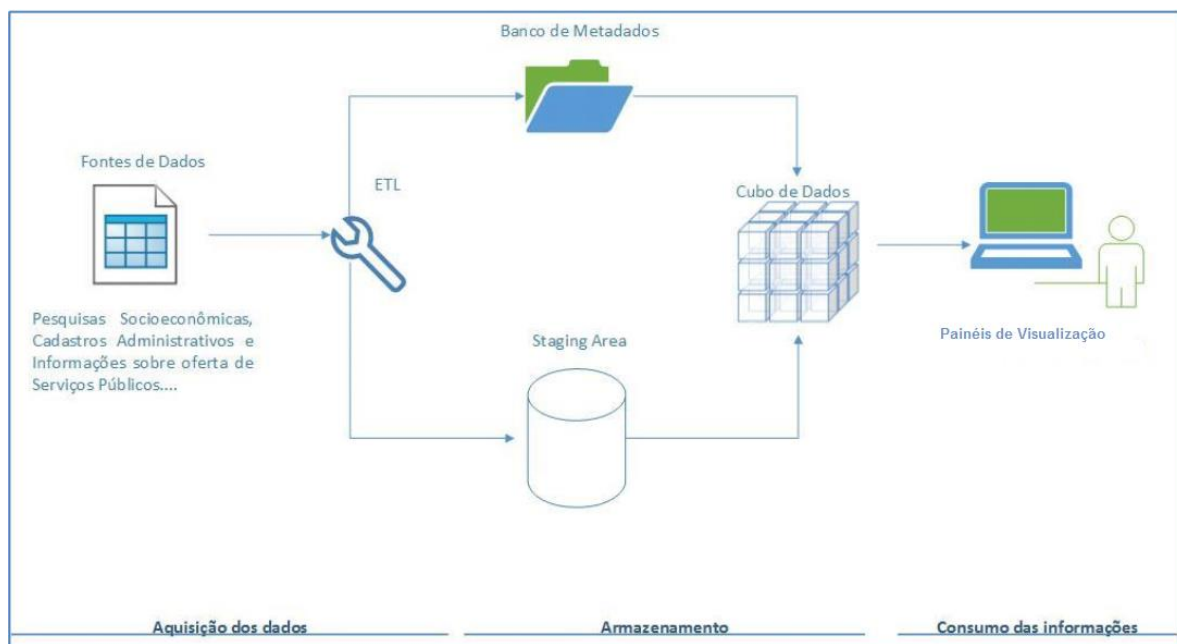
**Staging Areas** – Local de armazenamento dos dados brutos, para serem utilizados entre as etapas de armazenamento. Detalhes dessa etapa estarão descritos no Produto 6 – Processo automático para extração de dados.

**Banco de Metadados** – Armazena os metadados das bases, ou seja, um dicionário das variáveis utilizadas na base e serve de fonte de consulta para utilização dessas.

**Cubo de dados** – Aqui as informações apresentam um formato multidimensional, por meio de uma ou mais dimensões<sup>1</sup> e uma ou mais métricas<sup>2</sup>.

A fase final de um processo de *Data Warehousing* (construção, manutenção e acesso de um *Data Warehouse*), é o consumo das informações, onde são construídas as visualizações de indicadores que vão auxiliar a análise das informações em questão.

Figura 1: Processo de Data Warehousing



<sup>1</sup> Dimensões são coleções de dados descritivos distintos que irão classificar, definir e esclarecer as informações.

<sup>2</sup> Métricas são as medidas brutas e de simples composição, como por exemplo: valores e quantidades.



## 2.2 Nível de Estruturação das Fontes de dados

Como vimos anteriormente, a maior parte das fontes de dados, usadas até o momento, estão estruturadas de formas diferentes. Portanto, será adotado nesse projeto, o modelo de classificação das bases de dados utilizado pela SEPPIR, com os seguintes critérios:

1. Existência de um órgão responsável pela produção, consolidação e divulgação das informações.
2. Disponibilização das informações em formato digital, organizados a partir de uma estrutura onde cada registro corresponde a uma linha de dados e onde os campos estão organizados posicionalmente ou por algum tipo de caractere separador (ponto e vírgula, espaço em branco, pipes, etc).
3. Disponibilização dos metadados das bases, na forma de dicionários de variáveis, que possibilite o entendimento do conteúdo.

A partir desses critérios, podemos dividir as bases em dois grupos: Alta Estruturação/Organização – as bases de dados atendem a todos os critérios listados; e Baixa Estruturação/Organização – as bases de dados não atendem a pelo menos um dos critérios.

*Tabela 1 - Alta Estruturação/Organização*

Base de dados	Órgão responsável	Disponibilização dos Microdados	Disponibilização dos Metadados
<b>Pesquisa Nacional por Amostra de Domicílios (PNAD)</b>	IBGE	SIM	SIM
<b>Censo da Educação Superior</b>	INEP	SIM	SIM
<b>Estatísticas eleitorais do Tribunal Superior Eleitoral</b>	TSE	SIM	SIM
<b>Sistema de Informações de Mortalidade – SIM</b>	MS	SIM	SIM

<b>Censo demográfico</b>	IBGE	SIM	SIM
<b>Sistema de Vigilância de Violências e Acidentes (VIVA)</b>	MS	SIM	SIM

*Tabela 2 - Baixa Estruturação/Organização*

<b>Base de dados</b>	<b>Órgão responsável</b>	<b>Disponibilização dos Microdados</b>	<b>Disponibilização dos Metadados</b>
<b>Sistema de Informações de Agravos de Notificações - SINAN</b>	MS	Em análise	Em análise
<b>Central de Atendimento à Mulher – Ligue 180</b>	SPM	Em análise	Em análise
<b>Disque Direitos Humanos (Disque 100)</b>	SDH	Em análise	Em análise

Portanto, quanto mais consolidada é a política pública que dá origem às informações, mais alta é a organização e estruturação das bases. Deve-se mencionar, também, que os critérios apresentados não levam em consideração a qualidade das informações existentes nas bases de dados analisadas.

Outro ponto importante é que nos últimos anos o Governo vem se esforçando para melhorar a qualidade das bases de dados, por meio de padronizações das informações produzidas, tais como, a arquitetura e-PING (Padrão de Interoperabilidade de Governo Eletrônico), que define um conjunto mínimo de princípios, políticas e especificações técnicas para a utilização da tecnologia da informação e a INDA (Infraestrutura Nacional de Dados Abertos), que define um conjunto de padrões, procedimentos e mecanismos de controle para a disseminação e compartilhamento de dados e informações públicas, de forma complementar ao e-PING.

### **2.3 Mapeamento das Fontes de Dados**

Conforme descrito no Produto 2 dessa consultoria, foram escolhidos temas de trabalho e o rol de indicadores para cada um deles. A partir disso, a SPM fez a indicação das bases de

dados que seriam utilizadas no sistema de informação, nessa primeira versão, como apresentado abaixo:

#### **a) Pesquisa Nacional por Amostra de Domicílios (PNAD)**

**Instituição Responsável:** Instituto Brasileiro de Geografia e Estatística

**Tema:** Família (População), Grupos populacionais específicos (População), Trabalho, Trabalho remunerado (Trabalho), Outras formas de trabalho (Trabalho), Rendimento, despesa e consumo, Estatísticas multidomínio, População, Condições de vida, pobreza e desigualdade (Estatísticas multidomínio), Educação, Nupcialidade (População), Características gerais da população (População), Habitação, Sociedade da Informação (Estatísticas multidomínio), Saúde, Componentes da dinâmica demográfica e estatísticas vitais (População), Proteção social, Justiça e segurança.

**Periodicidade:** Anual

**Repositórios:**

<http://www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad2014/microdados.shtm>

**Unidade(s) de Coleta:** Domicílio e Pessoa

**Técnica(s) de Coleta:** CAPI - Entrevista pessoal assistida por computador

**Nível de Divulgação:** Brasil, Grandes Regiões, Unidades da Federação e nove Regiões Metropolitanas (Belém, Recife, Fortaleza, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo, Curitiba e Porto Alegre).

**Objetivo:** Produzir informações para análise do desenvolvimento socioeconômico do domicílio.

**Principais Variáveis Investigadas:** As principais variáveis investigadas são as características de habitação, demográficas gerais (sexo, raça/cor, idade), migração, fecundidade feminina, educação e trabalho e rendimento.

#### **b) Censo da Educação Superior**

**Instituição Responsável:** Diretoria de Estatísticas Educacionais (DEED)/INEP/MEC

**Tema:** Educação Superior

**Periodicidade:** Anual

**Repositórios:** <ftp://ftp.inep.gov.br/microdados>

**Unidade(s) de Coleta:** Instituição, Recursos Humanos (Função Docente, Funcionário Técnico-Administrativo), Curso, Candidato, Vaga, Ingresso, Matrícula, Concluinte, Função Docente.

**Técnica(s) de Coleta:** Censitária

**Nível de Divulgação:** País, Regiões, Estados, Municípios e Instituição de Ensino Superior.

**Objetivo:** Produzir informações detalhadas sobre a situação atual e as grandes tendências da educação superior no país.

**Principais Variáveis Investigadas:** Graduação presencial, graduação à distância, cursos sequenciais de formação específica, cursos sequenciais de formação específica, cursos sequenciais de complementação de estudos, pessoal técnico-administrativo, pessoal docente, dados financeiros, infraestrutura, instalações, equipamentos e outros recursos institucionais e área total, área construída e área de laboratórios (em m<sup>2</sup>) – nas IES federais.

### **c) Estatísticas eleitorais do Tribunal Superior Eleitoral**

**Instituição Responsável:** Tribunal Superior Eleitoral

**Tema:** Eleições brasileiras

**Periodicidade:** Bianual

**Repositórios:** [www.tse.jus.br/eleicoes/repositorio-de-dados-eleitorais](http://www.tse.jus.br/eleicoes/repositorio-de-dados-eleitorais);

<http://www.tse.jus.br/eleicoes/estatisticas>

**Unidade(s) de Coleta:** Eleitorado, Candidatos e resultados das eleições

**Técnica(s) de Coleta:** Dados dos tribunais regionais eleitorais

**Nível de Divulgação:** Nacional

**Objetivo:** Divulgar os dados brutos das eleições desde 1945 para pesquisadores que tem interesse em analisar os dados de eleitorados, candidatura, resultados e prestação de contas das eleições.

**Principais Variáveis Investigadas:** Estatística de candidatura (cargo, grau de instrução, partido e outros), Resultados (candidatos eleitos entre outros) e estatísticas de eleitorado.

### **d) Sistema de Informações de Agravos de Notificações - SINAN**

**Instituição Responsável:** Coordenação Geral de Doenças Transmissíveis/Secretaria de Vigilância à Saúde (SVS)/Ministério da Saúde (MS).

**Tema:** Saúde

**Periodicidade:** Contínuo

**Repositórios:** Não disponível

**Unidade(s) de Coleta:** Episódio do Agravado

**Técnica(s) de Coleta:** Não se aplica

**Nível de Divulgação:** Nacional, com detalhamento no nível estadual e municipal, segundo o local de residência ou de notificação.

**Objetivo:** Registrar, transmitir e disseminar os dados gerados pelo Sistema de Vigilância Epidemiológica, para apoiar a análise de informações de vigilância epidemiológicas das doenças de notificação compulsória. Também é objeto de notificação, os casos suspeitos ou confirmados de violências domésticas, sexuais e outras violências.

**Principais Variáveis Investigadas:** Notificação (mês, ano, estado e município), paciente (sexo, escolaridade, raça, urbano e rural, estado e município de residência), agravado (data dos primeiros sintomas, município, estado e país da infecção. Outras variáveis são coletadas, dependendo do tipo de agravado.).

#### **e) Sistema de Informações de Mortalidade – SIM**

**Instituição Responsável:** Secretaria de Vigilância à Saúde (SVS)/Ministério da Saúde

**Tema:** Saúde

**Periodicidade:** Contínuo

**Repositórios:** <ftp://ftp.datasus.gov.br/dissemin/publicos/SIM/>

**Unidade(s) de Coleta:** Óbito

**Técnica(s) de Coleta:** Registros Administrativos

**Nível de Divulgação:** Nacional, com detalhamento no nível estadual e municipal. Em uma versão mais completa permite obter informações a níveis intramunicipais.

**Objetivo:** Obter dados sobre mortalidade para auxiliar os gestores públicos na análise de situação, planejamento e avaliação das ações dos programas na área de saúde.

**Principais Variáveis Investigadas:** Certificado de óbito (número e tipo de óbito); Falecido (local de nascimento, data de nascimento, idade, sexo, raça/cor, estado civil, escolaridade, ocupação, distrito e município de residência); Óbito (data, local, estabelecimento de saúde, município); Mãe (em caso de morte fetal ou de criança com menos de 1 ano); Causa primária da morte (CID), causas secundárias; e prováveis circunstâncias de morte não natural.

#### **f) Central de Atendimento à Mulher – Ligue 180**

**Instituição Responsável:** Secretaria de Enfrentamento à Violência contra as Mulheres/ SPM

**Tema:** Denúncia de violência contra mulher

**Periodicidade:** Semestral

**Repositórios:** Não disponível

**Unidade(s) de Coleta:** Relatos e Denúncias de Violência contra as Mulheres

**Técnica(s) de Coleta:** Central de atendimento disque 180

**Nível de Divulgação:** Balanço do Disque 180 – nacional. Banco de dados 180- Interno

**Objetivo:** A Central de Atendimento à Mulher, criada em 2005 pela Secretaria de Políticas para as Mulheres, é um serviço de utilidade pública que funciona 24 horas por dia, nos 7 dias da semana, orientando as mulheres em situação de violência sobre seus direitos, informando a respeito dos serviços disponíveis para atendê-las e, desde março de 2014, recebendo e encaminhando denúncias de casos de violência contra as mulheres para as autoridades competentes. A partir das ligações feitas à Central, constitui-se uma base de dados, preenchidas pelas atendentes, com o perfil das mulheres atendidas e dos tipos de atendimento.

**Principais Variáveis Investigadas:** Tipo de crime (ameaça; assédio moral - trabalho; assédio sexual - trabalho; atentado violento ao pudor; calúnia; cárcere privado; dano emocional/diminuição da autoestima; difamação; estupro; exploração sexual; homicídio; injúria; lesão corporal grave; lesão corporal gravíssima; lesão corporal leve; perseguições; tentativa de homicídio; violência patrimonial); Tipo de violência (cárcere privado; violência física; violência moral; violência patrimonial; violência psicológica; violência sexual); Outras variáveis relacionadas à violência e sua caracterização (trata-se de caso de violência doméstica; relação com a/o agressor/a; tempo da relação; há quanto tempo ocorre a violência; frequência com que ocorre a violência; qual o risco que se percebe; se depende financeiramente da/o agressor/a); Variáveis de descrição do perfil da pessoa atendida (sexo; escolaridade; idade; cor ou raça).

#### **g) Disque Direitos Humanos (Disque 100)**

**Instituição Responsável:** Ouvidoria Nacional de Direitos Humanos/ Secretaria de Direitos Humanos

**Tema:** Registro de Denúncias de atentados a direitos humanos

**Periodicidade:** Mensal

**Repositórios:** Não disponível

**Unidade(s) de Coleta:** Relatos e Denúncias de violação de direitos humanos

**Técnica(s) de Coleta:** Registro administrativo

**Nível de Divulgação:** Nacional

**Objetivo:** O Disque Direitos Humanos – Disque 100 é um serviço de atendimento telefônico gratuito, que funciona 24 horas por dia, nos 7 dias da semana, que recebe denúncias e oferece orientações a respeito de violações de direitos humanos. A base de dados resultantes dos registros dos atendimentos pelos/as atendentes permite o monitoramento dos atendimentos e da efetividade da proteção oferecida à/ao cidadã/o. Os dados são coletados para Brasil, Unidades Federativas e municípios.

**Principais Variáveis Investigadas:** Cadastro do demandante (nome, nome social, nome da mãe, data de nascimento, idade, faixa etária, estado civil, ocupação, telefone, e-mail, CEP, endereço, UF, sexo, escolaridade, orientação sexual, identidade de gênero, cor ou raça, pessoa com deficiência/doença mental, pessoa em situação de rua); Cadastro da vítima; Cadastro da pessoa suspeita; Cadastro de pessoa jurídica; Tipo de violência (negligência, violência psicológica, violência sexual, exploração do trabalho infantil, discriminação, abuso financeiro e econômico, violência institucional, outras violações); e Relação da/o agressor/a com a vítima.

#### **h) Censo demográfico**

**Instituição Responsável:** Instituto Brasileiro de Geografia e Estatística – IBGE

**Tema:** População

**Periodicidade:** Decenal

**Repositórios:**

[http://www.ibge.gov.br/home/estatistica/populacao/censo2010/resultados\\_gerais\\_amostra/resultados\\_gerais\\_amostra\\_tab\\_uf\\_microdados.shtm](http://www.ibge.gov.br/home/estatistica/populacao/censo2010/resultados_gerais_amostra/resultados_gerais_amostra_tab_uf_microdados.shtm)

**Unidade(s) de Coleta:** Domicílio; Pessoa.

**Técnicas de Coleta:** Censo e por Amostragem Probabilística

**Nível de Divulgação:** Nacional

**Objetivo:** Contar os habitantes do território e identificar as suas características, obtendo informações para a definição de políticas públicas e tomada de decisões. Constitui a única fonte de referência sobre a situação de vida da população nos municípios e em seus recortes internos, como distritos, bairros e localidades, rurais ou urbanas.

**Principais Variáveis Investigadas:** Situação Urbana e Rural; Características do Domicílio; Emigração internacional, Sexo, Idade, Cor ou Raça, Etnia ou Povo a que pertence e Língua falada só para indígenas, Religião ou Culto, Registro de Nascimento, Deficiência Física ou Mental, Migração interna e Imigração internacional, Educação, Deslocamento para estudo, Nupcialidade, Características do Trabalho e do Rendimento, Deslocamento para trabalho, Fecundidade e Mortalidade.

#### **i) Sistema de Vigilância de Violências e Acidentes (VIVA)**

**Instituição Responsável:** Ministério da Saúde

**Tema:** Violências e Acidentes

**Periodicidade:** Anual

**Repositórios:** [http://svs.aids.gov.br/bases\\_vigitel\\_viva/dispbd\\_vivacont.php](http://svs.aids.gov.br/bases_vigitel_viva/dispbd_vivacont.php)

**Unidade(s) de Coleta:** Violência doméstica, sexual, e/ou outras violências interpessoais e autoprovocadas.

**Técnicas de Coleta:** Registro Administrativo

**Nível de Divulgação:** Nacional

**Objetivo:** Coletar dados e gerar informações sobre violências e acidentes para subsidiar políticas em saúde pública, direcionadas a esses agravos, além de possibilitar um diagnóstico mais sensível da situação, especialmente, de casos não reconhecidos e não incorporados aos sistemas de informações sobre mortalidade (SIM) e internações (SIH).

**Principais Variáveis Investigadas:** Tipo de atendimento (acidentes, violências); Pacientes (sexo, idade, cor ou raça, deficiência, localização do domicílio, parte do corpo atingida); Ocorrência (intencionalidade, tipos de acidente/violência/agressão, relação com o trabalho, uso de álcool ou drogas, local da ocorrência, evolução do atendimento); Autor/a (relação com a/o autor/a, sexo da/o autor/a).

Outras fontes de dados (registros administrativos da SPM) foram utilizadas na construção da ferramenta de visualização dos painéis, tais como listagem de serviços especializados de atendimento à mulher em situação de violência (ex: Centro Especializado de Atendimento à Mulher) e listagem de mecanismos institucionais (ex: Conselho Estadual dos Direitos da Mulher e Secretaria de Estado de Políticas para Mulheres).



### 3. CONCLUSÃO

As fontes de dados apresentadas nesse produto fazem parte de uma amostra para a implementação do Projeto-piloto, contudo passarão por aprovação interna e poderão sofrer alterações durante a construção do Sistema de Informação da SPM.

## 4. BIBLIOGRAFIA

**Carvano, Luiz Marcelo Ferreira** – Design do sistema de monitoramento – Área de Métodos Quantitativos -Setembro, 2013.

**SPM** -Relatório Anual Socioeconômico da Mulher–RASEAM- [http://www.spm.gov.br/central-de-conteudos/publicacoes/publicacoes/2015/livro-raseam\\_completo.pdf](http://www.spm.gov.br/central-de-conteudos/publicacoes/publicacoes/2015/livro-raseam_completo.pdf), Março, 2015.

## ÍNDICE DE TABELAS, GRÁFICOS, MAPAS E FIGURAS

Figura 1	Processo de Data Warehousing
Tabela 1	Alta Estruturação/Organização
Tabela 2	Baixa Estruturação/Organização

## ABREVIATÖES

CEP	Código de Endereçamento Postal
CID	Classificação Internacional de Doenças
DEEP	Diretoria de Estatísticas Educacionais
DW	<i>Data Warehouse</i>
E-PING	Padrão de Interoperabilidade de Governo Eletrônico
IBGE	Instituto Brasileiro de Geografia e Estatística
IES	Instituição de Ensino Superior
INDA	Infraestrutura Nacional de Dados Abertos
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
MEC	Ministério da Educação
MS	Ministério da Saúde
PNAD	Pesquisa Nacional por Amostra de Domicílios
SEPPIR	Secretaria de Políticas de Promoção da Igualdade Racial
SIM	Sistema de Informações de Mortalidade
SINAM	Sistema de Informações de Agravos de Notificações
SVS	Secretaria de Vigilância à Saúde
UF	Unidade Federativa