



## **IMPLANTAÇÃO DE PROCEDIMENTOS DE GESTÃO DE QUALIDADE DE DADOS DE CADASTRO TÉCNICO NA DIRETORIA METROPOLITANA**

### **RESUMO**

O BDG – Banco de Dados Geográficos da Sabesp representa todos os ativos georreferenciados da empresa e serve de base para o SIGNOS – Sistema de Informações Geográficas no Saneamento e uma série de outras aplicações na empresa. A gestão e governança desses dados é fundamental para a qualidade das análises efetuadas com esses dados espaciais, que por sua vez são indispensáveis para uma série de processos de negócio da empresa.

**PALAVRAS-CHAVE:** SIG, Cadastro, Governança de dados

### **INTRODUÇÃO**

O SIGNOS (Sistema de Informações Geográficas no Saneamento), aplicação de SIG (Sistema de Informações Geográficas) corporativa da Sabesp, foi implantado em dezembro de 2005, na área da Diretoria Metropolitana. Foi o produto final do Projeto GIS, cuja história na companhia remonta a meados da década de 90 e que, se começou a se tornar realidade em 2003, com sua inclusão no escopo das atividades da segunda fase do Programa de Despoluição do Tietê - PDT.

Em termos funcionais, o SIGNOS foi desenvolvido na plataforma GE Smallworld®, tendo tido diversas funcionalidades especialmente desenvolvidas e customizadas para o negócio da Sabesp, permitindo a edição gráfica do cadastro de redes (substituindo as ferramentas CAD) consultas espaciais, geração de mapas temáticos, relatórios, exportações de dados, etc.

Com ele a Sabesp, na área da Diretoria Metropolitana, passou a contar com um banco de dados geográfico (BDG) contendo informações georreferenciadas de seus ativos de negócio, tais como os cadastros de redes de distribuição de água, de redes de coleta de esgotos, adutoras, coletores-tronco, interceptores, instalações, reservatórios, faixas de servidão e domínio, além do seu cadastro comercial de clientes. Todas estas informações foram lançadas em uma base cartográfica georreferenciada gerada a partir de um levantamento aerofotogramétrico, com uma área de restituição de mais de 2.500 Km<sup>2</sup>, executado em 2003 especialmente para esta finalidade.

O BDG é alimentado por diversos processos: empreendimentos, vendas, engenharia de operação de água e esgotos e etc. em suas 261 tabelas (objetos), cada uma delas com até 50 atributos. Vários desses objetos são alimentados por processos diferentes, operados por atores diferentes, tanto próprios quanto terceiros. Os dados precisam ser verificados em termos de aderência das informações ao modelo de dados do BDG, topologia, completude, precisão e checagens de negócio. Isso é feito por softwares especializados que executam testes e publicam seus resultados em arquivos e relatórios, de modo que as áreas da empresa que são responsáveis pelos dados possam atuar nos mesmos.

### **OBJETIVOS**

Os procedimentos de gestão de qualidade de dados de cadastro técnico tem como objetivo a melhoria contínua das informações do BDG, propiciando dados melhores para uso nos processos de negócio da empresa que deles dependam.

### **METODOLOGIA UTILIZADA**

Através de benchmarking em instituições de manutenção de bases de dados espaciais, como o IBGE, que gere a INDE – Infraestrutura Nacional de Dados Espaciais e a EMPLASA, que gere a IDE-SP – Infraestrutura de Dados Espaciais do Estado de São Paulo, e algumas empresas de utilities, que também possuem BDGs tão complexos quanto o da Sabesp, como CPFL e COMGAS, pode-se observar quais as melhores práticas do segmento.



Inicialmente foram criados os documentos de modelo de dados do BDG, para os segmentos de rede e cartografia. Esses documentos consolidam o modelo de dados do BDG, elencando seus objetos, atributos, joins, regras de preenchimento e de consistência de dados. O produto desta etapa foi a catalogação dos dados do BDG, sendo que ao todo foram elencados 180 objetos para rede e 81 para cartografia, totalizando 261 objetos (vide Figura 1).

Campo	Nome Interno	Tipo	Obs
ID			
Localização	localizacao	GEOM(Ponto)	S
Operado Por	responsavel	ENUM	S
Proprietario	proprietario	ENUM	
Tipo de Tampa	tipo_tampa	ENUM	S
Estado	estado	ENUM	
Caldeirado?	caldeirado?	ENUM	
Dimensões da Tampa (mm)	dimensoes	ASC(2)	
Comprimento da Caixa (m)	comprimento_caixa	FLD	
Altura da Caixa (m)	altura_caixa	FLD	
Largura da Caixa (m)	largura_caixa	FLD	
Diâmetro da Chamimé (mm)	diâmetro_chamime	INT	
Comprimento da Chamimé (m)	comprimento_chamime	FLD	
Drenagem	drenagem	ENUM	
Escada?	escada?	ENUM	
Tubo	tubo	ENUM	
Estado de Conservação	estado_conservacao	ENUM	
Origem da Informação	origem_informacao	ASC(6)	
Diâmetro do Conduto	diâmetro_conduto	INT	
Data de Instalação	data_instalacao	DATA	
Data de Entrada em Operação	data_entrada_operacao	DATA	
Data de Abandono/Desativação	data_abandono	DATA	
Documentos Externos	documentos_externos	JOIN	
Observações	observacoes	TEXT	
Referência	referencia	ASC(6)	
Código SEM	codigo_sem	S_SEM	
Coord UTM N (m)	coord_utm_n	DS_DOUBLE	
Coord UTM E (m)	coord_utm_e	DS_DOUBLE	
Histórico Carregamento	re_historico_carregamento	JOIN	
Inserido por	inserido_por	ASC(6)	
Data de Inserção	data_insercao	DATA	
Atualizado Por	atualizado_por	ASC(6)	
Última Atualização	ultima_atualizacao	DATA	

Figura 1: Documento gerado: Modelo de Dados

A seguir foi criado o documento “**Plano de Testes de Controle de Qualidade de Dados Espaciais**”, com o objetivo de descrever os testes e verificações (aderência das informações ao modelo de dados do BDG, topologia, completude, precisão e checagens de negócio) de cada objeto, dentre os mais significativos e relevantes para o negócio.

**Teste: TUB-02**

**Objetos:**  
Tubulação de Adução (a) (RE-06) e Tubulação de Distribuição (a) (RE-07)

**Descrição do Teste:**  
Objetos de Tipo de Seção Circular sem atributo "Diâmetro" preenchido, igual a zero ou com valores considerados pouco usuais.

**Gravidade do problema:**  
**Alta**

**Impacto do problema:**  
É um dos atributos mais importantes para caracterização do ativo, sua falta prejudica ações operacionais e análises.

**Solução:**  
Preenchimento do atributo, com base em documentação ou verificação em campo

**Critério de seleção**  
(Tubulação de distribuição e Tubulação de adução) onde tipo de seção = "circular", estado = "em operação" e diâmetro = "nulo" ou "vazio" ou = "0" ou < "32 (mm)" e seção = "circular"

**Forma de Cálculo:**  
Extensão calculada dos objetos dividida pela extensão total calculada com estado = "em operação"

**Fluxos responsáveis pelo teste:**  
OCORR\_EM\_DIAM\_TUB\_ADUCAO.fmw e OCORR\_EM\_DIAM\_TUB\_DISTR, localizado em: \\10.7.186.80\mapa\Geoinformacao\Areas\_interna\Geoportal\fmw\fluxos

Figura 2: Documento gerado: Plano de Testes de Controle de Qualidade de Dados Espaciais

Utilizou-se o software O FME Database Fixed, da empresa Safe Software, que é uma ferramenta de ETL (*Extract, Transform, Load*) que consiste em um construtor de fluxos de tarefas customizáveis que pode ser



aplicado à qualquer base de dados, como é o caso do BDG da Sabesp. Os fluxos tem a função de fazer verificações, análises, consistências e batimentos de dados, publicando seus resultados em qualquer formato.

Tomando como exemplo o objeto “Tubulação de distribuição de água”, para o qual foram concebidos oito testes:

- i. Geometrias inválidas (poliarcos);
- ii. Objetos de Tipo de Seção Circular sem atributo “*Diâmetro*” preenchido, igual a zero ou com valores considerados pouco usuais.
- iii. Objetos com atributo “*Material*” desconhecido ou preenchido com valores fora da lista de materiais ou com grafia diferente da definida (ACO, DEFOFO, FG, FOFO, PEAD, PRFV, AMIANTO, PVC, CE e DESCONHECIDO)
- iv. Objetos com atributo “*Data de Instalação*” não preenchida ou futura
- v. Objetos sem atributo “*Comprimento real*” preenchido
- vi. Objetos com sobreposição geométrica, total ou parcial
- vii. Objetos com sobreposição geométrica, total ou parcial com objeto de cartografia “*margem de rio*”
- viii. Objetos com seção não-circular sem as dimensões da seção

Cada teste requereu a construção de um fluxo de tarefas no FME (vide Figura 3) que a partir da BDG Sabesp, fez as checagens e gerou um planilha em Excel com os registros com problemas (vide Figura 4).

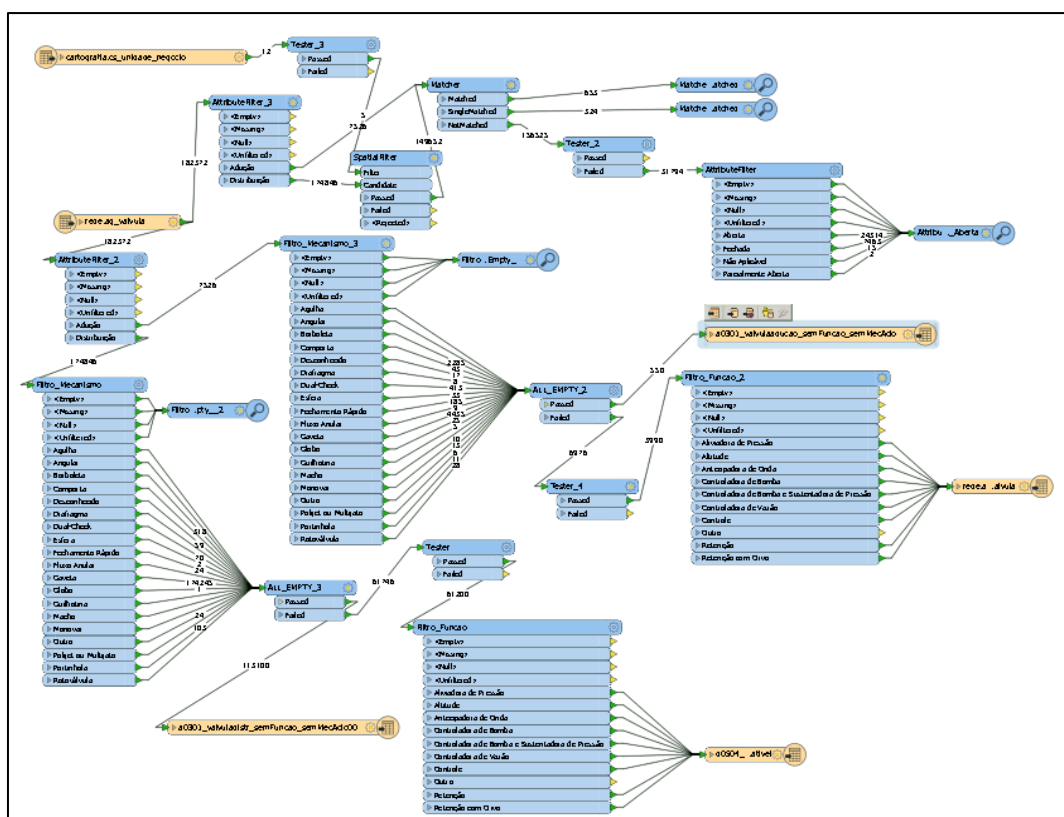


Figura 3: Exemplo de Fluxo de Tarefas para checagem de dados feito no FME



1	sigla	id	estado	diâmetro	comprimento_real	comprimento_calculado	material	_overlaps
2	MO	569306134	Em operação	75	0	0.278567765543682	FOFO	1
3	MO	569305991	Em operação	75	0	0.458802789878178	FOFO	1
4	MO	569306025	Em operação	100	0	0.756108457828637	FOFO	1
5	MO	615491267	Em operação	100		121.852.619.107.398	PVC	1
6	MO	7107991	Em operação	100		454.634.193.610.643	PVC	1
7	MO	615491049	Em operação	100		355.802.538.686.059	PVC	1
8	MO	7112833	Em operação	75		455.873.918.789.276	FOFO	1
9	MO	7113169	Em operação	75		408.107.106.039.579	FOFO	1
10	MO	7107943	Em operação	100		495.244.394.213.604	PVC	1
11	MO	7108129	Em operação	100		491.290.392.741.401	PVC	1
12	MO	7107796	Em operação	200		47.465.285.209.298	FOFO	1
13	MO	7103782	Em operação	32		170.799.578.307.601	PEAD	1
14	MO	7103029	Em operação	32		100.088.530.541.206	PEAD	1
15	MO	7103017	Em operação	32		276.162.180.069.401	PEAD	1
16	MO	7103806	Em operação	100		235.413.475.790.071	PVC	1
17	MO	615491426	Em operação	100		223.289.203.709.378	PVC	1
18	MO	171831534	Em operação	100		443.234.404.089.293	PVC	1
19	MO	615491334	Em operação	100		144.228.291.260.765	PVC	1
20	MO	615491314	Em operação	100		11.900.733.813.515	PVC	1
21	MO	615491116	Em operação	100		281.436.493.724.606	PVC	1
22	MO	615491311	Em operação	100		274.149.254.981.766	PVC	1

Figura 4: Exemplo de Resultado do Fluxo, em Excel

Ao todo foram criados 62 testes, sendo 21 para objetos de cartografia, 26 para objetos de água, 13 para objetos de esgoto e 2 para cadastro de ligações comerciais. Todos os fluxos foram documentados e consolidados em um documento.

Os testes são agendados para execução mensal no servidor e seus resultados publicados automaticamente. A Figura 5 traz um resumo do fluxo do processo:

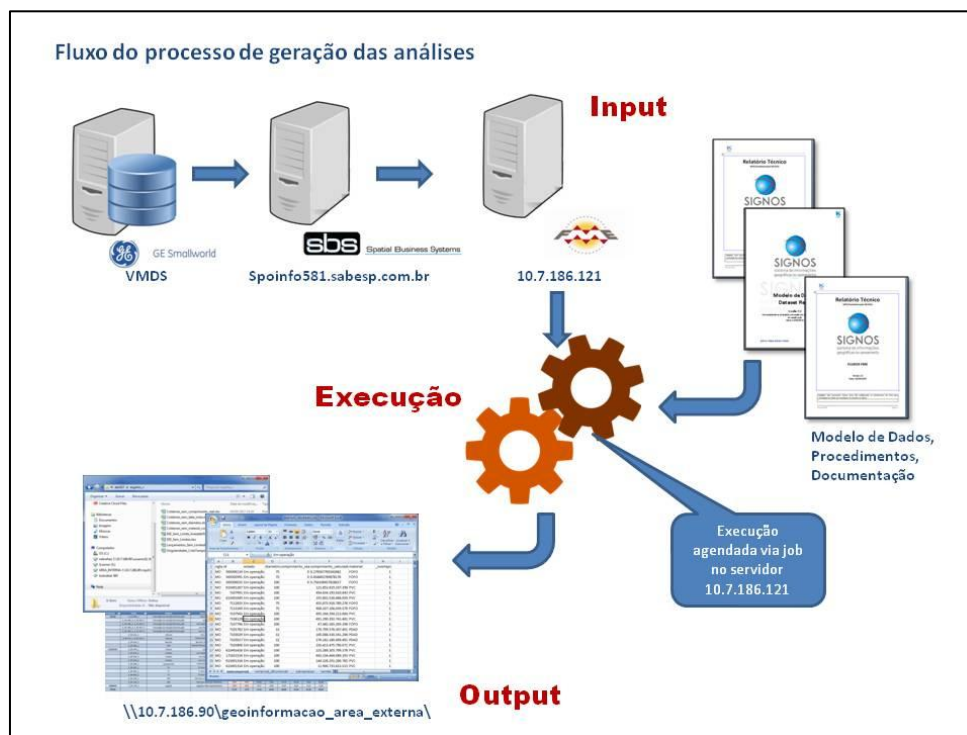


Figura 5: Resumo do fluxo de trabalho



## RESULTADOS OBTIDOS

Além dos documentos criados (Modelo de Dados e Plano de Testes de Controle de Qualidade de Dados Espaciais) foi criado um portal para publicação do resultado das análises na Intranet Sabesp (vide Figura 6). Dessa forma, todos os usuários da BDG podem acessar e tomar conhecimento tanto da porcentagem de reprovação de cada teste (separados por Unidade de Negócio) quanto acessar os documentos gerados (planilhas) com os registros com problemas (Figura 7)

Processo	Objeto	Teste	Descrição	Ano/Mês	% MA	% MT	% MC	% MN	% MS	% MO
CARTOGRAFIA	Eixo de Loquadouro...	EXK-02	Interage com Margem de Rio...	2018_03	0.00	0.00	0.00	0.00	0.00	0.00
CARTOGRAFIA	Eixo de Loquadouro...	EXK-03	Sem Atributos MaxEq MinEq...	2018_03	0.00	0.00	0.00	0.00	0.00	0.00
CARTOGRAFIA	Eixo de Loquadouro...	EXK-04	Sem Bairro...	2018_03	0.00	0.00	0.00	0.00	0.00	0.00
CARTOGRAFIA	Loquadouro	LOG-01	Loquadouro Sem Eixo...	2018_03	0.00	0.00	0.00	0.00	0.00	0.00
CARTOGRAFIA	Loquadouro	LOG-02	Loquadouro Sem Anotação...	2018_03	0.00	0.00	0.00	0.00	0.00	0.00
CARTOGRAFIA	Margem de Rio...	MIRG-01	Geometria Linha...	2018_03	0.00	0.00	0.00	0.00	0.00	0.00
CARTOGRAFIA	Parque	PQU-01	Irregularidade em Atributos...	2018_03	0.00	0.00	0.00	0.00	0.00	0.00
CARTOGRAFIA	Praca...	PRC-01	Irregularidade em Atributos...	2018_03	0.00	0.00	0.00	0.00	0.00	0.00
CARTOGRAFIA	Praca...	PRC-02	Sem Bairro...	2018_03	0.00	0.00	0.00	0.00	0.00	0.00
CARTOGRAFIA	Quadra	QDR-01	Quadra Sem Código do Município...	2018_03	0.00	0.00	0.00	0.00	0.00	0.00
CARTOGRAFIA	Quadra	QDR-02	Quadra Muito Pequena...	2018_03	0.00	0.00	0.00	0.00	0.00	0.00
CARTOGRAFIA	Quadra	QDR-03	Quadra Sem Marco Zero...	2018_03	0.00	0.00	0.00	0.00	0.00	0.00
CARTOGRAFIA	Quadra	QDR-04	Quadra com Geometria Irregular...	2018_03	0.00	0.00	0.00	0.00	0.00	0.00
CARTOGRAFIA	Quadra	QDR-05	Quadra Sem Anotação Principal...	2018_03	0.00	0.00	0.00	0.00	0.00	0.00
ESGOTO	Coletor...	CLT-01	sem Diâmetro...	2018_03	0.00	0.43	0.00	0.18	0.01	3.23
ESGOTO	Coletor...	CLT-02	com Material Desconhecido...	2018_03	0.00	0.00	0.00	0.01	0.04	0.01
ESGOTO	Coletor...	CLT-03	sem Data de Instalação...	2018_03	0.00	21.41	0.02	0.91	2.72	9.20
ESGOTO	Coletor...	CLT-04	sem Comprimento Real...	2018_03	0.00	1.21	0.00	8.68	0.05	4.27
ESGOTO	Coletor...	CLT-05	Diferença entre Compr Real e Calculado	2018_03	0.00	0.12	0.01	0.05	0.00	0.01
ESGOTO	Coletor...	CLT-06	Declividade Irregular...	2018_03	0.00	8.48	55.47	70.94	71.46	72.51
ESGOTO	Coletor...	CLT-07	interage com Margem de Rio...	2018_03	0.00	0.00	0.02	0.02	0.01	0.04
ESGOTO	Coletor...	CLT-09	Não Circular sem Dimensões de Seção	2018_03	0.00	0.00	0.00	0.00	0.00	0.00
ESGOTO	EEE	EEE-01	EEE Sem Lim de Propri ou de Influência...	2018_03	0.00	36.00	0.00	100.00	1.95	100.00
ESGOTO	Lançamento	LCT-01	Lançamento Sem Limite de Contribuição...	2018_03	0.00	0.00	19.93	98.28	0.71	80.24
ESGOTO	PI	PIE-01	PI Sem Cota de Tampa...	2018_03	0.00	0.00	0.00	2.32	9.72	1.61
ESGOTO	PV	PV-01	PV Sem Cota de Tampa...	2018_03	0.00	0.48	0.00	6.81	7.89	5.41
ESGOTO	TL	TL-01	TL Sem Cota de Tampa...	2018_03	0.00	0.00	0.00	5.59	13.01	2.51
VENDAS	Liqação	LIG-01	Liqações Não Especializadas...	2018_03	0.00	0.00	0.08	0.14	0.39	1.22
VENDAS	Liqação	LIG-02	Liqação Sem Endereço...	2018_03	0.00	0.00	0.18	1.58	1.65	0.98

Figura 6: Página na Intranet com a publicação do resultado de cada fluxo, por Unidade de Negócio.



Figura 7: Página na Intranet com a publicação dos arquivos para correção em xls.



## RECOMENDAÇÕES

Nesta era de *big data*, os dados são provavelmente o maior patrimônio de uma empresa. Sem eles, não há gestão e possivelmente nem empresa. Em se tratando de uma empresa de saneamento, onde 80% dos dados são georreferenciáveis, estamos falando em boa parte de dados espaciais. Esses dados vem de todo lugar, produzidos a todo o momento, por um grande número de atores, em uma dinâmica impossível de ser controlada. Logo, uma empresa de saneamento precisa forçosamente implantar uma política de governança de dados espaciais, que consiga minimamente organizar sua BDG, criar regras, procedimentos, padrões, metadados e controles de qualidade, publicar, disseminar essa política na corporação e principalmente capacitar a força de trabalho no que tange a seus conceitos.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. Safe Software - FME Database Fixed Manual
2. <http://www.idesp.sp.gov.br/>
3. <http://www.inde.gov.br/>
4. ABRAHÃO, Nagip – Governança de Dados em Saneamento – disponível em <https://www.linkedin.com/pulse/governan%C3%A7a-de-dados-em-gis-uma-necessidade-para-o-nagip-abrahao/>