

Considerações sobre a Produção da Base territorial do Censo de 2010

Rodrigo de Campos Macedo¹
Antonio Guarda²
Miriam Mattos da Silva Barbuda³
Wolney Cogoy de Menezes⁴

IBGE – Coordenação de Estruturas Territoriais
88010-300 Florianópolis SC
¹ rodrigo.macedo@ibge.gov.br
² antonio.guarda@ibge.gov.br
³ miriam.barbuda@ibge.gov.br
⁴ wolney.menezes@ibge.gov.br

Resumo: A base territorial é o arquivo gráfico que apóia o Censo e outras pesquisas amostrais realizadas pelo IBGE (Instituto Brasileiro de Geografia e Estatística). Ela inclui as malhas viária, hidrográfica, municipal, distrital, subdistrital, setorial (unidade territorial de coleta de dados), além de feições como quadras, faces (um dos lados de uma quadra), unidades domiciliares e algumas edificações. O IBGE a atualizou para a realização do Censo 2010 e continua atualizando-a a uma taxa de 10% ao ano. Para integrar esta base com os dados tabulares, é necessário associar os endereços às feições gráficas. O objetivo do presente texto é analisar alguns aspectos relacionados à produção da base territorial, incluindo considerações sobre a geocodificação e a associação com o cadastro de endereços.

Palavras chaves: geocodificação, cadastro, multi-resolução.

Abstract: The territorial basis is the system that integrates the graphic file to alphanumeric databases that support the census and other sample surveys conducted by IBGE (Brazilian Institute of Geography and Statistics). It includes the road, hydrographic, municipal, district, sub district, sector (territorial collection data unity) networks, and features such as blocks, faces (one side of a block), households and some buildings. The IBGE updates it to carry out the 2010 Census and continues on updating it at a rate of 10% per year. In order to integrate this database with tabular data, it is necessary to associate the address to features. The purpose of this paper is to examine some aspects related to the territorial base production, including considerations about geocodification and the association with the register address cadastre.

Keywords: geocoding, cadastre, multi-resolution.

1. Introdução

A base territorial é o arquivo gráfico que apóia o censo e outras pesquisas amostrais realizadas pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Ela inclui as malhas viária, hidrográfica, municipal, distrital, subdistrital e setorial, além de feições como quadras, faces, unidades domiciliares e algumas edificações. O IBGE a atualizou para a realização do censo de 2010 e continua atualizando-a a uma taxa de 10% ao ano. Para integrar esta base com os dados tabulares, é necessário associar os endereços às feições gráficas.

O objetivo do presente texto é analisar alguns aspectos relacionados à produção da base territorial, incluindo considerações sobre a geocodificação e a associação com o cadastro de endereços. Para isto, foi relatado o histórico da base territorial no IBGE, descrito seu processo de produção e realizado um estudo de caso, visando associar a listagem de endereços e o arquivo gráfico de Blumenau-SC.

2. Conceitos necessários

Todos os conceitos aqui descritos foram extraídos de IBGE, 2008.

O setor censitário é uma unidade territorial de coleta, ou seja, uma área constituída de um conjunto de domicílios onde serão aplicados os questionários ou protocolos de coleta de dados. Também é a unidade de controle cadastral formada por área contínua, integralmente contida em área urbana ou rural, cuja dimensão e número de domicílios ou de estabelecimentos agropecuários permitam ao pesquisador cumprir suas atividades dentro do prazo estabelecido para a coleta. Seu perímetro deve respeitar os limites territoriais legalmente definidos e os estabelecidos pelo IBGE para fins estatísticos. É definido, preferencialmente, por pontos de referência estáveis e de fácil identificação no campo, de modo a evitar erros de coleta.

Os setores são classificados em situação e tipo. O código de situação identifica a localização da área do setor em relação às áreas urbana ou rural. O código de tipo classifica o setor segundo o conceito que orientou sua delimitação.

A Tabela 1 descreve os códigos de situação.

Tabela 1: Códigos de situação dos setores censitários

Situação	Descrição	
1	Setor de Área Urbanizada de Cidade ou Vila	Urbana
2	Setor de Área Não Urbanizada de Cidade ou Vila	
3	Setor de Área Urbana Isolada	
4	Setor de Aglomerado Rural de Extensão Urbana	Rural
5	Setor de Aglomerado Rural Isolado - Povoado	
6	Setor de Aglomerado Rural Isolado - Núcleo	
7	Setor de Aglomerado Rural Isolado - Lugarejo	
8	Setor de Zona Rural Exclusive Aglomerado Rural	

Agglomerado rural é a localidade situada em área legalmente definida como rural e caracterizada por um conjunto de edificações permanentes e adjacentes, formando área continuamente construída e com arruamentos reconhecíveis ou dispostos ao longo de uma via de comunicação.

A Tabela 2 descreve os códigos de tipo.

Tabela 2: Códigos de tipo dos setores censitários

Tipo	Descrição	
0	Setor Comum	Não Especial
1	Setor Especial de Aglomerado Subnormal	Especial
2	Setor Especial de Quartéis, Bases Militares, etc.	
3	Setor Especial de Alojamento, Acampamentos, etc.	
4	Setor Especial de Embarcações, Barcos, Navios, etc.	
5	Setor Especial de Penitenciárias, Colônias Penais, Presídios, Cadeias, etc.	
6	Setor de Aglomerado Rural Isolado - Núcleo	
7	Setor Especial de Asilos, Orfanatos, Conventos, Hospitais, etc	
8	Setor Especial de Projetos de Assentamentos Rurais	

A base territorial é constituída por uma estrutura de mapeamento cartográfico (topográfico e cadastral) integrada a um sistema de cadastros, compondo um ambiente informacional gráfico e alfanumérico, em constante processo de inovação, visando a incorporação das características atualizadas do território nacional, relevantes para coleta censitária e demais pesquisas estatísticas. Ela contém informações de diversas áreas (estados, municípios, bairros, aglomerados, distritos, quadras, etc.), importantes para o recenseador na realização da coleta de dados e, posteriormente, na divulgação dos resultados das pesquisas.

A base territorial possui áreas de apuração e áreas de divulgação. Áreas de apuração são as áreas geográficas delimitadas nos mapas e cadastradas para servir de unidade espacial de apuração dos dados censitários. Essas áreas são isoladas em setores censitários específicos. Áreas de divulgação são áreas legalmente definidas para a divulgação dos dados estatísticos. São consideradas áreas de divulgação: unidades da federação, municípios, distritos, regiões administrativas, subdistritos e zonas, área urbana, área rural e área urbana isolada.

A malha viária é representada graficamente apenas nos setores com situação urbana e nos aglomerados rurais de extensão urbana, ou seja, nas situações 1, 2, 3 e 4. Sua representação ocorre através de feições do tipo linha, classificadas em duas camadas distintas: eixos de logradouro e faces. Os eixos de logradouro são as representações das vias, individualizadas pelas toponímias dentro dos setores censitários. As faces são as arestas das quadras e devem necessariamente ser geocodificadas. Possuem diversos campos, dentre os quais, destacam-se o CD_FACE e o CD_GEO.

O CD_FACE é um código de três dígitos que individualiza a face dentro de uma quadra. Já o CD_GEO é um código de 21 dígitos que individualiza a face em nível nacional, pois concatena os códigos de unidade federativa, município, setor, quadra e face.

Os canteiros centrais, rotatórias, rótulas, etc. auxiliam no reconhecimento de campo, porém, estas linhas não são faces. São apenas elementos de apoio ao mapa.

3. Histórico

No primeiro censo realizado pelo IBGE¹, em 1940, a melhor escala de informação (tanto a apuração como a divulgação) era a municipal. No censo de 1950, a divulgação de informação continuou sendo a municipal, mas foram implantados os setores censitários para a apuração. A partir de 1960, houve a divulgação por situação, ou seja, urbana e rural. No censo de 1970, houve a divulgação por situação e por tipo, ou seja, os setores especiais (quartéis, asilos, favelas, mocambos, alagados, etc.).

A partir do censo de 1980, o IBGE começou a representar a Divisão Político-Administrativa (DPA) sobre um mapeamento topográfico planimétrico sistemático, visando obter maior precisão e exatidão na geometria da DPA e da localização dos setores censitários. As dificuldades do mapeamento começam a aumentar em função do processo de urbanização do país, demandando mapas urbanos de melhor qualidade que os croquis que até então vinham sendo utilizados. O pensamento de que qualquer esboço de mapa com o intuito de apenas orientar o pesquisador em campo começou a mudar, gerando-se uma necessidade de cartografia cadastral.

Advindo o censo de 1991, foi criado um banco de dados alfanumérico que permitiu o controle e codificação de movimentação das estruturas territoriais em todos os níveis. Em conjunto com este banco de dados, foi necessária a revisão e atualização das plantas cadastrais utilizadas pelo IBGE.

Já para o Censo 2000, em função das demandas de limites, da localização dos elementos cartográficos com as devidas toponímias e da subordinação administrativa, foi necessário digitalizar a base territorial. Esta digitalização propiciou a comparabilidade entre as malhas censitárias, a crítica de consistência com o banco de dados e, a melhoria na compatibilidade entre os mapeamentos urbano e rural. Com isto, houve um aprimoramento dos bancos de dados existentes e a criação de outros, tais como o Banco de Estruturas Territoriais² (BET), a Base Operacional Geográfica³ (BOG) e o Cadastro de Segmento de Logradouros⁴ (CADLOG).

1 Quinto recenseamento nacional.

2 Cujas unidades territoriais mais detalhadas são as localidades.

3 Cujas unidades territoriais mais detalhadas são os setores censitários.

4 Visa localizar todos os segmentos de logradouros de um determinado setor censitário urbano.

Em 2010, o grau de aprimoramento do censo propiciou o registro das informações ao nível do endereço, possibilitando a criação do Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE). Para isto, o primeiro passo foi integrar os mapeamentos urbano e rural, já que um estava em escala cadastral e o outro em escala topográfica. O segundo passo foi a estruturação deste mapeamento unificado em um sistema de bancos de dados geoespaciais (SISMAP¹) integrados com os bancos existentes BET e BOG com o recém-criado CNEFE.

4. Procedimentos relacionados à produção da base territorial do censo 2010

O primeiro requisito para a produção da base territorial foi conjugar as duas situações (rural e urbana). A escala mais detalhada da situação rural era 1:50.000, manipulada em Desenho Assistido por Computador (CAD) em um aplicativo desenvolvido chamado SISCART. A situação urbana apresentava escalas variadas, mas em geral mais detalhadas que as utilizadas na situação rural, manipuladas em Sistema Geográfico de Informações (GIS) e em CAD, através de um aplicativo desenvolvido chamado MSU. A conjugação das duas tipologias requisitou a integração dos dados em diferentes escalas.

Esta integração foi realizada na malha setorial de 2007, ou seja, foi necessário desenhar a malha dos setores rurais e urbanos, sendo que nos setores urbanos deveriam ser representados os eixos de logradouro, quadras, faces, hidrografia e edificações, conforme a coleta do Censo de 2007. Após esta etapa, procedeu-se com as fases de desenho, avaliação e formação de um banco de dados. A Figura 1 esquematiza as etapas envolvidas na preparação dos dados.

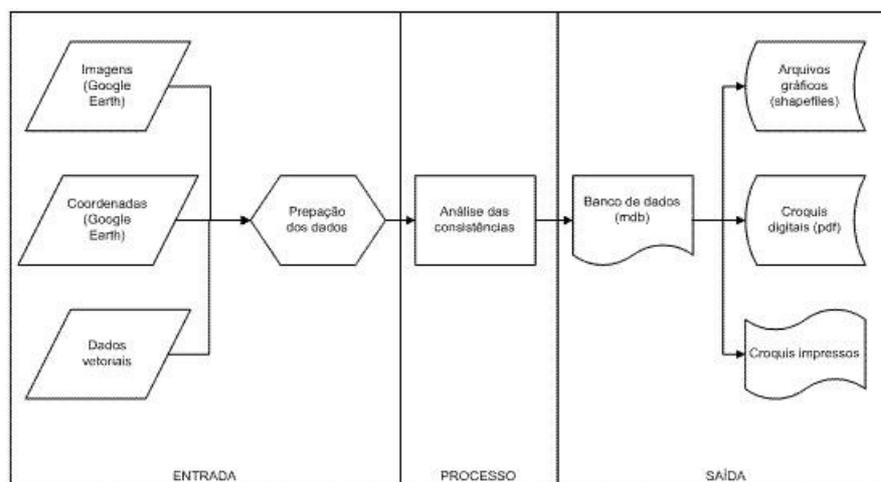


Figura 1: Esquema da produção da base territorial do censo de 2010

4.1. Preparação dos dados (ENTRADA)

Na fase de preparação dos dados (desenho) foram produzidas todas as feições necessárias à realização do censo de 2010: malhas municipal, distrital, subdistrital, setorial e viária (faces e eixos de logradouro).

As malhas municipal, distrital e subdistrital foram geradas a partir da malha setorial, através de fusão de setores. A produção da malha setorial requereu imagens e os vetores referentes aos eixos de logradouro. A produção da malha viária (eixos de logradouro e faces) requereu imagens. A codificação das faces levou em consideração as codificações anteriores (censos de 2000/2007), visando manter a comparabilidade entre as pesquisas.

A produção das malhas setorial e viária foi dividida em municípios com até 20.000 habitantes e municípios com mais de 20.000 habitantes.

1 Cuja unidade territorial mais detalhada é a face.

4.1.1. Municípios com até 20.000 habitantes

Para municípios com até 20.000 habitantes, a produção das malhas setorial foi realizada internamente, pelo IBGE. As imagens utilizadas (majoritariamente) foram as disponibilizadas através do *Google Earth*, que variam desde imagens com resolução espacial de 0,4m (*Geoeye*, por ex.) a 30m (*Thematic Mapper*, por ex.). Esta opção foi influenciada pela rápida e fácil disponibilização – características favoráveis à operacionalização do processo – porém desconsiderou alternativas existentes, tais como fotografias aéreas disponibilizadas por algumas prefeituras de Santa Catarina.

As imagens disponibilizadas no *Google Earth* requerem registro de posicionamento (georreferenciamento). As coordenadas utilizadas foram as disponibilizadas através do próprio *Google Earth*, porém devido à característica de agregação de diversos insumos de imagens, não há um erro padrão e/ou conhecido em relação ao posicionamento das coordenadas presentes nesta plataforma. Além disso, o *Google Earth* utiliza um modelo geodésico esferoidal (circunferóide), distinto dos modelos geodésicos presentes na maior parte dos *data*, em sua maioria elipsoidais. Isto requer uma avaliação e cálculo de conversões para utilizar as coordenadas presentes nesta plataforma para *data* usuais, como por exemplo, o SIRGAS2000, *datum* oficial do Brasil. As imagens foram georreferenciadas a partir da coleta de coordenadas (pontos notáveis, quando existentes) via *Google Earth* e posterior inserção destas coordenadas nas imagens a partir de um aplicativo¹. Além das dificuldades inerentes à seleção de pontos notáveis – cobertura de nuvem, dossel florestal, imagens de resolução espacial não compatível, etc. – acrescenta-se o desconhecimento da magnitude do erro posicional presente nas coordenadas obtidas via *Google Earth*.

A extração das feições de interesse foi realizada através de interpretação visual em tela (*overlay* digital), utilizando uma extensa gama de dados matriciais e vetoriais, incluindo levantamentos GPS. A oferta de dados variou de acordo com o município, mas a maior parte das imagens utilizadas foi disponibilizada através do *Google Earth*, georreferenciadas com coordenadas extraídas nesta mesma plataforma.

A pluralidade de insumos utilizados na produção da base dificulta a aderência entre as feições e certamente é um desafio aos profissionais envolvidos. Como exemplo, um setor censitário (polígono extraído a partir de uma imagem disponível no *Google Earth*), delimitado por um rio (linha extraída a partir da vetorização de uma folha topográfica 1:100.000) e um eixo de logradouro (linha obtida a partir do rastreamento GPS), podem apresentar problemas de aderência. O ajuste não é trivial, pois envolve erros posicionais nem sempre conhecidos, escalas e resoluções espaciais distintas.

4.1.2. Municípios com mais de 20.000 habitantes

Para municípios com mais de 20.000 habitantes, o IBGE contratou algumas empresas² através de licitação. Apesar de o termo de referência para a contratação deste mapeamento não ter sido analisado neste artigo, houve diversas queixas referentes aos dados entregues, principalmente em relação à falta de aderência entre as malhas setorial e viária, excessiva segmentação de faces e ausência de geocodificação ou mesmo geocodificação incorreta. Porém, estes dados foram adotados e, conseqüentemente, aprimorados.

4.2. Análise dos dados (PROCESSAMENTO)

Na fase de análise foram consolidadas todas as feições (de todo o Estado) e realizadas todas as consistências necessárias ao registro em banco de dados. Os dados dos municípios foram agrupados, formando-se um banco de dados para todo o Estado. As consistências consideradas foram as análises topológicas, espaciais e de atributos – as mesmas existentes para o registro em banco de dados. Os dados foram transferidos via FTP (Protocolo de Transferência de Arquivos) e armazenados em um banco de dados espaciais³.

4.3. Geração dos produtos (SAÍDA)

A partir do registro em banco de dados, foram gerados:

- 1 Quantum GIS.
- 2 Geoja Mapas Digitais e Soluções em Mapeamento para o trecho que engloba Santa Catarina.
- 3 Oracle Spatial 10.

- Croquis¹ – versão digital, para registro no Assistente Pessoal Digital (PDA) e versão plotada, ambos para apoio ao pesquisador;
- Dados gráficos² – versão digital, para registro no PDA, visando possibilitar as ações de atualização em campo.

Os croquis rurais foram gerados em um aplicativo chamado MSR e os croquis urbanos em um aplicativo chamado MSU.

O MSR permite selecionar Estado, município e setores rurais para geração de croqui. Além da seleção, o aplicativo recomenda a escala e o tamanho do papel. Caso não seja compatível, é possível ajustar a escala (até 1:25.000 no máximo) e o tamanho do papel. Ele utiliza as folhas topográficas (1:50.000 e 1:100.000) como “fundo” do croqui, em formato *raster*. Porém, há problemas relacionados à digitalização das folhas, posicionamento inadequado, etc.

O MSU apresenta as mesmas características do MSR, porém não é limitado a 1:25.000 e não utiliza as folhas topográficas como “fundo” do croqui. Porém, ele possui mais feições gráficas (malha viária e as faces).

Os croquis digitais (PDA) e impressos (plotados) devem ser essencialmente os mesmos, para não prejudicar o trabalho do pesquisador em campo. Além dos croquis, houve a geração de *shapfiles*. Estes foram gerados a partir do banco de dados, utilizando-se filtros de atributos e espaciais, visando garantir a consistência dos dados.

5. Estudo de Caso

Visando avaliar a geocodificação e a associação com o cadastro de endereços, foi realizado um experimento com os dados referentes ao município de Blumenau-SC.

5.1. Dados utilizados

A Tabela 3 apresenta os dados mais significativos para este experimento.

Tabela 3: Quantitativos dos setores censitários de Blumenau-SC por situação

Código de Situação dos Setores Censitários	Quantidade	
1	467	87,45%
2	30	5,62%
3	0	0,00%
4	3	0,56%
5	0	0,00%
6	0	0,00%
7	0	0,00%
8	34	6,37%
Total	534	100,00%

Os únicos setores passíveis de associação com o cadastro de endereços são os com código de situação 1, 3 e 4, representando 88,01% do quantitativo de setores do município. Isto ocorre porque não houve geocodificação nos setores de situação 2 e não há representação gráfica das faces nos demais setores.

Foram utilizadas a listagem de endereços (CNEFE) e as faces (SISMAP) de Blumenau-SC. A Tabela 4 compara os dois conjuntos de dados.

1 *Portable Document File* (.pdf).
2 *Shapefile* (.shp).

Tabela 4: Quantitativos das faces de Blumenau-SC por situação

Descrição	CNEFE	SISMAP
Faces codificadas como 000	1	9.015 (*)
Faces em setores com situação 1	12.679	12.460
Faces em setores com situação 4	3	3
Faces em setores com situação 2	500	3.099(**)
Faces em setores com situação 8	41	0(***)
Total associável	12.681 (95,90%)	12.463 (50,71%)

(*) Requer a transferência para outra camada, pois não são faces propriamente ditas. São canteiros e rotatórias, em sua maioria;

(**) Requer geocodificação;

(***) Requer a criação de algum tipo de representação gráfica.

Excepcionalmente, as faces referentes aos setores em situação 4 foram codificadas como 000, demandando geocodificação, tal como nos setores em situação 2.

5.2. Aplicativo utilizado

- i. Intergraph Geomedia Profissional 6.1.

5.3. Procedimentos

Dentre as diversas representações gráficas, o nível mais detalhado está presente na camada de faces, elegendo-a como ligação com os endereços (dado tabular). Dentre diversos campos existentes na camada de faces, foi utilizado o CD_GEO. Ele individualiza a face e está presente na listagem de endereços e no arquivo gráfico. Por este motivo, foi adotado para a associação entre a listagem de endereços e ao arquivo gráfico.

Os endereços foram agrupados por CD_GEO, ou seja, além do agrupamento das unidades visitadas, houve o agrupamento dos logradouros. As faces (tanto da listagem quanto do arquivo gráfico) foram agrupadas por situação de setor.

Após a seleção das faces associáveis, foi realizada associação 1:1 (*Join*) entre a listagem de endereços (agrupados por faces) e o arquivo gráfico, para cada situação de setor.

5.4. Resultados & Discussão

A Tabela 5 expressa as associações realizadas.

Tabela 5: Associações realizadas entre lista de endereços e dados gráficos (via CD_GEO)

	CNEFE		SISMAP	
CD_GEO associados	12.421	97,95%	12.421	99,66%
CD_GEO não associados	260	2,05%	42	0,34%

As 260 faces (CD_GEO) sem associação (listagem de endereços) provavelmente estão codificadas como 000 no arquivo gráfico. Após a geocodificação, este índice chegará aos 100%.

As 42 faces (CD_GEO) sem associação (arquivo gráfico) foram codificadas na pré-coleta e não confirmadas na coleta. Elas estão corretas e deveriam ser inseridas na listagem de endereços. Isto ocorreu por erro de interpretação do recenseador.

6. Considerações Finais e Recomendações

6.1. Considerações finais e recomendações sobre o processo de produção/manutenção da base territorial

O processo de produção/manutenção da base territorial para os censos e demais pesquisas domiciliares apresenta desafios enormes, principalmente por demandar elevado nível de detalhamento, simultaneamente com a abrangência de todo o território nacional.

Um dos maiores pontos de atenção é a escolha do insumo de imagem necessário para a extração destas feições: alta resolução espacial, abrangência nacional, características adequadas de precisão e exatidão posicional e pronta disponibilidade são requisitos complexos para serem unidos.

Especificamente para Santa Catarina, há um ortofotomosaico sendo entregue ao governo estadual. Este insumo apresenta características compatíveis com as necessidades de atualização da base territorial.

Quanto à extração das feições, recomenda-se automatizar parte do processo através de filtros e processamento digital de imagens, substituindo parte do desenho em tela por análise e correção.

6.2. Considerações finais e recomendações sobre os produtos oriundos da base territorial

Os aplicativos geradores de croquis (MSR e MSU) devem estar alinhados com a codificação da BOG (evitando redundância) e os critérios utilizados para a definição da escala e tamanho do papel devem ser divulgados. Recomenda-se que seja possível visualizar o croqui antes de sua impressão.

Em relação aos setores rurais, recomenda-se a substituição do *raster* da folha topográfica (fundo) pelos vetores da folha topográfica. O Centro de Informações de Recursos Ambientais e de Hidrometeorologia de Santa Catarina (CIRAM), em parceria com a Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina (EPAGRI), vetorizou todas as folhas topográficas. Estes vetores já estão disponíveis e registrados no banco de dados do IBGE.

6.3. Considerações finais e recomendações sobre a geocodificação e associação entre endereços e feições gráficas

Os diversos bancos de dados (CNEFE, SISMAP, BOG, BET, etc.) devem ser integrados e relacionados. Para isto, recomenda-se a padronização do CD_GEO como indexador entre as pesquisas. Além disso, recomenda-se:

- ii. Separar feições de linhas relacionadas aos elementos de apoio visual em campo – canteiros centrais, canteiros laterais, rotatória, rótulas, etc. Estas feições não são faces e podem ser integradas à camada “Outros”¹;
- iii. Geocodificar as faces ainda não codificadas, principalmente as pertencentes aos setores em situação 2 – Desta forma, torna-se possível a associação plena com a listagem de endereços nos setores com situação 1, 2, 3 e 4;
- iv. Incorporar a malha viária rural (eixos de logradouro rurais) – Desta forma, torna-se possível a associação com a listagem de endereços nos setores com situação 5, 6, 7 e 8;
- v. Incorporar as atualizações (toponímias e faces novas) realizadas pela Pesquisa Nacional de Amostras Domiciliares (PNAD).

7. Referências Bibliográficas

IBGE (Instituto Brasileiro de Geografia e Estatística). *Referências Técnicas para a Elaboração da Base Territorial – BT.* Rio de Janeiro. 2008. 24p.

1 Criada com a finalidade de apoio gráfico.