

Normalização e modelagem de dados geográficos: o caso do DER-PB

Thalles Ramon Pinheiro de Sousa ¹
Marcello Benigno Borges de Barros Filho ¹
Diego da Silva Valdevino ¹

¹ Instituto Federal de Educação, Ciência e Tecnologia da Paraíba - IFPB
Av. 1º de Maio, 720 - 58015-430 – João Pessoa - PB, Brasil
{thallesrps0, benigno.marcello, diego.valdevino}@gmail.com

Abstract. The conceptual modelling is a key part in the planning of database applications. The applications of geographic databases do not differ in their development process, taking into account also other factors, especially concerning the representation of spatial objects, abstraction of data and topological conditions between the entities of the study area. The Department of Highways of Paraíba (DER-PB), because it is a state authority, must follow standardization defined by the National Spatial Data Infrastructure (INDE), thus enabling the sharing and availability of geographic database between different user groups and public agencies, to allow maintenance and integrity of data structures. This article proposes to present methodology standardization and modelling of spatial data according to the guidelines of the INDE, using as a basis the data on highways and road works DER-PB. For the development of this work software were used: MySQL Workbench 6.3 for the construction of diagrams Entity Relationship and relational model; OMT-G Designer for the creation of the OMT-G model; and the PostgreSQL software with spatial extension PostGIS for the creation of the physical model of the BDG. Through the creation of semantic models Model ER and OMT-G proposed in this work, it was observed that they allow to organize, store and disseminate geospatial data in the DER-PB in addition to supporting the creation of new systems with geographical perspective, so incorporating aspects of multiple representations and properly adapted to the standards set by INDE.

Palavras-chave: normalization, INDE, spatial data modeling, OMT-G.

1. Introdução

Com a multiplicidade de geotecnologias existentes na atualidade é necessário que exista uma padronização dos dados geográficos, tanto na aquisição como no armazenamento para garantir o compartilhamento, interoperabilidade e disseminação, características básicas para a construção de uma Infraestrutura de Dados Espaciais (IDE). A GSDI (2009) descreve IDE como o conjunto relevante de tecnologias de base, políticas e os arranjos institucionais que facilitem a disponibilização e o acesso aos dados espaciais.

No mesmo sentido e com o objetivo de integrar os dados geoespaciais existentes nas diversas instituições brasileiras, o governo federal através do Decreto no 6.666, de 27/11/2008 instituiu a Infraestrutura Nacional de Dados Espaciais (INDE).

A INDE foi definida como:

O conjunto integrado de tecnologias; políticas; mecanismos e procedimentos de coordenação e monitoramento; padrões e acordos, necessário para facilitar e ordenar a geração, o armazenamento, o acesso, o compartilhamento, a disseminação e o uso dos dados geoespaciais de origem federal, estadual, distrital e municipal. (CONCAR, 2010).

A dificuldade de compatibilização para um padrão único de estrutura de dados geoespaciais e a preocupação em estabelecer novas normas para a cartografia digital trouxe à tona a necessidade da criação de um padrão para o intercâmbio de dados cartográficos digitais no âmbito das organizações governamentais, de modo a tornar o dado espacial produzido validado, em relação às regras topológicas, e estruturado segundo categorias e feições geográficas. (ET-EDGV, 2010).

Borba et al. (2008) esclarece que as Especificações Técnicas para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV) foram modeladas segundo o paradigma da orientação a

objetos (OO) utilizando a notação *Object Modeling Technique* (OMG), *Unified Modeling Language* (UML) e a extensão *Object Modeling Technique for Geographic Applications* (OMT-G), e tem por objetivo definir uma estrutura padrão de armazenamento dos dados geoespaciais vetoriais.

Davis Junior (2014) elucida OMT-G como um modelo de dados dotado de recursos para o projeto de bancos de dados e aplicações geográficas e tem por objetivo aumentar a capacidade de representação semântica daquele modelo e, portanto reduzindo a distância entre o modelo mental do espaço a ser modelado e o modelo de representação usual.

A estruturação de dados geoespaciais vetoriais a partir da ET-EDGV possibilita a padronização dos dados, facilita a identificação das topologias e das classes de feições vetoriais, auxiliando na produção de conhecimento dos elementos estudados.

A ET-EDGV (2010) conceitua modelo de dados como um conjunto de conceitos que podem ser usados para descrever a estrutura e as operações em banco de dados sendo também necessário para isto, construir uma abstração dos objetos e fenômenos do mundo real que seja adequada às finalidades pretendidas. Segundo Burrough e Frank (1995), os Sistemas de Informações Geográficas (SIG) e seus modelos de dados geográficos devem refletir a maneira como as pessoas veem o mundo.

Neste sentido, modelo de dados é um conjunto de conceitos e diferentes níveis de abstração que podem ser utilizados para descrever a estrutura lógica de um problema (alto nível) que fornece uma visão mais próxima do modo como os usuários visualizam os dados e a estrutura física (baixo nível) que fornece uma visão mais detalhada do modo como os dados estão realmente armazenados.

O presente trabalho tem por objetivo apresentar uma metodologia de normalização e modelagem de dados espaciais de acordo com os parâmetros de armazenagem e manipulação definidos pela ET-EDGV, usando como base os dados referentes às rodovias e obras rodoviárias do Departamento de Estradas e Rodagem do Estado da Paraíba (DER-PB), de forma a contribuir com a organização, preservação e integração de dados geoespaciais heterogêneos de acordo com as diretrizes estabelecidas pela INDE.

2. Materiais e Metodologia de Trabalho

2.1 Área de estudo

A base de dados de estudo deste trabalho (Figura 1) compreende as rodovias do Estado da Paraíba, administradas pelo DER-PB. O Estado da Paraíba possui 223 municípios conforme a malha digital municipal, situação em 2010 (IBGE, 2010).

A área de estudo compreende um total de 5.808 km de rodovias divididos em 599 trechos rodoviários, sendo 441 trechos de jurisdição estadual e 36 de classificação coincidente - construída e conservada pelo governo estadual, mas coincidente com a diretriz de rodovias presentes no plano rodoviário federal - controladas pelo DER-PB, além de 122 trechos de classificação federal, administradas pelo DNIT (Departamento Nacional de Infraestrutura e Transporte).

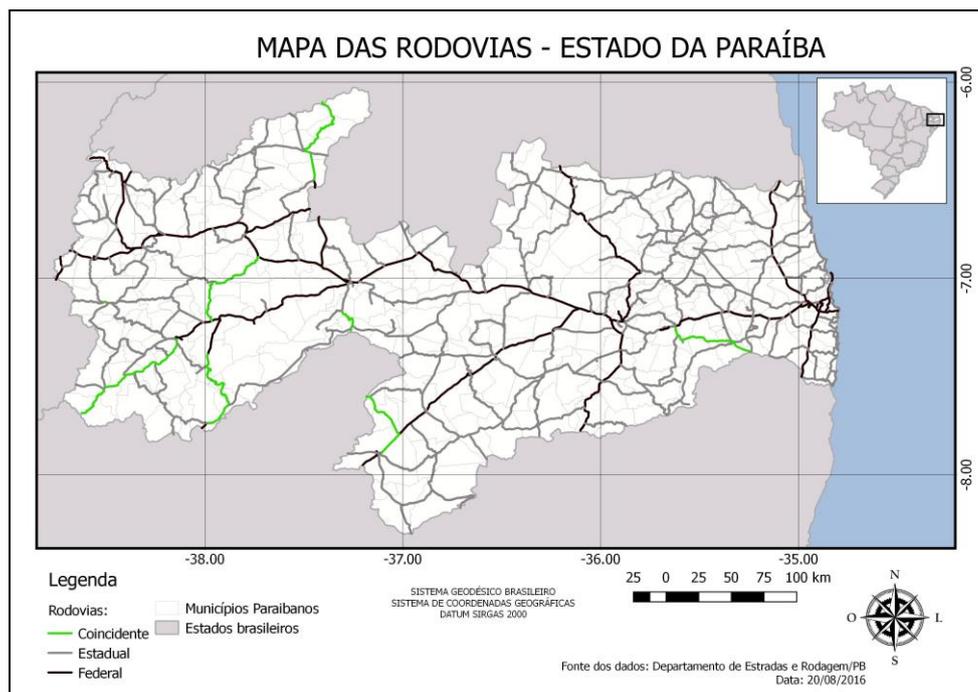


Figura 1. Mapa de localização das rodovias do Estado da Paraíba.

2.2 Materiais

Como materiais utilizados nesta pesquisa foram usados os dados referentes à área de estudo, informações das rodovias e obras rodoviárias em formato *shapefile* obtidas através do DER-PB, os programas para manipulação desses dados e equipamentos, os quais serão relacionados a seguir.

- Equipamentos computacionais:
 - Notebook Core i5 2.2 GHz, disco rígido de 1 TB, 8 GB de memória RAM e placa de vídeo de 2 GB.
- Programas computacionais:
 - MySQL Workbench 6.3;
 - PostgreSQL 9.5 e PostGIS 2.2;
 - OMT-G Designer.

2.3 Métodos

O fluxo de atividades para desenvolvimento da pesquisa pode ser visto na Figura 2. Conforme o fluxograma, a metodologia de trabalho consiste na análise dos *shapefiles* de rodovias e das obras rodoviárias, modelagem conceitual dos dados utilizando o modelo ER (Entidade Relacionamento), transformação do modelo ER para o relacional e normalização dos dados, criação do dicionário de dados dos atributos, transformação do modelo ER normalizado para o OMT-G e modelagem física e implementação do banco de dados geográficos (BDG).

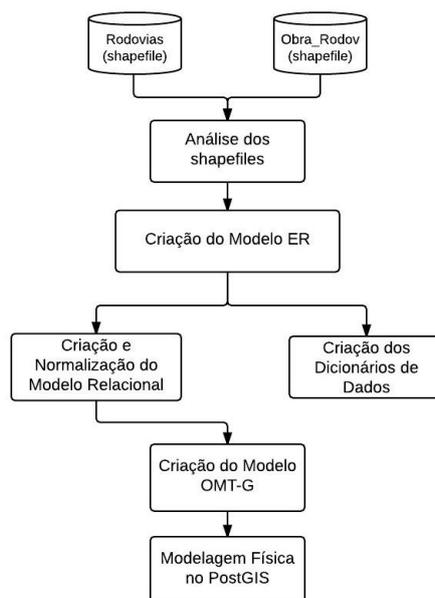


Figura 2. Fluxograma descrevendo a metodologia de trabalho adotada.

2.3.1 Análise dos shapefiles

A coleta de dados foi feita através da obtenção dos *shapefiles* de rodovias e obras rodoviárias, modelo ER e dicionário de dados cedidos pelo DER-PB. Posteriormente, foi realizada a análise desses dados e uma pesquisa referente aos atributos listados no dicionário de dados que teve como objetivo entender a base de dados disponibilizada para a pesquisa.

Através da análise do modelo ER presente no sistema de rodovias (Figura 3), verificou-se um modelo demasiadamente simples, onde apenas são identificadas duas entidades (Rodovias e Subtrechos), relacionadas entre si. Este modelo traz problemas de integridade, de implementação do modelo físico e provê uma possível redundância de dados referentes às rodovias.

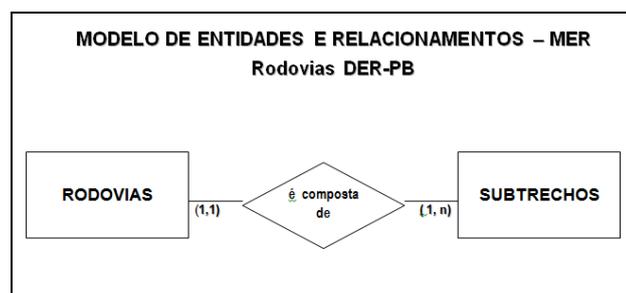


Figura 3. Modelo ER atual utilizado no sistema de rodovias do DER-PB.

A correta modelagem do sistema auxilia no correto desenvolvimento do banco de dados e evita que alterações sejam necessárias para corrigir erros de concepção provenientes de falhas durante a análise.

Dessa forma, foi possível verificar as características intrínsecas do objeto de estudo e as principais necessidades e dificuldades do DER-PB no monitoramento e mapeamento das entidades rodoviárias.

2.3.2 Criação do modelo ER

Nesta etapa do projeto, foi utilizado o MySQL Workbench 6.3 na construção do diagrama Entidade Relacionamento. A partir da análise dos dados de rodovias e das obras rodoviárias, pôde-se entender os principais relacionamentos e especificações do objeto de estudo, auxiliando assim na definição das entidades básicas da modelagem. Através do diagrama ER foi possível compreender o conjunto de dados de maneira mais fácil e usual.

Este modelo torna-se de fundamental importância para o sucesso na construção de um sistema. Portanto, é possível levantar as informações do mundo real, organizá-las de forma gráfica, ajustá-las eficientemente através da normalização dos dados e representar essa organização para validação através do modelo ER.

2.3.3 Modelo relacional e Normalização dos dados

A primeira fase da transformação do modelo ER para o modelo relacional consistiu na identificação das entidades fortes – sua existência independe de outras entidades – inclusão dos seus respectivos atributos simples e escolha de uma chave primária identificadora. Em seguida foram criadas as entidades fracas – são dependentes de outras entidades para existirem – contendo os seus atributos e foram inseridas chaves estrangeiras nas entidades nas quais exista alguma dependência, criando assim um relacionamento.

Subsequentemente, foram identificadas as entidades de relacionamento binário 1:N – relacionamento entre duas entidades que contém cardinalidade de um para muitos – e foi inserido como chave estrangeira no lado N, a chave primária de cada entidade identificada, relacionando as entidades.

Para os relacionamentos binários N:N – relacionamento entre duas entidades que contém cardinalidade muitos para muitos – é necessária a criação de uma nova entidade para realizar o relacionamento. Desta forma, as chaves primárias das entidades contidas no relacionamento são inseridas como chave estrangeira na nova entidade, combinando-as e formando assim, a chave primária da nova entidade. Também devem ser incluídos na nova entidade quaisquer atributos simples identificados neste tipo de relacionamento.

A etapa de normalização de dados foi iniciada pela transformação do modelo relacional não normalizado em um esquema relacional seguindo as regras da 1FN. Identificaram-se todas as tuplas multivaloradas nas entidades e realizou-se a divisão em outras tuplas, e em alguns casos, em novas entidades.

Posteriormente, foi necessário verificar nas entidades, se cada atributo dependia apenas de uma chave primária. Logo, todos os atributos embutidos em outras entidades foram decompostas em outras tabelas, a fim de evitar redundância de dados e aumento de desempenho e escalabilidade, conectando-as a partir de chave primária ou estrangeira.

A terceira etapa foi verificar se as entidades estavam condizentes com as regras da 3FN. Dessa forma, as relações não devem ter atributos que não pertençam a uma chave, não podendo haver dependência transitiva.

Estas etapas de normalização visaram principalmente reduzir a redundância de dados e possíveis anomalias. Assim, foi possível gerar um modelo de dados consistente após a adoção dos procedimentos de normalização.

2.3.4 Criação do dicionário de dados

Esta etapa está vinculada ao desenvolvimento do modelo ER normalizado e possui informações relevantes para as análises dos dados, identificando os atributos, relacionamentos, demais características e a descrição de cada elemento.

O dicionário de dados criado foi baseado no dicionário já existente e utilizado no DER-PB, seguindo as normativas e classificações definidas na categoria de Sistemas de Transportes – Subsistema Rodoviário (ET-EDGV, 2010).

2.3.5 Modelagem OMT-G

A primeira fase da transformação do diagrama ER normalizado para o modelo OMT-G consistiu identificar as entidades que contém primitivas geométricas e seus relacionamentos e assim, aumentar a capacidade de representação semântica da entidade.

Posteriormente, foi possível produzir o modelo utilizando o software OMT-G Designer, agregando às entidades suas restrições de integridade espacial, relacionamentos e tipologias geográficas.

Assim sendo, o modelo OMT-G permite aumentar a abstração dos dados geográficos e facilitar a compreensão espacial do objeto em estudo.

2.3.6 Modelo físico do BDG

Na construção do modelo físico do banco de dados, optou-se em utilizar um Sistema Gerenciador de Banco de Dados (SGBD) para auxiliar na tarefa de construção. Foi escolhido o PostgreSQL 9.5 com a extensão geográfica PostGIS 2.2, bem como a linguagem Structured Query Language (SQL) para implementação das tabelas.

A estruturação da criação das tabelas através de um script SQL foi realizada a fim de relacionar as informações contidas no modelo OMT-G, seguindo os relacionamentos apresentados e identificando as tipologias geográficas e as restrições de integridade espaciais semânticas e topológicas.

3. Resultados

Nesta seção estão apresentados como resultados os modelos de dados criados a partir da análise dos dados do sistema de rodovias do DER-PB. O diagrama ER criado será apresentado a seguir na Figura 4.

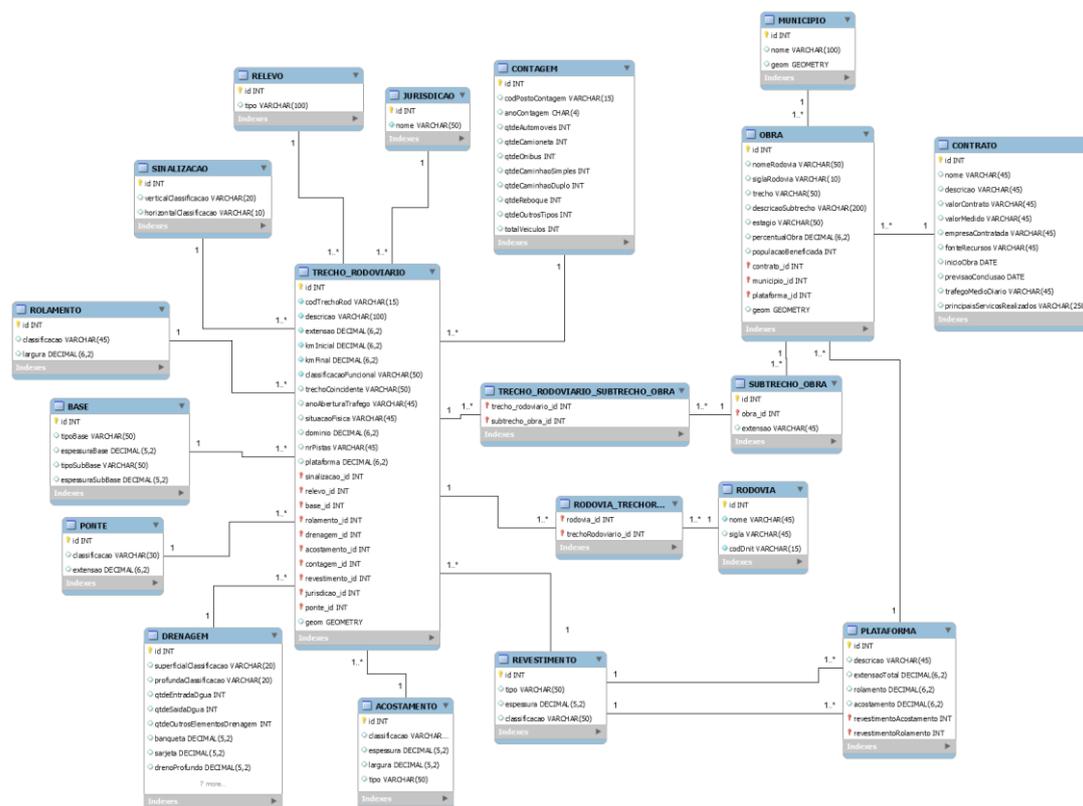


Figura 4. Modelo OMT-G do sistema de Rodovias do DER-PB.

O processo de modelagem e normalização dos dados resultou na transformação do modelo ER com duas entidades gerais em um novo modelo com dezenove entidades, representando assim um aumento da capacidade semântica e organizacional do modelo de dados, auxiliando na implementação do BDG.

A normalização não gera um modelo de dados perfeito, mas elimina diversas deficiências e redundância de dados. O modelo relacional normalizado serviu para corrigir o modelo ER e, além disso, auxiliou a criação do modelo OMT-G e identificação das topologias e representações geográficas.

Para a realização da modelagem conceitual dos dados geoespaciais vetoriais, fez-se imprescindível à adoção do modelo OMT-G baseado na transformação do Modelo ER normalizado, otimizando assim a modelagem da geometria e a topologia dos dados geográficos e oferecendo suporte ao entendimento das estruturas topológicas encontradas no SIG e seus relacionamentos espaciais.

O modelo do banco de dados geográficos OMT-G aderente ao subsistema de Rodovias da ET-EDGV, referente às classes de feição e as classes agregadoras, se encontra ilustrada na Figura 5.

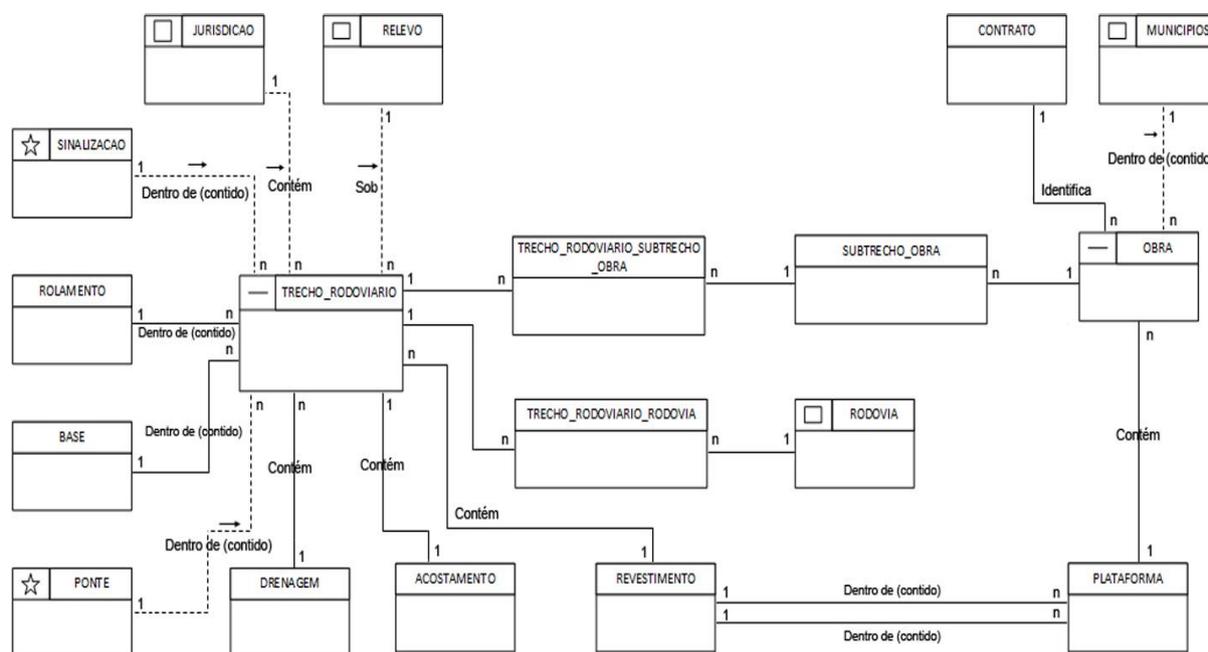


Figura 5. Modelo OMT-G do sistema de Rodovias do DER-PB.

A partir da análise e levantamento dos *shapefiles* rodovias e das obras rodoviárias, bases de dados deste trabalho, foram identificadas dezenove entidades, sendo oito delas, classes com primitivas geométricas de pontos, linhas e polígonos.

A modelagem de dados espaciais é um procedimento complexo, devido ao fato destes modelos necessitarem considerar muitas informações, abstrair o objeto de estudo e enriquecer a carga semântica, requerendo assim, domínio das técnicas a serem empregadas, da natureza diferenciada dos SIG bem como da realidade da aplicação.

4. Conclusões

O objetivo proposto para o trabalho, a respeito da modelagem, normalização e desenvolvimento de um banco de dados geográficos baseado na categoria de informação Sistema de Transportes – Rodovias, da ET-EDGV 2.1.3, foi alcançado plenamente e com sucesso. O modelo proposto nesta pesquisa atende as necessidades básicas propostas pela ET-EDGV.

Através da modelagem de dados pode-se delimitar o problema estudado e dividi-lo em vários problemas menores, restringindo a atenção a um único aspecto por vez até chegar à solução. Além disso, é possível elencar algumas vantagens na utilização do modelo de dados

como a possibilidade de projetar bancos de dados confiáveis, eficientes e com fácil escalabilidade; unificação de conceitos e linguagem para que não haja divergência no entendimento dos usuários; e a criação um repositório único para todos os dados do objeto de estudo.

A normalização não gera um modelo de dados perfeito, mas elimina diversas deficiências e redundância de dados. O modelo relacional normalizado serviu para corrigir o modelo ER e, além disso, auxiliou a criação do modelo OMT-G e identificação das topologias e representações geográficas.

Esta metodologia de BDG servirá para a organização, armazenamento e disseminação de dados geoespaciais dentro do DER-PB, como também, subsidiar a criação de novos sistemas com perspectiva geográfica.

O método adotado neste trabalho pode ser utilizado por qualquer instituição que pretenda realizar a padronização de dados em conformidade com a ET-EDGV, favorecendo assim a cooperação entre sistemas e o intercâmbio de dados, a não redundância através da normalização e a adição de restrições de integridade espaciais semânticas e topológicas através do modelo de dados OMT-G.

Diante do exposto, pode-se concluir que a implementação de um BDG torna-se imprescindível no âmbito dos sistemas de geoinformação, principalmente no que diz respeito à padronização de estruturas de dados e disponibilização da informação geográfica de forma aberta.

Referências bibliográficas

Borba, R. L. R.; Mota, G. L. A.; Brito, J. L. N. S. **Metodologia para Representação da Estrutura de Dados Geoespacial Vetorial da Mapoteca Nacional Digital em Bancos de Dados Geográficos Relacionais**. In: Biblioteca online do Instituto Nacional de Pesquisas Espaciais (INPE), 2008. Disponível em: <http://mtc-m16c.sid.inpe.br/col/sid.inpe.br/mtc-m16c/2015/11.13.18.05/doc/proceedings_geoinfo2008.35_40.pdf> Acesso: 17 ago. 2016.

Burrough, P.A.; Frank, A.U. **Concepts and paradigms in spatial information: are current geographical information systems truly generic?** International Journal of Geographical Information Systems, Londres, v.9, n.2, p.101-116, 1995.

CONCAR. **Normas Técnicas da Cartografia Nacional**. Brasília, 1984. Disponível em: <<http://www.concar.ibge.gov.br/detalheDocumentos.aspx?cod=8>>. Acesso em: 09 ago. 2016.

Davis Junior, C. A. **Object Modeling Technique for Geographic Applications - OMT-G**. 2014. Disponível em: <<http://homepages.dcc.ufmg.br/~clodoveu/DocuWiki/doku.php?id=omtg>> Acesso em: 20 set. 2016.

ET-EDGV. **Especificação Técnica para a Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV)**. Edição 2.1.3. Rio de Janeiro. 2010.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Malha digital municipal do Brasil: situação 2010**. Rio de Janeiro: IBGE, 2010.

GSDI - Global Spatial Data Infrastructure. **Spatial Data Infrastructure Cookbook**. 2009.