

CONTROLE DE QUALIDADE DE DADOS GEOESPACIAIS NA FASE DE REAMBULAÇÃO: ABORDAGEM DA COMPLETUDE

Joel Borges dos Passos¹, Leandro Luiz Silva de França¹, Ana Cláudia Bezerra de Albuquerque
Borborema de Andrade¹ e Rodrigo Wanderley de Cerqueira¹

¹Diretoria de Serviço Geográfico do Exército – DSG, 3º Centro de Geoinformação – Olinda-PE – CEP: 53.240-650
{joel.passos, franca.leandro, cerqueira.rodrigo}@eb.mil.br, anacborborema@gmail.com

RESUMO

O presente trabalho apresenta a metodologia para avaliação da completude de um banco de dados geográficos oriundo da fase de reambulação do Projeto de Mapeamento da Bahia. A completude foi avaliada comparando-se os dados coletados na fase de reambulação com bases cartográficas de referência (homologadas). Na avaliação foram quantificadas as porcentagens de itens em excesso, itens não previstos e itens ausentes. O resultado dessa avaliação é importante para afirmar que o produto da fase de reambulação executada pela Diretoria de Serviço Geográfico (DSG), após as devidas correções, atende aos requisitos de completude para o Conjunto de Dados Geoespaciais Vetoriais (CDGV) preconizados pela Especificação Técnica para Controle de Qualidade de Dados Geoespaciais (ET-CQDG).

Palavras-chave — Reambulação, avaliação de qualidade, completude, ET-CQDG.

ABSTRACT

This work presents the methodology to assess the completeness of a geographic database after the reambulation phase of the Bahia Mapping Project. The completeness was evaluated by comparing the collected data in the reambulation with reference cartographic bases (homologated). In the evaluation, the percentages of items in excess, items not foreseen and items absent were quantified. The result of this evaluation is important to affirm that the product of the reambulation phase executed by the Geographic Service Bureau (DSG), after corrections, meets the completeness requirements for the Vector Geospatial Data Set (CDGV) recommended by the Technical Specification for Geospatial Data Quality Control (ET-CQDG).

Keywords — Reambulation, quality assessment, completeness, ET-CQDG.

1. INTRODUÇÃO

Para retratar a qualidade dos dados geoespaciais é essencial a existência de normas e especificações técnicas. No âmbito

nacional, o plano de ação da Infraestrutura Nacional de Dados Espaciais (INDE) prevê uma série de normas, padrões e especificações para mapeamento.

Uma dessas especificações que auxilia o produtor de informações geográficas na avaliação da qualidade do produto é a ET-CQDG, a qual foi elaborada pela DSG e está alinhada com a ISO 19157:2013.

Essas normas estabelecem os parâmetros mínimos de qualidade que devem ser adotados para atender as demandas de mapeamento de acordo com o Sistema Cartográfico Nacional (SCN) [1].

Devido à evolução dos procedimentos de geração, atualização e controle de qualidade dos dados geográficos, os processos de produção cartográfica vêm passando por grandes mudanças. Dentre esses processos, uma das fases que se adaptou ao emprego de bancos de dados geográficos e ao desenvolvimento de novas ferramentas de SIG foi a Fase de Reambulação [2].

Essa fase tem por finalidade a execução do trabalho de campo para a coleta de topônimos, informações e dados relativos aos acidentes naturais e artificiais do terreno e a confirmação da correspondência entre as feições que foram interpretadas pelo operador e a verdade no terreno, alimentando um banco de dados com todas as instâncias das classes previstas nas especificações vigentes, na busca da interoperabilidade [2].

Para que haja uma adequada integração entre as informações coletadas nessa fase, é necessário que a qualidade do dado geográfico seja levada em conta. Para isso, em sua etapa final, os dados coletados passam por um processo de inspeção denominado de revisão.

Essa subfase visa detectar inconsistências no processo de reambulação, seja na aquisição ou preenchimento de atributos das feições, indicando todas as necessidades de correção. O processo de revisão da reambulação é dividido nos seguintes elementos de qualidade: acurácia posicional, acurácia temática, completude e consistência lógica.

A DSG, como órgão normatizador e de referência na área da cartografia, tem como prática realizar o controle de qualidade ao final de cada etapa do processo de produção. Nesse contexto, este trabalho tem como objetivo apresentar o processo de controle de qualidade da completude da etapa de produção da reambulação, considerando um Conjunto de Dados Geoespaciais Vetoriais (CDGV).

2. MATERIAIS E MÉTODOS

2.1. Área de estudo, equipamento, normas e especificações técnicas

A área de estudo desse trabalho compreende os objetos instanciados no banco de dados (feições), estruturados conforme a modelagem da Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV) versão 2.1.3, da carta topográfica de Mapa Índice (MI) 1785-2, reambulada no Estado da Bahia.

Os softwares e aplicativos utilizados durante as inspeções de qualidade foram: QGIS versão 2.18 e *PostgreSQL* com sua extensão espacial *PostGIS*.

Já os materiais técnicos e de consulta foram: a Especificações Técnicas para Aquisição de Dados Geoespaciais Vetoriais (ET-ADGV 2.1.3), a ET-EDGV 2.1.3, a ET-CQDG 1ª edição, a Metodologia de Reambulação 2ª versão e as notas técnicas e instruções específicas relativas ao projeto.

Das notas técnicas destacam-se as Tabelas de Aquisição e Preenchimento dos Atributos, que indicam como deve ser realizada a classificação das categorias, definindo as classes que deverão ser adquiridas e os atributos de preenchimento obrigatório.

Esses procedimentos visam atender as exigências do cliente, indo de encontro com o conceito de qualidade que corresponde à conformidade às especificações e à capacidade de um produto ou serviço satisfazer as necessidades do cliente [3].

2.2. Completude

A análise da completude, segundo a ISO 19157:2013, consiste em verificar a presença e a ausência de informações no conjunto de dados. Assim, constata-se o excesso ou a omissão de feições em relação aderência ao modelo de dados e a especificação de aquisição desses dados, neste caso a ET-EDGV e ET-ADGV, respectivamente.

A ET-CQDG e ET-ADGV preconizam quais feições deverão estar presentes nos produtos em determinadas escalas. Caso alguma feição que deveria ser representada não esteja (seja omitida), tem-se um caso de deterioração da qualidade do produto em relação à completude por omissão. Caso contrário, se uma feição que não deveria ser representada naquela escala estiver presente no produto, tem-se uma perda na qualidade do produto em relação à completude por excesso.

Estes subelementos da completude (excesso e omissão) podem ser quantificados em porcentagem [4].

As medidas relativas ao subelemento excesso são: **Porcentagem de itens em excesso** e **Porcentagem de itens não previstos**. A primeira medida quantifica a proporção de objetos que foram incluídos no conjunto avaliado, mas que não existem no universo de discurso, em relação à quantidade de objetos que deveriam estar presentes. A segunda quantifica a proporção de objetos não previstos no

conjunto de dados em relação à quantidade total de objetos. Objetos não previstos correspondem às geometrias duplicadas e aos objetos que não deveriam ter sido adquiridos na escala do produto [4].

Já a medida relativa ao subelemento omissão é a **Porcentagem de itens ausentes** que quantifica a proporção de objetos ausentes no conjunto avaliado em relação à quantidade de objetos que deveriam estar presentes [4].

2.3. Metodologia de avaliação da qualidade

O processo de avaliação de qualidade pode ser aplicado a uma etapa do processo de produção ou ao produto cartográfico finalizado [5]. Este trabalho foi desenvolvido para apresentar a avaliação de completude da fase de reambulação.

De acordo com [5] as inspeções de completude demandam a elaboração do plano de amostragem, para estabelecer um parâmetro de aprovação ou reprovação do produto cartográfico. Nesse plano, o nível de inspeção, que discrimina a quantidade relativa de inspeção, influencia no tamanho da amostra (n) e os níveis de Limite de Qualidade Aceitável (LQA) influenciam no número de aceitação (Ac) e rejeição (Re) para aprovação ou reprovação do lote (N).

O lote estabelecido neste trabalho é definido como um produto, ou seja, um conjunto de dados e refere-se aos elementos das classes instanciadas no banco de dados da folha MI 1785-2. No banco foi observada uma população (N) de 6.501 feições geográficas.

Recomenda-se o nível de inspeção II para aplicação inicial da avaliação e um (LQA) de 4,0% [5]. Nesse sentido o plano de amostragem com um lote (N) de 6501 feições, precisa de uma amostra de 200 feições e um nível de aceitação (Ac) de, no máximo, 14 registros inconsistentes para a aprovação deste lote [4] [6].

Porém para o plano de amostragem do banco de dados oriundos da reambulação o valor para o tamanho da amostra (n) sempre será o tamanho da população (N), logo, será realizada a inspeção completa orientada por feições geográficas. A Tabela 1 mostra o plano de amostragem definido para o banco do MI 1785-2 e o quantitativo de feições de todas as classes instanciadas no banco.

Plano de amostragem resumido				
Classe	Qtd de feições (N)	n° max. de violaç LQA 4%	tamanho da amostra (n)	Método de inspeção
38 classes da ET-EDGV	6501	0 (zero)	6501	completa

Tabela 1. Resumo do plano de amostragem do MI 1785-2

2.3. Revisão da Reambulação

A revisão da fase de reambulação consiste nas ações a serem executadas pelo revisor para garantir a consistência dos dados e a existência de todas as informações necessárias

para que o produto final atinja os padrões mínimos de qualidade de acordo com a ET-CQDG [2].

Adota-se como taxa de omissão e excesso para essa fase, o valor de 0%, para o conjunto de dados geoespaciais. Esse valor é fixado tendo em vista que esse produto ainda está em fase de produção e em condições de correções de todas as inconsistências encontradas.

Durante a revisão essas inconsistências são computadas e apontadas através de um arquivo *SqLite*, onde existe uma camada de pontos na qual o revisor insere os detalhamentos das inconsistências que precisam de ajustes ou correções.

A revisão é realizada por categoria de informação e ao final indicam-se quais foram as classes de feições aprovadas ou reprovadas no conjunto de dados espaciais.

As inconsistências inseridas nesse arquivo são os resultados do enquadramento das feições de acordo com uma base de dados de referência, tendo como objetivo verificar se uma feição inspecionada está em concordância com a classe de referência constante na base de dados.

Para o Projeto de Mapeamento da Bahia a base de dados de referência consiste nos insumos oficiais referentes área de trabalho [2]. Esses insumos são: Carta Topográfica Matricial na escala de 1:100.000, imagem ortorretificada de voos fotogramétricos, além de dados geográficos de órgãos estaduais e federais. Ressaltando-se que esses insumos foram devidamente homologados e conhecidos seus metadados.

3. RESULTADOS E DISCUSSÃO

A inspeção relativa à completude do banco de dados geográficos correspondente à folha de MI 1785-2 resultou em um arquivo de ocorrência com 89 inconsistências em 15 classes de um total de 39 classes instanciadas.

A Tabela 2 e 3 mostram os resultados da primeira análise que tem como parâmetro o excesso de objetos no banco de dados em relação às bases de referências.

Para o cálculo da medida de **Porcentagem de itens em excesso** no banco de dados avaliado, contou-se a quantidade de objetos presentes nas classes desse banco ($n1$) que não possuem correspondente na classe de referência (tamanho n). O resultado da medida é a proporção $n1/n$ expressa como porcentagem (Tabela 2).

ACURÁCIA DA COMPLETUDE/EXCESSO				
Classes da ET-EDGV	Número de ocorrências na base de referência (n)	Número de elementos do banco de dados	itens em excesso (n1)	Porcentagem de itens em excesso (n1/n)
dep_abast_agua	46	43	2	4,35%
edif_agrop_ext_veg	227	224	2	0,88%
edif_habitacional	1892	1890	1	0,05%
área_edificada	3	3	1	33,33%

Tabela 2. Medida porcentagem de itens em excesso.

A Figura 1 mostra um caso de excesso detectado na classe *depósito de abastecimento*. Observa-se que na figura da sobreposição das feições com a referência (Figura 1-b)

existe um objeto no banco de dados que não encontra correspondência na referência (Figura 1-a).



Figura 1. a) Base de referência (universo de discurso) e b) Sobreposição do banco de dados com a referência.

No cálculo da medida de **Porcentagem de itens não previstos** no banco de dados avaliado foram verificados os objetos duplicados (geometria e atributos coincidentes) e os objetos com dimensões menores que as regras de aquisição e, portanto, não previstos para ser coletado para escala de mapeamento. O resultado da medida é a proporção de erros sobre o tamanho total do conjunto, que é 6501 feições, expressa como porcentagem (Tabela 3).

classes da ET-EDGV	Número de objetos do banco de dados	itens não previstos		Porcentagem de itens não previstos (n1+n2/n)
		Duplicado (n1)	Aquisição Inválida (n2)	
		edif_habitacional	1890	
fonte_dagua	12	-	12	0,18%
campo_quadra	10	-	1	0,02%
trecho_rodoviario	306	3	-	0,05%
veg_caatinga	112	-	1	0,02%
veg_cultivada	89	-	2	0,03%
Total de objetos do banco de dados (n)				6501

Tabela 3. Medida porcentagem de itens não previstos.

Analisando a Tabela 3 observa-se que em cinco classes houve feições que não era previsto a aquisição. Por exemplo, elementos da classe *fonte d'água* somente serão adquiridos na escala de 1:25.000, porém como a escala de trabalho é 1:50.000 os 12 elementos dessa classe foram detectados como erro de excesso, aquisição inválida.

Para a classe de *edificação habitacional* o excesso foi detectado nos elementos com atributo "situação física" preenchido como "destruída", tendo em vista que pelas Tabelas de Aquisição e Preenchimento dos Atributos edificações habitacionais destruídas não são coletadas, por isso são elementos de aquisição inválida.

Já os elementos das classes *campo quadra*, *vegetação cultivada* e *vegetação caatinga* apresentaram dimensões de área menor que o previsto para a escala de trabalho conforme a ET-ADGV. Por exemplo, para a classe *campo quadra* ser representada por polígono é necessário uma área de 2.500m², porém no banco de dados estava com 2.370m², logo é detectado um erro de excesso.

A Tabela 4 mostra o resultado da segunda análise estabelecida diante da ausência (omissão) dos elementos da base de referências em relação ao banco de dados.

O cálculo da medida de **Porcentagem de itens ausentes** no banco de dados avaliado, contou-se a quantidade de objetos presentes no universo de discurso (base de referências) que não possuem correspondente no banco de dados. O resultado da medida é a proporção entre os objetos ausente ($n1$) e a quantidade total de feições no universo de discurso (n) expressa como porcentagem (Tabela 4).

ACURÁCIA DA COMPLETUDE/OMISSÃO				
Classe da ET-EDGV	Número de feições na base de referência (n)	Número de objetos do banco de dados	Itens ausentes ($n1$)	Porc. de itens ausentes ($n1/n$)
edif_abast_agua	28	25	3	10,71%
dep_abast_agua	46	43	4	8,70%
edif_agrop_ext_veg	227	224	1	0,44%
edif_habitacional	1892	1890	13	0,69%
deposito_geral	235	226	9	3,83%
area_edificada	3	3	1	33,33%
campo_quadra	13	10	3	23,08%
nome_local	19	17	2	10,53%
trecho_rodoviario	311	306	8	2,57%
trecho_drenagem	62	58	3	4,84%
massa_dagua	3	2	1	33,33%
trecho massa d'agua	4	3	1	25,00%

Tabela 4. Medida porcentagem de itens não previstos.

A Figura 2-a mostra o caso de omissão detectado na classe *trecho drenagem*. Percebe-se que das três feições na referência duas não encontra correspondência com objetos do banco de dados.

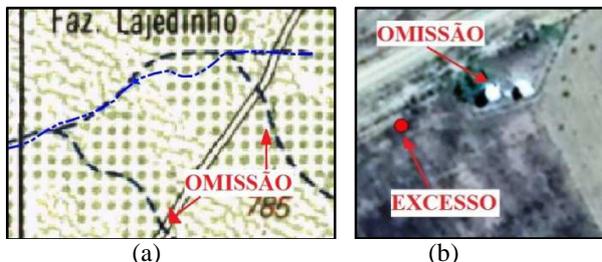


Figura 2. a) Sobreposição do objeto com a Carta Topográfica b) Omissão e excesso de elemento em função do deslocamento.

A Figura 2-b mostra a inspeção da completude da classe *depósito de abastecimento*. Percebe-se que houve um excesso na representação de uma caixa d'água no banco de dados, bem como uma omissão, em função da sua localização equivocada. Salienta-se que esta ocorrência apresentando deslocamento dentro dos valores previstos para a classe "A" do Padrão de Exatidão Cartográfica da Planimetria dos Produtos Cartográficos Digitais (PEC-PCD) que é de 14m (escala 1:50.000) deve ser sinalizada como erro de acurácia posicional.

Outros problemas de completude ocorrem quando se classifica erradamente um objeto (por exemplo, um curral

ser classificado como uma edificação habitacional). Esse tipo de ocorrência pode ser considerado como dois erros de completude: uma omissão (curral) e um excesso (edificação habitacional). Porém nesse projeto essas ocorrências são consideradas como um único erro temático (acurácia da classificação).

Conforme apresentado no plano de amostragem, o nível de conformidade para as taxas de excesso e omissão consideradas na fase de reambulação é de 0%. Desta forma, as classes que apresentaram os casos de excesso e omissão são reprovadas, devendo-se ser corrigidas ou ajustadas pelo reambulador.

4. CONCLUSÕES

A avaliação da qualidade na fase de reambulação é uma etapa essencial no processo de produção cartográfica, principalmente quando se refere ao elemento completude. A etapa de revisão da completude deve ser realizada por operadores experientes e que tenham domínio tanto das especificações técnicas gerais (ET-EDGV, ET-ADGV e ET-CQDG) quanto das peculiaridades do projeto.

Desta forma, é possível garantir que todas as exigências para o propósito de aplicação dos dados coletados e as necessidades do cliente estão sendo atendidas.

5. REFERÊNCIAS

- [1] Dalmolin, Q. e Leal, E.M. "Análise da qualidade posicional em bases cartográficas geradas em CAD". *Boletim de Ciências Geodésicas*, Curitiba, Universidade Federal do Paraná - UFPR, v. 7, n. 1, p. 21-40, 2001.
- [2] Passos, J.B. e França, L.L., "Processo de reambulação no mapeamento topográfico", *Revista brasileira de geomática*, Curitiba, v. 6, n. 2, p. 119-138, abr/jun.2018.
- [3] Servigne, S.; Lesage, N. e Libourel, T., "Quality components, standards, and metadata", Devillers, R.; Jeansoulin, R. (Eds.), *Fundamentals of spatial data quality*. [s.l.] ISTE, 2006. p.179-210.
- [4] DSG – Diretoria do Serviço Geográfico, "Norma da Especificação Técnica para Controle de Qualidade de Dados Geoespaciais (ET-CQDG) – 1ª Edição", Brasília, 2016. 94
- [5] Brasil, "Avaliação da qualidade de dados geoespaciais" IBGE, *Coordenação de Cartografia*, Rio de Janeiro, IBGE, 2017, 96p.
- [6] ISO, "ISO 19157:2013, Geographic information - Data quality", *International Organization for Standardization (ISO)*, Geneva, IDO, 2013. 146p