Qualidade de metadados geoespaciais conforme a ET-PCDG

Felipe de Carvalho Diniz

Diretoria de Serviço Geográfico – DSG
1ª Divisão de Levantamento
Rua Cleveland 250 – 90850-240 – Porto Alegre - RS, Brasil
diniz@dsg.eb.mil.br

Abstract. The presidencial degree number 6.666, from november 27 of 2008 estabilish the Infraestrutura Nacional de Dados Espaciais do Brasil (INDE). Acording to this degree, INDE is an integrated set of technologies, policies, coordination and standards necessary to promote the generation, storage, access and dissemination of geospatial data produced by public agencies of executive power. One of the responsibilities was the creation of the Action Plan for Implementation of INDE (CONCAR, 2010), a document that defines the activities, goals, and responsible for the implementation of an SDI in Brazil. In this document is designed the Brazilian Army with assignment of preparing the Technical Specification for Geospatial Products Data Sets (ET-PCDG), which, in addition to setting standards for products to become of official use in INDE, defines the group metadata for each type of product provided. It has also been appointed to the Brazilian Army the assignment of preparing the Technical Specification for Quality Control of Geospatial Product Data Set (ET-CQPCDG), which defines the procedures for the evaluation of the quality of cartographic products and their metadata. The objective of this work is the evaluation of 49 metadata for the vector 1:250,000 sheets of southern Brazil contained in BDGEx in accordance with ISO 19115:2003, the profile MGB, and ET-PCDG. The results where that the metadata evaluated had only errors that do not compromise the quality, verifying that the metadata contained in BDGEx are of a high reliability.

Palavras-chave: INDE, PCDG, Perfil MGB, metadados, Metadata, controle de qualidade, quality control

1. Introdução

O Decreto Presidencial nº 6.666, de 27 de novembro de 2008 (Brasil, 2008) institui a Infraestrutura Nacional de Dados Espaciais do Brasil (INDE). De acordo com este decreto, a INDE é o conjunto integrado de tecnologias, políticas, coordenação e padrões, necessário para promover a geração, o armazenamento, o acesso e disseminação de dados geoespaciais produzidos por órgãos públicos do Poder Executivo. Este decreto define conceitos, estabelece diretrizes, prazos e responsabilidades para a criação da INDE.

Uma das responsabilidades foi a criação do Plano de Ação para Implantação da INDE (CONCAR, 2010), um documento que define as atividades, metas e responsáveis para a implantação de uma IDE no Brasil. Nesse documento é adotado como o padrão de metadados brasileiro o Perfil de Metadados Geoespaciais do Brasil (Perfil MGB), baseado na norma ISO 19115:2003. Nesse documento também é designada ao Exército Brasileiro à atribuição de confeccionar a Especificação Técnica para Produtos de Conjuntos de Dados Geoespaciais (ET-PCDG), que, além de definir os padrões para os produtos se tornarem de uso oficial na INDE, define o grupo de metadados para cada tipo de produto previsto. Também foi designada ao Exército Brasileiro a atribuição de confeccionar a Especificação Técnica para Controle de Qualidade de Produto de Conjunto de Dados Geoespaciais (ET-CQPCDG), que define os procedimentos para a avalição da qualidade dos produtos cartográficos e de seus metadados.

No plano de ação da INDE também determina a criação do SIG-Brasil, o geoportal que servirá de ponto de acesso para as informações geoespaciais e seus metadados. Esse portal reúne as informações dos diversos órgãos fornecedores de dados geoespaciais oficias, sendo um deles o Exército Brasileiro.

O Exército Brasileiro possui o Bando de Dados Geográfico do Exército (BDGEx), geoportal onde são disseminados os dados geoespaciais internamente, para o próprio exército. O

BDGEx se tornará nó da INDE, sendo seus dados também podendo ser acessados pelo SIG-Brasil.

O objetivo do corrente trabalho é a avaliação de 49 metadados referentes às folhas vetoriais 1:250.000 da região sul do Brasil contidas no BDGEx de acordo com a norma ISO 19115:2003, o Perfil MGB, a ET-CQPCDG, e a ET-PCDG.

1.1 Metadados

Os metadados, de acordo com CONCAR (2009), podem ser definidos como um conjunto de informações a respeito dos dados espaciais que visam permitir, localizar, descobrir e avaliar uma informação geográfica.

Em virtude do grande número de instituições envolvidas na produção e distribuição de dados geoespaciais no Brasil é necessário que os dados estejam respeitando um conjunto de padrões e normas que irão garantir a interoperabilidade em sistemas diversos, viabilizando o compartilhamento dos dados entre diferentes organizações e instituições.

De acordo com CONCAR 2009 os metadados do Perfil MBG devem conter seções específicas visando:

- Identificar o produtor e a responsabilidade técnica da produção;
- Padronizar a terminologia;
- Garantir a interoperabilidade;
- Viabilizar a integração de informações;
- Identificar a qualidade dos dados;
- Viabilizar a divulgação dos dados.

O Perfil MGB define a formato XML como o padrão para utilização dos metadados. Tal escolha foi feita, pois o XML, conforme o Padrão de Interoperabilidade do Governo Eletrônico (e-PING, 2011), é o padrão primário de intercâmbio de dados para todos os sistemas do poder público. A utilização de XML implica em utilizar XML Schema como gramática para validar as instâncias (Goldberg, 2009).

1.2 Produtos de Conjunto de Dados Geoespaciais

Para criar especificações de produtos consistentes com as normas ISO 19100 foi criada a ISO 19131. Ela define o que deve ser abordado nas especificações de produtos, que tipos de fenômenos geográficos deve ser abordado pelo conjunto de dados, como esse conjunto é representado. Na Especificação Técnica dos Produtos de Conjunto de Dados Geoespaciais (ET-PCDG) foram definidos os principais produtos conforme a necessidade do Brasil utilizando a ISO 19131. Cada um dos produtos previstos na ET-PCDG necessitam de um conjunto de informações que são peculiares a esse tipo de produto, assim para cada produto foi definido o conjunto de metadados conforme o Perfil MGB.

Os metadados previstos na norma ISO foram reagrupados, visando facilitar o entendimento e o cadastramento, em três grupos:

- **Metametadados** – apresenta informações sobre o preenchimento dos próprios metadados;
- **Informações de Identificação** – apresenta metadados responsáveis pela acurada identificação do produto ou feição;
- **Características Técnicas** – agrega pacotes da ISO 19115 considerados como de caráter técnico ou correlacionados, incluindo as etapas de produção e os recursos, como por exemplo, os insumos, metodologias, especificações técnicas e o controle de qualidade segundo se preconiza na Especificação Técnica para Controle de Qualidade

de Produtos de Conjunto de Dados Geoespaciais Vetoriais.

2. Metodologia

Para avaliar a qualidade dos metadados é necessário definir os elementos de qualidade pertinentes aos metadados conforme especifica a ISO 19113 *Geographic Information Quality Principles* e ISO 19114 *Geographic Information Quality Evaluation Procedures*.

A Especificação Técnica para Controle de Qualidade de Produto de Conjunto de Dados Espaciais (ET-CQPCDG), ainda em desenvolvimento, define os elementos, subelementos e medidas para avaliação da qualidade dos metadados.

Para os metadados são definidos os seguintes elementos e subelementos:

- **Completude - Comissão**: Preenchimento de elementos não previstos para o preenchimento para aquele produto segundo a ET-PCDG
- **Completude – Omissão**: Não preenchimento de elementos obrigatórios segundo ao Perfil MGB
- **Consistência Lógica – Consistência de Formato**: Consistência da sintaxe XML empregada no metadado.
- **Consistência Lógica – Consistência de Domínio**: Consistência com o dicionário de Codelists do Perfil MGB
- **Acurácia Temática – Preenchimento dos Metadados**: Correto preenchimento dos elementos dos metadados.

2.1 Adaptação dos XML *Schemas*

O Perfil MGB apresenta cardinalidades distintas da ISO 19115. Para a validação conforme o Perfil MGB os Schemas XML foram alterados.

Na Figura 1 está apresentado como exemplo as mudanças nos elementos de MD_Metadata, e na figura 2 as mudanças nos relacionamentos de MD_Metadata

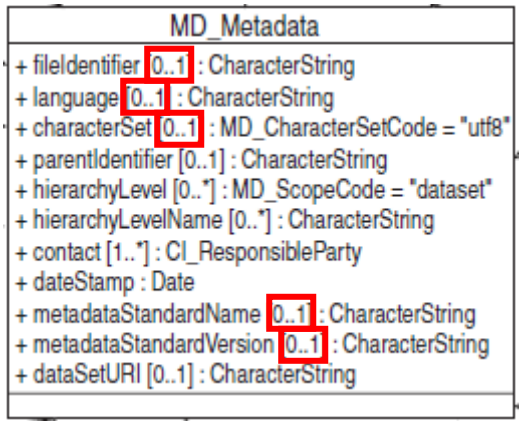

Figura 1. Modificações nos elementos de MD_Metadata
Fonte: Adaptado de ISO 19115:2003

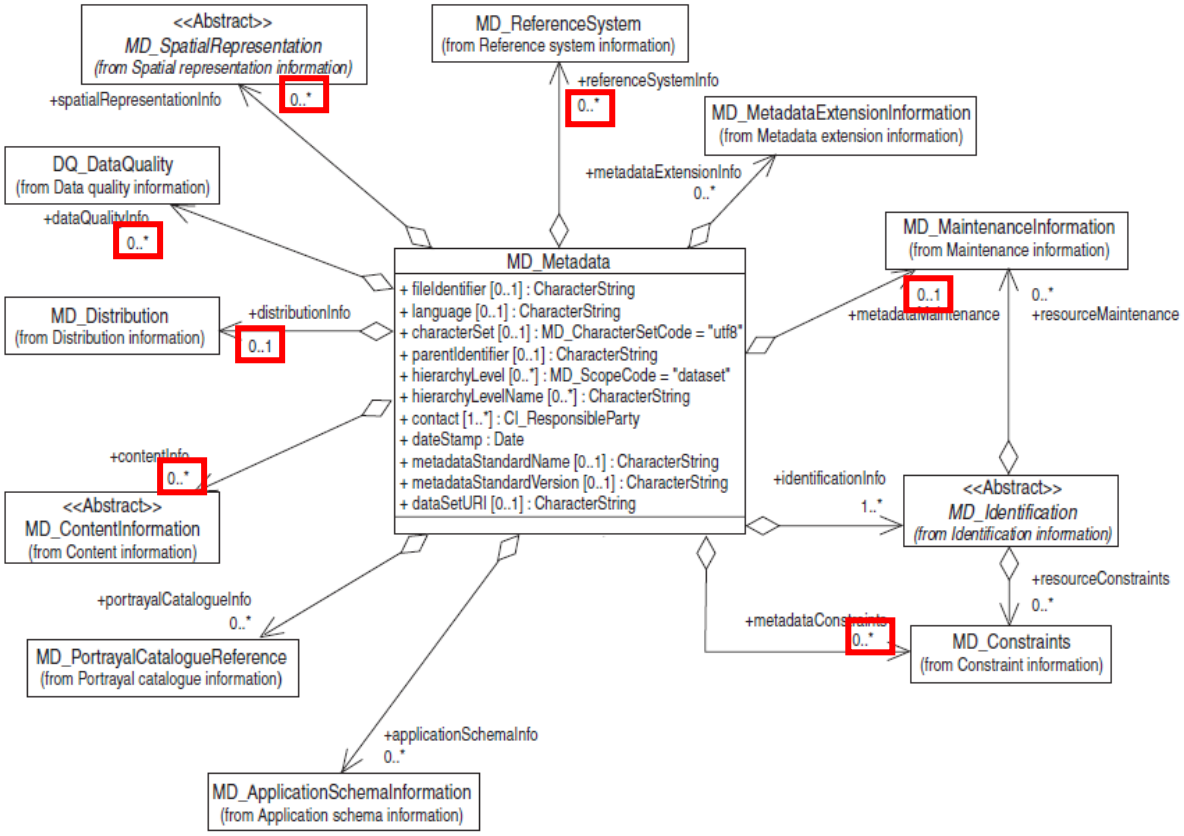

Figura 2. Modificações nos relacionamentos de MD_Metadata
Fonte: Adaptado de ISO 19115:2003

2.2 Dados para avaliação

Para avaliação dos metadados conforme a metodologia descrita na seção anterior foram utilizados 49 metadados de Conjunto de Produtos de Dados Geoespaciais Vetoriais das cartas 1:250.000 pertencentes a região sul do Brasil. Tais cartas e seus metadados foram obtidos no Bando de Dados Geográficos do Exército.

Também foram obtidos os insumos utilizados para a confecção do produto, os controles de produção das seções de vetorização e validação.

Também puderam ser verificadas as metodologias de trabalho da DSG, principalmente a metodologia de Preparo do Fluxo de Trabalho, que padroniza o preenchimento de metadados no Exército Brasileiro.

2.3 Validação dos metadados

Para a validação de metadados foi utilizado o Software Safe FME 2012, utilizando o Transformer XMLValidator.

A validação foi feita em 5 partes:

- ¬ Validação da syntax XML
- ¬ Validação do Schema ISO 19115
- ¬ Validação do Schema ISO 19115 adequado para o Perfil MGB
- ¬ Validação do Dicionário de CodeLists
- ¬ Verificação do Preenchimento

Na validação da sintaxe somente se examina erros quanto à utilização da linguagem XML, os erros mais comuns são tags incorretas, que faltando fechamento ou abertura.

A validação conforme a ISO 19115 foi feita utilizando os XML Schemas Definition Language oficiais da ISO.

A validação conforme o Perfil MGB foi feita em duas partes. A primeira utilizando o XML Schema da ISO 19115 com as cardinalidades modificadas para adequar ao Perfil MGB. A segunda parte foi feita utilizando as restrições apresentadas nas Tabelas 3,4,5 quanto a elementos de valores de preenchimento obrigatório, quanto a restrições de Codelists, e quanto se a elementos, fora aos apresentados na tabela, que foram preenchidos nos metadados.

A validação dos dicionários de CodeLists foi feita implementando um XML Schema com todos os codelists presentes nas tabelas 3,4,5 e verificado com os elementos destacados nas tabelas.

Na Figura 3 está um extrato da Tabela FME utilizada para realizar o processo.

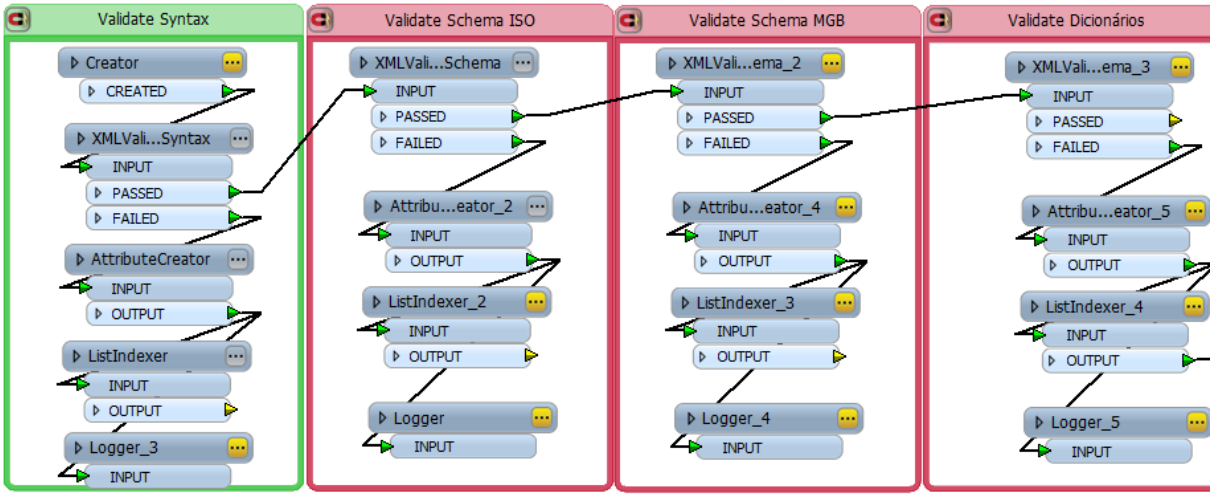

Figura 3. Extrato Validador FME

A verificação do preenchimento foi feita manualmente utilizando a carta vetorial correspondente ao metadado, os insumos utilizados no processo, os controles de produção das seções que realizaram processos sobre as cartas e utilizado a Metodologia de Preparo do Fluxo de Trabalho, a metodologia interna de uso do Exército Brasileiro para preenchimento dos metadados.

3. Resultados

Utilizando a metodologia descrita na seção anterior foram retornados 441 erros, sendo 49 deles duplicados em 2 processos distintos, o que dá um valor absoluto de 392 erros. Como forma avaliados 49 metadados foi obtido uma média de erro de 8 erros por metadado. Na figura 8 está apresentada a distribuição dos erros de validação encontrados.

Figura 9 – Quantidade de erros por elemento de qualidade

Erros de Sintaxe XML: Em todos os arquivos foram encontrados tags sem preenchimentos. Tais erros não comprometem a integridade do arquivo, porém deixam o arquivo XML "sujo". Na tabela 6 está um extrato de alguns dos tags vazios.

Tabela 6. Extrato Schema CI_Contact_Type

```
<gmd:source>
    <gmd:LI_Source>
<gmd:description>
    <gco:CharacterString></gco:CharacterString>
</gmd:description>
<gmd:sourceCitation>
    <gmd:CI_Citation>
        <gmd:title>
            <gco:CharacterString></gco:CharacterString>
        </gmd:title>
        <gmd:date>
        <gmd:CI_Date>
            <gmd:date>
                <gco:Date></gco:Date>
            </gmd:date>
```

- **Erros em relação ao XML Schema ISO 19115:** Em todos os arquivos foi encontrado o erro 44950 missing 'T' separator in dateTime value. O campo Data de Conclusão do processo (dataQualityInfo - DQ_DataQuality – lineage - LI_Lineage – processStep - LI_ProcessStep – dateTime), é do tipo dateTime, sendo necessário um valor no formato *yyyy-mm-dd***T***hh***:***mm***:***ss*. O valor informado nos metadados era no formato *yyyy-mm-dd***.**

- **Erros em relação ao XML Schema Perfil MGB:** Além dos erros encontrados também na validação do XML Schema da ISO 19115 foi verificado o não preenchimento do elemento Declaração (dataQualityInfo - DQ_DataQuality – lineage - LI_Lineage - statement). O elemento Declaração é de preenchimento condicional, caso os pacotes Source (dataQualityInfo - DQ_DataQuality – lineage - LI_Lineage – source) e ProcessStep (dataQualityInfo - DQ_DataQuality – lineage - LI_Lineage –

processStep) não estejam preenchidos e o Escopo (dataQualityInfo - DQ_DataQuality – scope - DQ_Scope – level - MD_ScopeCode) for Conjunto de Dados (Dataset) ele é de preenchimento obrigatório.

- **Erros em relação a Codelists:** Não foram encontrados erros em relação ao dicionário de codelists.

- **Erros em relação ao Preenchimento:** Em todos os arquivos foi verificado um erro no preenchimento do elemento Nome do Esquema de banco de dados (applicationSchemaInfo - MD_ApplicationSchemaInformation – name - CI_Citation – title), que foi preenchido como "BDOO Gothic 2.1.3" porém o preenchimento correto é "BDOO Gothic 2.1.0"

.

4. Conclusão

Apesar da presença de erros nos metadados avaliados os erros encontrados não comprometem a qualidade dos metadados.

Todos os erros em relação a Consistência de Formato são puramente estéticos da linguagem XML. Os erros de acurácia temática englobaram os erros pela falta do separador T, que não comprometem a informação, pois a DSG não registra o horário de conclusão dos processos, somente o dia, a solução para esse erro seria padronizar como T00:00:00 o horário de execução por exemplo.

Conforme os resultados obtidos na validação pode-se verificar que os metadados contidos no Banco de Dados Geográficos do Exército são de alta confiabilidade, demonstrando que o BDGEx, em relação a metadados, está pronto para se tornar um nó da Infraestrutura Nacional de Dados Espaciais.

Todos os problemas encontrados nos metadados avaliados serão comunicados a Diretoria de Serviço Geográfico para que sejam sanados.

Referências Bibliográficas

BRASIL (2008). Decreto nº 6.666, de 27 de novembro de 2008. Institui, no âmbito do Poder Executivo federal, a Infra-Estrutura Nacional de Dados Espaciais (INDE), e dá outras providências. 2008. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2008/Decreto/D6666.htm>. Acesso em: 20 jun. 2012.

Comissão Nacional de Cartografia – CONCAR (2009). Perfil de Metadados Geoespaciais do Brasil. Disponível em: <http://www.concar.gov.br/ arquivo/Perfil_MGB_Final_v1_homologado.pdf>. Acesso em: 20 jun. 2012.

Comissão Nacional de Cartografia – CONCAR (2010). Plano de Ação para Implantação da Infraestrutura Nacional de Dados Espaciais. Disponível em: <http://www.concar.gov.br/arquivo/PlanoDeAcaoINDE.pdf>. Acesso em: 20 jun. 2012.

Devillers, R.; Stein, A.; Bédard, Y.; Chrisman, N.; Fisher, P. and Shi, W. (2010). "Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities." Transactions in GIS, Blackwell Publishing Ltd, , 14, 387-400

EuroGeographics (2007). Guidelines for Implementing the ISO 19100 Geographic Information Quality Standards in National Mapping and Cadastral Agencies. Disponível em:
< http://www.eurogeographics.org/documents/Guidelines_ISO19100_Quality.pdf>. Acesso em: 07 julho 2012.

e-PING, Padrão de Interoperabilidade de Governo Eletrônico (2010). Disponível em:
<http://www.governoeletronico.gov.br/acoes-e-projetos/e-ping-padroes-deinteroperabilidade>. Acesso em: 20 junho 2012.

Exército Brasileiro – Diretoria de Serviço Geográfico – DSG. **Especificação Técnica para Controle de Qualidade de Produtos de Conjuntos de Dados Geoespaciai**s, 2012.

Exército Brasileiro – Diretoria de Serviço Geográfico – DSG. **Especificação Técnica para Produtos de Conjuntos de Dados Geoespaciais**, 2012.

Exército Brasileiro – Diretoria de Serviço Geográfico – DSG. **Metodologia de Preparo do Fluxo de Trabalho,** 2012.

Exército Brasileiro – Diretoria de Serviço Geográfico – DSG. **Metodologia para Construção dos Bancos de Dados Regionais de Disseminação**, 2012.

Goldberg, K.. **XML – Guia prático visual**. Alta Books, 2009.

ISO. **ISO 19113:2002. Geographic information - Quality principles**. International Organization for Standardization (ISO), 2002.

ISO. **ISO 19114:2003. Geographic information - Quality Evaluation Procedures**. International Organization for Standardization (ISO), 2003.

ISO. **ISO 19115:2003. Geographic information - Metadata**. International Organization for Standardization (ISO), 2003.

ISO. **ISO 19118:2005. Geographic Information – Encoding**. International Organization for Standardization (ISO), 2005.

ISO. **ISO 19131:2007. Geographic information – Data Product Specifications**. International Organization for Standardization (ISO), 2007.

ISO (2007). "ISO 19139:2007. Geographic information – Metadata XML Schema Implementation." International Organization for Standardization (ISO).