

Materialização do Perfil de Metadados Geoespaciais do Brasil em esquema XML derivado da ISO 19139

Andre Pinto Pascoal¹
Rafael Barboza de Carvalho²
Emerson Magnus de Araújo Xavier³

¹Companhia de Comando da 9ª Região Militar – Cia Cmdo 9ª RM
Av. Duque de Caxias, 1628, Amambai - Campo Grande - MS
andre.pascoal89@gmail.com

²3ª Divisão de Levantamento – 3ª DL
Av. Joaquim Nabuco, 1687 - Olinda - PE
rafael_topografia@hotmail.com

³Diretoria de Serviço Geográfico – DSG
Quartel-General do Exército, Bloco F, 2º Pav, SMU - Brasília - DF
emerson@dsg.eb.mil.br

Abstract. Metadata are descriptive information about some data. This kind of information may indicate relevant data properties, like geographic extension, spatial or radiometric resolution, acquisition date, responsible party, and even if the product have suffered some processing. The Brazilian Geospatial Metadata Profile (Perfil de Metadados Geoespaciais do Brasil - Perfil MGB) was created to normalize the metadata development for Brazilian geospatial products and is based on the ISO 19115:2003 standard, published by the International Organization for Standardization (ISO). This profile defines which elements must be included in a metadata instance that characterizes a product. However, it does not indicate how to store these metadata in an interoperable structure, like a file format. To address this issue, this work presents a set of XML schemas, based on the ISO 19139:2007 standard, in order to implement the Perfil MGB's concepts in a widely-used Web-enabled format - the XML (eXtensible Markup Language). The objective is to permit encoding metadata registries and validate them according current specifications. Data producers and users can take advantages from this approach by using a common format. The obtained results corroborate this solution by correctly identifying validation errors in 50 metadata instances randomly downloaded from geospatial data producers Web sites. Nevertheless, the results also point out the complexity in to maintain procedures of geometadata management, since none of metadata used in the experiment were valid against any of tested grammars.

Palavras-chave: metadata, interoperability, ISO 19115, XML, metadado, interoperabilidade.

1. Introdução

O crescimento da Internet e a disponibilização de diversos serviços, inclusive geográficos, e o seu uso em larga escala, transformou a Web em uma importante opção para o intercâmbio de dados geográficos. Porém, essa facilidade de compartilhamento pode causar problemas quando a informação geográfica não é acompanhada de informações pertinentes ao dado. Essas informações são necessárias para a sua correta utilização, além de prover a confiabilidade necessária. O acesso a esses dados geográficos, provenientes de fontes distintas, tem contribuído com a difusão da informação geográfica nos diversos níveis de usuários, sendo aplicados em diferentes áreas do conhecimento humano.

Atualmente torna-se necessário padronizar as informações que deverão acompanhar o dado, facilitando seu uso. A informação do dado é chamada de metadado, elemento central deste trabalho. Como a Organização da Nações Unidas (ONU, 2007) destaca, os metadados são elementos prioritários de padronização para se viabilizar a interoperabilidade semântica entre sistemas de informação governamentais. Portanto, conhecer as características do dado geográfico é fundamental para que se possa ter um melhor aproveitamento da informação e para que haja interoperabilidade entre os diversos sistemas.

Nesta direção, o Governo Brasileiro edita o Decreto nº 6.666, de 27 de novembro de 2008, que institui a Infraestrutura Nacional de Dados Espaciais (INDE). A INDE tem por finalidade catalogar, integrar e harmonizar os dados geoespaciais existentes nas instituições do Governo Brasileiro, obrigatório para as instituições públicas e opcional para as privadas (Brasil, 2008). Este decreto estabelece em seus artigos 3º e 6º que o compartilhamento e disseminação dos dados geoespaciais e seus metadados é obrigatório para todos os órgãos e entidades do Poder Executivo federal. É de competência da Comissão Nacional de Cartografia (CONCAR) homologar os padrões para a INDE e as normas para a Cartografia Nacional.

A fim de atender à demanda de padronização de metadados surgida com a INDE, a CONCAR elabora o Perfil de Metadados Geoespaciais do Brasil (Perfil MGB), baseado na norma ISO 19115:2003, publicada pela International Organization for Standardization (ISO). Um perfil de metadados é um conjunto básico necessário de elementos que retratam as características dos produtos geoespaciais de uma determinada comunidade e garante sua identificação (ISO, 2003).

O Perfil MGB é um conjunto estruturado de elementos básicos e necessários, que retratam as características dos produtos geoespaciais brasileiros garantindo sua identificação, avaliação e utilização consistente (CONCAR, 2009). Porém, esta norma nacional não define como codificar esses metadados em arquivos.

A arquitetura Padrões de Interoperabilidade de Governo Eletrônico (e-PING) é editada pela Secretaria de Logística e Tecnologia da Informação (SLTI) e regulamenta as atividades de interoperabilidade e governo eletrônico no Brasil (SLTI, 2011). Esta arquitetura adota o eXtensible Markup Language (XML) como padrão primário de intercâmbio de dados para todos os sistemas do setor público.

A norma ISO 19139 tem por objetivo codificar os conceitos da ISO 19115 em XML (ISO, 2007). Esta norma define uma codificação de metadados geoespaciais em esquemas XML derivado da ISO 19115:2003.

Neste contexto, este trabalho propõe materializar os conceitos definidos no Perfil MGB em esquemas XML baseados na norma ISO 19139 a fim de definir um formato de arquivo para disseminação dos metadados geoespaciais do Brasil. Esta solução permite codificar e validar os metadados dentro das normas estabelecidas pela CONCAR.

O restante deste artigo está estruturado da seguinte forma. A segunda seção traz a fundamentação teórica, com os conceitos e teorias que apoiam o trabalho. Na terceira seção é apresentada a materialização do Perfil MGB em esquema XML. O experimento que valida a solução proposta encontram-se na quarta seção. Por fim, a quinta e última seção traz a discussão dos resultados e as conclusões deste trabalho.

2. Fundamentação teórica

O advento da cartografia digital trouxe novas necessidades no registro das informações sobre a produção de dados geoespaciais. Assim surgem os metadados, que segundo CONCAR (2009) podem ser definidos como o conjunto de informações a respeito dos dados espaciais que visam permitir, localizar, descobrir e avaliar uma informação geográfica.

Segundo Brasil (2008), metadados são o conjunto de informações descritivas sobre os dados, incluindo as características de produção, qualidade, armazenamento. Esses “dados sobre dados” são essenciais para identificar, documentar, integrar e disponibilizar os produtos geoespaciais descritos. Desta forma, os metadados permitem descrever as características particulares da informação geográfica tais como: escala, autor, data, projeção, nome do mapa, localização etc., permitindo assim que o usuário conheça as características técnicas de um determinado produto.

Os metadados também são importantes para os produtos geoespaciais derivados de imagens de sensoriamento remoto. Perguntas como “qual a extensão geográfica da imagem?”,

“qual a resolução espacial?” e “qual a data de aquisição?” podem ser respondidas com o uso de um metadado específico. Os estudos de Manso et al. (2004) e Holler et al. (2011) sinalizam para esta necessidade.

2.1 Perfil de Metadados Geoespaciais do Brasil (Perfil MGB)

Segundo Brasil (2008) cada órgão e entidade do poder público federal deve disseminar e compartilhar os metadados geoespaciais. Desta forma faz-se necessário a utilizar normas na confecção dos metadados, pois sendo gerados por órgãos diferentes, porém seguindo a mesma norma, garantir-se-á a interoperabilidade.

A finalidade do Perfil MGB é estabelecer um padrão nacional para estruturação dos metadados. Foi elaborado com a participação de representantes dos órgão produtores de dados geoespaciais do Brasil. O perfil foi inteiramente baseado na norma internacional ISO 19115 (ISO, 2003). O Perfil MGB possui dois níveis de conformidade: sumarizado e completo. O sumarizado define basicamente os elementos no núcleo da ISO 19115, enquanto o completo abrange boa parte da norma internacional.

A ISO 19115:2003 utiliza a linguagem UML (Unified Modeling Language) para representar as seções dos metadados. É uma norma muito ampla, possui cerca de 400 elementos sendo oito obrigatórios para perfis derivados deste padrão. É um padrão internacional adotado por diversos países, tais como Portugal, Espanha, Estados Unidos, Canadá e Peru. Esta norma permite definir perfis e extensões para campos específicos de aplicações (CONCAR, 2009).

Na construção do Perfil MGB foi executada uma análise do padrão ISO 19115, a fim de identificar as seções, entidades e elementos que descrevem as características dos dados geoespaciais produzidos no Brasil. Dessa forma, o perfil define um subconjunto dos elementos disponíveis na ISO 19115, altera a cardinalidade de alguns elementos, e define algumas listas de códigos.

2.2 ISO 19139:2007 e esquemas XML

A ISO 19139:2007 define um conjunto de esquemas XML para metadados definidos na ISO 19115:2003. Seu objetivo é definir um formato de arquivos para metadados geoespaciais que seguem a norma ISO 19115:2003. Esses esquemas permitem estruturar e validar os arquivos XML dos metadados em conformidade com a norma (ISO, 2007).

O formato XML permite que as instâncias de metadados possam trafegar pela Internet inclusive em serviços Web geoespaciais, como previsto na especificação do Catalogue Services – Web (Nebert et al., 2007).

Mantido desde 1998 pelo World Wide Consortium (W3C), o XML é uma linguagem de marcação flexível e simples. Apresenta como características principais ser uma linguagem baseada em texto, ter separação do conteúdo da formatação, ser simples e fácil de ser interpretado e ter a possibilidade de criação de *tags* sem limitações, facilitando o intercâmbio de dados pela Internet (Goldberg, 2009).

Um documento XML pode ser considerado bem formado se estiver de acordo com o que prescreve as normas (Bray et al., 2006). Este documento pode ser válido desde que obedeça a algumas normas descritas na sua gramática. O XML Schema Definition (XSD) é um formato de gramática para XML (Fallside e Walmsley, 2004). A ISO 19139 é descrita sob a forma de esquemas XML construídos sob a especificação de Thompson et al. (2004).

3. Materialização do Perfil MGB em esquema XML

Esta seção apresenta a metodologia desenvolvida neste trabalho para materializar os conceitos do Perfil MGB (sumarizado e completo) em esquemas XML derivados da ISO 19139. Como o Perfil MGB é uma visão da ISO 19115, este trabalho representa esta visão da ISO 19139, que codifica os elementos da ISO 19115. Os esquemas XML da ISO 19139 são

usados como base para as adaptações necessárias à realidade nacional.

A adaptação ao Perfil MGB sumarizado exigiu poucas mudanças nos esquemas da ISO 19139. A Tabela 1 apresenta os elementos que foram alterados. Basicamente foi alterada a cardinalidade de alguns elementos da ISO em relação ao Perfil sumarizado (Sum).

Tabela 1. Cardinalidade dos elementos alterados para o Perfil MGB sumarizado.

Entidade	Elemento XML	Cardinalidade	
		ISO	Sum
Responsável	MD_Metadata.identificationInfo>MD_DataIdentification.pointOfContact	0..n	1..n
Categoria Temática	MD_Metadata.identificationInfo>MD_DataIdentification.topicCategory	0..n	1..n
Formato de Distribuição	MD_Metadata.distributionInfo>MD_Distribution.distributionFormat	0..n	1..n
Sistema de Referência	MD_Metadata.referenceSystemInfo	0..n	1..n
Status	MD_Metadata.identificationInfo>MD_DataIdentification.status	0..n	1..n

3.1 Perfil MGB completo

Esta subseção descreve alguns dos procedimentos necessários para criar os esquemas XML para o Perfil MGB completo.

O MD_Metadata é o elemento raiz de uma instância de metadados geoespaciais. Ele contém informações das entidades, define e estabelece hierarquias dos metadados de um produto.

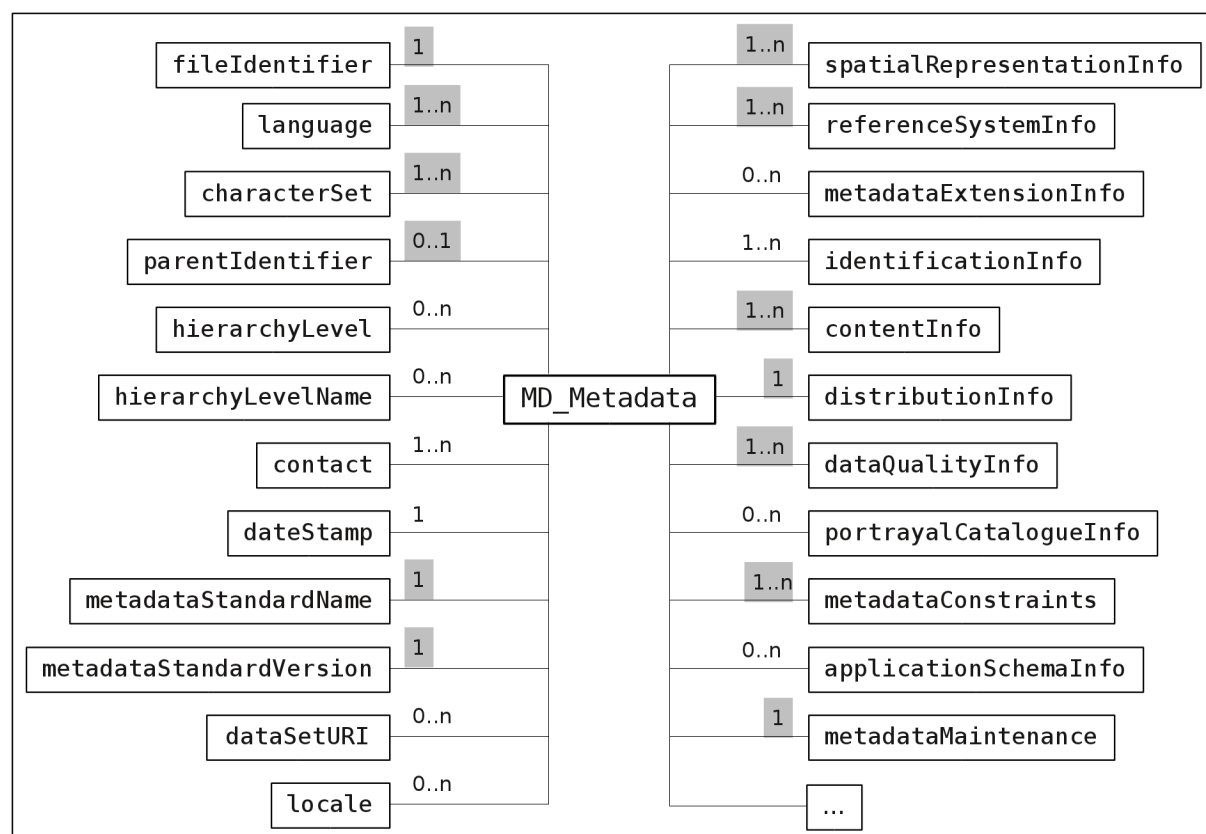


Figura 1. Componentes do MD_Metadata com suas cardinalidades.

Para a construção do esquema XML para o Perfil MGB completo foi efetuada uma análise de todos os elementos descritos no perfil para as adequações necessárias no esquema da ISO 19139. Uma alteração frequente nesta atividade foi alterar as cardinalidades dos elementos obrigatórios e opcionais. Um exemplo encontra-se na Figura 1, que apresenta o elemento raiz com seus componentes e suas respectivas cardinalidades. As alterações neste nível estão destacadas em cinza.

Na análise realizada verifica-se, por exemplo, que o MD_Metadata apresenta um componente chamado “language”, que conforme a norma ISO 19115 encontra-se com cardinalidade 0..1 (opcional). Porém o Perfil MGB determina que a cardinalidade exigida é 1..n (um ou muitos). A análise também indicou que não foi necessário criar nenhum elemento novo no esquema, sendo os elementos disponíveis na ISO suficientes para adaptá-la às necessidades do perfil.

A Seção “Identification” é um dos componentes obrigatórios do MD_Metadata e precisou sofrer várias alterações para se adequar ao perfil nacional. Esta seção possui vários elementos: Citação, Resumo, Objetivo, Créditos, Status, Ponto de Contato, Palavras Chaves Descritivas, Pré-visualização Gráfica e Informação de Dados Agregados. Um exemplo das adaptações necessárias ocorreu no elemento Ponto de Contato (Point of Contact), que na norma ISO 19115:2003 aparece no MD_Identification com cardinalidade opcional e no Perfil MGB com cardinalidade obrigatória.

Outras mudanças foram necessárias, como a adaptação dos valores do tipo “MD_TopicCategoryCode” na seção “Identification” da ISO. Este tipo de valor é apresentado na ISO como uma enumeração, e não uma lista controlada. Mas o Perfil MGB redefine os valores para uma lista controlada nacional, e adiciona outros valores possíveis ao tipo original. Face o exposto, este tipo de elemento precisou ser alterado no esquema XML original da ISO 19139.

O Perfil MGB define diversas listas controladas para preencher valores nos metadados. Porém estas listas não requerem nenhuma mudança no esquema XML.

4. Experimento: validando metadados em arquivos XML

O experimento para validar a solução proposta foi dividido em duas fases. Na primeira fase foi realizado um experimento controlado, visando verificar a validade dos esquemas propostos. Na segunda etapa foram utilizadas 50 instâncias de metadados obtidos na Internet de produtores de dados nacionais, onde se espera que o Perfil MGB seja aplicado. Para a validação dos arquivos XML contra os esquemas foi utilizado o software Apache Xerces (Apache, 2012).

O experimento controlado consistiu em verificar a validade dos esquemas XML gerados a partir dos esquemas da ISO 19139. Foram elaborados dois tipos de esquemas: um para o Perfil MGB sumarizado e outro pro completo. Duas instância de metadados em XML foram preparadas para as duas situações. Estes arquivos foram revisados com base nas definições do perfil para evitar qualquer dúvida. A validação ocorreu sem problemas, permitindo verificar a validade dos esquemas nesta situação bem definida.

Na segunda fase do experimento, os 50 arquivos XML de metadados obtidos na Web foram validados contra três esquemas XML diferentes: ISO 19139:2007, Perfil MGB sumarizado e completo. Nenhum dos 50 metadados avaliados estava válido em qualquer dos esquemas XML utilizados. Foram encontrados aproximadamente 1500 erros em média para cada esquema utilizado. A Tabela 2 sumariza os erros mínimo, médio e máximo e a quantidade total de erros de validação encontrados para os 50 arquivos contra os três esquemas. Pelo menos um erro foi encontrado em cada arquivo, o que o torna inválido contra a gramática considerada. A Tabela 2 permite inferir que à medida que o esquema vai ficando mais restrito, ou seja, com mais regras, a quantidade de erros identificados tende a aumentar.

Tabela 2. Erros de validação encontrados para os três esquemas.

Esquema	Erro mínimo	Erro médio	Erro máximo	Total
ISO 19139	1	28	80	1415
Sumarizado	2	29	80	1446
Completo	5	33	84	1637

Foram encontrados diversos tipos de erros de validação, sendo os mais comuns: (1) elemento não permitido, (2) elemento não definido, (3) atributo não definido, (4) valor inválido, e (5) elemento não encontrado. Esses cinco tipos de erros correspondem a mais de 99% do total encontrado. A Figura 2 apresenta a proporção de cada tipo de erro encontrado no experimento para o Perfil MGB sumarizado, juntamente com dois exemplos para os erros mais comuns.

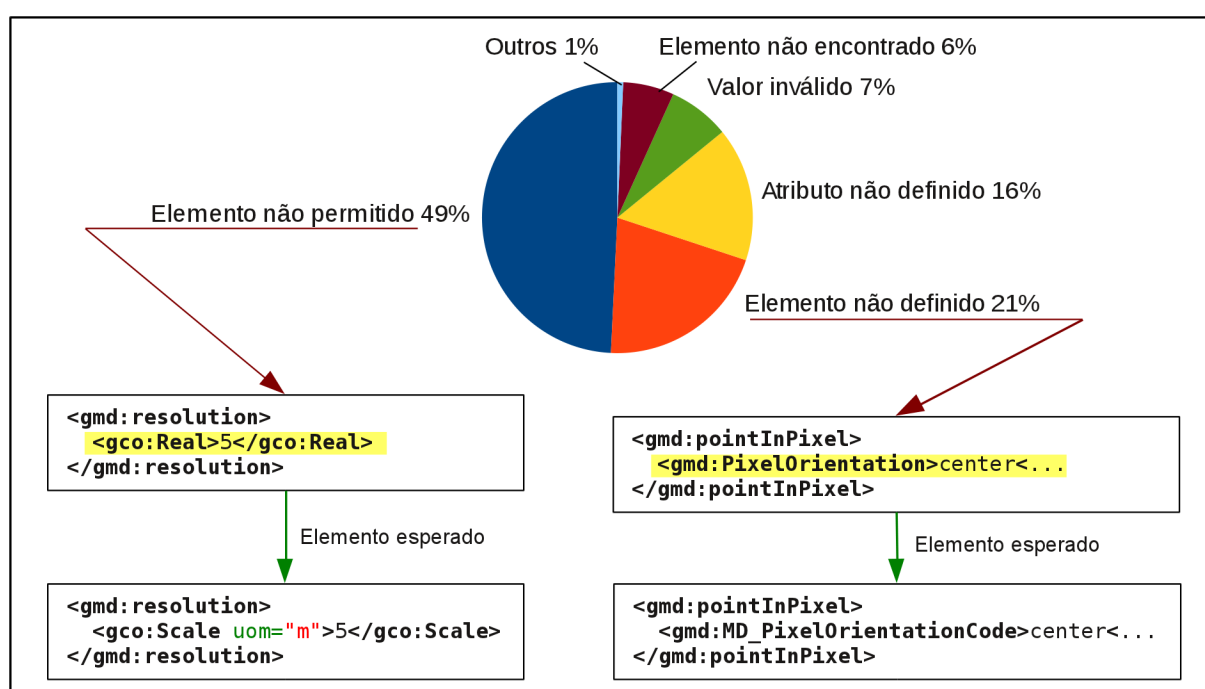


Figura 2. Proporção de erros encontrados para o Perfil sumarizado, com exemplos.

Os erros de tipo 1 ocorrem quando o esquema define um elemento mas no XML aparece outro, que existe no esquema mas não é o previsto. Por exemplo, o esquema requer um elemento “Boolean” numa certa posição mas no documento XML está um “CharacterString”. Os erros de tipo 2 ocorrem quando são utilizados elementos que não existem no esquema. Similar aos erros de tipo 3, só que neste são usados atributos não definidos. Estes dois últimos constituem erros graves de validação, pois elementos não definidos no esquema também não o são em especificações técnicas, impossibilitando identificar seu significado ou objetivo. Um erro de tipo 4 comum é usar um número real quando o esquema espera um inteiro. As faltas de um elemento obrigatório são classificadas como erros de tipo 5.

Outra análise possível dos resultados deste experimento envolve a comparação do número de elementos de um metadado com os erros de validação identificados no documento que o materializa. Quanto mais elementos descritivos do produto no metadado, maiores as chances de ocorrerem erros de validação. O gráfico na Figura 3 apresenta os resultados para os erros de validação proporcionais ao número de elementos XML nos 50 arquivos avaliados para os

esquemas da ISO 19139:2007 (identificado como *ISO*), Perfil MGB sumarizado (*Sum*) e completo (*MGB*).

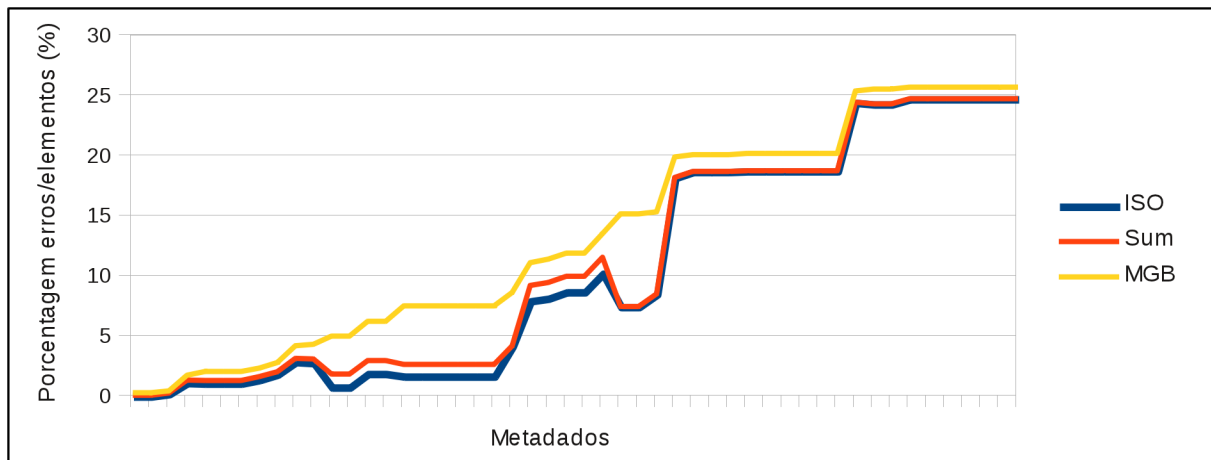


Figura 3. Taxa de erro na validação para os três esquemas considerados.

A Figura 3 mostra que apenas cerca de 20% dos arquivos analisados apresentou uma taxa de erro na validação inferior a 5%. Alguns arquivos apresentaram uma taxa de erro bem inferior à média encontrada no experimento, variando de 7 erros no esquema da ISO até 12 no perfil completo, para mais de 2200 elementos descritivos. Porém, outros metadados codificados em XML tiveram uma taxa de erro bem superior à média, chegando a mais de 25%. Isso significa que a quarta parte do arquivo possui elementos que não são válidos para nenhum dos esquemas constantes do experimento.

5. Discussão dos resultados e conclusões

A solução apresentada neste trabalho consiste em materializar o Perfil de Metadados Geoespaciais do Brasil (Perfil MGB), tanto na forma sumarizada como na completa, em esquemas XML derivados da norma ISO 19139:2007. O objetivo é produzir um mecanismo para que as instituições produtoras de dados geoespaciais possam estruturar e validar os metadados dos seus produtos de acordo com as normas vigentes. A comunidade nacional de geoinformação também se beneficia desta solução ao poder contar com uma estrutura de padronizada e pronta para a Web. Os experimentos realizados apontam para a validade desta abordagem ao definir esquemas XML para o Perfil MGB capazes de identificar falhas em metadados que devem seguir o perfil nacional.

Este artigo tem como foco principal a interoperabilidade sintática na produção e disseminação dos metadados de geoinformação. O uso do esquema XML pode garantir que um usuário receba um metadado com a completude definida para as necessidades brasileiras. A principal contribuição deste trabalho é apresentar e discutir uma solução para o intercâmbio de arquivos de metadados geoespaciais que seguem o Perfil MGB.

As limitações dessa solução vêm da complexidade em montar uma estrutura de produção e manutenção de metadados geoespaciais, realidade até há pouco distante da cultura nacional. Definir uma estrutura para os arquivos não significa, necessariamente, que as informações codificadas estejam corretas. Porém, garantir uma estrutura comum é mais um passo em direção à qualidade.

Os resultados do experimento indicam essa complexidade posto que nenhuma das 50 instâncias de metadados, obtidas aleatoriamente em *sites* de produtores de geodados brasileiros, era válida contra nenhuma das três gramáticas utilizadas. Estes resultados apontam para a necessidade de adequação às regras definidas no Perfil MGB.

Referências Bibliográficas

Apache. **Xerces-C++ XML Parser**. Apache Software Foundation, 2012. Disponível em: <<http://xerces.apache.org/xerces-c/>>. Acesso em: 16 nov. 2012.

Brasil. **Decreto Presidencial 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. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2008/Decreto/D6666.htm>. Acesso em: 15 nov. 2012.

Bray, T.; Paoli, J.; Sperberg-McQueen, C. M.; Maler, E.; Yergeau, F. **Extensible Markup Language (XML) 1.0**. 4th ed. World Wide Web Consortium (W3C), 2006. Disponível em: <<http://www.w3.org/TR/2006/REC-xml-20060816/>>. Acesso em: 16 nov. 2012.

CONCAR. **Perfil de Metadados Geoespaciais do Brasil**. Comissão Nacional de Cartografia (CONCAR), 2009. Disponível em: <http://www.concar.ibge.gov.br/arquivo/Perfil_MGB_Final_v1_homologado.pdf>. Acesso em: 15 nov. 2012.

Fallside, D. C.; Walmsley, P. **XML Schema Part 0: Primer**. 2nd ed. World Wide Web Consortium (W3C), 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-xmlschema-0-20041028/>>. Acesso em: 16 nov. 2012.

Goldberg, K. H. **XML – Guia prático visual**. Rio de Janeiro: Alta Books, 2009. 280 p.

Holler, W. A.; Custódio, D.; Niemann, R. S.; Rôvere, S. D. Proposta de organização e visualização de grande quantidade de imagens orbitais através de web services. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 15., 2011, Curitiba. **Anais...** São José dos Campos: INPE, 2011. Artigos, p. 8924-8930.

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

ISO. **ISO 19139:2007. Geographic information - Metadata - XML schema implementation**. International Organization for Standardization (ISO), 2007.

Manso, M. A.; Noguera-Iso, J.; Bernabé, M. A.; Zarazaga-Soria, F. J. Automatic Metadata Extraction from Geographic Information. In: AGILE Conference on Geographic Information Science, 7., 2004, Heraklion, Greece. **Proceedings...** p. 379-385.

Nebert, D.; Whiteside, A.; Vretanos, P. A. **OpenGIS® Catalogue Services Specification**. Version 2.0.2, OGC 07-006r1. Open Geospatial Consortium (OGC), 2007. 218 p. Disponível em: <http://portal.opengeospatial.org/files/?artifact_id=20555>. Acesso em: 17 nov. 2012.

ONU. **Libro Blanco de Interoperabilidad de Gobierno Electrónico para América Latina y el Caribe**. Versión 3. Organização da Nações Unidas (ONU), 2007. Disponível em: <<http://www.cepal.org/ddpe/publicaciones/xml/2/37352/Libroblanco.pdf>>. Acesso em: 16 nov. 2012.

SLTI. **e-PING - Padrões de Interoperabilidade de Governo Eletrônico**. Secretaria de Logística e Tecnologia da Informação do Ministério do Planejamento (SLTI), 2011. Disponível em: <<http://www.governoeletronico.gov.br/biblioteca/arquivos/documento-da-e-ping-versao-2012/>>. Acesso em: 16 nov. 2012.

Thompson, H. S.; Beech, D.; Maloney, M.; Mendelsohn, N. **XML Schema Part 1: Structures**. 2nd ed. World Wide Web Consortium (W3C), 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-xmlschema-1-20041028/>>. Acesso em: 16 nov. 2012.