

LINGUAGENS DE MARCAÇÃO ESPECÍFICAS POR DOMÍNIO E METADADOS DESCRITIVOS: funções para a descoberta de recursos científicos¹

DOMAIN-SPECIFIC MARKUP LANGUAGES AND DESCRIPTIVE METADATA: their functions in scientific resource discovery

Marcia Lei Zeng
School of Library and Information Science,
Kent State University, Kent, Ohio, USA
mzeng@kent.edu

RESUMO

Enquanto os metadados tiveram grande foco em publicações, projetos e iniciativas dos profissionais da informação durante as últimas duas décadas, um número significativo de linguagens de marcação específicas por domínio também se desenvolveram paralelamente a uma taxa equivalente aos padrões de metadados; mas ainda não recebem atenção comparável. Esse artigo discute as funções desses dois tipos de abordagens na descoberta de recursos científicos e aponta papéis potenciais e complementares por meio de abordagens de interoperabilidade apropriadas.

PALAVRAS-CHAVE: Metadado. Linguagem de marcação. Recurso científico.

1 DESENVOLVIMENTO DE PADRÕES DE METADADOS E DE LINGUAGENS DE MARCAÇÃO ESPECÍFICAS POR DOMÍNIO

Os metadados são definidos como “informação estruturada que descreve, explica, localiza, ou então, facilita a recuperação, o uso ou o gerenciamento de uma fonte de informação” (NISO, 2004)². Muitos padrões de metadados foram criados por uma variedade de comunidades. Os exemplos incluem:

- a) Padrões de metadados aplicados em muitas áreas do assunto e de recursos:
 - *Dublin Core Metadata Element Set (DCMES)* – Conjunto de Elementos de Metadados Dublin Core;
 - *Dublin Core Metadata Terms – Termos de Metadados Dublin Core*;
 - *Electronic Theses and Dissertations Metadata Standard (ETD-MS)* – Padrão de Metadados de Teses e Dissertações Eletrônicas;
 - *Learning Object Metadata (LOM)* – Metadados de Objetos de Aprendizagem;
 - *Metadata Object Description Schema (MODS)* – Esquema de Descrição de Objetos de Metadados.
- b) Padrões de metadados em áreas científicas:
 - *ADN (ADEPT/DLESE/NASA) Metadata Framework – Estrutura de Metadados ADN* – para a comunidade do sistema educacional da Terra;



Esta obra está licenciada sob uma [Licença Creative Commons](https://creativecommons.org/licenses/by/4.0/).

DOI 10.5007/1518-2924.2010v15nesp2p164

¹ Tradução de Gleisy R. B. Fachin, Departamento de Ciência da Informação, Universidade Federal de Santa Catarina – gleisy@cin.ufsc.br.
Revisão de Maria de Cléofas Faggion Alencar, Embrapa Meio Ambiente – cleo@cnpma.embrapa.br.

² NISO. *Understanding Metadata*. Bethesda, MD: NISO Press. <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>.
Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf., Florianópolis, n. esp., 2º sem. 2010. ISSN 1518-2924.

- *Content Standards for Digital Geospatial Metadata (CSDGM)* – Padrões de Conteúdo para Metadados Geoespaciais Digitais;
- *Darwin Core* – um padrão para descrever objetos contidos em coleções de espécies da história natural e bases de dados de observação de espécies;
- ISO/TS 19115:2003 *Geographic information Metadata – Metadados de Informação Geográfica*.

Durante a evolução da nossa era digital, a XML (*Extensible Markup Language* - Linguagem de Marcação Extensível) – desenvolvida por um grupo de trabalho em XML, anteriormente formado sob os auspícios da *World Wide Web Consortium (W3C)* em 1996 (continuamente atualizado) – se transformou na metalinguagem oficial de muitas comunidades. Simplificando, a XML “descreve uma classe de objetos de dados chamados documentos XML e, parcialmente, descreve o comportamento de programas de computador que os processam”³. As marcações codificam uma descrição do *layout* de armazenagem do documento e a estrutura lógica. A XML fornece um mecanismo para impor restrições no *layout* de armazenamento e na estrutura lógica. O W3C não criou uma linguagem de marcação baseada no domínio, ao contrário, deixou para os criadores específicos de cada domínio fazerem isso. Assim, as linguagens de marcação específicas por domínio também foram desenvolvidas por várias comunidades, particularmente em áreas científicas. Por exemplo:

a) Linguagens de marcação específicas por domínio de alto nível:

- MathML (*Mathematical Markup Language – Linguagem de Marcação Matemática*) – contém dois conjuntos de etiquetas: uma para a apresentação de fórmulas; outra para o significado de expressões matemáticas. Essa condição fornece uma base para a inclusão de expressões matemáticas em publicações na *Web*;
- GML (*Geography Markup Language – Linguagem de Marcação para Geografia*) – uma gramática XML para expressar características geográficas;
- CML (*Chemical Markup Language – Linguagem de Marcação para Química*) – concentrada em “moléculas” (entidades distintas representadas por uma fórmula e geralmente uma tabela de ligações), essa linguagem mantém uma hierarquia para compostos de moléculas, reações e estruturas/seqüências macromoleculares;
- MatML (*eXtensible Markup Language for Materials Property Data – Linguagem de Marcação Extensível para Dados de Propriedades de Materiais*);
- MML (*Medical Markup Language – Linguagem de Marcação Médica*);
- [entre outras]

b) Outras linguagens de marcação especializadas ou perfis de aplicação que estenderam ainda mais as linguagens de marcação de alto nível. Utilizando a GML como um exemplo, mais de duas dúzias de padrões foram ou estão sendo desenvolvidos, como:

- GML (*Geography Markup Language – Linguagem de Marcação para Geografia*):
 - ✓ TWML (*Tsunami Warning Markup Language – Linguagem de Marcação para Aviso de Tsunamis*) - Uma linguagem baseada em padrões para Comunicados de Tsunamis;
 - ✓ *Canadian Road Markup Language* – Linguagem de Marcação para Estradas Canadenses (para o Arquivo da Rede de Estradas);
 - ✓ CityGML;
 - ✓ CWML (*Cyclone Warning Markup Language – Linguagem de Marcação de Aviso de Ciclones*);

³ Extensible Markup Language (XML) 1.0 (Fifth Edition), W3C Recommendation 26 November 2008. <http://www.w3.org/TR/REC-xml/>.
Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf., Florianópolis, n. esp., 2º sem. 2010. ISSN 1518-2924.

- ✓ dwGML (*Digital Weather Geography Markup Language – Linguagem de Marcação para a Geografia de Condições Meteorológicas Digitais*);
- ✓ GDF-GML (*Geographic Data Files – Arquivos de Dados Geográficos*);
- ✓ GML 3.1.1 – Esquema de aplicação para produtos de observação da Terra;
- ✓ *GPlates Markup Language – Linguagem de Marcação GPlacas* (representação de dados geológicos em um contexto de placas tectônicas);
- ✓ SoTerML (*Soil and Terrain Markup Language – Linguagem de Marcação para Solos e Regiões*);
- ✓ [E, aproximadamente, 20 outras]⁴.

A Figura 1 ilustra as relações de metalinguagens e as linguagens de marcação específicas que são desenvolvidas para tipos particulares de recursos ou para domínios de assuntos específicos.

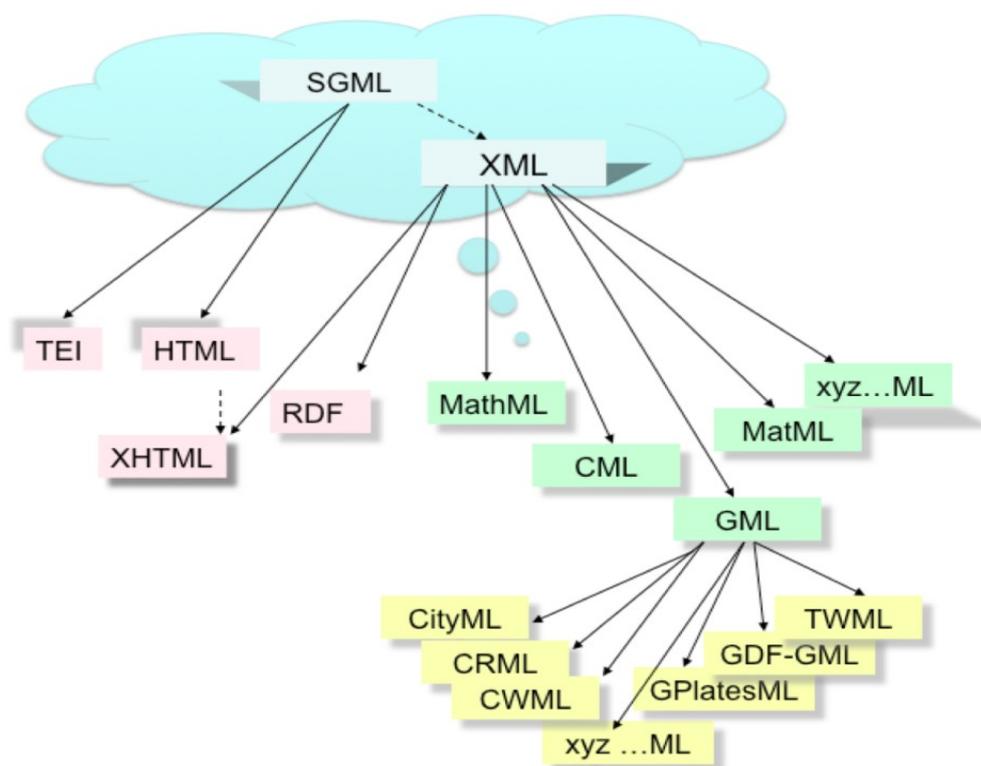


Figura 1 – Metalinguagem e Linguagens de Marcação Específicas (uma lista representativa).

Fonte: A autora.

O crescimento, tanto dos padrões de metadados como de linguagens de marcação específicas por domínio nas áreas científicas, reflete a grande necessidade de representação de dados mais estruturados e padronizados para fins de descoberta. Eles são desenvolvidos para fins de descoberta específicos.

2 MELHORIA DA FUNÇÃO DE DESCOBERTA POR MEIO DA DECOMPOSIÇÃO DE ESTRUTURA E CONTEÚDO

Tanto para padrões de metadados ou padrões de linguagem de marcação, todas as metodologias esforçam-se para descrever efetivamente recursos de informação com o

⁴ GML Application Schemas and Profiles. <http://www.ogcnetwork.net/node/210>.

propósito da descoberta. Se olharmos para os recursos de informação propriamente ditos, o que há para ser descoberto pode estar em diferentes níveis atômicos. A maioria dos recursos de informação (por exemplo, uma base de dados, uma dissertação, um artigo ou um manual) pode ser entendida como *containers* de informação, os quais possuem estruturas internas que são do interesse dos usuários. As coleções e repositórios digitais utilizam, principalmente, o nível de descrição dos *containers* de informação com propósito de organizar, acessar, usar e reusar o conhecimento científico. As diferenças residem em algumas coleções e repositórios digitais distintos que possuem granularidade maior de alguns *containers* do que de outros.

O conceito de “decomposição” foi primeiramente usado e testado no projeto Green's Functions Research and Education Enhancement Network - Rede de Melhoramento da Educação e Pesquisa, da Função de *Green* (GREEN, presidido pelo Dr. Greg Shreve, do Instituto de Linguística, da Universidade de *Kent State* (*Kent State University Applied Linguistics Institute*, onde este autor também tem a função de Co-Investigador Principal). A Função GREEN é uma técnica analítica amplamente utilizada, aplicada em muitas disciplinas e em muitas aplicações, para estudar um espectro maior de diferentes fenômenos, variando entre precipitação de reservatórios de gás e óleo até a deformação termoelástica em *chips* de silicone⁵ (SHREVE; ZENG, 2003). O projeto GREEN é um dos projetos financiados pela National Science Foundation (NSF) – Projeto National Science, Mathematics, Engineering, and Technology Education Digital Library (NSDL) de 2001 a 2004. As discussões neste artigo estão baseadas nos estudos iniciais do projeto⁶ e representam também explorações metodológicas e funcionais mais profundas de metadados e linguagens de marcação.

2.1 Decomposição de estrutura

Com o intuito de revelar a riqueza de conteúdos de um recurso, geralmente, a primeira tentativa é descrever **a estrutura dos componentes** em nível granular maior. Em um registro de dados estruturado, por exemplo, os componentes consistiriam de elementos detalhados: título, resumo ou tabela de conteúdos, índice, referências, imagens contidas, ilustrações, tabelas, glossário, e assim por diante. Os padrões de metadados comumente denotam elementos específicos para algum dos componentes. Os exemplos a seguir são do *DCMI Metadata Terms – Termos de Metadados DCMI*; cujos prefixos são ”dt” (o *qname* do *namespace* DC-15) ou “dct” (para os *namespace* dos Termos *DCMI*, que também incluem os elementos DC-15). Esses elementos são: “*dc:title*”, “*dct:abstract*”, “*dct:bibliographicCitation*”, “*dct:references*” e “*dct:tableOfContents*”. Outros padrões de metadados têm métodos similares para decompor as estruturas. Uma abordagem arbitrária é desmembrar esses componentes e tratá-los como recursos individuais com descrições de metadados individuais. Por exemplo, podemos criar um registro para cada capítulo importante de uma dissertação, além do registro para a dissertação como um todo. Os componentes são ligados ao todo por um elemento de “relação” (por exemplo, “*dct:hasPart*” e “*dct:isPartOf*”). Com essa abordagem, os metadados são adaptados para tratar as tarefas de descrição dos componentes, enquanto apresentam relações estruturais (em vez de relações semânticas) entre dois ou mais componentes.

⁵ SHREVE, Gregory M.; ZENG, Marcia Lei. Integrating Resource Metadata and Domain Markup in an NSDL Collection. In: *DC-2003: Proceedings of the International DCMI Metadata Conference and Workshop*, Sep. 28-Oct. 2, 2003, Seattle, Washington: p. 223-229.

Disponível em: http://www.siderean.com/dc2003/604_paper62.pdf.

⁶ *Ibid.*

Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf., Florianópolis, n. esp., 2º sem. 2010. ISSN 1518-2924.

2.2 Decomposição de conteúdo

Outra tentativa é expor o(s) tópico(s) que um recurso de informação discute ou coisas que ele representa por meio da decomposição do **conteúdo** de um recurso. Para isso, os metadados e linguagens de marcação têm níveis de efetividade diferentes.

2.2.1 Metadados de “assunto” melhorados

Um elemento universal de quase todos os padrões de metadados é o “assunto”. Outros elementos relacionados a “assunto” estão incluídos em maior ou menor quantidade nesses padrões, como por exemplo (com o prefixo denotando o *namespace* deles) “*dc:spatial*”, “*dc:temporal*”, “*vra:culturalContext*”, “*vra:stylePeriod*”, “*vra:material*”, “*lom:taxon*”, “*lom:keywords*”, “*lom:purpose*” e assim sucessivamente.

Os perfis de aplicação de metadados e os novos esquemas desenvolvidos baseados em padrões de metadados gerais podem adicionar mais requisitos, de modo a aumentar e assegurar os elementos de assunto específicos por domínio. Um exemplo do Esquema de Metadados da National Library of Medicine – Biblioteca Nacional de Medicina (NLM): “*dc:subject*” é estendido ao “*DC.Subject.MeSH*” e ao “*NLMDC.Subject.NLMClass*”. O MeSH (Medical Subject Headings – Cabeçalhos de Assunto Médico) e as Classificações NLM contém conceitos e classes de conceitos altamente específicos. Os elementos de metadados dedicados permitirão acesso consistente e sistemático aos recursos em áreas médicas e relacionadas.

Entretanto, existem dois problemas com essas práticas: 1) os elementos de metadados relacionados a assunto são muito restritos nos conjuntos de elementos de metadados; 2) para um “*container*”, uma ocorrência de metadados está sempre na posição de um substituto. Tomando como exemplo uma dissertação em Engenharia de Materiais: mesmo que um registro de metadados proporcione acesso detalhado em relação ao assunto, por exemplo, dez palavras-chave específicas, que fornecem nomes de materiais ou problemas discutidos na dissertação, esse registro só satisfaz a tarefa de “apontar uma direção” em um esforço de descoberta. A abordagem de decomposição de estrutura mencionada previamente pode trazer melhor acesso ao assunto, porque os tópicos diferentes de cada capítulo da dissertação emergem quando o capítulo possui um registro de metadados descritivos individual. Ainda assim, a descrição dos metadados é para o *container*.

2.2.2 Metadados orientados por domínio altamente enumerativos

Os esquemas de metadados ou perfis de aplicação apontados para áreas científicas estenderam seus elementos além da descrição de um *container* e enfatizam a descrição específica orientada por domínio de assunto. Aqui, podemos utilizar dois padrões para demonstrar tais tendências. Note que, no texto que segue, as categorias de elementos estão incluídas entre chaves {...} e elementos entre aspas “...”.

A Estrutura *ADN*⁷ foi construída sobre as especificações e as melhores práticas do projeto *Alexandria Digital Earth Prototype* (ADEPT) - *Protótipo Terra Digital de Alexandria*, da *Digital Library for Earth System Education* (DLESE) - *Biblioteca Digital para a Educação do Sistema Terra*), *Joined Digital Library* (NASA) - *Biblioteca Digital Participativa*. Das

⁷ ADN Framework. <http://www.dlese.org/Metadata/adn-item/index.php>.
Enc. Bibli. R. Eletr. Bibliotecon. Ci. Inf., Florianópolis, n. esp., 2º sem. 2010. ISSN 1518-2924.

nove categorias (*{Administration}*, *{Creator-cataloger}*, *{Educational}*, *{General}*, *{Geospatial}*, *{Relation}*, *{Rights}*, *{Technical}* e *{Temporal}*), a categoria *{Geospatial}* é uma das mais específicas criadas com o propósito de descoberta. “Especialistas em Biblioteconomia, em Engenharia de *softwares* e Geociência da *Alexandria Digital Library - Biblioteca Digital de Alexandria*, a *Colorado School of Mines - Escola de Minas do Colorado* e o Centro do Programa DLESE participaram do desenvolvimento da estrutura dos conceitos geoespaciais, por um período de 18 meses”⁸. Os elementos capturados dessa categoria *{Geospatial}* formam uma lista impressionante: “*Overarching bounding box*”, “*Detailed geometries*”, “*Elevation*”, “*Planet or body*”, “*Place-name tied to coordinates*”, “*Place-name not tied to coordinates*”, “*Event name tied to coordinates*”, “*Event name not tied to coordinates*”, “*Objects in space*”, “*Coordinate system*”, “*Vertical and horizontal datum*”, “*Projection*”, e “*Latitudes and longitudes*”. Para materiais educacionais e, especialmente quando o nível granular é elevado, essa estrutura de metadados parece ter funcionado com sucesso para a Biblioteca Digital DLESE (um recurso comunitário de geociência que apoia o ensino e aprendizado sobre o Sistema Terrestre).

O segundo exemplo são os termos de metadados do *Darwin Core*, que apresenta os termos de metadados (isto é, elementos) em nove categorias principais: *{Record-level Terms}*, *{Occurrence}*, *{Event}*, *{Location}*, *{GeologicalContext}*, *{Identification}*, *{Taxon}*, *{ResourceRelationship}* e *{MeasurementOrFact}*. Mais de cem termos de metadados contribuem com esse esquema, por exemplo, *{Event}* tem 15 termos; *{dcterms:Location}* possui mais de 40 termos; *{geologicalContextID}* tem 18 termos; e *{MeasurementOrFact}* tem nove termos. O que uma categoria inclui são elementos muito específicos que devem ser usados para descrever muitas características dos conteúdos em um recurso. Tendo a categoria *{Taxon}*, alguns termos são para as legendas “*scientificName*”, “*acceptedNameUsage*”, “*parentNameUsage*”, “*originalNameUsage*”, “*nameAccordingTo*”, “*namePublishedIn*”, “*vernacularName*”, “*nomenclaturalCode*”, e outros são para a taxonomia “*higherClassification*”, “*kingdom*”, “*phylum*”, “*class*”, “*order*”, “*family*”, “*genus*”, “*subgenus*”, “*specificEpithet*”, “*infraspecificEpithet*”, “*taxonRank*”, “*verbatimTaxonRank*”. Esse padrão já foi utilizado, com sucesso, como base para outros perfis de aplicação, como o utilizado pelo conhecido Sistema de Informação Biogeográfica do Oceano (OBIS)⁹. O OBIS publica dados em nome de cientistas de agências governamentais, museus, universidades, companhias comerciais e organizações não-governamentais¹⁰. O esquema OBIS é uma extensão do padrão *Darwin Core*, Versão 2. Quando as questões são enviadas para os contribuintes de dados distribuídos, o portal e o conjunto de dados do OBIS utilizarão esses elementos (campos) para transferir as necessidades de informações e os resultados.

2.2.3 Linguagens de marcação específicas por domínio

As linguagens de marcação têm como ponto de partida a função de revelar os conteúdos de um recurso. Usando um padrão de linguagem de marcação, os elementos úteis em um recurso científico, como as fórmulas matemáticas, as propriedades dos materiais e os compostos químicos são marcados e preparados para indexação e recuperação. A Figura 2 é uma ilustração criada com base no esquema MatML¹¹. Ele mostra que:

a) As informações contidas pelo elemento “*Material*” é compartimentalizada em cinco elementos principais:

1. O elemento “*BulkDetails*” contém uma descrição do material completo;

⁸ ADN Framework Geospatial Overview, 2003, <http://www.dlese.org/Metadata/adn-item/geospatial.php>.

⁹ OBIS. The Ocean Biogeographic Information System, <http://www.iobis.org/>.

¹⁰ OBIS. About the data. The Ocean Biogeographic Information System, http://www.iobis.org/tech/#_Toc164083855.

¹¹ *MatML Overview*. <http://www.matml.org/>.

Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf., Florianópolis, n. esp., 2º sem. 2010. ISSN 1518-2924.

2. O elemento “*ComponentDetails*” contém uma descrição de cada componente do material completo (útil para sistemas de materiais complexos, como compostos ou soldas);
 3. O elemento “*Metadata*” contém descrições de dados encontrados no documento;
 4. O elemento “*Graphs*” codifica gráficos bidimensionais;
 5. O “*Glossary*”.
- b) Pode-se usar, ainda, o elemento “*BulkDetails*” para encontrar os sub-elementos, alguns dos quais têm seus próprios sub-sub-elementos, como mostrado no exemplo “*Characterization*”:

BulkDetails

- * Nome
- * Classe
- * Sub-classe
- * Especificação
- * Caracterização
 - Fórmula
 - Composição Química
 - Composição da Fase
 - Detalhes Dimensionais
 - Notas
- * Fonte
- * Forma
- * Detalhes de Processamento
- * Dados de Propriedade
- * Notas

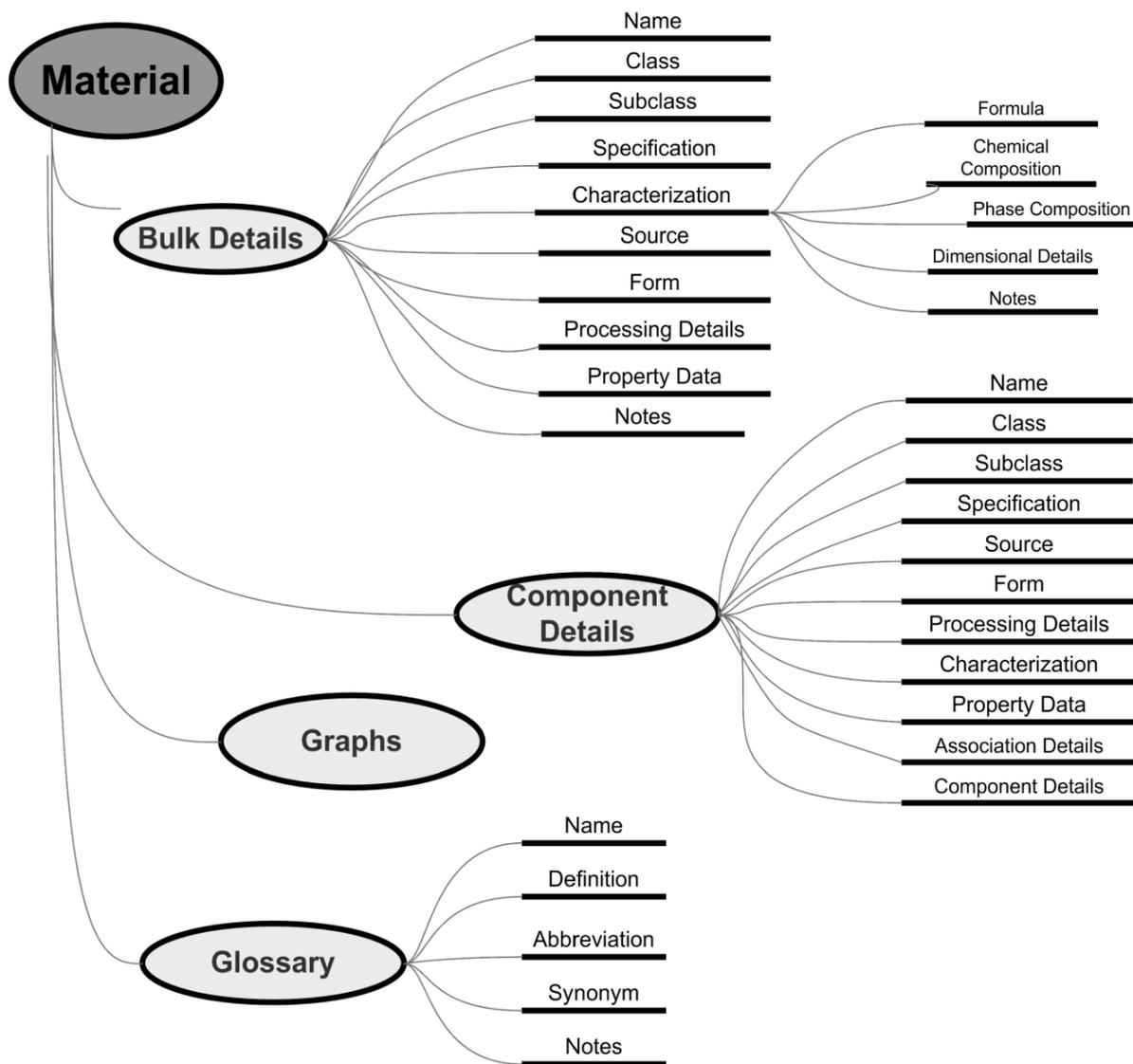


Figura 2 - Ilustração dos elementos MatML, baseado no esquema MatML¹².
Source: Smith e Zeng (2008, p. 198).

Se tomarmos novamente o exemplo da dissertação em Engenharia de Materiais e usarmos os elementos definidos pelo MatML, os conteúdos da dissertação (ou de cada capítulo) são marcados e facilmente descobertos de acordo com todas essas propriedades específicas.

3 MAXIMIZAÇÃO DA FUNÇÃO DE DESCOBERTA

Não há dúvida de que tanto as descrições de um *container* do recurso de informação quanto o seu conteúdo são necessários para as descobertas de recursos de materiais científicos. Essas duas abordagens são complementares e deveriam ser utilizadas de modo integrado. A seguir estão as metodologias propostas baseadas em um experimento anterior¹³ e em estudos contínuos:

¹² SMITH, Terence; ZENG, Marcia Lei. Semantic Tools to Support the Use and Construction of Concept-based Learning Spaces. In: *E-Learning for Geographers*. REES, Philip; MACKAY, Louise; FILL, Karen; DURHAM, Helen (eds.). Hersey, Pennsylvania: Idea Group, 2008, p. 185-203.

¹³ Shreve; Zeng, 2003, *op. cit.*

3.1 Extensão de um esquema de metadados com uma categoria específica por domínio

Em um perfil de aplicação, uma categoria adicional de elementos específica por domínio é acrescentada. Os elementos nessa categoria são de (ou baseados em) um padrão de linguagem de marcação.

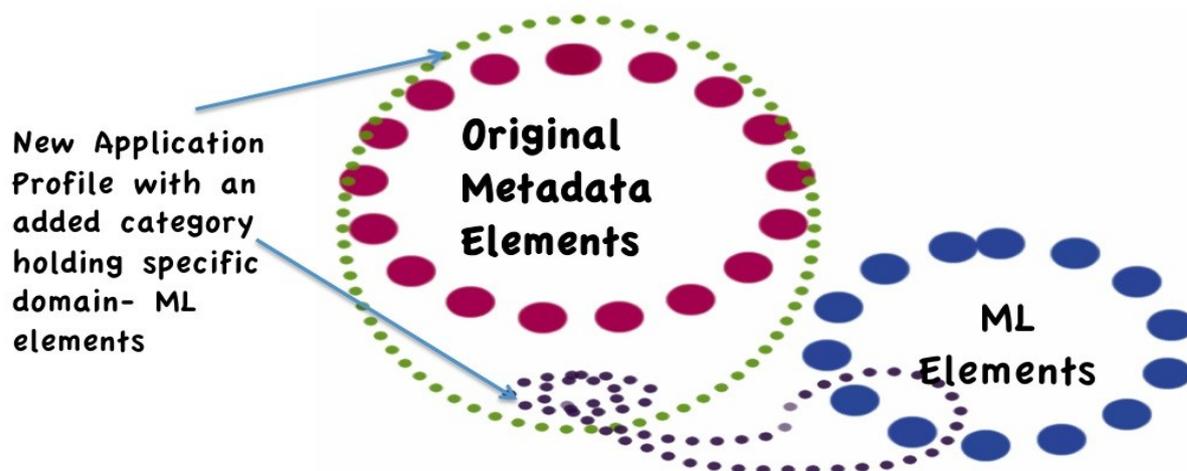


Figura 3 - Ilustração de uma metodologia que estende um esquema de metadados com uma categoria específica por domínio.

Fonte: Autora.

Na seção anterior, o exemplo do ADN mostrou como categorias orientadas por assunto podem ser integradas a uma estrutura de metadados. No perfil de aplicação LOM do projeto GREEN¹⁴, experimentamos estender as nove categorias para dez, adicionando a categoria *{Materials}*, que contém elementos selecionados, definidos pelo esquema MatML DTD ou XML. Por exemplo, originalmente um documento possuiria a descrição do metadados LOM mais aqueles na categoria *{Materials}*:

```
Title: Boundary Element Analysis of Bimaterials Using Anisotropic Elastic Green's Functions  
Identifier: http://www.boulder.nist.gov/div853/greenfn/pdfiles/jbwshop0.PDF  
Taxon: Anisotropic Elastic Solids  
Keywords: anisotropic solids, Kelvin solution, copper-nickel system, boundary integral equations, elastic constants, multilayer materials, matrix function  
... ..  
Materials:  
Bulk material: Copper-nickel multiplayer  
Component: Cu-Ni; Co-Cr; Fe-GaAs  
Processing: The materials are fabricated by depositing alternating layers of thin-film materials such as Cu-Ni, Co-Cr, and Fe-GaAs.  
... ..
```

Nesse caso, não importa se foi utilizado o LOM ou outro perfil de aplicação de metadados, as descrições para a categoria adicionada podem ser completadas separadamente no decorrer do fluxo de trabalho. O processo é mais próximo da indexação de assunto, com mais propriedades e valores específicos por domínio. O resultado ainda é um substituto de um

¹⁴ Shreve; Zeng, 2003, *op. cit.*

recurso, porém agora fornece informações mais detalhadas sobre o conteúdo do recurso que o registro de metadados original. Isso pode ajudar o usuário a decidir se vale a pena obter e ler o recurso.

A decisão se a categoria é obrigatória ou opcional terá um impacto no decorrer do fluxo de trabalho e no carregamento. Se não for obrigatória, somente um conjunto seletivo de recursos teria realmente incorporado esses elementos e fornecido valores em seus registros. Conseqüentemente, o sistema que dá acesso a essa coleção digital deveria ser cuidadoso quando fornecer a navegação e a filtragem na opção materiais do conjunto de recursos incompleto.

3.2 Utilização do(s) elemento(s) de “relação” para estabelecer *links* para marcação externa

Em quase todos os esquemas de metadados, há normalmente um elemento ou um grupo de elementos de “relação”, como por exemplo: “*dc:relation*”, “*lom:relation*” e, mais especificamente, para relações administrativas, como: “*dct:isVersionOf*”, “*dct:replaces*” ou para relações estruturais: “*dct:hasPart*”, “*dct:isPartOf*”, “*lom:relation.kind*”, e “*lom:relation.source*”. Embora não especificados exclusivamente, os valores não literais podem ser usados com qualquer desses elementos. Isso significa que um recurso externo relacionado com um identificador identificável pode ser conectado dessa maneira.

Esse método consiste em empregar o(s) elemento(s) de “relação” para ligar a um registro que é gerado de acordo com um padrão de linguagem de marcação ou com todo um documento cujas legendas de marcação estão incorporadas ao texto completo.

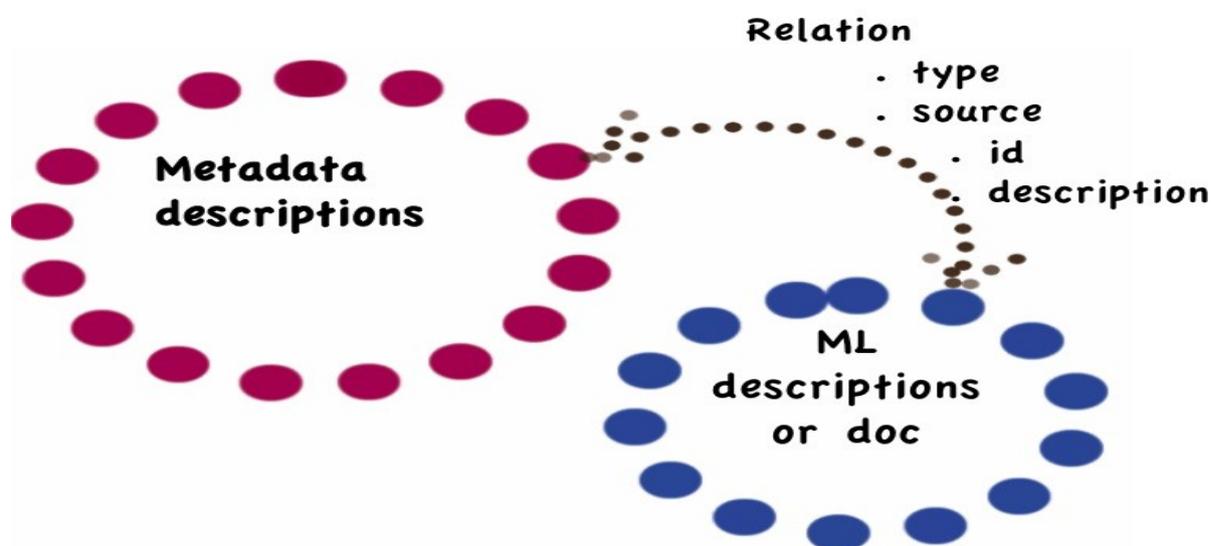


Figura 4 – Ilustração de uma metodologia que usa o(s) elemento(s) de “relação” para estabelecer *links* para marcação externa.

Fonte: Autora.

Para criar esse tipo de registro de marcação é preciso grande conhecimento do assunto. E incorporar as legendas de marcação no texto completo é um processo ainda mais sofisticado (veja exemplos fornecidos pelo *website* MatML, em: <http://www.matml.org/examples.htm>). Como resultado, os registros ligados possuem alto valor na revelação dos conteúdos de um recurso de informação.

As decisões podem ser baseadas em uma avaliação de quais conteúdos do documento descrito devem ser marcados. Entretanto, as duas descrições devem ser coordenadas no momento em que o registro de metadados é criado. Caso contrário, isso levará o dobro do tempo: será preciso voltar para criar as ligações e isso deve ser verificado para assegurar que o recurso em mãos é aquele descrito pelo registro existente.

3.3. Combinação de descrições de metadados e de marcação através de um terceiro esquema

Similar à montagem de um quebra-cabeça, os tipos diferentes de elementos de metadados (descritivo, administrativo, técnico, de uso e de preservação) e de diferentes esquemas, vocabulários e aplicações podem ser combinados entre operações. Um registro de metadados (constantemente considerado uma unidade básica nas profissões da informação) deve ser visto como um conjunto de descrições. Por esta razão, a combinação de descrições de metadados deveria ser racional e executável.

O esquema METS – *Metadata Encoding and Transmission Standard - Padrão de Transmissão e Codificação de Metadados* “é um padrão para codificar metadados descritivos, administrativos e estruturais relativos a objetos em uma biblioteca digital”¹⁵. Um registro METS pode conter sete seções principais: Cabeçalho, Mapa Estrutural, Metadados Administrativos, Metadados Descritivos, Estrutura de Ligação, Seção de Arquivo e Comportamentos, como ilustrado na Figura 5, a seguir:

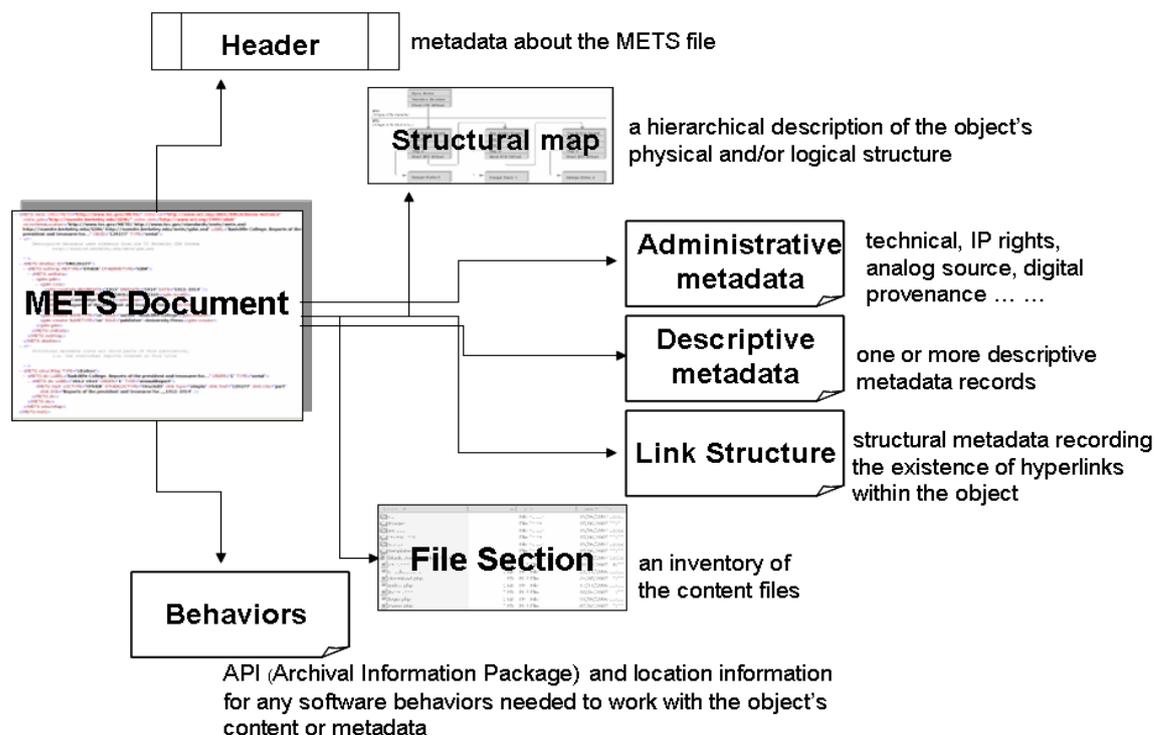


Figura 5 – Ilustração dos componentes do METS, baseado em McDonough, 2006: www.loc.gov/standards/premis/ICDAT2006/pages/Slide15_gif.htm¹⁶.

Fonte: Compilado em Zeng e Qin, 2008, p. 200.

¹⁵ *Metadata Encoding and Transmission Standard (METS)*. <http://www.loc.gov/standards/mets/>.

¹⁶ ZENG, Marcia Lei; QIN, Jian. *Metadata*. New York, NY: Neal-Schuman, 2008.

Enc. Bibli.: R. Eletr. Bibliotecon. Ci. Inf., Florianópolis, n. esp., 2º sem. 2010. ISSN 1518-2924.

Dessas sete seções, a seção de metadados descritivos em um registro METS pode apontar para metadados externos ao documento METS ou pode conter os metadados descritivos adaptados internamente. O METS permite o reuso dos metadados descritivos tanto por meio da inclusão desses em um novo registro quanto com o fornecimento de um apontador para um registro externo. Em outras seções, como o mapa estrutural e a estrutura de ligações, é teoricamente possível e realizável combinar descrições de metadados com as descrições de marcação (sejam documentos ou registros). Essa metodologia é ilustrada pela Figura 6:

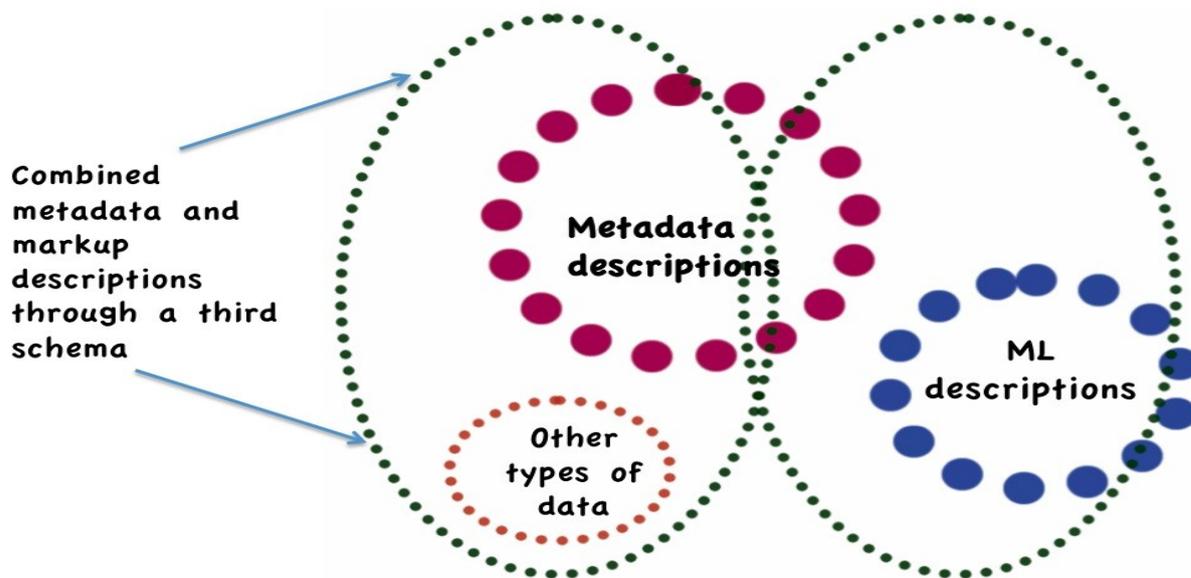


Figura 6 – Ilustração da metodologia que combina descrições de metadados e de marcação através de um terceiro esquema.

Fonte: Autora.

O emprego de um terceiro esquema como fundação para agregar características de descrição de metadados e de registros (ou documentos) de marcação aumenta consideravelmente a consistência, e, portanto, garante também a interoperabilidade. Se não estiver utilizando METS, definir um esquema RDF com o mesmo objetivo pode ser igualmente eficiente. Isso pode ser usado para criar documentos que indicam todas as relações de composição, decomposição, combinação e recombinação para recursos originais ou novos.

Somando-se aos resultados da combinação, essa abordagem permite, também, integrar descrições geradas por máquinas e por pessoas. As descrições existentes podem ser reusadas para outros projetos apropriados. Simultaneamente, a qualidade dos metadados pode ser melhorada por meio de metadados recombinantes. De modo geral, os registros integrados podem ser gerados para melhorar o acesso e o compartilhamento. Essa abordagem requererá um esforço adicional para planejar, testar e organizar o fluxo de trabalho e, portanto, é um processo mais complicado e envolverá mais etapas do que aquelas descritas nas Seções 3.1 e 3.2.

CONCLUSÃO

Os padrões de metadados e linguagens de marcação, bem como aplicações, continuarão sendo desenvolvidos, cada um em sua própria direção, facilitando a descoberta adicional dos

recursos científicos. Eles têm despendido esforços tremendos e gerado resultados notáveis nas últimas duas décadas. As linguagens de marcação específicas por domínio parecem ter se atrasado conceitualmente por serem consideradas um recurso complementar aos metadados. As descrições de metadados por si só, e no nível mais comum de hoje, não criam uma coleção rica, granular, associativa e recombinante que os cientistas gostariam de ver em uma coleção digital. É necessário convergência, especialmente em áreas onde os metadados e os esforços de marcação se sobrepõem.

Os tópicos discutidos nesse artigo sugerem uma agenda de pesquisa ambiciosa, particularmente nas áreas de inter-relacionamentos e interoperabilidade. A autora espera chamar a atenção de toda comunidade pesquisadora com o objetivo de encontrar mais oportunidades de colaboração experimental.

REFERÊNCIAS

ADN Framework Geospatial Overview, 2003. Disponível em: <http://www.dlese.org/Metadata/adn-item/geospatial.php>.

NISO. Understanding Metadata. Bethesda, MD: NISO Press, 2004. Disponível em: <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>.

OBIS. The Ocean Biogeographic Information System. About the data. 2009. Disponível em: http://www.iobis.org/tech/#_Toc164083855.

SHREVE, Gregory M.; ZENG, Marcia Lei. Integrating Resource Metadata and Domain Markup in an NSDL Collection. In: DC-2003: Proceedings of the International DCMI Metadata Conference and Workshop, Sept. 28-Oct. 2, 2003, Seattle, Washington, p. 223-229, 2003. Disponível em: http://www.siderean.com/dc2003/604_paper62.pdf.

SMITH, Terence; ZENG, Marcia Lei. Semantic Tools to Support the Use and Construction of Concept-based Learning Spaces. In: E-Learning for Geographers. REES, Philip; MACKAY, Louise; FILL, Karen; DURHAM, Helen (eds.). Hersey, Pennsylvania: Idea Group, 2008, p. 185-203.

ZENG, Marcia Lei; QIN, Jian. Metadata. New York, NY: Neal-Schuman, 2008.

Padrões de Metadados Mencionados no Artigo:

ADN (ADEPT/DLESE/NASA) Metadata Framework. Disponível em: <http://www.dlese.org/Metadata/adn-item/>.

Content Standards for Digital Geospatial Metadata (CSDGM). Disponível em: http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf.

Darwin Core. Disponível em: <http://rs.tdwg.org/dwc/>.

Dublin Core Metadata Element Set. Disponível em: <http://dublincore.org/documents/dces/>.

Dublin Core Metadata Terms. Disponível em: <http://dublincore.org/documents/dcmi-terms/>.

Electronic Theses and Dissertations Metadata Standard (ETD-MS). Disponível em: <http://www.ndltd.org/standards/metadata>.

International Standards Organisation. ISO/TS 19115:2003 Geographic information – Metadata. Disponível em: http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020.

Learning Object Metadata (LOM). Disponível em: <http://ltsc.ieee.org/wg12/index.html>.

Metadata Object Description Schema (MODS). Disponível em: <http://www.loc.gov/standards/mods/>.

National Library of Medicine (NLM) Metadata Schema. Disponível em: <http://www.nlm.nih.gov/tsd/cataloging/metafilenew.html>.

The Ocean Biogeographic Information System (OBIS) schema, version 1.1. Disponível em: <http://www.iobis.org/tech/provider/schemadef1.html>.

Padrões de Linguagem de Marcação Mencionados no Artigo

Canadian Road Markup Language (for Road Network File). Disponível em: <http://www.ogcnetwork.net/node/225>.

Chemical Markup Language (CML). Disponível em: <http://cml.sourceforge.net/>.

CityGML. Disponível em: <http://www.ogcnetwork.net/gml-citygml>.

Cyclone Warning Markup Language (CWML). Disponível em: <http://www.ogcnetwork.net/node/397>.

Digital Weather Geography Markup Language (dwGML). Disponível em: <http://www.ogcnetwork.net/node/226>.

Geography Markup Language (GML). Access: <http://www.opengeospatial.org/standards/gml>.

GML 3.1.1 Application schema for Earth Observation products. Disponível em: <http://www.ogcnetwork.net/node/380>.

GML-Geographic Data Files (GDF-GML). Disponível em: <http://www.ogcnetwork.net/node/218>.

GPLates Markup Language. Disponível em: <http://www.ogcnetwork.net/node/216>.

Mathematical Markup Language (MathML). Disponível em: <http://www.w3.org/TR/MathML/>.

MatML, XML for Materials Property Data. Disponível em: <http://www.matml.org/>.

Medical Markup Language (MML) Specification 3.0. Disponível em: http://www.medxml.net/E_mml30/MMLV3Spec.pdf.

SoTerML (Soil and Terrain Markup Language). Disponível em: <http://www.ogcnetwork.net/node/618>.

Tsunami Warning Markup Language (TWML) – Draft. Disponível em:
<http://www.ogcnetwork.net/node/378>.

ABSTRACT

While metadata has been a strong focus within information professionals' publications, projects, and initiatives during the last two decades, a significant number of domain-specific markup languages have also been developing on a parallel path at the same rate as metadata standards; yet, they do not receive comparable attention. This essay discusses the functions of these two kinds of approaches in scientific resource discovery and points out their potential complementary roles through appropriate interoperability approaches.

KEYWORDS:

Metadata. Markup languages. Scientific resource.

Originais recebidos em: 26/07/2010

Traduzidos em: setembro de 2010

Aceito para publicação em: 29/09/2010