

## Dublin Core e MARC 21 : um estudo de correspondência de elementos de metadados

Maria das Dores Rosa Alves<sup>1</sup>, Marcia Izabel Fugisawa Souza<sup>1</sup>

<sup>1</sup>Embrapa Informática Agropecuária - Caixa postal 6014 - Campinas, SP - Brasil

(maria, marcia)@emnpia.embrapa.br

*Abstract. The Information Technology has been making possible the development of applications and methodologies of descriptive analysis, such as the metadata standards, allowing new practical approaches for the organization of the information, mainly electronic. The adoption of metadata standard ensures the interoperability among applications as well as the sharing of data among systems. Based on this hypothesis, a study on the equivalence between the Dublin Core standard and the MARC 21 format was done, in order, not only, to provide subsidies for the development of data conversion tools derived from the cataloguing, but also to establish the relation among those standard, wich allows the reuse of metadata set in the same organization.*

*Resumo. A Tecnologia da Informação tem possibilitado o desenvolvimento de aplicações e metodologias de análise descritiva, como é o caso dos padrões de metadados, permitindo novas abordagens práticas para a organização da informação, principalmente em meio eletrônico. A adoção de padrão de metadados propicia a interoperabilidade entre aplicações e o compartilhamento de dados entre sistemas. Baseado nesse pressuposto, foi realizado um estudo da correspondência entre o padrão Dublin Core e o formato MARC 21, visando não apenas fornecer subsídios para o desenvolvimento de ferramentas de conversão de dados oriundos da catalogação, mas também sistematizar as relações entre esses padrões, de modo a propiciar o reuso de conjuntos de metadados na mesma organização.*

### 1. Introdução

O sucesso no compartilhamento de recursos e aplicações requer o estabelecimento de padrões de tratamento de dados e informações [Comitê... 2005]. Baseado nesse pressuposto, propõe-se a analisar as relações de correspondência entre o formato MARC 21 e o padrão Dublin Core.

O estabelecimento e a adoção de formatos, padrões e normas, no âmbito da representação bibliográfica são essenciais ao intercâmbio de informações. Este estudo pretende contribuir para a sistematização das relações existentes entre os dois padrões, visando não apenas subsidiar um possível reaproveitamento de catalogações de outros sistemas dentro da Embrapa, mas também tornar possível a expansão dos repositórios digitais de informação, e com isso, racionalizar o trabalho do bibliotecário catalogador, propiciando, ainda, economia de recursos.

A Empresa Brasileira de Pesquisa Agropecuária - Embrapa desenvolveu a Agência de Informação Embrapa, um sistema de organização da informação tecnológica voltado aos interesses do agronegócio, disponível na internet, para acesso instantâneo e na íntegra a recursos eletrônicos. Entre outros benefícios, a Agência de Informação Embrapa dedica-se a agregar valor à informação e ao conhecimento, antes de disponibilizá-los. Para isso, adota padrões de tratamento da informação, compatíveis com o ambiente *web*, com vistas à integração de dados e à interoperabilidade com outros sistemas e mecanismos de busca.

A Agência utiliza o padrão Dublin Core (DC) para o tratamento descritivo dos recursos de informação. O DC provê uma base para interoperabilidade semântica entre outros formatos mais complexos já consolidados na forma de padrões internacionais de descrição de informação eletrônica, no caso aqui tratado e objeto deste estudo o formato MARC 21 [Souza et al. 2005].

## 2. Metadados

A utilização de metadados na organização de recursos eletrônicos vem ao encontro da necessidade crescente de descobrir e disponibilizar informações na Internet e nas intranets. O uso de padrões de metadados possibilita aos sistemas de informação e de gestão do conhecimento a integração e o compartilhamento de recursos e aplicações.

Pode-se definir metadados como dados codificados e estruturados que descrevem a característica de recursos de informação, sejam eles produtos ou serviços. Elementos como autor, título, assunto são exemplos de metadados e podem ser usados para descrever tanto um livro em um catálogo de uma biblioteca *on line* ou não, quanto para descrever uma *home page*, uma base de dados ou qualquer outro recurso eletrônico em ambiente *web* [Souza et al. 2000].

## 3. Interoperabilidade

Os metadados são importantes na organização, gestão, recuperação da informação digital, principalmente, e na descoberta de recursos, ou seja, possibilitam a busca de recursos de informação por critérios relevantes, identificação de recursos, agrupamento de recursos similares, diferenciação de recursos não similares e a obtenção de informação de localização [Gonçalves 2005]. A descrição de recursos por meio de metadados permite que eles sejam compreendidos por programas, sendo pois, essenciais à interoperabilidade entre aplicações e ao compartilhamento de dados entre sistemas.

A capacidade de bases de dados trocarem e compartilharem documentos, consultas e serviços, usando diferentes plataformas de hardware e software, estrutura de dados e interfaces, é chamada de interoperabilidade. Através dessa troca e compartilhamento são realizadas interações entre sistemas. A principal característica para o sucesso dessas interações é a consistência, a qual é alcançada através do uso de padrões.

O objetivo da interoperabilidade é disponibilizar serviços coerentes aos usuários, a partir de componentes tecnicamente diferentes e gerenciados por organizações diferentes [Chaves 2002]. Isto requer cooperação em três níveis: técnico, de conteúdo e organizacional. No nível técnico se incluem os formatos, protocolos, sistemas de segurança, etc., para que as mensagens possam ser trocadas; no nível de conteúdos estão os dados e metadados, incluindo os protocolos semânticos, usados para a interpretação da informação; no nível

## II Workshop de Bibliotecas Digitais

organizacional se enquadram as regras básicas para o acesso, a preservação de coleções e os serviços (pagamentos, autenticações, etc.).

No âmbito deste estudo são considerados o nível técnico, nos aspectos que cobrem o formato, e o nível de conteúdo de dados e metadados para troca de informações entre sistemas.

### 4. Dublin Core

O Dublin Core é um padrão de metadados, composto por 15 elementos, planejado para facilitar a descrição de recursos eletrônicos. Souza et al. (2000) destacam que as principais características deste padrão são a simplicidade na descrição dos recursos, o entendimento semântico universal (dos elementos), o escopo internacional e a extensibilidade (o que permite adaptações às necessidades adicionais de descrição).

O DC pode ser inserido em uma página HTML (*Hypertext Markup Language*) e utiliza a linguagem XML (*eXtensible Markup Language*). Adota a sintaxe do RDF (*Resource Description Framework*), possui um conjunto de 15 elementos básicos, apresentados no Quadro 1, que podem ser implementados livremente para atender as necessidades de cada usuário [Dublin Core Metadata Initiative, 2006], e, ainda é um formato padrão adotado para efetuar a interoperabilidade entre outros formatos.

Elementos	Descrição
Título	Nome dado aos recursos
Criador	Entidade originalmente responsável pela criação do conteúdo do recurso
Assunto	Tema do conteúdo do recurso. Pode ser expresso em palavras-chaves e/ou categoria. Recomenda-se o uso de vocabulários controlados
Descrição	Relato do conteúdo do recurso. Exemplos: sumário, resumo e texto livre
Publicador	Entidade responsável por tornar o recurso disponível
Colaborador	Entidade responsável pela contribuição intelectual ao conteúdo do recurso
Data	Data associada a um evento ou ciclo de vida do recurso
Tipo	Natureza ou gênero do conteúdo do recurso. Exemplos: texto, imagem, som, dados, <i>software</i>
Formato	Manifestação física ou digital do recurso. Exemplos: html, pdf, ppt, gif
Identificador	Referência não-ambigua (localizador) para o recurso dentro de dado contexto
Fonte	Referência a um recurso do qual o presente é derivado
Idioma	Língua do conteúdo intelectual do recurso
Relação	Referência para um recurso relacionado
Cobertura	Extensão ou escopo do conteúdo do recurso. pode ser temporal e espacial
Direitos autorais	Informação sobre os direitos assegurados dentro e sobre o recurso

Quadro 1 – Elementos metadados Dublin Core

### 5. Formato MARC 21

MARC é um acrônimo de MACHine-Readable Cataloging: um conjunto de padrões para identificar, armazenar, e comunicar informações bibliográficas em formato legível por máquina, de forma que diferentes computadores e programas possam reconhecer, processar e estabelecer pontos de acesso dos elementos que compõem a descrição bibliográfica [Library of Congress, 2006c].

O formato MARC organiza as informações de forma a serem lidas pelo computador e possibilita a descrição bibliográfica de diferentes tipos de documentos, utilizando uma

estrutura de campos fixos e variáveis, subcampos e indicadores. Todo registro MARC é dividido logicamente em campos. Há um campo para a informação de autor, um para título e assim por diante. Como a descrição de cada campo é muito extensa para ser definida dentro do registro, foram definidas *tags* (etiquetas) de três dígitos numéricos para identificar cada campo.

O MARC 21 surgiu em 1999, oriundo da fusão do USMARC (utilizado pela Library of Congress) e do CAN/MARC (utilizado pela National Library of Canada) [Library of Congress, 2006c]. No Quadro 2 estão descritos os campos básicos do formato MARC 21.

Campos	Descrição
0XX	Informações de controle, números e códigos
1XX	Autoria (nome pessoal, entidade, evento)
2XX	Titulos, edição, imprensa
3XX	Descrição física
4XX	Série
5XX	Notas
6XX	Entradas de assunto
7XX	Entradas secundárias (nome pessoal, entidade, evento, título)
8XX	Entradas secundárias de série
9XX	Uso local

Quadro 2 – Campos básicos do formato MARC 21

#### 6. Correspondência entre Dublin Core e formato MARC 21

Dados consistem de breves descrições padronizadas, seguindo elaborados códigos de regras. Cada padrão possui diferentes formas de especificação dos metadados utilizando diferentes linguagens de representação.

Todo elemento DC possui um valor. Cada elemento tem ocorrência ilimitada, e para diferenciar o valor de cada ocorrência, são utilizados qualificadores, os quais podem ter um identificador, chamado esquema e ou modificador. Ambos servem para informar aos usuários e programas como interpretar o valor naquela ocorrência do elemento [Souza et al. 2002]. Esses qualificadores permitem ampliar o grau de especificidade dos dados a serem descritos, contribuindo para melhoria da satisfação do usuário.

O formato MARC 21 utiliza campos fixos de controle que contém informações codificadas usadas no processamento dos registros. Outras informações codificadas fazem parte de listas designadas pelo próprio formato, como códigos para Idiomas, País de publicação e Área geográfica [Library of Congress, 2006b]. Todos os dados contidos nos campos, a partir do campo 010, se subdividem em subcampos, os quais são precedidos por um delimitador e um código de subcampo. O delimitador é formado por um símbolo (aqui representado pela barra invertida “\”) que separa os diferentes subcampos. Os indicadores (Ind.) complementam esses dados e imprimem ‘valor’ ao conteúdo dos campos. Indicadores são números de um dígito e aparecem em duas posições logo após o número do campo.

Tomando por base essas observações foi possível identificar as correspondências existentes entre o formato MARC 21 e o padrão Dublin Core, apresentadas de forma breve e sucinta no Quadro 3. A versão completa do Quadro 3 está disponível para consulta no endereço <[http://doesagencia.enptia.embrapa.br/informatica/MARC21\\_DC\\_WDL\\_short\\_Quadro.pdf](http://doesagencia.enptia.embrapa.br/informatica/MARC21_DC_WDL_short_Quadro.pdf)>.

## II Workshop de Bibliotecas Digitais

<b>Dublin Core - Elementos</b>	<b>Formato MARC 21 – Campos / Subcampos / Indicadores</b>
Título	041 la, lh; 130 la, vl, ln lp; 210 la, 240 la, vl; 242 la, ly; 245 la, lb, lh, ln, lp; 246 la; 440 la, ln, lp, lv; 490 la, lv; 546 la; 740 la, ln, lp;
Criador	100 e 700 subcampos la, lb, lc, le, lq, lu; 720 la, le; 110 e 710 subcampos a, lb, ln, lp, lu; 111 e 711 subcampos la, lc, ld, le, ln, lp, lu; 720 la, le;
Descrição	041 lb, lf; 500 la; 501 la; 502 la; 505 la, lg, lt; 520 la, lb Ind.1 = #, 2 e 3
Assunto	050 la; 080 la; 082 la; 650 la Ind 2 = Ø, 3, 4, lb, lx; 653 la
Cobertura	043 la, lc; 045 la; 522 la; 650 ly, lz; 651 la, ly; 752 la, ld
Publicador	260 la, lb
Colaborador	Não há correspondência
Tipo	Campo fixo de controle 006 posição 00-17; Campo de fixo de controle 008 posição 18-34; Campo fixo de descrição física 007
Identificador	020 la; 022 la; 520 lu; 856 lu Ind 1 = 4 e Ind 2 = 2
Datas	260 lc, lg; 533 ld; Campo fixo de controle 008 posição 07-10
Formato	300 la, lc; 340 la, lb, le, li; 533 le; 856 lq Ind 1 = 4 e Ind 2 = 2
Idioma	Não há correspondência
Fonte	786 la, lb, lc, ld, lh, lk, ll, lo, lz, lx
Relação	440 la, ln, lp, lv, lx; 490 la, lv, lx; 510 la, lb, lc, lx; 530 la, lb, lc, lu; 538 la, lu; 760, 773, 774, 775, 776, 780, 785, 786, 800, 810, 811, 830 e seus respectivos subcampos relacionados
Direitos autorais	506 la, lu; 540 la, lu

**Quadro 3 – Correspondência do padrão Dublin Core da Agência de Informação Embrapa e o formato MARC 21**

### 7. Resultados e discussão

Este estudo apresenta as características do padrão de metadados Dublin Core na Agência de Informação Embrapa e do formato MARC 21, bem como seus campos, subcampos e elementos correspondentes, visando principalmente sua utilização para subsidiar o desenvolvimento de ferramentas que permitam a conversão e a integração de dados.

Além da análise de correspondência entre ambos (MARC 21 e DC), é possível constatar que, de uma forma geral o padrão Dublin Core destaca-se pela sua simplicidade e interoperabilidade [Library of Congress, 2006a, 2006d].

Apesar das diferenças e da complexidade de descrição dos metadados no formato MARC 21, pode-se verificar que seu conjunto de descritores deve conter apenas informações apropriadas e suficientes para descrever os dados. A informação contida nesse conjunto, não apenas deve ser compreendida por qualquer pessoa, mas também ser compilada e/ou interpretada pelas novas tecnologias para tratamento da informação, de modo a servir de subsídio a sistemas de buscas e recuperação de informações.

Cabe à instituição, que tem como política o compartilhamento de dados, estabelecer os campos que serão utilizados ou reaproveitados, e fazer a customização das ferramentas de catalogação, de acordo com as necessidades do seu sistema.

### 8. Considerações

A integração e o compartilhamento de dados entre sistemas, bem como a interoperabilidade entre aplicações, além de desejáveis são viáveis, notadamente em relação a sistemas que utilizam o padrão de metadados Dublin Core e o formato MARC 21, para descrição de dados.

Um fator importante a considerar refere-se à qualidade dos dados gerados a partir de padrões de consenso internacional, que segundo Gonçalves (2005) inclui os aspectos de precisão, fidelidade, critérios de seleção, generalizações, consistência dos dados, definições utilizadas e metadados sobre as fontes de dados; tendo ainda como vantagem a dinamização do processo de catalogação de recursos eletrônicos e a eliminação de esforços repetitivos e redundantes, além de permitir que os usuários finais sejam plenamente satisfeitos quanto à qualidade da informação recuperada.

### 9. Referências

- Chaves, M. S. *Padrões em bibliotecas digitais*. Porto Alegre: PUC-RS, 2002. 54 p.
- Comitê Executivo de Governo Eletrônico. *e-PING padrões de interoperabilidade de governo eletrônico - documento de referência versão 1.5*. [S.l.], 2005. 66 p.
- Dublin Core Metadata Initiative. *Dublin Core metadata element set, version 1.1: reference description*. [S.l.], 2004. Disponível em: <<http://dublincore.org/documents/dces/>>. Acesso em: fev. 2006.
- Gonçalves, J. I. *Estudo de gestão de dados não convencionais baseada em metadados para o ambiente de dispositivos móveis*. 2005. 78 p. Dissertação (Mestrado) - Informática, Universidade Federal do Paraná, Curitiba.
- Library of Congress. Network Development and MARC Standards Office. *Dublin Core/MARC/GILS Crosswalk*. Disponível em: <<http://www.loc.gov/marc/dccross.html>> Acesso em: maio 2006a.
- Library of Congress. Network Development and MARC Standards Office. *MARC Standards*. Disponível em: <<http://www.loc.gov/marc/>>. Acesso em: maio 2006b.
- Library of Congress. Network Development and MARC Standards Office. *MARC21 concise format of bibliographic data*. Disponível em: <<http://www.loc.gov/marc/bibliographic/ecbdhome.html>>. Acesso em: maio 2006c.
- Library of Congress. Network Development and MARC Standards Office. *MARC to Dublin Core crosswalk*. Disponível em: <<http://www.loc.gov/marc/marc2dc.html>> Acesso em: maio 2006d.
- Souza, M. I. F., Alves, M. D. R., Santos, A. D. dos, Cintra, M. A. M. de U., Oliveira, M. J. de. *Guia para descrição de recurso eletrônico no Sistema Agência utilizando Dublin Core*. Campinas: Embrapa Informática Agropecuária, 2002. 44 p. (Documentos, 11).
- Souza, M. I. F., Moura, M. F., Santos, A. D. dos. *Estudo comparativo entre os metadados da Agência de Informação Embrapa e do Acervo Documental do AINFO*. Campinas: Embrapa Informática Agropecuária, 2005. 10 p. (Comunicado Técnico, 66).
- Souza, M. I. F., Vendrusculo, L. G., Melo, G. C. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. *Ciência da Informação*, v. 29, n. 1, p. 93-102, abr. 2000.