

# FERRAMENTA DE CONVERSÃO DE METADADOS DE RECURSOS ELETRÔNICOS DO PADRÃO DUBLIN CORE PARA O PADRÃO RDF

Arlindo L. **Boiça Neto**<sup>1</sup>; Leandro H. de M. **Oliveira**<sup>2</sup>

Nº 0803081

## Resumo :

Atualmente a descrição de recursos eletrônicos como vídeos, áudios, imagens e textos são feitos por meio de metadados, que provêm informações sobre seu conteúdo e relacionamentos. Os padrões Dublin Core (DUBLIN CORE, 2008) e RDF (W3C, 2008) são usados para descrever esses recursos, organizando seus metadados e tornando-os visíveis aos motores de busca e sistemas de recuperação. Desde seu lançamento (em 2000) o Sistema Gestor de Conteúdo da Agência de Informação Embrapa<sup>1</sup>, um ambiente web para gestão da informação tecnológica que reúne ferramentas para a organização, estruturação, tratamento e recuperação da informação, tem usado o padrão Dublin Core para representar seus recursos. Entretanto, a conversão desses recursos para o padrão RDF se fez necessária, principalmente, para atender as expectativas de interoperabilidade entre sistemas, intercâmbio de recursos e manutenção de padrões e formatos. Este trabalho apresenta uma ferramenta que realiza a conversão entre os padrões Dublin Core e RDF no âmbito do Gestor de Conteúdo. Seu desenvolvimento foi possível pelo mapeamento paralelo dos atributos referentes aos dois padrões, seguido da implementação em linguagem Java e a aplicação de testes e validações com o usuário final. Os recursos eletrônicos foram convertidos corretamente e seguem as recomendações estabelecidas pela W3C2 que regula os dois padrões. Considerando os fins de seus objetivos, o resultado alcançado torna possível a disponibilização dos recursos eletrônicos da Embrapa em dois formatos diferentes, amplamente aceitos e recomendados, aumentando assim sua visibilidade e possibilitando a difusão de conhecimento, inerente a todos os recursos disponibilizados.

1. Estagiário na EMBRAPA/Informática Agropecuária - Graduação em Engenharia da Computação, PUC/Campinas, SP, ✉ arlindoboica@gmail.com
2. Supervisor: Pesquisador, EMBRAPA/Informática Agropecuária, Campinas, SP.

---

1 Empresa Brasileira de Pesquisa Agropecuária, empresa do Governo Federal ligada ao Ministério da Agricultura, Pecuária e Abastecimento.

2 World Wide Web Consortium – [www.w3c.com](http://www.w3c.com)

## **Abstract:**

Currently the description of electronic resources such as videos, audio and text are made through metadata, which provides information on the content and the relationships of these electronic resources. Patterns Dublin Core and RDF are used in order to facilitate the description of resources making them more visible to search engines and recovery systems. Currently the system manager of information content of the agency Embrapa (Empresa Brasileira de Pesquisa Agropecuária) uses the standard Dublin Core to represent their resources, but there is a need of conversion of these resources for the RDF standard. This conversion is necessary for greater interoperability between systems, exchange of resources, maintenance of standards and formats and is a recommendation of the entity that developed the standard, the W3C. Therefore it was necessary to the creation of a tool developed in Java, to make the conversion between the patterns through a mapping of attributes relating to the standards and Dublin Core Rdf. After the creation was necessary for some tests that the results were validated, the verification was made on the site where the W3Schools is a tool that validates electronic resources. The resources were translated correctly by following the recommendations of the two governing bodies standards, can thus be used in the system manager of information Embrapa. The work was very important so that resources could be converted and also for the learning of XML and functioning of metadata.

## **Introdução**

A representação e descrição dos recursos eletrônicos, tais como áudio, vídeo, imagens e textos podem ser feitos por meio de metadados. Dentre os tipos de formatos existentes para descrição dos metadados, estão o Dublin Core e o RDF (Resource Description Framework). O padrão Dublin Core é um esquema para descrição de metadados que favorece a representação de recursos eletrônicos, tornando-os mais visíveis aos motores de busca e sistemas de recuperação. Esse padrão descreve objetos digitais como vídeos, sons, imagens, sites e texto por meio de arquivos XML. A organização DCMI (Dublin Core Metadata Initiative), é uma organização dedicada a desenvolver esses tipos de padrão para descrever fontes que tornem mais fácil a utilização de recursos eletrônicos. De acordo com a DCMI (DUBLIN CORE, 2008), o padrão Dublin Core contém 15 elementos que descrevem todas as características do recurso, sendo eles:

1. Title: Tipicamente, um título será o nome pelo qual o recurso é formalmente conhecido;
  2. Creator: A entidade responsável em primeira instância pela existência do recurso;
  3. Subject: Tópicos do conteúdo do recurso;
  4. Description: Uma descrição do conteúdo do recurso;
  5. Publisher: Uma entidade responsável por tornar o recurso acessível;
  6. Contributor: Uma entidade responsável por qualquer contribuição para o conteúdo do recurso;
  7. Date: Uma data associada a um evento do ciclo de vida do recurso;
  8. Type: A natureza ou gênero do conteúdo do recurso;
  9. Format: A manifestação física ou digital do recurso;
  10. Identifier: Uma referência não ambígua ao recurso, definida num determinado contexto;
  11. Source: Uma referência a um recurso de onde o presente recurso possa ter derivado;
  12. Language: A língua do conteúdo intelectual do recurso;
  13. Relation: Uma referência a um recurso relacionado;
  14. Coverage: A extensão ou alcance do recurso;
  15. Rights: Informação de direitos sobre o recurso ou relativos ao mesmo;
- (DUBLIN CORE, 2008)

O padrão RDF também é um esquema de descrição de metadados em arquivos XML, utilizado para representação de recursos na Web. O padrão RDF é uma recomendação da World Wide Web Consortium (W3C) para descrição de recursos, e possui 3 componentes: 1) **recurso**: qualquer recurso eletrônico representado por arquivo digital que tenha uma URI (Uniform Resource Identification); isso inclui, por exemplo, as páginas WEB, 2) **propriedade**: atributos que descrevem as características de um recurso, por exemplo *author* ou *title*, e 3) **indicação**: uma combinação de um recurso com uma propriedade e seu valor (W3C, 2008).

*O Sistema Gestor de Conteúdo da Agência Informação Embrapa utiliza o padrão Dublin Core para a representação dos recursos eletrônicos, esse sistema reúne ferramentas para a organização, estruturação, tratamento e recuperação da informação com qualidade e precisão. Possibilita, ainda, a publicação automática de conteúdos organizados em websites. Portanto existe a necessidade da realização da conversão dos recursos para uma maior interoperabilidade entre sistemas, intercâmbio de recursos e manutenção de padrões e formatos.*

Este trabalho apresenta um conversor automático dos recursos eletrônicos vinculados ao Sistema Gestor de Conteúdo, originalmente descritos no padrão Dublin Core para o padrão RDF. Seu principal objetivo é gerar a representação em padrão RDF para cada recurso eletrônico armazenado e descrito em Dublin Core. Vale ressaltar que tal conversão corresponde a uma recomendação da W3C e da DCMI para interoperabilidade e intercâmbio de recursos eletrônicos, bem como para manutenção dos formatos e padrões recomendados.

## **Material e Métodos**

Para a criação deste conversor, a primeira tarefa realizada foi a identificação dos atributos que representam as propriedades dos recursos eletrônicos utilizados em cada padrão. Após sua identificação, um mapeamento entre os atributos de cada padrão foi criado, de maneira que as associações e correspondências entre os mesmos pudessem ser mapeadas.

A partir deste mapeamento, o próximo passo foi realizar as consultas nos bancos de dados do Gestor de Conteúdo para colher e armazenar todos os recursos eletrônicos armazenados. Para esta tarefa foram utilizados os programas DBVizualizer (site do DBVizualizer aqui) e o PGAdmin (site do PGAdmin aqui) através da linguagem SQL.

Uma estratégia utilizada para a conversão, foi a criação de um modelo RDF padrão, onde os recursos em Dublin Core utilizariam este modelo para facilitar a conversão dos padrões. Esse modelo tem as conversões das informações do banco de dados que estão no padrão Dublin Core, e as respectivas conversões para o modelo RDF em um arquivo XML. Nesta atividade foi necessário estudar e analisar vários documentos que ajudaram na compreensão dos dados contidos no banco de dados, para depois se criar um documento XML. A criação desses modelos de conversão foi realizada com base nas leituras e observações dos documentos e manuais da DCMI e da W3C (DUBLIN CORE, 2008) e em (W3, 2008). Adicionalmente, outra fonte de informações e estudos foram livros didáticos que ensinam a linguagem XML, e também alguns artigos que foram obtidos na Embrapa.

Esse modelo padrão em RDF foi utilizado para que uma aplicação em Java lesse as informações dos recursos eletrônicos descritos em Dublin Core armazenados no banco de dados e fizesse a conversão para RDF, gerando assim, para cada recurso, um arquivo XML do modelo especificado anteriormente. Esse arquivo XML tem os seguintes itens: *recursodc*, *atributo\_publicador*, *atributo\_colaborador*, *atributo\_criador*, *categoria*, *cobertura*, *descricao*, *direitoautoral*, *grupoidentificador*, *palavra\_chave*, *relação* e *título*, e

cada um deles tem vários atributos, tais como: idioma, ordem, valor, esquema, formato do recurso, data da criação, data de modificação e outros.

Para codificação da ferramenta, foi utilizado o ambiente de desenvolvimento Eclipse (site do Eclipse aqui) instalados no ambiente Linux na distribuição Kubuntu versão 8.04. A implementação da ferramenta foi flexível e para cada tipo de informação do banco de dados, foi criada uma classe, e com isso se conseguiu recuperar as informações e inseri-las no campo correspondente ao modelo RDF.

## Resultados e Discussão

Tendo como entrada o mapeamento dos atributos dos dois padrões e um determinado recurso eletrônico armazenado no banco de dados, o conversor gera como resultado um arquivo XML com extensão (rdf) representando o mesmo recurso no padrão RDF. Esse arquivo começa com uma definição dos atributos Dublin Core, representados pelas letras “dc” no campo description rdf: about e também uma referência que representa a função que ira ser convertida do padrão especificado para o padrão RDF. Essas definições são representadas a seguir baseando-se em DEITEL et al. (2003), XML (2008) e W3SCHOOLS (2008):

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:vcard="http://www.w3.org/2001/vcard-rdf/3.0#"
  xmlns:j.0="http://purl.org/dc/terms/"
  xmlns:dc="http://purl.org/dc/elements/1.1/" >
  <rdf:Description rdf:about="">
    <dc:format></dc:format>
    <dc:type></dc:type>
    <dc:title rdf:nodeID="A12"/>
    <dc:language></dc:language>
    <dc:creator rdf:nodeID="A7"/>
    <dc:publisher rdf:nodeID="A5"/>
    <dc:date>2000</dc:date>
    <dc:subject rdf:nodeID="A1"/>
    <dc:description rdf:nodeID="A10"/>
    <dc:identifier rdf:nodeID="A6"/>
```

```

    <dc:coverage rdf:nodeID="A9"/>
    <dc:relation rdf:nodeID="A11"/>
    <dc:rights rdf:nodeID="A8"/>
    <dc:contributor rdf:nodeID="A13"/>
    <dc:source></dc:source>
</rdf:Description>

```

As funções `rdf:description` são utilizadas para fazer a conversão do padrão Dublin Core para o padrão RDF, cada uma representando um tipo de informação do recurso. Dessa forma, cada um dos 15 elementos do padrão Dublin Core podem ser representados através de um caractere A, seguido de um valor numérico. Um exemplo de como essa conversão foi realizada, é representada através da informação *grupo identificador* baseando-se em DEITEL et al. (2003), XML (2008) e W3SCHOOLS (2008) :

```

<rdf:Description rdf:nodeID="A6">
    <vcard:ordem>1</vcard:ordem>
    <vcard:esquemaidentificador>3</vcard:esquemaidentificador>
    <vcard:valoridentificador>1415.0980</vcard:valoridentificador>
    <vcard:formatorecurso>6</vcard:formatorecurso>
    <vcard:idiomarecurso>PT</vcard:idiomarecurso>
    <vcard:notas></vcard:notas>
    <vcard:tamanhoarquivo>0</vcard:tamanhoarquivo>
    <vcard:duracaovideo>null</vcard:duracaovideo>
    <vcard:datacriacao>2003-01-01</vcard:datacriacao>
    <vcard:esquemadatacriacao>3</vcard:esquemadatacriacao>
    <vcard:datamodificacao>null</vcard:datamodificacao>
    <vcard:esquemadatamodificacao>null</vcard:esquemadatamodificacao>
    <vcard:dataaproximada>null</vcard:dataaproximada>
    <vcard:esquemadataaproximada>null</vcard:esquemadataaproximada>
</rdf:Description>

```

Dentro de cada campo, são representadas as informações referentes ao grupo identificador de acordo com a especificação de cada campo, um exemplo pode ser representado da seguinte maneira: `<vcard:datacriacao>2003-01-01</vcard:datacriacao>` Portanto, todas as informações referentes ao grupo identificador vão ser representadas dessa maneira. Caso não exista nenhuma informação, o campo referente a informação ficará representado pela palavra 'null'. Dessa maneira, o recurso que estava no formato Dublin Core, é convertido para o padrão RDF.

As principais dificuldades encontradas para a realização deste trabalho foram a compreensão das informações do banco de dados, e a conversão dos padrões, já que existem várias regras que devem ser seguidas para que a mesma seja realizada corretamente. Esta dificuldade foi superada pelas pesquisas realizadas nos livros e na internet, bem como pelos estudos empenhados para estabelecer a diferença entre os dois padrões.

## **Conclusão**

Diante do objetivo proposto, podemos concluir que este trabalho tornou possível a conversão de todos os recursos eletrônicos atualmente armazenados no Sistema Gestor de Conteúdo da Agência Informação Embrapa<sup>3</sup>, do padrão Dublin Core para o padrão RDF. Esse fato, aliado a possibilidade de disponibilização dos recursos em dois padrões diferentes, amplamente aceitos e recomendados pela comunidade, dá à Embrapa maior visibilidade, pois a insere no âmbito de compartilhamento de recursos eletrônicos na Web e melhora a difusão de conhecimento inerente a cada recurso eletrônico disponível.

## **Referências Bibliográficas**

DEITEL, H.M.; DEITEL, P.J; NIETO, R. *XML: Como programar*. Editora Bookman. 2003. 974p.

DEITEL, H.M.; DEITEL, P.J. *Java, como programar*. 6ª Edição. São Paulo: Editora Pearson Education do Brasil, 2005. 1110p.

DUBLIN CORE : Disponível em <<http://www.dublincore.org>>. acesso em: 03 jun.2008.

MENDES, A. *Programando com XML*. Campus. 2004. 332p.

W3C: Disponível em <<http://www.w3.org/RDF>>. acesso em: 20 jun.2008.

W3SCHOOLS: Disponível em <<http://www.w3schools.com>>. acesso em: 10 jun.2008.

XML: Disponível em <<http://www.xml.com>>. acesso em: 05 jun.2008.

---

3 Atualmente, o Gestor de Conteúdo tem por volta de 4000 recursos eletrônicos disponíveis.