

Comunicado Técnico

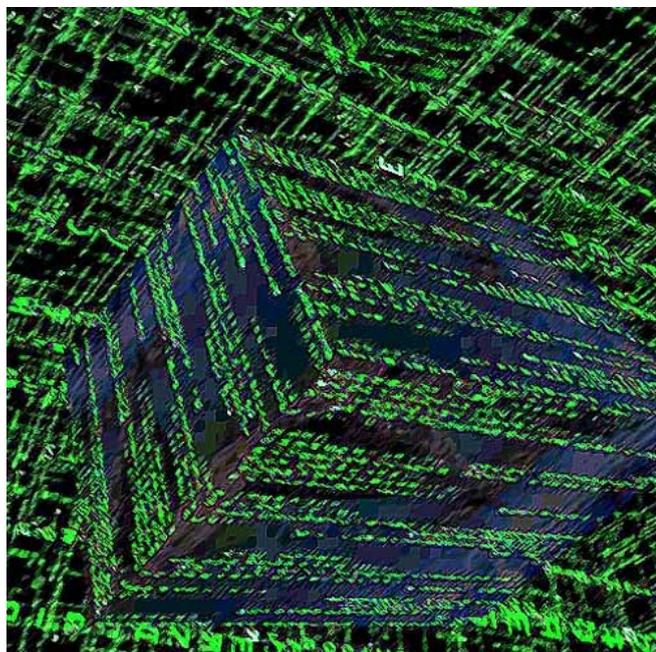
66

Março, 2005
Campinas, SP

ISSN 1677-8464

Estudo Comparativo entre os Metadados da Agência de Informação Embrapa e do Acervo Documental do AINFO

Marcia Izabel Fugisawa Souza¹
 Maria Fernanda Moura²
 Adriana Delfino dos Santos³



A Embrapa Informática Agropecuária é a principal responsável pela concretização de duas importantes iniciativas no âmbito da organização da informação: o AINFO e a Agência de Informação Embrapa.

O AINFO é uma experiência bem sucedida de *software* para automação e gestão de bibliotecas, não apenas na Embrapa, mas no país. É adotado como ferramenta-padrão para o gerenciamento de acervos e de atividades técnicas de bibliotecas no sistema Embrapa e em diversas instituições públicas brasileiras (Embrapa Informática Agropecuária, 2004). Desenvolvido há 13 anos, suas versões vêm passando por constantes melhorias e evoluções. A função principal do AINFO é o gerenciamento de bases de dados bibliográficas e cadastrais. Sua principal base de dados é o Acervo Documental e os tipos principais de materiais bibliográficos que podem ser aqui cadastrados são: livros, teses, folhetos, CD-ROM, normas técnicas, materiais não-convencionais, artigos em eventos, *proceedings*, relatórios técnicos, *software*, fotografias, imagens de satélite, documentos eletrônicos, etc.

A Agência de Informação Embrapa é uma experiência mais recente, desenvolvida para a organização, gerenciamento e disponibilização de informação tecnológica, com veiculação na mídia eletrônica, via Internet (Moura & Santos, 2002; Souza, 2002). A Agência é um portal *web*, organizado sob o

prisma de cadeias produtivas, de modo a transformar a informação gerada pela pesquisa em capital intelectual, à medida que agrega valor aos conteúdos temáticos (Evangelista et al., 2003). A Agência pode ser resumida em uma infra-estrutura de disponibilização de informação para portal *web*, cuja função primordial é oferecer acesso remoto instantâneo a recurso eletrônico, na íntegra. Para garantir esse acesso é recomendado que todo recurso eletrônico catalogado pela Agência possua uma cópia armazenada em seu repositório, desde que devidamente autorizada por seus proprietários intelectuais. Os recursos eletrônicos na Agência compreendem textos, vídeos, figuras, fotos, eventos, bases de dados, mapas, imagens de satélites, *sites*, etc.

Tanto o AINFO como a Agência, em síntese, visam a organização da informação, com o objetivo de facilitar o acesso, quer seja local ou remoto, quer seja acervo bibliográfico (técnico-científico) ou informação referencial (tecnológica), em mídia eletrônica ou impressa. Entretanto, os dois sistemas adotam formatos de metadados distintos para a descrição dos recursos de informação. O AINFO utiliza formato proprietário, formado por um conjunto de campos bibliográficos, com regras de descrição baseadas nas "Normas de Referenciação Bibliográfica na Embrapa" (Embrapa, 2001). O sistema Agência adota o formato de metadados Dublin Core (Dublin Core Metadata Initiative,

¹ M.Sc. em Biblioteconomia, Técnico de Nível Superior da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP. (e-mail: marcia@cnptia.embrapa.br).

² M.Sc. em Engenharia Elétrica, Pesquisadora da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP. (e-mail: fernanda@cnptia.embrapa.br).

³ M.Sc. em Engenharia Elétrica, Pesquisadora da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP. (e-mail: adriana@cnptia.embrapa.br).

2004b), composto por 15 elementos de descrição, amparado na AACR2 (Ribeiro, 2003) e tesouros para indexação de assuntos.

Partindo desse contexto, foi realizada uma análise conjunta dos metadados adotados (campos da base de dados Acervo Documental do AINFO – doravante denominada AD/AINFO e elementos do Dublin Core da Agência – DC/Agência), a fim de identificar possíveis relações de similaridade e de complementaridade entre ambos. Buscou-se, com isso, subsidiar uma desejável integração de ferramentas de catalogação de recursos de informação, que atenda tanto às exigências atuais do AD/AINFO e do DC/Agência, como de quaisquer sistemas que venham a ser implantados na Embrapa, no futuro. Esta análise está fundamentada na especificação da versão 1.0 da Agência e no AINFO 2001.

Dessa forma, os objetivos deste trabalho são:

- 1) analisar as relações de similaridade e complementaridade existentes entre os formatos de metadados AD/AINFO e DC/Agência;
- 2) elaborar uma proposta de integração e/ou unificação das ferramentas de catalogação, tomando por base a análise das inter-relações existentes entre AD/AINFO e DC/Agência, preservando as especificidades de cada sistema.

Metodologia

Para facilitar a visualização e posterior análise das relações de similaridade e complementaridade entre os formatos de metadados, foram elaboradas tabelas com os metadados AD/AINFO e DC/Agência. A confecção dessas tabelas visou ilustrar a amplitude e abrangência dos metadados de ambos os sistemas, além de dar idéia da diversidade de recursos de informação que podem ser descritos por eles.

Outra tabela foi elaborada contendo os campos do AD/AINFO, os elementos do DC/Agência e uma coluna com comentários sobre a existência ou não de inter-relação entre os metadados. Os atributos de qualidade, como esquema e modificadores, também tiveram sua importância destacada na análise, conforme apontamento feito nos comentários sobre a ocorrência ou não de similaridade.

Sistematizadas as informações mostradas nas tabelas, procedeu-se a uma discussão dos dados com membros da equipe da Agência para considerações e sugestões. Essa discussão convergiu para a elaboração de uma proposta de unificação da ferramenta de catalogação com capacidade de abrigar especificidades dos distintos sistemas, sem qualquer prejuízo para ambos.

Uso do Dublin Core na Agência

A necessidade de catalogar recursos na Agência de Informação Embrapa foi, desde seu início, tema de inúmeras discussões entre os componentes da equipe do Projeto Agência.

O formato Dublin Core (DC) foi exaustivamente discutido no

âmbito do referido Projeto, e a decisão pelo seu uso selou o entendimento da equipe técnica acerca da importância da adoção de um conjunto de metadados que realmente viesse a agregar valor ao conteúdo da Agência.

Os metadados DC estão embutidos nas páginas HTML da Agência, possibilitando que os robôs de busca na *web* as localizem e reconheçam seus metadados. Isso possibilita a recuperação de informações mais precisas pelos usuários desses robôs, sem que eles necessitem estar acessando diretamente a Agência para encontrar as informações por ela disponibilizadas.

O DC tem vantagem sobre os demais formatos por ter sido criado sob os auspícios da W3C (World Wide Web Consortium, 2004c), e ser recomendado como padrão de uso para Internet, pois reflete necessidades específicas da *web*. Além de tratar recursos eletrônicos de forma simplificada, 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.

Além disso, a adoção do DC na infra-estrutura tecnológica da Agência preparou-a para futuramente promover o intercâmbio de metadados com outras aplicações *web*, já que existe linguagem específica estabelecida pelo W3C para expressar os metadados Dc – a chamada RDF/XML (Dublin Core Metadata Initiative, 2004c). A sintaxe do RDF - *Resource Description Framework* está baseada na linguagem de marcação XML (World Wide Web Consortium, 2004b), também estabelecida por aquele consórcio para intercâmbio de dados (World Wide Web Consortium, 2004a) e pelo governo brasileiro, como pode ser observado na versão 0 da arquitetura e-PING – Padrões de Interoperabilidade de Governo Eletrônico (Comitê Executivo do Governo Eletrônico, 2004).

Do ponto de vista de formato de metadado, há um espectro de três faixas que se torna sucessivamente mais rico em termos de estrutura e completude (Dempsey & Heery, 1997). Na faixa 1, o formato de registro de informação é simples e proprietário – a indexação é do tipo *full text*. Na faixa 2, o formato de registro é estruturado e adota padrões emergentes - a indexação é do tipo "por campo".

Na faixa 3, o formato descritivo de registro é completo e adota padrões internacionais. Nesta terceira faixa, os metadados podem ser usados não só para localizar e encontrar recurso, como também podem atuar na documentação de objetos ou coleções de objetos - a sua criação e manutenção requerem conhecimento de especialista – a sua indexação é por *elaborate tagging*. Por *elaborate tagging* entende-se uma rotulação semanticamente mais completa e melhor qualificada; os procedimentos de busca se utilizam desses rótulos de um modo mais elaborado e complexo, por exemplo: título, tipo, nome de autor principal e veículo – vários metadados para melhor qualificar as respostas da busca, diminuindo a razão entre a precisão e a revocação, isto é, a razão entre o número de documentos recuperados, que são relevantes para a busca especificada, e o número de documentos que realmente são relevantes para a busca especificada e foram de fato recuperados (Han & Kamber, 2001).

O formato DC classifica-se na faixa 2, pois inclui dados que contém descrição suficientemente completa para permitir que um usuário acesse um recurso potencialmente útil ou de seu interesse, sem ter que recuperá-lo ou conectá-lo; os dados são estruturados e suportam busca do tipo "por campo".

Elementos metadados Dublin Core

A seguir é apresentada a relação de metadados DC (Tabela 1), composta de 15 elementos (Dublin Core Metadata Initiative, 2004a), utilizados na Agência.

Tabela 1. Elementos metadados DC/Agência.

<i>Elemento</i>	<i>Descrição</i>
Título	Nome dado ao recurso através do qual ele é formalmente conhecido; permite a inclusão de título alternativo, traduzido e de série
Criador	Entidade originalmente responsável pela criação do conteúdo do recurso; permite várias ocorrências com vários modificadores (nome pessoal, endereço pessoal, afiliação, nome corporativo, endereço corporativo)
Descrição	Relato do conteúdo do recurso; possibilita inclusão de texto livre, sumário e resumo
Assunto	Tema do conteúdo do recurso; palavras-chaves e categoria – oferece opções de uso de tesouros
Cobertura	Extensão ou escopo do conteúdo do recurso; pode ser temporal e espacial; expande o tratamento de assunto e permite o uso de tesouros
Publicador	Entidade responsável por tornar o recurso disponível; permite várias ocorrências com vários modificadores (nome pessoal, endereço pessoal, nome corporativo, endereço corporativo, afiliação)
Colaborador	Pessoa ou entidade responsável pela contribuição intelectual ao conteúdo do recurso; complementa e amplia a descrição do recurso; permite várias ocorrências com vários modificadores (nome pessoal, endereço pessoal, nome corporativo, endereço corporativo, afiliação)
Tipo	Natureza ou gênero do conteúdo do recurso, por exemplo, texto, imagem, som, dados, <i>software</i> , etc.; agrega valor à descrição do recurso eletrônico
Identificador	Referência não-ambígua (localizador) para o recurso dentro de dado contexto; na Agência, é composto pelo agrupamento Localizador + Data + Formato + Idioma, com combinações entre vários esquemas e ocorrências
Data	Data associada a um evento ou ciclo de vida do recurso; permite inclusão de data de criação, de última modificação e data aproximada (recursos eletrônicos)
Formato	Manifestação física ou digital do recurso, por exemplo, html, pdf, doc. ppt, gif, xls, etc.; agrega valor à descrição do recurso eletrônico
Idioma	Língua do conteúdo intelectual do recurso
Fonte	Referência a um recurso do qual o presente é derivado
Relação	Referência para um recurso relacionado; agrega valor à descrição do recurso eletrônico
Direitos autorais	Informação sobre os direitos assegurados dentro e sobre o recurso; agrega valor à descrição do recurso eletrônico

Fonte: Dublin Core Metadata Initiative (2004a).

Metadados do Acervo Documental do AINFO

O AINFO emprega a palavra "campo" (Lenk & Antunes, 2002) para designar o termo denominado "metadado", adotado na Agência e neste documento. O conjunto de metadados AD/AINFO, de acordo com Dempsey &

Heery (1997), classifica-se na faixa 1, cujo formato de registro de informação é simples e proprietário, com indexação *full-text*.

Na base de dados AD/AINFO há 33 campos, como mostra a Tabela 2.

Tabela 2. Campos da base de dados AD/AINFO.

<i>Elemento</i>	<i>Descrição</i>
Identificador	código de identificação do material (gerado automaticamente após a inclusão de um novo registro)
Data corrente	data corrente de inclusão do registro
Instituição	Sigla da Unidade (Instituição)
Origem	Origem do material bibliográfico (ex.: compra, doação, permuta)
Tipo	Tipo do material bibliográfico (ex.: folheto, livro, mapa, tese, etc.)
Classificação	Código de classificação do documento para localização nas estantes
Cutter	Código de autor para localização nas estantes
Código do documento	Código de identificação para evitar duplicação de registro
Autoria	Autores individuais ou coletivos, indicação dos responsáveis pela publicação
Afiliação	Local de trabalho dos autores conforme aparece no documento
Título e subtítulo	Título do documento
Título original	Título original
Edição	Número de edição
Ano de publicação	Ano de publicação do documento
Fonte/Imprensa	Imprensa ou referência bibliográfica
Volume	Número de volume
Exemplar	Número de exemplar
Páginas	Número de páginas do documento
Descrição física	Descrição física do documento
Série	Série do documento
ISBN	International Standard Book Number
Idioma	Idioma do documento
Notas	Complementação das informações do documento (ex. Tese, Editado por..., Trabalho apresentado na...)
Incluir Acervo Documental	Indicador de inclusão no Acervo Documental da Embrapa
Conteúdo	Descrição do conteúdo do documento (resumo/abstract ou sumário)
Palavras-chaves	Palavras-chaves do documento
URL	Endereço eletrônico do texto completo do documento
Observações	Observações (para outras informações não técnicas)
Identificador	Código de identificação complementar para cada volume ou exemplar do documento
Registro	Número de registro (tombo) do documento
Data receb.	Data de recebimento do material
Status	Estado do material: disponível ou não, para empréstimo
Imprime etiqueta	indicador de impressão de etiqueta de lombada e de código de barra

Fonte: Lenk & Antunes (2002).

Resultados Obtidos

Inter-relações entre os Metadados Dublin Core e Acervo Documental do AINFO

Os dados da Tabela 3 estão acompanhados de comentários que visam facilitar a avaliação sobre a presença ou não de similaridade e complementaridade, o que contribui para o

melhor entendimento da correspondência entre os metadados DC/Agência e AD/AINFO.

Tabela 3. Relações de Similaridade (S) e Complementaridade (C) dos metadados DC/Agência versus AD/AINFO.

<i>DC/Agência (elementos)</i>	<i>AD/AINFO (campos correspondentes)</i>	<i>Comentários</i>
Título	Título e subtítulo + Título original + Série	o elemento Título (DC/Agência) é similar ao campo Título e subtítulo (AD/AINFO). (S) o elemento Título (DC/Agência) abriga os campos Título original e Série (AD/AINFO), via modificadores. (S) o elemento Título (DC/Agência) permite a inclusão de Título alternativo, via modificador – não previsto no AD/AINFO. (C)
Criador	Autoria + Afiliação	o elemento Criador (DC/Agência) é similar ao campo Autoria (AD/AINFO). (S) o elemento Criador (DC/Agência) abriga o campo Afiliação (AD/AINFO), via modificador. (S) o elemento Criador (DC/Agência) permite várias ocorrências, além de combinações entre vários modificadores, como: nome pessoal, endereço pessoal, afiliação, nome corporativo, endereço corporativo. (C)
Descrição	Conteúdo	o elemento Descrição (DC/Agência) é similar ao campo Conteúdo (AD/AINFO). (S)
Assunto – Palavra-chave	Palavras-chaves	o elemento Assunto – Palavra-chave (DC/Agência) é similar ao campo Palavras-chaves (AD/AINFO). (S) o elemento Assunto – Palavra-chave (DC/Agência) preconiza o uso de tesouros, via esquema, com opção de tesouros para indexação de assunto. (C)
Assunto – Categoria	Não possui metadado correspondente	o elemento Assunto-Categoria (DC/Agência) não possui correspondente no AD/AINFO (C) o elemento Assunto-Categoria (DC/Agência) preconiza o uso de tesouros, via esquema, para a categorização de assunto. (C)
Cobertura	Não possui metadado correspondente	o elemento Cobertura (DC/Agência) expande o tratamento de assunto, via modificadores (temporal e espacial). Não possui correspondente no AD/AINFO. (C) o elemento Cobertura (DC/Agência) preconiza o uso de tesouros, via esquema, para o tratamento complementar de assunto. (C)
Publicador	Fonte/Imprensa	o elemento Publicador (DC/Agência) é similar ao campo Fonte/Imprensa (AD/AINFO). (S) o elemento Publicador (DC/Agência) permite várias ocorrências, além de vários modificadores, como: nenhum, nome pessoal, nome corporativo, endereço pessoal, endereço corporativo, afiliação – não previstos no AD/AINFO. (C)
Colaborador	Não possui metadado correspondente	o elemento Colaborador (DC/Agência) permite complementar e ampliar a descrição do recurso, em várias ocorrências, via modificadores: nenhum, nome pessoal, endereço pessoal, endereço corporativo, afiliação. Não possui correspondente no AD/AINFO. (C)
Tipo	Não possui metadado correspondente	o elemento Tipo (DC/Agência) permite caracterizar o recurso eletrônico quanto à natureza ou gênero de seu conteúdo, agregando valor à sua descrição. Não possui correspondente no AD/AINFO. (C)

Continuação da Tabela 3.

Identificador	ISBN + URL	<ol style="list-style-type: none"> o elemento Identificador (DC/Agência) não possui correspondente no AINFO. (C) o elemento Identificador (DC/Agência) é exclusivo para descrever recurso eletrônico, agrupando localizadores + Data + Formato + Idioma, sob várias ocorrências. (C) elemento Identificador (DC/Agência) é abrangente, pois inclui esquemas: URI, ISBN, ISSN, FPI e nenhum; complementa-se com os campos ISBN e URL do AD/AINFO (C) o elemento Identificador (DC/Agência) permite ampliar a descrição de datas típicas de recursos eletrônicos (criação, última modificação e aproximada), combinadas a vários esquemas. (C) o elemento Identificador (DC/Agência) permite ampliar a descrição do recurso com a indicação do formato, agregando valor à descrição. (C)
Data	Ano de publicação	<ol style="list-style-type: none"> o elemento Data (DC/Agência), apesar de similar a Ano de publicação (AD/AINFO), é mais abrangente, ampliando a descrição com modificadores (data de criação, data da última modificação e data aproximada) e esquemas (nenhum, ano, ano/mês, ano/mês/dia). (S + C)
Formato	Descrição física	<p>o elemento Formato (DC/Agência) apresenta similaridade com o campo Descrição física do AD/AINFO, no tocante às informações de identificação e forma dos documentos que são descritos. (S)</p> <p>o elemento Formato (DC/Agência) permite ampliar a descrição do recurso eletrônico, agregando valor à descrição. (C)</p>
Idioma	Idioma	<p>o elemento Idioma (DC/Agência) é similar a Idioma (AD/AINFO). (S)</p> <p>o elemento Idioma (DC/Agência) permite combinações com modificadores e esquemas do grupo Identificador. (C)</p>
Fonte	Notas	<ol style="list-style-type: none"> elemento Fonte (DC/Agência) é similar a Notas (AD/AINFO). (S)
Relação	Não possui metadado correspondente	<ol style="list-style-type: none"> elemento Relação (DC/Agência) não possui correspondente no AD/AINFO. (C) <p>o elemento Relação (DC/Agência) aponta possível relacionamento do item que está sendo descrito com outro recurso, agregando valor à descrição. (C)</p> <p>o elemento Relação (DC/Agência) permite várias ocorrências, sob diferentes modificadores e esquemas. (C)</p>
Direitos autorais	Não possui metadado correspondente	<p>o elemento Direitos autorais (DC/Agência) não possui correspondente no AD/AINFO. (C)</p>

Discussão

A análise das relações de similaridade e complementaridade existente entre os metadados adotados pelo AD/AINFO e pela DC/Agência, tomando por base os comentários feitos na Tabela 3, sugere que é possível adotar uma única ferramenta

dois sistemas, o que evitaria a duplicação de esforços e a dispersão de recursos.

Os metadados AD/AINFO e DC/Agência reunidos na Tabela 4, a seguir, formam o conjunto de elementos que devem ser considerados na especificação dessa nova ferramenta de *software* para a catalogação.

Tabela 4. Conjunto de elementos metadados AD/AINFO - DC/Agência.

<i>Fusão de Metadados AD/AINFO e DC/Agência</i>	<i>Descrição</i>	<i>Preenchimento AD/AINFO</i>	<i>Preenchimento DC/Agência</i>
Identificador (AD/AINFO)	Código de identificação do material	X (gerado automaticamente)	
Data corrente (AD/AINFO)	Data corrente de inclusão do registro	X (obrigatório)	
Instituição (AD/AINFO)	Sigla da Unidade	X (obrigatório)	
Origem (AD/AINFO)	Origem do material bibliográfico	X (obrigatório)	
Classificação (AD/AINFO)	Código para localização nas estantes	X	
Cutter (AD/AINFO)	Código de autor para localização nas estantes	X	
Código documento (AD/AINFO)	Código do documento	X	
Autoria/Afiliação + Criador (AD/AINFO-DC/Agência)	Autoria/criador do recurso + afiliação	X	X
Título e subtítulo/Título original + título traduzido + Série (AD/AINFO-DC/Agência)	Título e subtítulo + título original + título de série do recurso	X (título é obrigatório)	X (obrigatório)
Edição (AD/AINFO)	Número da edição	X	
Ano de publicação + Data (AD/AINFO-DC/Agência)	Ano + Data de publicação do recurso	X (obrigatório)	X
Fonte/Imprensa/Local publicação + Publicador (AD/AINFO-DC/Agência)	Fonte/Imprensa + Local publicação + publicador	X (obrigatório)	X
Colaborador (DC/Agência)	Colaborador intelectual do recurso		X
Volume (AD/AINFO)	Indicação do volume do material bibliográfico	X	
Exemplar (AD/AINFO)	Número do exemplar do material bibliográfico	X	
Páginas (AD/AINFO)	Número de páginas do material bibliográfico	X	
Descrição física + Formato (AD/AINFO-DC/Agência)	Descrição física do material bibliográfico + formato do recurso	X	X
Nota + Fonte (AD/AINFO-DC/Agência)	Informações adicionais e fonte do recurso	X	X
Conteúdo + Descrição (AD/AINFO-DC/Agência)	Descrição do conteúdo do recurso	X	X
Palavras-chaves + Assunto (AD/AINFO-DC/Agência)	Palavras-chaves do recurso	X	X (obrigatório)
Categoria-Assunto (DC/Agência)	Categoria de assunto do recurso		X (obrigatório)
Cobertura (DC/Agência)	Cobertura do recurso		X
Tipo (AD/AINFO)	Tipo do material bibliográfico	X	
Tipo (DC/Agência)	Tipo do recurso eletrônico		X
ISBN/URL + Grupo Identificador (AD/AINFO-DC/Agência)	ISBN e endereço eletrônico do recurso	X	X
Idioma (AD/AINFO-DC/Agência)	Idioma do documento/recurso	X	X
Relação (DC/Agência)	Indicação de relação do recurso		X

Continuação Tabela 4.

Direitos autorais (DC/Agência)	Direitos autorais do recurso		X
Incluir no Acervo Documental (AD/AINFO)	Indicação de inclusão no acervo da Embrapa	X	
Observações (AD/AINFO)	Observações gerais sobre o documento	X	
Identificador AD/AINFO)	Código de identificação do material	X	
Registro AD/AINFO)	Número de registro do material na biblioteca (tombo)	X	
Data receb. AD/AINFO)	Data/recebimento do material bibliográfico	X	
Status (AD/AINFO-DC/Agência)	Estado do recurso na base de dados	X (gerado automaticamente)	X (gerado automaticamente)
Imprime etiqueta (AD/AINFO)	Impressão de etiqueta de lombada	X	

Considerações Finais

Concebidos sob demandas e exigências diferenciadas no campo da organização e acesso à informação, os dois sistemas – cada qual ao seu tempo – refletem necessidades bastantes distintas. De um lado, o AD/AINFO privilegiando o acesso à descrição de recursos de informação, em formato de referências bibliográficas; de outro, a Agência privilegiando o acesso a conteúdos de informação, em formato eletrônico, na íntegra e na Internet. Apesar dessas especificidades, ambos os sistemas apresentam entre si traços de similaridade e complementaridade, que vêm motivando a análise e a proposição de uma desejável integração de ferramentas e procedimentos.

De acordo com Pacheco & Kern (2001), a convivência de sistemas não integrados acarreta problemas, entre os quais: redundância de dados e inconsistência de dados redundantes; dispêndios de recursos que não agregam valor na transferência de informação de um sistema para outro; perda semântica na tradução, redigitação de dados.

A necessidade de integração de sistemas é reconhecida também por bibliotecários – usuários internos – como pelos técnicos desenvolvedores do AINFO e da Agência. A integração traria benefícios diretos aos usuários dos dois sistemas, evitando a duplicação de esforços dos bibliotecários e eliminando redundâncias indesejáveis. Como vantagem ter-se-ia a dinamização do processo de catalogação de recursos de informação e a eliminação de esforços repetitivos e redundantes (Bartolo et al., 2004).

As aplicações legadas de organização da informação, como AD/AINFO, CGPE - Cadastro Geral de Publicações Embrapa (Embrapa Informação Tecnológica, 2004), DC/Agência e outros devem tratar e exportar informações e dados. Para isso, deve-se dispor de uma linguagem comum que contemple os metadados mencionados na Tabela 4. A solução ideal seria a criação de uma ontologia de metadados de recursos eletrônicos para a Agência e o Acervo Documental do AINFO, que integre essas diferentes bases de metadados. Sugere-se que essa ontologia seja o ponto de partida para a integração de outros sistemas de informação

da Empresa, que demandem o uso metadados para descrever recursos eletrônicos, conforme ilustra a Fig. 1.

Ontologia semelhante foi desenvolvida e implantada com sucesso no Brasil, visando a integração de bases de informação e conhecimento no âmbito da Ciência e Tecnologia; trata-se da Linguagem de Marcação da Plataforma Lattes (LMPL), descrita por Pacheco & Kern (2001).

Caso as aplicações estejam em ambiente *web*, sugere-se utilizar o RDF/XML, para implementação dessa ontologia, pois o mesmo é uma linguagem que expressa metadados que são intercambiáveis em plataforma *web* (Dublin Core Metadata Initiative, 2004c). Deve-se considerar o fato de já existir uma recomendação do Dublin Core Metadata initiative para expressar DC em RDF/XML. Para ambiente não-*web*, sugere-se a utilização do formato MARC 21 (Estados Unidos, 2003) e o protocolo Z 39.50 (Estados Unidos, 2004).

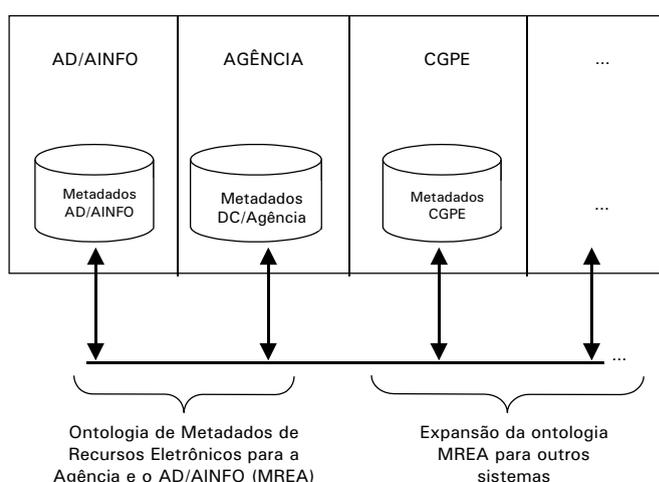


FIG. 1. Esquema de uso da Ontologia MREA visando a integração das bases de metadados e seu potencial de expansão para outros sistemas de informação da Embrapa.

Como o AD/AINFO já importa e exporta dados no formato MARC 21, uma solução de curto prazo, bastante simplificada e direcionada à questão da troca de dados entre

ele e a Agência, pode ser obtida com a integração de um conversor de Dublin para MARC 21, de domínio público denominado *d2m : Dublin Core to MARC converter* (BIBSYS, 1998). Esse código é implementado em linguagem de programação C, mas poderia ser facilmente adaptado à Agência através da construção de uma interface no padrão JNI (Java Native Interface), exemplificada na experiência relatada por Cruz (2001). No entanto, é necessário também obter o caminho inverso, ou seja, do MARC 21 para o Dublin Core; não há *software* de domínio público prontamente disponível, entretanto, existe uma tabela de equivalências já definida – a MARC to Dublin Core Crosswalk –, desenvolvida pela Library of Congress (Estados Unidos, 2001), e que pode ser utilizada na construção do *software*.

No entanto, num prazo maior, desenvolvendo-se uma única ferramenta de catalogação, que integre o AD/AINFO e Agência pode-se:

- facilitar o fluxo de trabalho e evitar esforços redundantes;
- contribuir para a expansão dos repositórios digitais de informação na Embrapa;
- permitir a identificação e extração de um conjunto mínimo de metadados estruturados passíveis de serem reconhecidos e lidos por um maior número de robôs de busca.

Referências Bibliográficas

- BARTOLO, L. M.; LOWE, C. S.; FENG, L. Z.; PATTEN, B. MatDL: integrating digital libraries into scientific practice. *Journal of Digital Information*, v. 5, n. 3, art. 297, Aug. 2004. Disponível em: <<http://jodi.ecs.soton.ac.uk/Articles/v05/i03/Bartolo/>>. Acesso em: 29 set. 2004.
- BIBSYS. *d2m - Dublin Core to MARC converter*. 1998. Disponível em: <<http://www.bibsys.no/meta/d2m/>>. Acesso em: 10 out. 2004.
- COMITÊ EXECUTIVO DE GOVERNO ELETRÔNICO. *e-PING padrões de interoperabilidade de governo eletrônico - documento de referência versão 0*: parte II especificação - dos componentes da e-PING. [S.l.], 2004. 64 p. Disponível em: <http://www.governoeletronico.gov.br/Governoeletronico/publicacao/down_anexo.wsp?tmp.arquivo=E15_24115_1e-ping_minuta_v0_31052004_consulta.pdf>. Acesso em: 7 out. 2004.
- CRUZ, S. A. B. da. *Aplicação da JNI - Java Native Interface - na construção da ferramenta ServCLIPS*. Campinas: Embrapa Informática Agropecuária, 2001. 57 p. (Embrapa Informática Agropecuária. Documentos, 9). Disponível em: <<http://www.cnptia.embrapa.br/modules/tinycontent3/content/2001/DOCUMENTO9int.pdf>>. Acesso em: 7 out. 2004.
- DEMPSEY, L.; HEERY, R. *DESIRE - Development of a European Service for Information Research and Education*: specification for resource description methods. Part 1. A review of metadata: a survey of current resource description formats, version 1.0. [Bath: UKOLN], 1997. 92 p. Disponível em: <<http://ukoln.ac.uk/metadata/desire/overview/overview.pdf>>. Acesso em: 7 out. 2004.
- DUBLIN CORE METADATA INITIATIVE. *Dublin Core element set, version 1.1*: reference description. Disponível em: <<http://purl.org/dc/documents/rec-dces-19990702.htm>>. Acesso em: 20 set. 2004a.
- DUBLIN CORE METADATA INITIATIVE. *Dublin Core Metadata Initiative (DCMI)*. Disponível em: <<http://dublincore.org/>>. Acesso em: 20 set. 2004b.
- DUBLIN CORE METADATA INITIATIVE. *Expressing qualified Dublin Core in RDF/XMI*. Disponível em <>. Acesso em: 7 out. 2004c.
- EMBRAPA. Manual de informação e documentação: referência bibliográfica da Embrapa. *Boletim de Comunicações Administrativas*, Brasília, DF, v. 27, n. 46, p. 5-99, out. 2001.
- EMBRAPA INFORMAÇÃO TECNOLÓGICA. *Cadastro geral de publicações da Embrapa*. Disponível em: <<http://www.sct.embrapa.br/cgp/>>. Acesso em: 15 out. 2004.
- EMBRAPA INFORMÁTICA AGROPECUÁRIA. *AINFO 2001 versão 1.3*. Disponível em: <<http://www.cnptia.embrapa.br/modules/tinycontent0/index.php?id=28>>. Acesso em: 20 set. 2004.
- ESTADOS UNIDOS. Library of Congress. Network Development and MARC Standards Office. *MARC standards*. 2003. Disponível em: <<http://www.loc.gov/marc/marc.html>>. Acesso em: 14 out. 2004.
- ESTADOS UNIDOS. Library of Congress. Network Development and MARC Standards Office. *Z39.50 maintenance agency page*. 2004. Disponível em: <<http://www.loc.gov/z3950/agency/>>. Acesso em: 14 out. 2004.
- ESTADOS UNIDOS. Library of Congress. Network Development and MARC Standards Office. *MARC to Dublin Core crosswalk*. 2001. Disponível em: <<http://www.loc.gov/marc/marc2dc.html>>. Acesso em: 14 out. 2004.

EVANGELISTA, S. R. M.; SOUZA, K. X. S. de; SOUZA, M. I. F.; BRAGA, S. A. C.; LEITE, M. A. de A.; SANTOS, A. D. dos; MOURA, M. F. Gerenciador de conteúdos da Agência Embrapa de Informação. In: INTERNATIONAL SYMPOSIUM ON KNOWLEDGE MANAGEMENT - ISKM = SIMPÓSIO INTERNACIONAL DE GESTÃO DO CONHECIMENTO, 2003, Curitiba. **Anais...** Curitiba: Pontifícia Universidade Católica do Paraná, 2003. p.1-12. CD-ROM.

HAN, J. KAMBER, M. **Data mining concepts and techniques**. San Diego, CA: Academic Press, 2001. 550 p.

LENK, L. M.; ANTUNES, J. F. G. **AINFO 2001**: manual *online*. Campinas: Embrapa Informática Agropecuária, 2002. 292 p. (Embrapa Informática Agropecuária. Documentos, 10). Disponível em: <<http://www.cnptia.embrapa.br/modules/tinycontent5/Content/manuais/ainfo.pdf>>. Acesso em: 20 set. 2004.

MOURA, M. F., SANTOS, A. D. dos. **Ferramentas automatizadas para a manutenção e acesso a dados das Agências de Informação da Embrapa**: apostila de treinamento. Campinas: Embrapa Informática Agropecuária, 2002. 78 p. (Embrapa Informática Agropecuária. Documentos, 17).

PACHECO, R. C. dos S.; KERN, V. M. Uma ontologia comum para a integração de bases de informações e conhecimento sobre ciência e tecnologia. **Ciência da Informação**, Brasília, DF, v. 30, n. 3, p. 56-63, set./dez. 2001. Disponível em: <<http://www.scielo.br/pdf/ci/v30n3/7287.pdf>>. Acesso em: 10 out. 2004.

RIBEIRO, A. M. de C. M. **Catálogo de recursos bibliográficos pelo AACR2R 2002**: Anglo-American Cataloguing Rules. 2. ed. rev. Brasília, DF: Ed. do Autor, 2003. v. 1.

SOUZA, K. X. S. de. Agência de Produtos e Serviços de Informação. In: WORKSHOP O AGRONEGÓCIO NA SOCIEDADE DA INFORMAÇÃO - AGROSOFF - 2002, 2002, Brasília. **Anais...** Brasília, DF: Sociedade da Informação, 2002. Trabalho apresentado.

WORLD WIDE WEB CONSORTIUM. **Extensible markup language (XML)**. Disponível em: <<http://www.w3.org/>>. Acesso em: 20 set. 2004a.

WORLD WIDE WEB CONSORTIUM. **RDF primer - W.C recommendation 10 February 2004**. Disponível em: <<http://www.w3.org/TR/REC-rdf-syntax/>>. Acesso em: 7 out. 2004b.

WORLD WIDE WEB CONSORTIUM. **W3C - World Wide Web Consortium**. Disponível em: <<http://www.w3.org/>>. Acesso em: 20 set. 2004c.

Comunicado Técnico, 66

Ministério da Agricultura, Pecuária e Abastecimento



Embrapa Informática Agropecuária
Área de Comunicação e Negócios (ACN)
Endereço: Caixa Postal 6041 - Barão Geraldo
13083-970 - Campinas, SP
Fone: (19) 3789-5743
Fax: (19) 3289-9594
e-mail: sac@cnptia.embrapa.com.br

1ª edição on-line - 2005

Todos os direitos reservados.

Comitê de Publicações

Presidente: Marcos Lordello Chaim (*presidente em exercício*)
Membros Efetivos: Carla Geovana Macário, Ivanilde Dispatto, José Ruy Porto de Carvalho, Luciana Alvim Santos Romani, Marcia Isabel Fugisawa Souza, Suzilei Almeida Carneiro (*secretária*)
Suplentes: Carlos Alberto Alves Meira, Eduardo Delgado Assad, Maria Angelica Andrade Leite, Maria Fernanda Moura, Maria Goretti Gurgel Praxedis

Expediente

Supervisor editorial: Ivanilde Dispatto
Normalização bibliográfica: Marcia Izabel Fugisawa Souza
Editoração eletrônica: Área de Comunicação e Negócios