



ASSOCIAÇÃO PAULISTA DE BIBLIOTECÁRIOS

ENSAIOS APB

METADADOS
(Revisão de Literatura)

Marta Lígia Pomim Valentim

Ensaio APB, n. 86

APB - ASSOCIAÇÃO PAULISTA DE BIBLIOTECÁRIOS - APB

**METADADOS
(Revisão de Literatura)**

Marta Lígia Pomim Valentim

Ensaio APB, n. 86

APB - ASSOCIAÇÃO PAULISTA DE BIBLIOTECÁRIOS - APB

**METADADOS
(Revisão de Literatura)**

Marta Lígia Pomim Valentim

Ensaio APB, n. 86

**São Paulo
Janeiro
2001**

METADADOS revisão de literatura

Marta Lígia Pomim Valentim (1)

O formato eletrônico estará cada vez mais presente no cotidiano, seja intra ou extra muros. Muitas informações surgem, atualmente, no formato eletrônico e assim são disponibilizadas.

No final da década de sessenta existia uma dificuldade em organizar e disseminar informações coletadas de diferentes origens em um único arquivo eletrônico. Por isso, surgiu naquela época o Common Communication Format (CCF) "com o objetivo de facilitar a comunicação de dados bibliográficos entre os setores da comunidade de informação especializada" (Rowley, 1994, p.81).

Rowley (1994, p.81) afirma:

"os serviços de indexação e resumos (mas não todos) adotam o manual de referência United Nations Information System in Science and Technology (UNISIST), que prescreve seus próprios designadores de conteúdo para as descrições bibliográficas de vários tipos de material. Esses dois importantes formatos definem, organizam e identificam os elementos de dados de modo diferente e se baseiam em diferentes conjuntos de códigos".

Com a evolução das tecnologias da informação e, principalmente, com a rede Internet e com a World Wide Web, houve necessidade de desenvolver modelos de compatibilização entre diferentes formatos: o conceito de metadados.

Na opinião de Takahashi

"metadados são dados que descrevem outros dados, em uma definição sem maior rigor técnico. Por exemplo, em um formulário qualquer que tenha sido preenchido com os dados de um indivíduo, há tipicamente, em cada campo do formulário, uma informação explicitando que tipo de dados deve ser ali descrito: por exemplo, "NOME COMPLETO", "ENDEREÇO" etc. Essa informação é um metadado (...) O conceito de metadados tem adquirido importância crescente na medida em que mais e mais conteúdos completos são armazenados em computadores e transmitidos via redes: informações cartográficas, grandes bases de dados em textos livres em diversas línguas etc. O pro-

¹ Doutora em Ciência da Informação e Documentação pela ECA/USP; autora do livro "O Custo de Informação Tecnológica"; organizadora e uma das autoras do livro "O Profissional da Informação"; professora da Universidade Estadual de Londrina.

cessamento adequado dessas grandes massas de dados passa pelo processamento dos metadados entremeados aos dados, que impõem estrutura e inteligibilidade aos mesmos" (2000, p.59).

O conceito de metadados exposto por Souza et al.:

"são descrições de dados armazenados em banco de dados, ou como é comumente definido dados sobre dados a partir de um dicionário digital de dados. Esse dicionário de dados normalmente é utilizado para organizar os metadados. Ele poderá conter uma seção descrevendo, numa visão geral, como os dados são subdivididos em arquivos, que campos de registros se relacionam e possuir tópicos tais como: convenções adotadas em sua definição (...) O metadado é estruturado com elementos de descrição do conteúdo dos dados" (1997, p.94-95).

Ainda segundo Souza et al. é importante que haja padrões de conteúdo, para que o intercâmbio de dados em meio eletrônico tenha qualidade, assim como também possibilite uma recuperação da informação mais efetiva. A definição de padrões contribui para especificar o formato dos dados e quais informações são necessárias para que o acesso e a recuperação da informação seja eficaz.

Dentre as vantagens dos metadados apontadas pelos autores estão:

- √ *"Estabelecimento de padrões de dados diante da heterogeneidade de informações contidas na rede;*
- √ *Facilidade na definição da linguagem de consulta;*
- √ *Facilidade e maior precisão na recuperação das informações desejadas;*
- √ *Troca de informações entre aplicações e entre organizações" (Souza et al., 1997, p.97).*

A importância dos modelos de metadados está na padronização e transferência de dados, no custo da armazenagem e sua recuperação, diante da heterogeneidade de tipos documentais existentes, assim como na linguagem utilizada, otimizando desta forma, os recursos de informação disponíveis em meios eletrônicos.

Sherry L. Vellucci fez uma revisão na literatura sobre metadados, cobrindo desde o segundo semestre de 1996 até o início de 1998. Sua revisão enfocou o escopo e limitações do termo metadado, relato de projetos, definição e conceituação de metadado, categorias e funções de metadados, modelos de metadados, elementos que compõem a geração de metadados, formatos e linguagens de marcação de metadado, domínios específicos de metadados e esforços cooperativos para o desenvolvimento de metadados.

Observa Vellucci (1998, p.191) que a comunidade de catalogação descreve recursos de informação não eletrônicos de forma detalhada e estabelecida através de regras fixas de descrição, tradicionalmente usa as condições dos dados bibliográficos para também catalogar esse tipo de infor-

mação, ou seja, o metadado. As diferentes "áreas" do Código de Catalogação Anglo-Americano exemplificam a abordagem aqui adotada. Quando os métodos de organização da informação da área da ciência da informação e biblioteconomia convergem para o ambiente eletrônico, o termo metadado passa a ter a conotação de descrever dados sobre dados eletrônicos e passa a ser usado dessa forma.

Dempsey e Heery apud Vellucci (1998, p.191) definem metadado como dados que descrevem os atributos de um recurso ou dados associados com objetos que aliviam os usuários potenciais de ter que conhecer de antemão a existência ou características desses recursos. Os autores prepararam o projeto Development of a European Service for Information on Research and Education (DESIRE) que estabelece três categorias de metadados:

1. Dados desestruturados - extraídos automaticamente de fontes através de pesquisa de indexação;
2. Dados estruturados de forma básica - quando apresentam campos de pesquisa;
3. Dados altamente estruturados - com códigos de identificação/localização, descrição da especialidade e domínio específico (Vellucci, 1998, p.189).

A vantagem do uso do termo metadado, para Caplan apud Vellucci (1998, p.191), está na neutralidade do termo e, também, por não ser pejorativo, associado, por exemplo, ao termo tradicional "registro de catalogação" que tem uma imagem negativa em ambientes fora da biblioteconomia.

A definição do termo metadado pela World Wide Web Consortium apud Vellucci (1998, p.191) é "informação compreendida por máquina sobre objetos Web".

Vellucci (1998, p.192) incorporou componentes de cada definição apresentada e definiu metadado como: "dados que descrevem atributos de um recurso/fonte, caracteriza suas relações, contribui para a sua localização e uso efetivo e existe no ambiente eletrônico".

Esses atributos identificam e descrevem a informação propriamente dita, além de contribuírem para a identificação e acesso de forma rápida e eficaz, nos diferentes recursos eletrônicos disponíveis na Internet.

Muitos definem metadado, como dado sobre dado. Esta definição está apoiada em alguns aspectos como a facilidade de identificação e de recuperação, justamente pelas marcações geradas espontaneamente pela maioria dos modelos de metadados, ou seja, alguns tipos de metadados são gerados automaticamente dependendo da linguagem utilizada.

Para Lagoze et al. apud Vellucci (1998, p.192) o termo metadado extrapola o contexto de identificação/localização, bem como de descrição e recuperação e se estende até as áreas de gerenciamento e uso desses recursos, como, por exemplo:

- 1) termos e condições dos dados;
- 2) dados administrativos;
- 3) avaliação de satisfação de uso dos dados;
- 4) dados de proveniência;
- 5) encadeamento ou relação de dados;
- 6) dados estruturais.

A descrição da informação através de um padrão de marcação estabelece com a máquina um reconhecimento imediato de conteúdos básicos e complexos de um documento eletrônico. Como exemplo, o artigo eletrônico de Sherry L. Vellucci "Options for organizing electronic resources: the coexistence of metadata será demonstrado utilizando um modelo de metadados (Figura 1).

Figura 1 Exemplo de Metadado

VELLUCCI, SHERRY L. 1997. Options for organizing electronic resources: The Coexistence of Metadata. Bulletin of the American Society for Information Science. 1997 October/November; 24(1). ISSN: 0095-4403. Available WWW: <http://www.asis.org/Bulletin/Oct-97/vellucci.htm>.

Figure 1. Normal Citation Data for Online Journal Article

```
<html>
<head>
<title></title>
<meta name="DC.Title" content="Options for Organizing Electronic Resources">
<meta name="DC.Title.Alternative" content="The Coexistence of Metadata">
<meta name="DC.Creator.PersonalName" content="Vellucci, Sherry L.">
<meta name="DC.Creator.PersonalName.Address" content="velluccs@stjohns.edu">
<meta name="DC.Subject" scheme="LCSH" content="Metadata">
<meta name="DC.Subject" scheme="LCSH" content="Cataloging of computer network
resources">
<meta name="DC.Subject" scheme="LCSH" content="World Wide Web (Information retrieval
system)">
<meta name="DC.Subject" scheme="DDC" content="005.721">
<meta name="DC.Subject" scheme="LCC" content="QA76.9">
<meta name="DC.Description" content="Article discusses organizing Internet resources in library,
scholarly, and Internet communities and the need to layer, exchange, and translate data among
various metadata formats.">
<meta name="DC.Publisher" content="American Society for Information Science">
<meta name="DC.Publisher.Address" content="asis@asis.org">
<meta name="DC.Date" scheme="ISO8601" content="1997-10">
<meta name="DC.Type" content="Text.Article">
<meta name="DC.Format" scheme="IMT" content="text/html">
<meta name="DC.Identifier" content="http://www.asis.org/Bulletin/Oct-97/vellucci.htm">
<meta name="DC.Source" scheme="ISSN" content="0095-4403">
<meta name="DC.Language" scheme="ISO639-1" content="en">
<meta name="DC.Relation" content="Bulletin of the American Society for Information Science,
vol.24, n.1, October/November 1997">
<meta name="DC.Rights" scheme="URL" content="http://www.asis.org/policies/web-policy.html">
<meta name="DC.Date.X-MetadataLastModified" scheme="ISO8601" content="1998-10-03">
<Link Rel=schema.dc HREF="http://purl.org/metadata/dublin_core_elements">
</head>
```

FONTE: Vellucci, S. L. - 1998 - p.193

A estrutura apresentada na figura acima, mostra de que forma o metadado é constituído em um documento eletrônico (Anexo 1). Por isso é possível localizá-lo, reconhecer seus conteúdos, identificar responsabilidade de autoria etc.

O reconhecimento desses conteúdos pode ser realizado também por uma pessoa, desde que conheça a linguagem de marcação utilizada. A descrição da informação pode ser formatada de acordo com as necessidades de quem está criando o documento eletrônico ou simplesmente ser automática através de um modelo de metadados.

A linguagem de marcação e descrição, por exemplo, utilizando HyperText Markup Language (HTML), é composta de descritores que podem ser caracterizados em:

- a) elementos necessários - tipo de documento, título do documento, indica o corpo do documento, indicação do cabeçalho do documento etc.;
- b) Cabeçalhos, estilo de tipos e descritores de parágrafos - indica a quantidade de cabeçalhos, tipo de fonte, formatos de parágrafos, indicações de início e término de parágrafos;
- c) Códigos de localização - como Uniform Resource Locators (URL) e HyperText Transfer Protocol (HTTP), File Transfer Protocol (FTP), Protocolo Gopher, Protocolo Telnet etc.;
- d) Descritores - como correio eletrônico (e-mail), Wide Area Information Server (WAIS) etc.;
- e) Âncoras e relações - âncora de *hiperlink*, *links*, *hiperlinks*, etc.;
- f) Listas - Lista ordenada, numérica, tipo diretório, tipo dicionário etc.;
- g) Texto e Tabelas - texto formatado, pré-formatado, tabelas, alinhamentos, colunas, linhas etc.;
- h) Inclusão de imagens - inclui arquivo de imagem, alinhamento de imagem; *hiperlink*, etc.;
- i) Formulários - criados automaticamente, programa para dados de formulário, tipo, etc.;
- j) Geral - traços, código de caracteres numéricos, linhas de comentários, etc.;
- k) Caracteres diferenciados - acentuação, etc. (Evans, 1996).

Essas marcações são encontradas em todos os tipos de informação eletrônica disponibilizadas através de recursos eletrônicos como a rede Internet. Modelos de metadados diferentes são usados para informação textual, imagem e som.

Com relação aos modelos de metadados muitos autores os definem da seguinte forma:

- √ devem ser simples, curtos e fáceis de entender e usar (Hakala et al. apud Vellucci, 1998, p.194).

- √ podem ser descritos agregados de imagens ou textos, como coleções, como únicas entidades ou como ambos. Determinar o nível apropriado de descrição e distingui-los entre si para assegurar uma taxa inteligível aceitável nos resultados de pesquisa não é fácil (Cromwell-Kessler apud Vellucci, 1998, p.194-95).
- √ deve haver um esforço no desenvolvimento de modelos de metadados, buscando uma solução para o problema acima mencionado (Cromwell-Kessler; Daniel e Lagoze apud Vellucci, 1998, p.195).

Lange e Winkler apud Vellucci (1998, p.188) comparam os prós e contras da catalogação tradicional com as novas estruturas de metadados e propõem a utilização de alguns modelos/normas de tratamento de metadados como: a) Text Encoding Initiative Headers (TEI); b) Encoded Archival Description (EAD); c) Federal Geographic Data Committee (FGDC). Esses modelos de metadados serão mencionados mais detalhadamente neste artigo.

Heery apud Vellucci (1998, p.189) compara cinco modelos de metadados: a) IAFA²; b) MARC³; c) TEI; d) Dublin Core⁴; e) Uniform Resource Characteristics -- um relatório preliminar --, do Resource Organization and Discovery in Subject-Based Services (ROADS). Heery considera o desenvolvimento de modelos e padrões para metadados na Internet, fundamental na resolução ou diminuição dos problemas de descrição dos metadados. Todos os modelos elencados por Heery utilizam critérios idênticos como: consistência; facilidade de criação; conteúdo; associação de protocolos Internet.

Três modelos de criação e uso de metadados são propostos por Weibel apud Vellucci (1998, p.195):

- modelo embutido - os metadados são formulados pelo criador do documento eletrônico e residem no recurso que é descrito utilizando, por exemplo, a linguagem HTML;
- modelo agenciado - uma agência cria, coleta e gerencia registros de metadados referentes aos recursos separadamente;
- modelo filtrado - onde uma agência distinta gerencia os diversos registros de metadados, de uma variedade de fontes que usam estruturas diferentes, traçando um padrão comum para as diferentes fontes e dados, exatamente como o modelo Dublin Core.

A preocupação com modelos estruturais de metadados começa com um dos maiores grupos de trabalho da Internet, o The Internet Engineering Task Force (IETF), eles desenvolvem e pesquisam a arquitetura e operação de dados na Internet e criaram uma estrutura denominada RFC 1807 que propu-

² The Internet Anonymous FTP Archive (IAFA)

³ Machine-Readable Cataloguing (MARC)

⁴ Dublin Core Metadata Element Set (DC)

nha a definição de um formato para troca de descrição bibliográfica de relatórios técnicos através do *e-mail* (Lasher e Cohen apud Vellucci, 1998, p.196).

Outra estrutura de metadados utilizada pelo IAFA / Whois++ é o FTP, modelo usado para acesso de arquivos de informação via Internet. Os grupos de trabalho IAFA e IETF projetaram modelos para facilitar o acesso efetivo de arquivos de informação descrevendo os conteúdos e serviços disponíveis em um arquivo particular FTP (Heery apud Vellucci, 1998, p.196).

O modelo estrutural de metadados do IAFA é simples texto em ASCII⁵ com os elementos de dados definidos como atributos e, para isso, usa uma estrutura de arquivo plano que permite o encadeamento de *links* de forma limitada.

Historicamente o primeiro formato de intercâmbio bibliográfico foi desenvolvido na década de sessenta pela Library of Congress (LC) e pela British Library; denominado MARC ele contemplava até 61 elementos de dados. Dele surgiram algumas propostas variantes desse formato como: a) UNIMARC; b) OCLCMARC; c) USMARC; d) UKMARC; e) IberMARC; f) Canadian MARC; g) BRMARC. Todos esses formatos são baseados na norma International Standard Organization (ISO) 2709, sobre intercâmbio de dados bibliográficos em meio eletrônico (Rowley, 1994, p.77; Vellucci, 1998, p.195-197).

Na opinião de Vellucci, o registro MARC é um metadado, baseado no Anglo American Cataloguing Rules (AACR), modelo de descrição de dados tradicional na área de biblioteconomia, criado pela American Library Association (ALA). O MARC visa o acesso público de catálogos *online* (OPACs⁶) através da Internet (1998, p.196).

Uma variação do MARC, o USMARC, desenvolvido pela Online Computer Library Center's (OCLC), através do Projeto InterCat⁷, tem objetivo similar, isto é, o acesso a registros bibliográficos via Internet, visando a cooperação de dados, utilizando um padrão/modelo de estrutura de dados.

Muitos projetos surgiram (Figura 2) com o objetivo de organizar e estabelecer padrões para o acesso e a recuperação de informação eletrônica de forma rápida e eficaz. A tendência é que os diferentes padrões/modelos de metadados sejam cada vez mais compatíveis e, talvez, no futuro, as tecnologias de informação inteligentes reconheçam qualquer tipo de modelo de metadados independente de sua organização e de seu objetivo.

⁵ American Standard Code for Information Interchange (ASCII)

⁶ Online Public Access Catalogs (OPACs)

⁷ A Catalog of Internet Resources (InterCat)

Figura 2
Projetos de Padronização da Descrição de Metadado

Projeto	Responsável	Ano de Criação
Catriona II ⁸	Strathclyde University e Napier University (Escócia/UK)	1996
CORC ⁹	Online Computer Library Center (OCLC)	1996
Dublin Core	OCLC e NCSA ¹⁰	1995
InterCat	Online Computer Library Center (OCLC)	1993

Muitos autores enfatizam a importância do campo 856¹¹ do formato MARC, imprescindível para a localização da informação na rede Internet, uma vez que contém o URL e permite *links* hipertextuais diretos entre os registros MARC, assim como os formatos derivados e outros recursos eletrônicos disponíveis na rede (Dillon e Jull; Woodward; Vellucci apud Vellucci, 1998, p.197).

O Catriona é um projeto de pesquisa que surgiu da necessidade de duas universidades escocesas organizarem suas informações/recursos eletrônicos, de forma que a armazenagem, o acesso e a recuperação respeitassem um padrão de qualidade aceitável.

O CORC baseado no InterCat, ambos da OCLC, também é um sistema que visa a organização, armazenagem, acesso e recuperação de metadados com qualidade. Esse sistema tem uma preocupação com relação à maximização de conhecimento e compartilhamento de recursos, uma vez que seu objetivo é a cooperação de dados.

Num esforço internacional para estabelecer um modelo que possibilite uma cooperação de metadados surge o Dublin Core, pela primeira vez no evento OCLC/NCSA Metadata Workshop, realizado em março de 1995. Atualmente, existem 150 participantes no projeto, distribuídos em 20 países e, dentre o desenvolvimento da pesquisa, baseado no protocolo Z39.50, estabeleceram 15 elementos de descrição para metadados.

Os elementos de descrição do DC são:

- a) *"Título - o nome (identificação) dado ao documento ou recurso eletrônico;*
- b) *Autor/Criador - pessoas físicas ou jurídicas responsáveis pelo conteúdo intelectual do documento/recursos eletrônico;*
- c) *Assunto/Palavras-Chave - conteúdo temático do documento ou recurso eletrônico pode ser expressado por palavras-chave, índices, etc.;*

⁸ Cataloguing and Retrieval of Information Over Networks Applications (Catriona)

⁹ The Cooperative Online Resource Catalog (CORC)

¹⁰ National Center for Supercomputing Applications (NCSA)

¹¹ O campo 856 do formato MARC, contém informação da localização eletrônica (URL)

- d) *Descrição* - também referente ao conteúdo temático do documento/recurso eletrônico, pode ser expressado através de resumos quando for conteúdo textual através de descrição - Document-like Objects (DLO) -, quando conteúdo 'midiático';
- e) *Editor* - instituição, setor ou departamento responsável pela publicação/disponibilização do documento/recurso eletrônico;
- f) *Outros Colaboradores* - pessoas físicas ou jurídicas que contribuíram com o conteúdo intelectual do documento/recursos eletrônico;
- g) *Data* - a data em que o documento/recurso eletrônico foi publicado/disponibilizado naquele formato;
- h) *Tipo do Recurso* - gênero do documento/recurso eletrônico, ou seja, tipologias documentais e tipologias 'midiáticas';
- i) *Formato* - a identificação física do documento/recurso eletrônico, como por exemplo, a linguagem utilizada HTML;
- j) *Identificação do Recurso* - código utilizado para identificar o documento/recursos eletrônico, como por exemplo a URL;
- k) *Fonte* - identificação do qual se originou o documento/recurso eletrônico;
- l) *Língua* - idioma do conteúdo intelectual do documento/recurso eletrônico;
- m) *Relação* - estabelecimento de possíveis relações com outros documentos/recursos eletrônicos;
- n) *Cobertura* - características espaciais ou temporais do conteúdo intelectual do documento/recurso eletrônico;
- o) *Direito Autoral* - informações sobre os direitos de propriedade ou intelectuais sobre o documento/recurso eletrônico" (Bearman et al. 1999, p.6-12; Weibel, 1995, p.3; Weibel, 1997, p.5).

"As principais características do padrão DC são a simplicidade na descrição dos recursos, entendimento semântico universal (dos elementos), escopo internacional e extensibilidade (o que permite sua adaptação às necessidades adicionais de descrição)" (Souza, M. et al., 2000, p.94; Cromwell-Kessler, 1997, p.2; Vellucci, 1998 p.208).

Vários esforços de diferentes grupos de pesquisa e comissões de trabalho resultaram em publicações, relatórios, artigos científicos etc. Dentre eles, pode-se citar o Dublin Core Literature, publicação que introduz os interessados no universo Dublin Core. A revisão elaborada por Thiele categoriza a literatura em cinco grupos: a) procedimentos e relatórios de vários seminários; b) guia e mapeamento do Dublin Core para outros sistemas de metadados; c) relação de outros padrões com o Dublin Core; d) biblioteca digital e projetos de metadados; e) artigos diversos (Vellucci, 1998, p.209).

Ainda sobre os esforços cooperativos para entender e usar o modelo Dublin Core, Powel apud Vellucci descreve três modelos de gerenciamento de metadados através de sites na Web e descreve ferramentas como

Server Side Include (SSI), esquema para inserir o metadado Dublin Core nas páginas da Web (1998, p.210).

O Spectrum Cataloging Markup Language (SCML), linguagem de marcação de metadados é uma ferramenta de conversão de metadados que pode extrair metadados da estrutura de registros eletrônicos disponíveis na Internet, implementando extensões do Dublin Core e gerando novos registros no formato Dublin Core (Godby e Miller apud Vellucci, 1998, p.210).

Ainda em relação à conversão de metadados existe a possibilidade de incorporar os elementos de dados do Dublin Core à Apple Computer's Metadata Content Format (MCF) e usá-los com os elementos de dados da Apple como se fossem uma ferramenta 'HotSauce'. Da mesma forma, existe a possibilidade de os metadados Dublin Core, inseridos na linguagem HTML, serem convertidos em registros MARC (Knight apud Vellucci, 1998, p.210).

O projeto InterCat serviu para aumentar a consciência dos bibliotecários de catalogação em relação à organização dos recursos informacionais na Internet e o MARC, como um dos primeiros formatos de metadados, começou a preocupar-se com o conteúdo textual desses metadados (Vellucci, 1998, p.198).

Para Sun apud Vellucci alguns problemas envolvem os formatos baseados no MARC, dentre os quais, por exemplo, o acesso aos registros eletrônicos chineses, uma questão essencialmente ligada aos caracteres, ou seja, o vernáculo da língua chinesa necessita de softwares paralelos de conversão, para que o acesso realmente possa ocorrer. Outros autores contribuem para esta discussão e sugerem a adoção/uso de outros tipos de modelos de metadados como o TEI e Dublin Core que gerenciam este problema (Brugger; Gaynor; Hillman; Jeng; Palowitch e Horowitz; Vellucci; Younger apud Vellucci, 1998, p.199).

A catalogação de *Web sites* consiste primeiramente nos *links* com outros recursos da Internet, os *Sites* devem ser pensados como uma forma moderna de apresentação de registros bibliográficos (Beall; Johns; Neumeister; Sun apud Vellucci, 1998, p.198).

Da mesma maneira que o MARC formata a estrutura de um registro informacional e o codifica para as OPACs, linguagens de marcação formatam estruturas informacionais e as codificam na Internet. No entanto, o formato MARC distingue o registro bibliográfico propriamente dito dos códigos que o descrevem. Linguagens de marcação são usadas para codificar os dados descritivos, assim como o próprio objeto. Isso permite ao metadado ser inserido dentro do próprio documento. Atualmente a linguagem de marcação mais usada é a Standard Generalized Markup Language (SGML) (Vellucci, 1998, p.199).

A termo Document Type Definition (DTD) refere-se à implementação de SGML. Muito usado como um padrão para codificar texto, utiliza linguagens HTML e SGML, através da definição de vários elementos de dados para

tipos específicos de documentos. DTD define os tipos de documentos e suas estruturas, especificando quais elementos são requeridos em um determinado tipo de documento, bem como que elementos podem estar presentes no documento. A estrutura de um documento pode ser inteiramente marcada e pode ser conferida novamente pelo DTD (Library of Congress; Gaynor apud Vellucci, 1998, p.199).

O TEI também utiliza a linguagem SGML e desenvolve guias para a preparação e o intercâmbio de recursos eletrônicos, incluindo tanto os textuais quanto os não-textuais. A descrição pode ser mantida em um arquivo separado, como também pode ser incluída no próprio recurso eletrônico (Sperberg-McQueen, Burnard apud Vellucci, 1998, p.200).

Giordano apud Vellucci também explica que o TEI descreve informação bibliográfica e não-bibliográfica e possibilita a identificação, recuperação e análise de máquina da informação eletrônica. A linguagem de marcação do TEI tem quatro componentes: 1) descrição do arquivo; 2) descrição codificada; 3) perfil da descrição; 4) revisão da descrição. O TEI pode variar em tamanho e complexidade, pois depende da natureza do projeto e da quantidade de informação que precisa ser codificada (1998, p.200).

Heery, apud Vellucci, enfatiza a flexibilidade do TEI, uma vez que é o único que possui a capacidade de ajustar o contexto do metadado, podendo ser aplicado a qualquer recurso de eletrônico, isto é, estruturados ou não-estruturados (1998, p.201).

Alguns grupos de trabalhos ligados à comunidade bibliotecária examinam a possibilidade da conversão automática de formatos TEI em formatos MARC, realizado através de um programa de conversão, desenvolvido pela University of Virginia (Palowitch, Horowitz; Vellucci apud Vellucci, 1998, p.201).

Existem mais de 60 projetos utilizando o TEI Guidelines para textos completos, dentre os quais: o Dunlop, processo de recuperação direcionado ao conteúdo textual, que usa o TEI com três objetivos: 1 - descrever o conteúdo textual; 2 - descrever o texto de forma individual; 3 - descrever o texto como um registro bibliográfico único (Vellucci, 1998, p.201).

O Encoded Archival Description (EAD), outro tipo de SGML DTD, foi desenvolvido mais recentemente para a comunidade de arquivologia, respondendo à crescente demanda de acesso à informação na Internet, com o objetivo de expandir os registros de dados no formato MARC (Thibodeau et al. apud Vellucci, 1998, p.202).

O Computer Interchange of Museum Information (CIMI DTD), adotou a linguagem SGML para troca entre diferentes formatos. Uma parte do projeto denominado Cultural Heritage Information Online (CHIO) desenvolveu o SGML DTD com o objetivo de exibir o catálogo, sendo o primeiro DTD desenvolvido para gerar informação de museus (Burnard, Light; Dempsey, Heery apud Vellucci, 1998, p.204).

The Visual Arts Data Service (VADS), um dos componentes do Arts and Humanities Data Service (AHDS), produziu um modelo específico para a área, extraído de normas de artes visuais, museus e informação cultural. O CIMI DTD é a soma de doze normas direcionadas a materiais visuais e, a partir de uma revisão, elaboraram aproximadamente 35 pontos de acesso básicos para registros SGML (Gill et al. apud Vellucci, 1998, p.205).

Government Information Locator Service (GILS) é um sistema de documentos e informações governamentais de metadados e que inicialmente foi considerado especificamente para essa aplicação. O GILS foi implementado para ajudar o setor público a identificar: a) localização; b) acesso e distribuição de informação governamental às agências governamentais; c) reduzir a duplicação de informação nas diversas agências governamentais (Adam, Thibodeau; Andrews, Duhon; Moen apud Vellucci, 1998, p.205). Observa Christian, apud Vellucci, que o GILS teve uma grande expansão em nível internacional, mudando até mesmo o sentido inicial da sigla GILS para Global Information Locator Service (1998, p.206).

Geospatial Metadata, desenvolvido pela Federal Geographic Data Committee, é apoiado na norma da FGDC. Esse modelo de metadado é entendido como uma ajuda aos usuários, de forma que o usuário consiga determinar os dados existentes, a qualidade dos dados, as condições de acesso e como deve ser o processamento dos dados (Dempsey, Heery; Allen apud Vellucci, 1998, p.206).

Domaratz apud Vellucci, descreve o desenvolvimento do National Spatial Data Infrastructure e o National Geospatial Data Clearinghouse, enfatizando o acesso e a recuperação de dados. Descreve sete características informacionais do Spatial metadados:

- 1) Identificação da informação;
- 2) Qualidade dos dados da informação;
- 3) Dados de localização da informação organizada;
- 4) Localização da informação de referência;
- 5) Informação / atributo da entidade/instituição;
- 6) Distribuição da informação;
- 7) Informação de referência do metadado (1998, p.207).

O Alexandria Project foi designado para explorar problemas relacionados à distribuição geográfica de informação de referência. The Alexandria Digital (ADL) é o centro do projeto e utiliza uma combinação de formato USMARC e FGDC. A coleção da ADL possui uma variedade de dados geoespaciais junto de seu metadado, e inclui somente metadados de uma variedade de fontes de referência geográficas (Frew et al.; Smith apud Vellucci, 1998, p.208).

Os esforços realizados para o desenvolvimento de uma arquitetura de metadados que seja flexível e converta as diferentes estruturas de elementos de dados sinalizam alguns avanços com o próprio Dublin Core. The Warwick Framework é uma arquitetura que agrega múltiplos elementos de

metadados e permite que a descrição tenha um denominador comum, como ocorre com o Dublin Core e o MARC, e habilita o acesso para cada metadado isoladamente ou para todo o pacote de metadados (Lagoze; Lagoze et al.; Thiele; Kahn, Wilensky; Dempsey, Weibel; Hakala et al. apud Vellucci, 1998, p.210-211).

O desenvolvimento de uma extensão do Warwick Framework chamada Distributed Active Relationships (DARs) que opera num contexto da Warwick Framework Catalog possui um mecanismo de expressão relacionando os pacotes de metadados dentro da arquitetura do Warwick Framework. Duas implantações do DARs ocorreram: uma delas é usada no Flexible and Extensible Digital Object Repository Architecture (FEDORA) e a segunda é empregada no Resource Description Framework (RDF) (Daniel, Lagoze apud Vellucci, 1998, p.211).

O RDF é a arquitetura de metadados mais recente em desenvolvimento. Como no Warwick Framework, o RDF provê uma infra-estrutura que dá suporte ao processamento automatizado de metadados junto aos diversos recursos eletrônicos da Internet e contém a descrição e relação de expressão num *Web site*, numa *Home Page*, ou numa biblioteca digital. Serve como modelo de dados representando metadados; a possibilidade da sintaxe de expressão e transporte do metadado vai maximizar a independência da interoperabilidade desenvolvida pelos servidores e clientes Web. A linguagem de marcação utilizada pelo RDF é restrita à eXtensible Markup Language (XML) (World Wide Web Consortium apud Vellucci, 1998, p.211).

Vellucci finaliza a revisão afirmando que, enquanto os princípios gerais da organização, acesso e recuperação da informação, mais especificamente dos metadados, não forem direcionados para esforços cooperativos que resultem em um modelo flexível, aplicado a diferentes estruturas de dados, haverá disputa para impor modelos. Acredita que o desenvolvimento de ferramentas de extração automática, geração e conversão da descrição de metadados deve continuar buscando uma nova estrutura que requererá uma aplicação mais ampla de registros de dados, desenvolvendo-se em sistemas de recuperação verdadeiramente eficientes.

Nesse sentido, os modelos de metadados são fundamentais para contribuir para a qualidade das bases de dados eletrônicas, uma vez que esses modelos são a base da descrição/representação do documento e da descrição/representação dos conteúdos dos documentos.

As bases de dados devem contemplar as diferentes estruturas textuais com que trabalham, isto é, todos os tipos documentais existentes em uma base de dados devem ser contemplados integralmente no modelo de dados da base. As que utilizam modelos de metadados adequados às estruturas textuais ou tipologias documentais, representam a informação com maior adequação e completude. Portanto, pressupõe-se que o resultado de pesquisa obtido também será mais adequado e completo.

A organização da informação, neste caso as representações descritiva e temática, deve ser aplicada buscando a interação e a conectividade de dados. As redes eletrônicas exigem uma postura mais cooperativa e globalizada, visando o grande fluxo de informação, bem como os serviços disponíveis em formatos eletrônicos. As estruturas e os diferentes conteúdos informacionais existentes nas redes não são estáticos, assim como o seu público usuário. Por essa razão, existe a necessidade dos profissionais da informação trabalharem de forma integrada, objetivando organizar de forma eficiente o caos informacional eletrônico.

Referências Bibliográficas

BEARMAN, D. et al. A common model to support interoperable metadata. Progress report on reconciling metadata requirements from the Dublin Core and INDECS/DOI Communities. D-Lib Magazine, v.5, n.1, Jan. 1999. Disponível em: <<http://www.dlib.org/dlib/january99/bearman/01bearman.html>>. Acesso em: 12 set. 2000.

CORMWELL-KESSLER, W. 2000. Dublin Core metadata in the RLG information landscape. D-Lib Magazine, Dec. 1997. Disponível em: <<http://www.dlib.org/dlib/december97/cormwell-kessler.htm>>. Acesso em: 12 set. 2000.

DORNER, D. 2000. Gazing into the future: cataloguing in the 21st century. IN: CATALOGUING SPECIAL INTEREST GROUP AT THE 1997 NZLIA CONFERENCE, 1997, Wanganui. Anais eletrônicos... Wanganui, New Zealand: 1997. Disponível em: <http://www.vuw.ac.nz/dlis/publins/futcat.htm> Acesso em: 16 set. 2000.

HARTLEY, J., SYDES, M., BLURTON, A. 1996. Obtaining information accurately and quickly: are structured abstracts more efficient? Journal of Information Science, v.22, n.5, p.349-356, 1996.

LEE, H. 1995. Justifying database normalization: a cost/benefit model. Information Processing & Management, v.31, n.1, p.59-67, 1995.

RITTBERGER, M., RITTBERGER, W. 1997. Measuring quality in the production of databases. Journal of Information Sciences, v.23, n.1, p.25-37, 1997.

ROWLEY, J. 1994. Informática para bibliotecas. Brasília: Briquet de Lemos, 1994. 307p.

SOUZA, I. F. et al. 2000. Metadados para descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. Ciência da Informação, v.29, n.1, p.93-102, jan./abr. 2000.

SOUZA, T. B. de et al. 1997. Metadados: catalogando dados na Internet. Tran-sinformação, v.9, n.2, p.93-105, maio./ago. 1997.

SPINK, A. 1995. Term relevance feedback and mediated database searching: implications for information retrieval practice and systems design. Information Processing & Management, v.31, n.2, p.161-171, 1995.

TAKAHASHI, T. (Org.). 2000. Sociedade da informação no Brasil: livro verde. Brasília: MCT, 2000. 195p.

VELLUCCI, S. L. 1998. Metadata. Annual Review of Information Science and Technology (ARIST), Medford, v.33, p.187-222, 1998.

_____. 1997. Options for organizing electronic resources: the coexistence of metadata. Asis Bulletin, Oct., 1997. Disponível em: <<http://www.asis.org/Bulletin/Oct-97/vellucci.htm>>. Acesso em: 12 jun. 2000.

WEIBEL, S. 1997. The Dublin Core: a simple content description model for electronic resources. Asis Bulletin, Oct., 1997. Disponível em: <<http://www.asis.org/Bulletin/Oct-97/weibel.htm>>. Acesso em: 12 set. 2000.

_____. 1995. Metadata: The foundations of resource description. D-Lib Magazine, Jul. 1995. Disponível em: <<http://www.dlib.org/dlib/July95/07weibel.html>>. Acesso em: 12 set. 2000.

ENSAIOS APB

Coordenação editorial: Oswaldo Francisco de Almeida Júnior

- 86 - VALENTIM, Marta Lígia Pomim. Metadados (Revisão de Literatura). Jan. 2001.
- 85 - CÔRTE, Adelaide Ramos e. A Biblioteconomia frente às inovações tecnológicas. Dez. 2000.
- 84 - CUNHA, Miriam Vieira da. O Profissional da Informação: Formação e Mercado de Trabalho - 3 (Revisão de Literatura - Abordagens Institucional e Educativa e Funções). Nov. 2000.
- 83 - CUNHA, Miriam Vieira da. O Profissional da Informação: Formação e Mercado de Trabalho - 2 (Revisão de Literatura - Abordagem Funcional). Out. 2000.
- 82 - CUNHA, Miriam Vieira da. O Profissional da Informação: Formação e Mercado de Trabalho - 1 (Revisão de Literatura). Set. 2000.
- 81 - SILINGOVSKI, Regina Rita Liberati. O software "Database Marketing" como instrumento na tomada de decisões na administração da informação. Ago. 2000.
- 80 - MARCHIORI, Patrícia Zeni. Navegar é preciso: como entender a estrutura de busca na Web. Jul. 2000
- 79 - MOLOGNI, Micherle. Programa INFOINDEX: a agilidade no trabalho de classificação e indexação. Jun. 2000.
- 78 - TOMAËL, Maria Inês et al. Fontes de informação na Internet: acesso e avaliação das disponíveis nos sites das Universidades. Maio 2000.
- 77 - TOMAËL, Maria Inês et al. Critérios para avaliar fontes de informação na Internet. Abr. 2000.
- 76 - DUTRA, Miriam Regiane. A indústria da informação no Brasil: reflexões. Mar. 2000.
- 75 - VERGUEIRO, Waldomiro. Qualidade em serviços de informação: o foco no cliente. vol. 2. Fev. 2000.
- 74 - VERGUEIRO, Waldomiro. Qualidade em serviços de informação: o foco no cliente. vol. 1. Jan. 2000.
- 73 - SOUZA, Samuel R. M. de. Como fazemos as coisas por aqui? Bibliotecários e Cultura Organizacional. Dez. 99.
- 72 - PEREIRA, Enidélci A. Zaquia et al. Agentes de Tecnologia: uma experiência de estágio na área de informação e gerência do Curso de Biblioteconomia da UEL. Nov. 99
- 71 - VALENTIM, Marta Lígia Pomim. A atividade de investigação em Ciência da Informação. Out. 99
- 70 - MARQUES, Eliana Maria. Biblioteca Pública no Brasil: sonho ou realidade? Set. 99
- 69 - FIERLI, Aglaé de Lima, CATARINO, Maria Elisabete. Classificação Decimal de Dewey em CD-ROM. Ago. 99.
- 68 - FREIRE, Bernardina M. Juvenal, PEREIRA, Raquel G., LIMA, Geysa F. C. de. Biblioteca volante em canteiro de obras: relato de uma experiência. Jul. 99.
- 67 - ALMEIDA JÚNIOR, Oswaldo Francisco de. Agorá informacional. Jun. 99.
- 66 - OLIVEIRA, Ana Lúcia Antunes de. A biblioteca escolar no processo de ensino-aprendizagem: uma experiência. Maio. 99.
- 65 - ALMEIDA, Elisângela Lino de. Conservação e acondicionamento de discos de vinil e fitas cassetes. Abr. 99.
- 64 - FIGUEIREDO, Nice. Automação das bibliotecas universitárias: a visão dos usuários. Mar. 99.
- 63 - BLATTMANN, Ursula, DUTRA, Sigrid Karin Weiss. Atividades em bibliotecas colaborando com a educação a distância. Fev. 99.
- 62 - LIMA, Vânia Mara Alves. Comunicação e representação documentária. Jan. 99.
- 61 - GOMEZ, Margarita Victoria. Educação e informática: caminho entrelaçado com a biblioteconomia. Dez. 98.
- 60 - BROWN, Doris R. O consórcio nas bibliotecas acadêmicas dos EUA. Nov. 98.
- 59 - MODESTO, Fernando. O bibliotecário e o mercado de trabalho: estratégias para o emprego. Out. 98.
- 58 - FIGUEIREDO, Nice. A automação das bibliotecas universitárias: resultado de pesquisa. Set. 98.
- 57 - SOUZA, Marta Alves de. Fontes de informação em Ciências Exatas: uma síntese. Ago. 98.
- 56 - SIQUEIRA, Maria das Graças. Ler é conquistar autonomia. Jul. 98
- 55 - COSTA, Márcia Betânia da. Implantação do movimento 5S's em unidades de informação. Jun. 98.
- 54 - GAUZ, Valeria. O bibliófilo José Mindlin: impressões de um encontro. Maio. 98.
- 53 - RUSSO, Mariza. Bibliotecas universitárias brasileiras: diretrizes para o próximo milênio. Abr. 98.
- 52 - FERREIRA, Margarida M. Estudo do formato para registro bibliográfico Marc: volume 2. Mar. 98.
- 51 - FERREIRA, Margarida M. Estudo do formato para registro bibliográfico Marc: volume 1. Fev. 98.
- 50 - FIGUEIREDO, Nice. Repensando a biblioteca universitária brasileira: como prosseguir - notas para um projeto de pesquisa. Jan. 98.
- 49 - FIGUEIREDO, Nice. Repensando a biblioteca pública brasileira: considerações em torno de resultados de pesquisa. Dez. 97.
- 48 - TOMAËL, Maria Inês. Informação e globalização: reflexos de uma nova era. Nov. 97.
- 47 - RECINE, Analúcia Viviani dos Santos. Análise de partituras. Out. 97.
- 46 - MODESTO, Fernando. O bibliotecário e o mercado de trabalho: alguns comentários. Set. 97.
- 45 - TÁLAMO, Maria de Fátima G. M. Linguagem documentária. Ago. 97.
- 44 - LIMA, Justino Alves. As entidades da biblioteconomia: uma tentativa de globalização e uma iniciativa de intervenção política. Jul. 97.
- 43 - BARRETO, Angela Maria. Conversas com quem gosta de informar. Jun. 97.

- 42 – FUJITA, Mariângela Spotti Lopes. *Elaboração de tesouros monolíngues com o programa TECER: considerações sobre o uso.* Maio 97.
- 41 – SANTOS, Jussara Pereira. *O ensino de biblioteconomia no Mercosul: propostas de integração e harmonização curricular.* Abr. 97.
- 40 – SMIT, Johanna W., MACAMBYRA, Marina M. *Tratamento de multimídia.* Mar. 97.
- 39 – LIMA, Justino Alves. *Mobilização para uma política de conservação e manutenção de acervos contra o agente biológico humano.* Fev. 97.
- 38 – SOUZA, Francisco das Chagas de. *O bibliotecário brasileiro e seu humanismo.* Jan. 97.
- 37 – ALMEIDA JÚNIOR, Oswaldo Francisco de. *Roubo, depredação de materiais e campanhas educativas em bibliotecas: proposta de um modelo de avaliação.* Dez. 96.
- 36 – FERREIRA, Sueli Mara S. P., KROEFF, Márcia S. *Referências bibliográficas de documentos eletrônicos: vol. 2.* Nov. 96.
- 35 – FERREIRA, Sueli Mara S. P., KROEFF, Márcia S. *Referências bibliográficas de documentos eletrônicos: vol. 1.* Out. 96.
- 34 – MARCHIORI, Patrícia Zeni. *Eram os deuses astronautas? ou São os bibliotecários, profissionais da informação?* Set. 96.
- 33 – MARTUCCI, Elisabeth Márcia. *Abordagem qualitativa de pesquisa em biblioteconomia: uma introdução.* Ago. 96.
- 32 – GUIMARÃES, José Augusto Chaves. *A Legislação profissional do bibliotecário.* Jul. 96.
- 31 – ARAÚJO, Eliany Alvarenga de. *Sociedade de informação: espaço da palavra onde o silêncio mora?* Jun. 96.
- 30 – BARTALO, Linete et al. *A importância da leitura na formação do professor.* Maio. 96.
- 29 – MODESTO, Fernando. *Combate ao vírus de computador na biblioteca.* Abr. 96.
- 28 – SOUZA, Marta Alves de. *Internet: a rede global.* Mar. 96.
- 27 – LANE, Sandra S., VAL, Marta R. S. Ribeiro do. *Preservação de acervos de bibliotecas: Parte II. Um modelo de programa local.* Fev. 96.
- 26 – LANE, Sandra S., VAL, Marta R. S. Ribeiro do. *Preservação de acervos de bibliotecas: Parte I. Degradação dos materiais.* Jan. 96.
- 25 – VERGUEIRO, Waldomiro C. S. *Gestão da Qualidade e Bibliotecas Públicas: o difícil caminho para as instituições brasileiras.* Dez. 95.
- 24 – SILVA, A. M. S., ALMEIDA, G. M. A. B., BELLUZZO, R. C. B. *O Plano de Gestão da Qualidade e sua implantação na rede de bibliotecas da UNESP: relato de uma experiência.* Nov. 95.
- 23 – SMIT, Johanna. *Algumas questões sobre os documentos audiovisuais em bibliotecas.* Out. 95.
- 22 – FARIA, Ivete Pieruccini. *Livro e leitura no Brasil: alguns aspectos acerca da entrada do impresso no país.* Set. 95.
- 21 – FUJINO, Asa. *A gestão da informação no processo de cooperação universidade-empresa: uma visão crítica.* Ago. 95.
- 20 – CÔRTE, Adelaide Ramos e. *Memória técnica.* Jul. 95.
- 19 – MODESTO, Fernando. *Apontamentos sobre a ergonomia na implantação e uso do computador na biblioteca.* Jun. 95.
- 18 – LIMA, Justino Alves. *Bibliotecas e bibliotecários: o perfil de um caso.* Maio 95.
- 17 – CARDIN, Tânia Maria Sanvezzo. *Lixo reciclável x incentivo à leitura: uma relação que deu certo no município de Ibiporã - PR.* Abr. 95.
- 16 – VALLS, Valéria. *O espaço do bibliotecário no gerenciamento de documentos do Sistema da Qualidade.* Mar. 95.
- 15 – ALMEIDA JÚNIOR, Oswaldo Francisco de. *Biblioteca pública: ambigüidade, conformismo e ação guerrilheira do bibliotecário.* Fev. 95.
- 14 – VALENTIM, Marta Lígia Pomim. *Leitura Técnica e seu Papel na Pesquisa & Desenvolvimento.* Jan. 95.
- 13 – ALMEIDA JÚNIOR, Oswaldo Francisco de. *O Espaço da Biblioteca: uma reflexão.* 1994.
- 12 – RIVA, Eliane Barbosa et al. *Terceira Idade: programa integrado.* 1994.
- 11 – TOMAZELLI, Angela M. et al. *Criança de Periferia não Lê: desmistificação.* 1994.
- 10 – SILVA, Helen de Castro et al. *Um espaço para a Fantasia.* 1994.
- 09 – LARROUDE, Rita Luisa et al. *Terceira Idade: relato de uma experiência, 1991-1992.* 1994.
- 08 – FERREIRA, Marta Nosé et al. *Projeto "Soma".* 1994.
- 07 – DIAS, Maria Cristina Santarém et al. *Alternativas para Contornar a Crise da Leitura: uma experiência do ônibus-biblioteca na cidade de São Paulo.* 1994.
- 06 – BARROS, Maria Helena T. C. de. *A Atuação da Biblioteca Escolar: relato de uma crise.* 1994.
- 05 – OLIVEIRA, Silas Marques de. *A Crise dos recursos Humanos em Bibliotecas.* 1994.
- 04 – MURGIA, Eduardo. *A Crise da Informação.* 1994.
- 03 – TAVARES, Maria Christina de Moraes. *Atuação da Biblioteca Infante-Juvenil.* 1994.
- 02 – MOSTAFA, Solange Puntel. *Balcão de Informações: o mercado emergente.* 1994.
- 01 – MELO, José Marques de. *Comunicação de Massa x Leitura.* 1994.