

**José Carlos Abbud Grácio**

**Metadados para a descrição de  
recursos da Internet: o padrão  
Dublin Core, aplicações e a  
questão da interoperabilidade**

Marília

2002

**José Carlos Abbud Grácio**

**Metadados para a descrição de recursos da  
Internet: o padrão Dublin Core, aplicações e  
a questão da interoperabilidade**

Dissertação apresentada no  
Programa de Pós-graduação em  
Ciência da Informação da  
Universidade Estadual Paulista  
"Julio de Mesquita Filho" -UNESP,  
Câmpus de Marília, para obtenção  
do título de Mestre.

Orientadora: Doutora Plácida Leopoldina Ventura  
Amorim da Costa Santos

Marília

À minha esposa Maria Cláudia, companheira em todos os momentos, mulher no sentido mais completo da palavra, pelo amor, pela paciência e pelo incentivo.

Aos nossos filhos Caio e Paula, de cujo convívio tive que me ausentar em alguns momentos para cumprir esta meta, razão de ser de nossas vidas, dos quais me orgulho de ser pai.

Aos meus pais Élio (em memória) e Olinda que, com seus esforços, incentivos e confiança, me deram educação e estudo para que eu pudesse ter uma profissão e uma família.

## **Agradecimentos**

Expresso meus sinceros agradecimentos a todas as pessoas, que, direta ou indiretamente, colaboraram para a realização deste trabalho, destacando de forma particular minha gratidão:

- a Deus, por tudo e todos que colocou em meu caminho para a realização deste trabalho;
- à professora Doutora Plácida Leopoldina Ventura Amorim da Costa Santos, não só uma orientadora, mas uma pessoa que me conduziu e orientou durante todo o tempo, contribuindo para o meu crescimento profissional e pessoal, e a quem aprendi a respeitar mais ainda depois dessa convivência, como pessoa e profissional séria e dedicada que é;
- aos docentes do curso de pós-graduação em Ciência da Informação da UNESP de Marília, pelos ensinamentos transmitidos, fundamentais para a conclusão da dissertação;
- aos professores Doutor José Augusto Chaves Guimarães e Doutor Marcos Luiz Mucheroni, pelas contribuições dadas na qualificação, de grande importância para a conclusão do trabalho;
- a todos os funcionários da Biblioteca da UNESP de Marília, profissionais competentes e dedicados, pela presteza e pela atenção demonstradas;
- a todos os funcionários da pós-graduação em Ciência da Informação da UNESP de Marília, pelo atendimento e pela gentileza com que sempre me receberam;
- à Profa. Yolanda Kinuyo Matsuda, pela revisão do texto.

GRACIO, José Carlos Abbud. *Metadados para a descrição de recursos da Internet: o padrão Dublin Core, aplicações e a questão da interoperabilidade*. 2002. 127f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília. 2002.

## RESUMO

Os avanços na Ciência e Tecnologia, particularmente na informática e nas telecomunicações, fizeram com que a Internet se tornasse uma das mais importantes fontes de informação. Ao mesmo tempo, a carência de dados adequados sobre os recursos disponíveis na Web leva à necessidade de criação de elementos descritivos que possam contribuir para a representação de tais recursos. Nesse sentido, a utilização de metadados parece ser a melhor solução para o problema. Metadados podem ser definidos como um conjunto de elementos que descrevem as informações contidas em um recurso, permitindo, assim, sua busca e recuperação. Desse modo, tem-se os metadados em, seu conceito, funcionamento, padrões existentes e formas de utilização, bem como as características e a estrutura do padrão Dublin Core para descrição de recursos na Web e, ainda, um estudo de diversos projetos que utilizam metadados. A análise dos projetos indica que o padrão Dublin Core, composto de um conjunto de 15 elementos, vem sendo amplamente utilizado por sua simplicidade de descrição, extensibilidade e interoperabilidade. Essa última característica é um requisito importante, considerando-se os diversos padrões de metadados existentes, pois a correspondência entre eles é fundamental para a troca de informações. Os resultados obtidos indicam que estudos de metadados conduzem à interoperabilidade entre os padrões existentes e sua representação por linguagens de marca, como a linguagem XML, juntamente com a arquitetura RDF.

Palavras-chave: Metadados. Dublin Core. Interoperabilidade. Descrição de recursos na Internet. Busca e recuperação de informação na Web.

GRACIO, José Carlos Abbud. *Metadata for the description of Internet resources: Dublin Core standard, applications and interoperability*. 2002. 127 p. Dissertation (Master Degree in Information Science) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília. 2002.

### **ABSTRACT**

Advances in Science and Technology, particularly in computing science and telecommunication, have made the Internet one of the most important sources of information. At the same time, the lack of adequate data about the resources available on the Web calls for the need to create descriptive elements which may contribute to represent such resources. In this sense, the development of metadata seems to be the best solution to the problem. Metadata are a set of elements designed to describe the information of a resource, thus allowing its search and retrieval. This work presents a discussion of metadata, how they work on the Web, the existing standards and their use, the concepts found in various domains, the characteristics and structure of the Dublin Core standard used to describe the Web resources, and also a study of several projects adopting metadata. The analysis of the projects shows that the Dublin Core standard, which is a 15-element set of descriptors, has been widely used for its simplicity, extensibility and interoperability. The latter characteristic is an important requirement when considering the various metadata standards currently in use, since their correspondence is fundamental to information exchange. The results obtained indicate that metadata studies lead to interoperability among the existing standards and their representation by markup languages like XML, as well as RDF architecture.

**KEY WORDS:** Metadata. Dublin Core. Interoperability. Description of Internet resources. Web information search and retrieval.

## SUMÁRIO

1 – Introdução	9
2 – Metadados	19
2.1 – Definições	21
2.2 – Estrutura	25
2.3 – Iniciativas	30
2.3 – Interoperabilidade	31
2.5 - Padrões existentes	35
3 - O padrão de metadados Dublin Core	39
3.1 - Histórico	40
3.2 - O “Dublin Core Metadata Initiative” (DCMI)	42
3.3 - O padrão de metadados Dublin Core (DC)	50
3.3.1 - Atributos	52
3.3.2 - Elementos do padrão DC	54
3.3.3 – Qualificadores para os elementos DC	67
3.4 – Representação do DC em HTML	80
4 - Projetos que utilizam Metadados	86
4.1 - Metodologia para seleção dos projetos	87
4.2 - Projetos selecionados	89
4.3 - Variáveis analisadas	99
4.3.1 - Quadro comparativo das variáveis selecionadas	100
4.4 - Análise dos projetos selecionados	102
5 – Considerações finais	111
Referências	119

**Lista de Quadros**

Quadro 1 - Algumas relações entre elementos e padrões de metadados (CROMWELL-KESSLER, 1998)	33
Quadro 2 – Qualificadores do padrão DC	69
Quadro 3 – Variáveis de análise dos projetos selecionados	101
Quadro 4 – Variável local dos projetos selecionados	103
Quadro 5 – Variável país dos projetos selecionados	104
Quadro 6 – Variável ano dos projetos selecionados	105
Quadro 7 – Variável área dos projetos selecionados	105
Quadro 8 – Variável tipo dos projetos selecionados	106
Quadro 9 – Variável padrão dos projetos selecionados	107

# **Capítulo 1**

## **Introdução**

Como profissional da computação, formado desde 1987 no curso de Bacharelado em Ciência da Computação da UNICAMP, e desde então trabalhando na área de informática, tenho me deparado constantemente com questões referentes ao tratamento da informação.

Essas questões e suas implicações na geração de conhecimento, juntamente com a experiência e o conhecimento adquiridos da área de informática, despertaram-me o interesse pela área da Ciência da Informação (CI).

Antes de entender os conceitos de informação e conhecimento, e suas implicações, é necessário compreender onde se situa a CI.

A preocupação com o tratamento da informação teve uma ênfase maior com o desenvolvimento científico e tecnológico ocorrido no período seguinte à 1ª Guerra Mundial, por volta de 1930, permeando o crescimento do capitalismo industrial da década de 30 e gerando um crescimento na utilização de informações de ciência e tecnologia (explosão da informação), base para o crescimento econômico da época.

Nesse contexto histórico, surgem no período de 1960 a 1970 os primeiros conceitos e definições de uma nova área, a Ciência da Informação, pautada na interdisciplinaridade, nos estudos de como tratar a informação e de como seria a atuação dos profissionais da área.

No Brasil, a CI foi introduzida em 1970 com o início do curso de mestrado, instituído pelo antigo Instituto Brasileiro de Bibliografia e Documentação (IBBD), atual Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), com mandato acadêmico na Universidade Federal do Rio de Janeiro (UFRJ).

Outros marcos importantes no Brasil foram a criação da Revista Ciência da Informação em 1972, as duas Reuniões Nacionais de Ciência da Informação, em 1975 e em 1979, e o 1º Encontro de Pesquisa

da Pós-graduação em Ciência da Informação em 1994 (PINHEIRO; LOUREIRO, 1995).

Com o nascimento de uma nova área, surgem os conceitos e definições para CI, voltados principalmente para o tratamento da informação.

Rees e Saracevic (1965), citado em Pinheiro e Loureiro (1995, p.43), definem CI como “o estudo dos fenômenos da comunicação e das propriedades dos sistemas de comunicação”, abordando com maior ênfase o processo de comunicação.

Borko (1968), citado em Pinheiro e Loureiro (1995, p.43), destaca que a CI “é uma disciplina que investiga as propriedades e comportamentos da informação, sua utilização, sua transmissão e seu processamento para armazenagem e recuperação ótimas”.

Foskett (1980) apresenta a CI como uma área interdisciplinar (Biblioteconomia, Computação, Psicologia, Comunicação e Lingüística), que tem a ver com os problemas da comunicação e da transferência do conhecimento organizado.

Le Coadic (1996) define CI como o estudo das propriedades gerais da informação (natureza, gênese e efeito), dos processos e sistemas de construção, comunicação e uso dessa informação.

Fica claro, então, que a informação é o objeto principal de estudo da CI, informação essa que tem como propósito gerar conhecimento nos indivíduos e nos espaços de sua convivência, proporcionando um instrumento modificador da consciência humana e da sociedade.

A informação possui, em seu conteúdo, competência para produzir conhecimento, mas é necessário que exista o acesso do indivíduo a essa informação e que ele possa entendê-la, pois não basta a informação estar disponível, se o indivíduo não possuir mecanismos que possibilitem o acesso a ela, e também não é possível adquirir

conhecimento se o indivíduo não possui conhecimento prévio para poder entender a informação acessada.

Vivemos em um mundo em que tudo pode ser traduzido em informação, ou seja, tudo é portador de informação, desde uma planta que informa sobre o ecossistema do local até um sinal de fumaça dos filmes de faroeste (SMIT, 2000).

Bouche (1998) define informação como uma forma que circula sobre o canal, do emissor ao receptor e não possui valor em si. Entretanto, as melhores definições são aquelas que relacionam a informação com a intenção de gerar conhecimento, nas quais ela é qualificada como um instrumento modificador da consciência do homem e de seu grupo (BARRETO, 1994).

Barreto (1999a, p.168) conceitua informação como “conjuntos significantes com a competência e a intenção de gerar conhecimento no indivíduo, em seu grupo e na sociedade”, conhecimento esse que tem como objetivo promover o desenvolvimento do indivíduo e da sociedade em que ele vive.

Para Le Coadic (1996), informação é o sangue da CI, ou seja, só interessa se circula e, sobretudo, se circula livremente, tendo como objetivo final o seu uso e os resultados que são obtidos por ela.

Santos e Sant´Ana (2002) conceituam informação como “um conjunto finito de dados dotado de semântica e que tem a sua significação ligada ao contexto do agente que a interpreta ou recolhe e de fatores como tempo, forma de transmissão e suporte utilizado”.

Podemos notar que a informação, para gerar conhecimento em um indivíduo, necessita circular, ser transmitida, chegar ao indivíduo em um intervalo de tempo aceitável e em um formato compreensível.

Para isso, a informação necessita ser registrada ou documentada. Desde a Antiguidade já se sentia a necessidade de armazenar as informações, para possibilitar o acesso adequado a elas. Daí

surtem bibliotecas, museus e arquivos, que se caracterizam por registrar, armazenar e disponibilizar as informações institucionalizadas para que indivíduos possam ter acesso a elas, bem como os profissionais da informação, responsáveis em desenvolver essas tarefas.

Mason (1990) enuncia que os objetivos que perseguem os profissionais da informação são o de disponibilizar a informação certa, da fonte certa, para o cliente certo, no momento certo, no formato certo e a um custo compatível.

Atualmente, as mudanças tecnológicas no armazenamento e na transmissão da informação, proporcionadas pela informática e pelo avanço das telecomunicações, estão alterando a relação dos profissionais da informação com a forma de tratar essa informação, bem como a relação dos usuários com o acesso à informação armazenada.

Os usuários atuais necessitam de novas informações e de novos elementos em suas pesquisas, isto é, que as bases de dados atuais possuam, além de textos, elementos como sons e imagens. Devido a essas mudanças tecnológicas, eles têm, algumas vezes, o acesso à informação em tempo real e interativo, mudando assim a relação de tempo e espaço. Mas, para que o acesso à informação estocada possa atender ao usuário na sua pesquisa, a informação deve ser tratada e representada, possibilitando a sua busca e recuperação.

Nesse sentido, o estudo da organização do conhecimento, que procura oferecer subsídios teóricos à questão do tratamento da informação, tem um papel importante na questão do acesso à informação (GUIMARÃES, 2001).

Barite (2001) define organização do conhecimento como o estudo que tem por objetivo específico otimizar a circulação do conhecimento nas sociedades, e uma de suas premissas básicas é a de que o conhecimento se obtém a partir da informação. Ao socializar-se, esse conhecimento obtido transforma-se novamente em informação.

A preocupação pela organização do conhecimento vem desde a Antiguidade, passando pela implantação de bibliotecas públicas no século XIX, pelo surgimento de instrumentos mais elaborados para indexação e hoje como uma preocupação efetiva.

A organização do conhecimento é relacionada inicialmente com o documento, posteriormente com a informação (conteúdo vinculado à informação) e atualmente com o paradigma do conhecimento (o que se almeja com a apropriação da informação) (GUIMARÃES, 2001).

Para Dahlberg (1995), o conhecimento é algo que não pode ser transferido, somente podendo ser adquirido por alguém através do seu próprio repensar, com as informações de que ele dispõe como um processo individual.

Barite (2001) define conhecimento como o processo intelectual ou emocional que realiza um indivíduo para entender um fenômeno do mundo exterior e compreender seu resultado, reafirmando ou removendo sua concepção de mundo.

Para Santos e Sant´Ana (2002), conhecimento pode ser conceituado como:

um conjunto de informações contextualizadas e dotadas de semântica inerente ao agente que o detém, seja a mente humana ou não, e seu conteúdo semântico se dará em função do conjunto de informações que o compõem e de suas ligações com outras unidades de conhecimento, e do processo de contextualização.

Podemos notar que o conceito de informação e de conhecimento depende do contexto em que estão sendo utilizados, apresentando assim variações conceituais. Barite (2001), sem citar a fonte dessa informação, informa que já foram encontradas mais de 500 definições para informação e quase 30 para conhecimento.

A explosão da quantidade de informação disponível, causada com o advento da escrita e da imprensa, e mais recentemente com os avanços da eletrônica, da informática e das telecomunicações, gerou um

grande fluxo de informações nos meios de comunicação, como a Internet, exigindo cada vez mais a utilização de computadores interligados em rede.

O computador, diferente da sua origem, quando sua principal tarefa era exclusivamente realizar cálculos matemáticos complexos, é hoje uma ferramenta importante no armazenamento, na organização, na recuperação e no intercâmbio de informações armazenadas em meio eletrônico (BAX, 2001).

A Internet é atualmente a expressão maior da utilização dos computadores e dos meios eletrônicos para o armazenamento, a busca e a recuperação de informações armazenadas em meio eletrônico. Para recuperar a informação armazenada na Internet e transformá-la em conhecimento, são utilizadas atualmente ferramentas de busca, que consistem em programas de computadores com bancos de dados que armazenam descritores de recursos disponíveis na Internet, como Yahoo, Google entre outros. Tais ferramentas não possuem mecanismos de busca iguais e, de acordo com a característica de cada uma, o número e a qualidade das informações recuperadas podem variar enormemente.

As ferramentas de busca do tipo diretório, que organizam as informações em categorias, realizam a indexação de documentos utilizando especialistas, gerando informações com mais conteúdo, mas tornam esse trabalho muito demorado. As ferramentas do tipo motores de busca, ao contrário, utilizam software (chamado robô de busca) para buscar automaticamente as informações na Web, tornando a indexação mais ágil, mas acabam gerando um número muito grande de informações, deixando algumas páginas da Internet fora do catálogo (CENDÓN, 2001).

Como a Internet e os recursos armazenados em meio eletrônico crescem de maneira exponencial, um dos principais problemas com o tratamento desse tipo de informação é a carência de dados descritivos do conteúdo dessa informação, pois é necessário torná-la

recuperável e disponível para o usuário, permitindo que este possa utilizá-la na geração de conhecimento e no seu desenvolvimento.

Essa carência leva à necessidade do desenvolvimento de elementos de descrição que possam representar o conteúdo dos recursos armazenados, permitindo sua busca e recuperação mais efetivas pelas ferramentas atuais.

Nesse sentido, a maioria dos profissionais que estudam o problema da descrição, busca e recuperação de informação apontam que a melhor solução para esse problema pode ser a criação de metadados para descrever os recursos contidos na Web, pois permite o tratamento da informação armazenada em computadores, transmitida pela rede e principalmente aquela disponível na Internet.

Como verificamos nos projetos analisados e na bibliografia levantada, a utilização de metadados é recente, principalmente no Brasil. Portanto, é importante saber o que são metadados, como podem ajudar a descrever um recurso, como são utilizados, que padrões de metadados existem, quais projetos estão utilizando metadados para a descrição de recursos e como o fazem.

O objetivo desta pesquisa é apresentar o padrão de metadados Dublin Core, o conceito sobre metadados, analisar a aplicação dos padrões existentes no contexto da busca e recuperação da informação e abordar a questão da interoperabilidade proporcionada pelo uso de metadados.

Apresentaremos os conceitos aplicados a metadados, mostrando sua estrutura e as discussões que se fazem sobre interoperabilidade entre os diferentes padrões de metadados.

Mostraremos como metadados são utilizados para descrever documentos eletrônicos, que estão em grande quantidade disponíveis na Internet, dando ênfase ao padrão DC, que foi especialmente criado para

descrever documentos eletrônicos disponíveis na Web, hoje uma grande fonte de informações.

Verificaremos também quais padrões de metadados estão sendo utilizados em alguns projetos selecionados de pesquisas realizadas em revistas científicas, e como estão sendo aplicados.

Ao final da pesquisa, verificaremos se o padrão de metadados DC pode ser apontado como parâmetro para descrição de recursos eletrônicos da Web.

Além do capítulo introdutório, que aponta os objetivos da pesquisa, a dissertação conta com a seguinte organização:

Capítulo 2 – Metadados: apresenta o conceito de metadados na visão de alguns autores e propõe uma definição; mostra como é estruturado um padrão de metadados para a descrição do conteúdo de uma informação, discute a questão da interoperabilidade entre padrões de metadados e descreve algumas iniciativas importantes nessa área.

Capítulo 3 – O padrão de metadados Dublin Core: apresenta um histórico da origem do padrão, suas características e sua estrutura para descrição de recursos na Web, ou seja, seus elementos e qualificadores.

Capítulo 4 – Projetos que utilizam metadados: apresenta uma análise de 13 projetos que utilizam metadados para a pesquisa e recuperação das informações armazenadas em meio eletrônico, com o objetivo de entender como os padrões de metadados estão sendo utilizados na prática.

Capítulo 5 – Considerações finais: apresenta considerações baseadas na pesquisa realizada, discutindo como metadados podem ser utilizados, e apontando trabalhos futuros que podem ser desenvolvidos na área de metadados.

Pretendemos, assim, abordar o tema metadados mostrando seu conceito, percorrendo seu histórico e analisando projetos que se

preocupam com padrões de descrição de documentos eletrônicos, com o objetivo de oferecer subsídios para responder às questões levantadas anteriormente.

## **Capítulo 2**

# **Metadados**

O rápido desenvolvimento da World Wide Web (WWW), o seu tamanho e o conseqüente aumento do número de recursos disponíveis na Internet levam à necessidade de ferramentas para o gerenciamento, a organização e o armazenamento dos recursos disponíveis em meio eletrônico.

Podemos entender como recurso toda informação armazenada eletronicamente que pode ser acessada e recuperada, independente do formato, como texto, imagem, som, vídeo, uma página da Web etc. Para que tenhamos condições de acesso a esses recursos, necessitamos de sistemas de informação bem planejados, visto que o objetivo principal dessa informação armazenada é o usuário.

As ferramentas de busca automáticas disponíveis na Internet recuperam grandes quantidades de informação, mas com pouca precisão.

Não é objetivo da pesquisa a discussão teórica do conceito de precisão da informação, mas podemos dizer que o conceito de precisão, para a avaliação de fontes eletrônicas na Internet, aproxima-se dos critérios de avaliação de fontes impressas, conhecidos pelos bibliotecários, ou seja, é a relação existente entre o número de documentos relevantes recuperados e o número total de documentos recuperados. Um dos problemas é a carência de parâmetros adequados na Internet, pois na Web fica difícil, ou quase impossível, medir o total de recursos existentes, devido à sua constante mudança e crescimento.

Para que uma pesquisa possa ter qualidade e aproximar-se do resultado desejado, os recursos disponíveis na Internet deveriam possuir dados suficientes para uma descrição mais adequada do seu conteúdo. Hoje em dia, vemos uma carência desses dados.

Podemos, por exemplo, utilizar os mecanismos de busca atuais para pesquisar um documento que contenha referências a um determinado autor, porém não é possível pesquisar apenas aqueles

documentos que foram escritos por esse autor. Isso mostra que o resultado da pesquisa traz um grande número de documentos não relevantes, levando o usuário a despender tempo para a procura dos documentos de interesse.

A utilização de metadados para descrever esses documentos permite realizar, entre outras coisas, uma pesquisa de documentos escritos pelo autor desejado, do tipo *autor = "nome do autor desejado"*, na medida em que fornece esse tipo de informação aos mecanismos de busca e conseqüentemente permite que o usuário possa realizar a pesquisa.

## **2.1 - Definições**

Comumente chamado de dados sobre dados, o termo metadados pode ser mais bem descrito como um conjunto de dados chamados de elementos, cujo número é variável de acordo com o padrão, e que descreve o conteúdo de um recurso, possibilitando a um usuário ou a um mecanismo de busca acessar e recuperar esse recurso. Esses elementos descrevem informações como nome, descrição, localização, formato, entre outras, que possibilitam um número maior de campos para pesquisas.

Podemos estabelecer a relação de um recurso com a sua descrição por metadados, como fazemos a relação dos documentos de uma Biblioteca com os seus registros no catálogo, com o objetivo organizar as informações contidas nos documentos e permitir sua recuperação.

O termo metadados possui um significado ou um conceito de acordo com o profissional e a área em que é utilizado, mas tem sempre como objetivo principal a descrição da informação para sua busca e recuperação.

Diversos conceitos são dados para explicar o significado de metadados:

- “dados sobre dados”: é a definição clássica e a mais comum de se encontrar. É citada em Madsen, Foog e Ruggles (1994), Klensin (1995) e Garcia (1999).
- “dados que descrevem elementos do recurso e também descrevem conjunto de dados e sistemas de gerenciamento de Banco de Dados”: citado em Madsen, Foog e Ruggles (1994).
- “dados sobre indexação e catalogação criados para ordenar e tornar mais acessível a informação”: definição dada por administradores de museus e bibliotecas, segundo Gilliland-Swetland (1998).
- “dados que servem para descrever grupos de dados que poderíamos chamar de objetos informatizados” ou “descrições estruturadas de um objeto informatizado”: citado por Gill (1998).
- “conjunto de dados usados para descrever e representar um objeto de informação”: citada por Ortiz-Repiso Jimenez (1999). Deve-se entender como objeto de informação os recursos como arquivos de imagens, áudio, vídeo e documentos eletrônicos.
- “documentação que descreve o dado armazenado”: citado por Barreto (1999b).
- “catalogação do dado ou descrição do recurso”: citado por Souza; Vendrusculo e Melo (2000).
- “dados que descrevem outros dados”: citado por Takahashi (2000) no livro Sociedade da Informação no Brasil: livro verde.

- “dados acerca de dados que estão no espaço digital”: citada por Rosetto (2002).

Creemos que metadados poderiam ser preliminarmente definidos como:

***Conjunto de elementos que descrevem as informações contidas em um recurso, com o objetivo de possibilitar sua busca e recuperação.***

O conjunto de elementos ou o conjunto semântico de campos representa o conteúdo do recurso descrito, ou seja, as informações que possibilitam identificar o que o recurso representa e o que ele contém. Esse conjunto pode ter um número de elementos variável de acordo com o padrão de metadados utilizado.

Os elementos devem conter dois tipos de informações:

- descritivas, ou seja, aquelas referentes às características explícitas do recurso, tais como título, data, formato, tipo etc;
- e temáticas, de conteúdo intelectual, ou seja, aquelas que expressam o conteúdo do recurso, tais como palavras-chave e referências cruzadas.

Preferencialmente, esses elementos devem seguir esquemas de codificação que sejam usuais e comuns, como o uso de vocabulário controlado, esquemas de classificação e formatos de descrição formais, permitindo que esses metadados possam ser trocados entre diferentes instituições.

Recurso, como já mencionamos anteriormente, é toda informação que pode ser armazenada em meio eletrônico, podendo estar apresentada como texto, imagem, som, vídeo, página da Web etc.

O recurso, tendo sido descrito por esse conjunto de elementos, possibilita a um usuário ou a um mecanismo de busca realizar

uma busca e obter uma resposta que o leve a recuperar esse recurso para sua utilização.

Assim, podemos concluir que um conjunto de elementos que possua informações para identificação do conteúdo e da descrição de um recurso, possibilitando sua busca e recuperação em meio eletrônico pode ser chamado de metadados.

Metadados cuidadosamente elaborados, seguindo padrões nacionais ou preferencialmente internacionais, permitem o tratamento adequado das informações contidas em um recurso, permitindo maior acessibilidade a ele, mantendo a relação entre um recurso digitalizado e o original, podendo descrever direitos e restrições do recurso e conservando-o disponível, independente das alterações de software e hardware (GILLILAND-SWETLAND, 1998).

A importância de descrever as informações contidas em um recurso cresce na proporção direta ao número de recursos descritos, como na WWW, que possui, atualmente, a maior coleção de recursos digitais do mundo e é a que mais cresce. O número de sites existentes, segundo a Web Growth Summary, cresceu de 130 em 06/93 para 650.000 em 01/97 (um crescimento de 500.000%) (GILL, 1998). No ano de 2001, o mecanismo de busca Google possuía cerca de 1,3 bilhões de sites cadastrados.

A Internet não surgiu com a preocupação de catalogar as informações contidas nela. Com seu crescimento, recentemente essa preocupação tem sido alvo de estudos, mesmo porque os instrumentos de busca atuais da Internet, apesar de robustos, não atendem de maneira satisfatória os usuários que os utilizam, pois não trazem toda a informação disponível na rede.

Gill, já em 1998, apontava que qualquer solução para busca e recuperação de recursos na Web passa pela utilização de um catálogo distribuído e afirmava:

a existência de descritores consistentes, exatos e bem distribuídos dos recursos da Web permitirá maior precisão na busca e uma classificação mais rigorosa dos resultados obtidos segundo sua relevância. (1998, p. 15)

Esses descritores, quanto mais exatos, melhor descreverão os recursos disponíveis na Web, pois darão mais informações do conteúdo do recurso, permitindo que o usuário tenha maior precisão e refinamento na sua pesquisa, principalmente na Internet, onde um refinamento maior pode significar a exclusão de uma grande quantidade de páginas da Web indesejáveis.

A utilização de metadados permite também estabelecer padrões de dados diante da heterogeneidade das informações disponíveis em rede, principalmente as da Internet (GARCIA, 1999). Esses padrões possibilitam uma melhor descrição dos recursos.

Os padrões de metadados, uma vez estabelecidos, permitem a troca de informações entre instituições que utilizam o mesmo padrão ou até mesmo entre aquelas que utilizam padrões diferentes. Isso é importante, pois além de diminuir o trabalho de descrição de recursos, permite que um usuário possa, em uma única pesquisa, buscar informações em diferentes instituições.

Uma outra característica que tem levado à utilização crescente de diversos padrões de metadados na Web é a simplicidade na descrição do recurso, possível na maioria das vezes de ser feita pelo próprio responsável pelo recurso, pois ele necessita saber quais são os elementos que compõem o padrão e definir sua estrutura.

## **2.2 - Estrutura**

Mais do que elementos descritivos de um recurso, metadados podem incorporar elementos que representem o contexto em que ele está inserido, como o tipo do recurso, sua localização, e seu

relacionamento com outros recursos. Possuem uma relação direta com o recurso, assim como um livro tem com seu registro em um catálogo de biblioteca.

Metadados podem ser utilizados para representar vários recursos, dependendo do domínio: em um provedor Web, para identificar e localizar páginas na Internet; na digitalização de imagens, para descrever a informação contida nelas; em dados eletrônicos, para descrever a informação contextual contida no documento eletrônico (GILLILAND-SWETLAND, 1998).

Segundo Gilliland-Swetland (1998), metadados podem ser divididos em 5 tipos, de acordo com os aspectos de sua funcionalidade em um sistema digital:

- 1- Administrativos: usados na gestão e administração de recursos de informação.
- 2- Descritivos: usados para descrever informação sobre recursos.
- 3- De conservação: relacionados com a conservação de recursos de informação.
- 4- Técnicos: relacionados com o funcionamento dos sistemas e o comportamentos dos metadados.
- 5- De uso: relacionados com o nível e o tipo do uso dos recursos de informação.

Gilliland-Swetland (1998) também define atributos-chave para os metadados:

- Fonte: metadados podem ser gerados no momento da criação do recurso ou posteriormente.
- Método de criação: metadados podem ser criados por mecanismos computadorizados, como robôs de busca, ou manualmente por indivíduos.

- Caráter: metadados podem ser criados por especialistas ou pelo responsável pelo recurso.
- Situação: metadados podem ser “estáticos”, que nunca mudam uma vez criados; “dinâmicos”, que podem mudar com o uso e a manipulação; de “larga duração”, para assegurar que o recurso siga sendo acessível e usual, e de “curta duração”.
- Estrutura: metadados podem ser estruturados, como Machine Readable Cataloguing (MARC) e DC, ou não estruturados.
- Semântica: metadados podem utilizar vocabulário controlado ou não para descrição de seus elementos.
- Nível: metadados podem ser relacionados com coleções ou com objetos individuais.

Nos atributos descritos acima, é importante ressaltar que metadados podem ser descritos por não especialistas, como o dono de um recurso ou o desenvolvedor de uma página na Web. Também podem ser descritos utilizando software de busca, normalmente chamados de “havester” ou robôs que, por não serem estáticos, permitem que os elementos descritivos possam ser modificados enquanto o recurso existir e estiver disponível.

Uma característica importante é que metadados podem ser representados no próprio recurso ou separadamente, como em um Banco de Dados separado do recurso.

Os metadados, quando representados dentro do recurso podem estar contidos em páginas Web desenvolvidas em “Standard Generalized Markup Language” (SGML), em “HyperText Markup Language” (HTML) ou em “Extensible Markup Language” (XML).

Essas linguagens de marcação surgiram com a necessidade de maior interação e troca de informações entre usuários e computadores,

gerada com o aumento do número de computadores, com as novas tecnologias de comunicação e com o surgimento da WWW (BAX, 2001).

O SGML surgiu em meados de 1970, sendo uma linguagem que permite definir qualquer conjunto de marcas, ou seja, é auto-descritiva. Um documento SGML carrega consigo sua própria especificação, o que é chamado de "Data Type Document" (DTD). Podemos dizer que SGML é uma linguagem para definir outras linguagens (BAX, 2001).

O HTML origina-se do SGML, mas com seu conjunto de marcas pré-definido, concebido com a função de organizar as informações a serem transferidas pela Web. Por já possuir essas marcas, o HTML se tornou o padrão mais utilizado na Web.

Na Web utiliza-se o protocolo "HyperText Transfer Protocol" (HTTP) para o transporte dos documentos descritos em HTML entre o servidor que contém a informação e o usuário.

Com o avanço da Internet e as necessidades cada vez maiores de interação com os usuários, como no caso do comércio eletrônico, surgiu o XML, resultado dos estudos de um grupo de especialistas do World Wide Web Consortium (W3C), que em 1996 propuseram uma linguagem mais simples que o SGML, mais flexível que o HTML, que permitisse a criação de marcas e fosse voltada para as necessidades atuais da Web.

Um dos objetivos do XML é indicar o que os dados significam, além de descrever como os mesmos devem ser mostrados,. Por essas características, o XML tem sido fonte de estudos e projetos na sua utilização com metadados.

O site do Dublin Core Metadata Initiative (DCMI), no endereço <http://dublincore.org/documents/2001/09/20/dcmes-xml/>, já contém uma DTD para descrição dos elementos do padrão DC em XML, o

que mostra a preocupação da utilização de XML com metadados (BECKETT; MILLER; BRICKLEY, 2001).

Bryan (1997), aponta XML como uma ferramenta ideal para o armazenamento de informações em banco de dados, já que possui a característica de não necessitar de uma aplicação específica de software para interpretar os dados, pois informações armazenadas em XML podem ser transferidas entre diversos ambientes de hardware e software, independente do surgimento de novas tecnologias de programação e de processamento de dados.

Qualquer alteração na estrutura dos dados em XML ou no seu conteúdo é simples de realizar e não necessita de alterações nas aplicações que tratam essas informações.

Essas características das linguagens de marca viabilizam a interoperabilidade entre diversos ambientes, possibilitando a troca de informações e o acesso distribuído.

Diversos mecanismos de busca automáticos, como Alta Vista, Infoseek e outros utilizam ferramentas que, além de obterem o conteúdo do recurso disponível na Internet, também obtêm as informações contidas na marca META da linguagem HTML.

Esse tipo de informação é importante quando tratamos principalmente de sons, imagens e quadros, pois esses documentos possuem pouco ou nenhum texto para ser indexado por esses mecanismos de busca. As informações contidas na marca META da linguagem HTML também são importantes para a descrição temática do recurso, pois permitem a representação do seu conteúdo intelectual que pode, em alguns casos, não estar explícita no conteúdo textual do item.

Nesse contexto dos avanços tecnológicos e das linguagens de marca, diversas iniciativas têm surgido e várias pesquisas estão sendo realizadas para se chegar a um padrão de metadados ideal para Web, mas apesar de todo esse esforço internacional, até o momento não se

conseguiu chegar a um consenso sobre o conteúdo e a estrutura adequados para os metadados que descrevem recursos da Web, talvez por conta da imensidão e da diversidade de informações que contém a Internet.

### **2.3 - Iniciativas**

Temos diferentes padrões de metadados descrevendo variados tipos de recursos, espalhados por diversos países, cada um utilizando um contexto e uma estrutura distinta.

A utilização de vocabulário controlado e de sistemas formais de classificação e catalogação tem sido adotada como uma maneira de diminuir as diferenças entre os padrões existentes e possibilitar uma correspondência entre eles.

Essa tem sido uma preocupação da área de Ciência da Informação, que se intensificou a partir do pós-guerra, com a questão da explosão da informação científica e tecnológica, dando origem, inicialmente, aos tesouros nas décadas de 60 e 70, como ferramenta de suporte à recuperação da informação.

Com o surgimento dos computadores e sua utilização pelas bibliotecas, os tesouros passam também a ser utilizados na recuperação automatizada.

Com o surgimento da Internet e sua explosão na década de 90, a preocupação passa a ser também com a organização lógica das informações disponíveis na grande rede e como transformá-las em conhecimento, ou seja, surge a necessidade de entender como se dá a organização do conhecimento na internet e como podemos descrever os recursos disponíveis na Web para sua utilização.

Iniciativas importantes para ajudar a resolver o problema da descrição de recursos na Web têm surgido, como a “Dublin Core Metadata Initiative” (DCMI), que envolve profissionais de diversas áreas e países trabalhando juntos nas discussões sobre como metadados podem ajudar a resolver esse problema, e o “Resource Description Framework” (RDF), uma aplicação de metadados em XML com o objetivo de facilitar a utilização de pacotes de metadados diversos.

No Brasil, temos o projeto Sociedade da Informação, que tem como objetivo principal o acesso e a inclusão de todos os brasileiros na chamada Era da Informação, e para isso conta com diversos profissionais da área de informação. Nesse sentido, metadados têm papel importante na identificação de conteúdos da rede (TAKAHASHI, 2000).

Com as diversas iniciativas e os vários projetos, surge a questão de qual padrão de metadados utilizar.

Cada domínio define suas necessidades e diante delas escolhe o padrão de metadados que melhor atende a elas. Mas, mesmo dentro de um mesmo domínio, podemos ter a utilização de padrões diferentes.

Para resolver o problema da troca de informação entre instituições que utilizam padrões de metadados diferentes, é necessário estabelecer a correspondência entre esses padrões.

## **2.4 - Interoperabilidade**

Os dados disponíveis na Web são produzidos, na sua maioria, de forma independente, acarretando problemas de heterogeneidade, além de estarem distribuídos geograficamente. Prover o compartilhamento destes dados para permitir o acesso a essas informações é o grande desafio.

Para resolver esse problema da heterogeneidade dos dados na Web, é necessário estabelecer mecanismos de integração que permitam aos usuários acessarem os dados de forma amigável e precisa, estabelecendo uma interoperabilidade entre as bases de dados e os próprios dados.

A heterogeneidade semântica é um obstáculo para prover interoperabilidade entre diversas fontes de dados, pois podemos ter diversas interpretações da mesma informação, causando conflitos de sinonímia (dados com o mesmo conteúdo semântico, mas com nomes diferentes) e de homonímia (dados com o mesmo nome, mas com conceitos diferentes) (MARINO, 2001).

Nesse sentido, metadados são um mecanismo que ajudariam a resolver o problema da interoperabilidade, pois permitem estabelecer um acesso uniforme aos dados disponíveis bem como possibilitam a definição do significado da informação.

Os dados podem estar descritos por um único padrão de metadados, o que representa uma facilidade na integração dos recursos disponíveis. Mas com o surgimento de diversos padrões de metadados e sendo a Internet muito vasta, normalmente esses dados estão descritos por padrões distintos.

Estudos sobre a correspondência entre padrões de metadados têm surgido como uma solução para a integração de dados disponíveis em fontes de informações distribuídas fisicamente e heterogeneamente, possibilitando estabelecer a interoperabilidade entre essas fontes e especialmente entre os padrões que as descrevem.

Para Cromwell-Kessler (1998), a correspondência entre os padrões é um dos segredos para que o usuário possa buscar as informações desejadas na Web, pois é estabelecida uma associação entre esses padrões, possibilitando a utilização em conjunto de diversos

padrões, permitindo que através de uma única interface de busca, o usuário possa pesquisar e acessar as informações desejadas.

Estabelecer interoperabilidade não é uma tarefa fácil, pois mesmo dentro da mesma área temos padrões distintos, estruturados de maneira diferente, o que dificulta estabelecer relação entre os elementos.

Um exemplo é o estudo realizado por Cromwell-Kessler (1998) para estabelecer correspondência entre padrões de metadados para descrição de informações sobre patrimônio cultural, como obras de arte e objetos de museu.

Foram estudados padrões diferentes, entre eles o de Categorias para descrição de obras de arte (CDWA), o Object ID, o CIMI Schema, o "Foundation for Documents of Architecture (FDA)", o VRA Core, o "Record Export for Art and Cultural Heritage (REACH)", o USMARC e o Dublin Core. Como resultado, foi montada uma tabela com a relação existente entre os elementos que fazem parte de cada padrão.

Segue abaixo, a título de ilustração, uma parte da tabela "Correspondências entre padrões para metadados" (CROMWELL-KESSLER, 1998), para mostrarmos a relação entre alguns elementos e alguns padrões:

CDWA	Object ID	FDA	VRA Core	REACH	USMARC	DC
Titles or names	Title	Group/Item Identification-Repository Title Group/Item Identification-Descriptive Title Group/Item Identification-Inscribed Title	W2. Title	Field #4: Object Name/Title	24Xa Title and Title-Related Information	Title

Quadro 1 – Algumas relações entre elementos e padrões de metadados (CROMWELL-KESSLER, 1998)

CDWA	Object ID	FDA	VRA Core	REACH	USMARC	DC
State					562c Copy and Version Identification Note-Version Identification	Description
Classification						Subject
Copyright/Restrictions	Copyright restriction	Internal Documentation Restriction			540a Terms Governing Use and Reproduction	Rights

Quadro 1 – Algumas relações entre elementos e padrões de metadados (CROMWELL-KESSLER, 1998) (cont.)

Nessa tabela podemos verificar que nem todos os elementos possuem relação, ou seja, diversos elementos estão representados em alguns padrões e não estão em outros, além do fato de que alguns elementos, apesar de possuírem relação, são descritos através de esquemas de representação diferentes.

A proposta da Biblioteca Digital Brasileira, descrita por Marcondes e Sayão (2001) trata a questão da interoperabilidade como uma das alternativas fundamentais para o sucesso da proposta, pois possibilita a consulta de uma única vez a todas as fontes de forma integrada e transparente, com rapidez e resultados consolidados.

Entre bibliotecas digitais, o conceito de integração e interoperabilidade consiste na “possibilidade de um usuário realizar buscas a recursos informacionais heterogêneos, armazenados em diferentes servidores na rede, utilizando-se de uma interface única sem tomar conhecimento de onde nem como estes recursos estão armazenados” (MARCONDES; SAYÃO, 2001).

Podemos verificar que o conceito de interoperabilidade, no caso da proposta da Biblioteca Digital Brasileira, é aplicado à troca de informações entre instituições que utilizam o mesmo padrão de metadados, no caso o padrão DC.

Nesse contexto, surgem as arquiteturas de metadados, com o objetivo de possibilitar a interoperabilidade entre padrões distintos, através de estruturas flexíveis que maximizem a troca de informações. Essas arquiteturas estabelecem suporte à codificação e ao transporte de metadados distintos. São exemplos a arquitetura Warwick, proposta em 1996 no 2º Workshop promovido pelo DCMI e o RDF, uma recomendação da W3C de 1999 que vem se destacando como a melhor solução de arquitetura, por ser ao mesmo tempo simples e abrangente (MARINO, 2001).

Outro aspecto importante na questão da interoperabilidade é a forma como os metadados podem ser codificados, e apesar das diversas formas, como o HTML, no contexto da Web a linguagem XML é aquela que permite melhor descrever as informações para a troca de metadados.

Para Kerhervé (1997) citado em Barreto (1999) e conforme visto nas pesquisas de Cromwell-Lessler (1998) e Marcondes e Sayão (2001), devido ao grande número de padrões existentes e em uso, as pesquisas com metadados devem ir em direção à interoperabilidade entre os padrões utilizados, pois estabelecer um padrão único parece algo pouco provável diante da enorme quantidade de informações disponíveis, principalmente na Web.

## **2.5 - Padrões existentes**

Um padrão de metadados pode ser descrito como um conjunto de elementos descritores que segue um determinado modelo de dados com o objetivo de descrever recursos de um domínio específico. Podemos entender como um modelo de dados um conjunto de conceitos e regras (BARRETO, 1999).

Podemos considerar que o primeiro padrão de metadados foi o MARC, criado na década de 60 nos Estados Unidos, com o objetivo de possibilitar a troca de registros bibliográficos e catalográficos entre bibliotecas, com o uso de computadores. Atualmente a versão original do padrão é denominada MARC21.

Os padrões criados para descrever informações bibliográficas e catalográficas, como o MARC, possuem um conjunto de elementos complexos e rígidos, necessitando de especialistas para a descrição desses elementos.

Diferente desses padrões, os padrões criados para a descrição de recursos disponíveis na Internet têm a característica de possuir um conjunto de elementos mais simples, em menor quantidade e flexíveis, o que facilita a descrição desses elementos pelo próprio autor do recurso ou por administradores de sites. Essa simplicidade é o segredo para sua rápida utilização na Web.

Durante a pesquisa, através da análise de diversos projetos que utilizam metadados, foram identificados diferentes padrões de metadados que estão sendo utilizados para descrever um recurso. Todos os padrões têm o objetivo de facilitar a busca e recuperação de recursos, mas cada um em uma área específica e variando suas informações de acordo com o tipo de recurso descrito, ou seja, de acordo com seu domínio.

Seguem abaixo alguns dos padrões de metadados identificados:

- Dublin Core (DC): padrão de dados para catalogação de recursos eletrônicos da WEB;
- Federal Data Geographic Committee (FGDC): trata de descrição de dados geo-espaciais;
- Machine Readable Cataloging (MARC): trata de dados de catalogação bibliográfica;

- Flexible Image Transport System (FITS): padrão criado pela União Internacional dos Astrônomos, para armazenar informações sobre imagens num cabeçalho de arquivo (PIMENTEL; OLIVEIRA, 2000).
- Object ID: lista de termos que definem a informação mínima essencial para poder seguir a pista de objetos de arte perdidos ou roubados (CROMWELL-KESSLER, 1998).
- Foundation for Documents of Architecture (FDA): padrão para informação de documentos de arquitetura.
- VRA: padrão para descrição de obras de arte e de suas cópias digitais (CROMWELL-KESSLER, 1998).
- Record Export for Art and Cultural Heritage (REACH): padrão para descrição de objetos de museus (coleções).
- Spatial archive and interchange format (SAIF): padrão para compartilhamento de dados espaciais e espaço-temporais (GARCIA, 1999).
- Global Information Locator System (GILS): padrão utilizado em informações governamentais (ROSETTO, 2002).
- Encoded Archival Description (EAD): padrão para inventário de arquivos (ROSETTO, 2002).

Dos padrões de metadados apresentados, o DC é o padrão mais citado, pois foi criado e desenvolvido com o objetivo de tratar recursos da Web, que hoje em dia é uma fonte de disponibilização de informações das mais procuradas por pessoas e instituições.

A criação do DCMI, uma iniciativa composta por profissionais de diversas áreas e de diversos países, contribuiu bastante para a difusão desse padrão, bem como da idéia de se utilizar metadados na descrição de recursos disponíveis na Web.

Sobre metadados, Gilliland-Swetland (1998, p.9) escreve que “metadados representa a pedra de Rosetta que nos permitirá decodificar os recursos e transformá-los em fontes de informação dentro dos sistemas digitais do século XXI”.

Como verificamos, metadados são uma das ferramentas fundamentais, e bastante explorada por pesquisadores de todo o mundo, para resolver o problema da descrição, busca e recuperação de recursos.

Como a busca e a recuperação tornam-se mais difíceis à proporção que o número de recursos a serem descritos aumenta, como o caso dos dados disponíveis na Internet, a utilização de metadados para descrição de recursos seria uma solução para os dados contidos na WWW.

No capítulo seguinte, descreveremos o padrão de metadados DC, mantido por diversos profissionais de vários países e que hoje constitui um dos padrões mais aceitos e utilizados do mundo. Verificaremos como é seu desenvolvimento, quais são os elementos e qualificadores que o compõem e suas características.

## **Capítulo 3**

# **O padrão de metadados Dublin Core**

### 3.1 - Histórico

O padrão de metadados DC tem sua raiz em Chicago, na 2ª Conferência Internacional sobre WWW em Outubro de 1994, onde Yuri Rubinsky, Stuart Weibel e Eric Miller, todos da "Online Computer Library Center" (OCLC) e Joe Hardin da "National Center for Supercomputing Applications" (NCSA) conduziram uma discussão em semântica e Web.

Essa discussão levou a NCSA e a OCLC a organizarem em 1995 um evento, chamado de "OCLC/NCSA Metadata Workshop", de que participaram mais de 50 pessoas com o propósito de discutirem como um conjunto semântico, para recursos baseados na Web, poderia ser extremamente útil para uma pesquisa e recuperação de recursos na Internet. O workshop foi precedido por uma lista de discussão com os participantes.

O objetivo principal desse workshop era chegar a uma definição de um conjunto mínimo de elementos para recursos da Web. Participaram desse evento profissionais de várias áreas: ciência da computação, bibliotecários, profissionais de serviços de informação "on-line", indexadores, catalogadores, profissionais envolvidos com tratamento de dados geo-espaciais, imagens, museus e arquivos etc.

Pretendia-se tratar o problema da catalogação de recursos da rede, com a adoção, a extensão ou a modificação de padrões existentes e de protocolos para facilitar sua recuperação e acesso, atendendo assim a várias comunidades que utilizassem metadados (DESAI, 1997).

Para se chegar a um consenso sobre esse conjunto mínimo de elementos, foram determinadas algumas suposições a serem seguidas na discussão (DESAI, 1997):

- os elementos deveriam descrever um documento como objeto (DLO);

- chegar a um conjunto de elementos comuns;
- os elementos deveriam suportar a recuperação dos recursos da rede;
- todos os elementos de metadados poderiam ser repetitivos;
- todos os elementos seriam opcionais;
- todos os elementos descreveriam o recurso por si próprios, com exceção do elemento fonte;
- os elementos deveriam descrever características intrínsecas do recurso;
- nenhuma suposição seria feita para dizer se o recurso era acessível pela rede ou especificamente eletrônico;
- elementos não incluídos não seriam necessariamente excluídos.

Muitos participantes reconheceram a dificuldade de trabalhar com padrões que possuíssem muitos elementos e procuraram determinar uma relação de elementos não exaustiva para descrever os recursos da rede, tornando-se consenso que a relação de elementos não deveria ser extensa.

Ao conjunto de elementos deram o nome de padrão de metadados DC, pois o evento se dava em Dublin, Ohio.

Outro resultado importante foi o surgimento do DCMI, que é uma organização dedicada a promover a difusão da adoção de padrões de metadados e o desenvolvimento de vocabulário especializado de metadados para descrever recursos que facilitem mais sistemas inteligentes de recuperação de informação (<http://www.dublincore.org/about/overview>).

### 3.2 - O “Dublin Core Metadata Initiative” (DCMI)

Criado a partir do 1º Workshop, o DCMI tem como principal missão criar mecanismos que facilitem a recuperação de recursos na Internet, utilizando-se de padrões de metadados.

Uma das principais características do DCMI é a abertura à participação de qualquer pessoa ou instituição interessada em buscar um consenso mínimo no desenvolvimento de vocabulários de metadados.

Possui profissionais envolvidos em diversas áreas, como Museus, Bibliotecas Digitais, Instituições Educacionais, Informática, Rede de Computadores, Publicação, entre outras. Também possui a participação de organizações como “Coalition for Networked Information”, “Die Deutsche Bibliothek”, “Distributed Systems Technology Centre”, “Joint Information Systems Committee”, “The Library of Congress”, “National Institute of Informatics”, “National Library of Australia”, “National Library of Canada”, “Helsinki University Library - National Library of Finland”, “National Computational Science Alliance”, “National Science Foundation”, “Online Computer Library Center” e “The UK Office for Library and Information Networking” (<http://www.dublincore.org/about/participants>).

Por essas características, seus membros estão distribuídos pelo mundo, destacando-se EUA, Reino Unido, Suécia, Dinamarca, Alemanha, Portugal, Austrália, Nova Zelândia, Japão, China etc.

Entre as atividades desenvolvidas pelo DCMI, destacam-se (<http://www.dublincore.org/about/overview>):

- Desenvolvimento de padrões de metadados para a recuperação de informações nos domínios da Internet;
- Definir estruturas para a interoperabilidade de conjuntos de metadados;

- Organizar workshops internacionais e organizar grupos de trabalhos para o desenvolvimento e a manutenção das recomendações do DCMI;
- Disponibilizar ferramentas, serviços e infraestrutura em metadados;
- Alcançar outras comunidades de metadados, desenvolvendo a interação entre elas e o DCMI.

O DCMI estruturou-se para exercer todas essas atividades, e hoje é composto de (<http://www.dublincore.org/about/organization>):

- um Conselho de Administradores: que aconselha a Diretoria nas questões estratégicas e de alocação de recursos financeiros, sendo seus membros de diversos países;
- uma Diretoria: composta por um Diretor Executivo e outro Administrativo, que supervisionam o gerenciamento e a coordenação das atividades dos Grupos de Trabalho;
- um Conselho Consultivo: composto pelos presidentes dos Grupos de Trabalho, dando consultoria à Diretoria;
- Grupos de Trabalho: criados de acordo com as necessidades e composto por voluntários, que contribuem na resolução de problemas específicos;

Essa organização, além de promover os workshops, também organiza listas de discussão e mantém um site na Internet sobre o DCMI e tudo o que acontece de novidade (<http://www.dublincore.org>).

Em relação à organização de workshops internacionais, com o objetivo de trocar experiências e informações, já foram realizados nove elementos até 2001. É marcante a interdisciplinaridade nas participações, fazendo com que exista nesses eventos uma troca importante de experiências, sendo esta uma das principais razões dos progressos alcançados na área de metadados.

Seguem abaixo os workshops realizados, com seus respectivos locais, datas, objetivos e resultados (<http://www.dublincore.org/workshops>):

### **1º: “OCLC/NCSA Metadata Workshop”**

Realizado em Dublin, Ohio EUA, entre 1 e 3 de março de 1995, teve como objetivo reunir profissionais de diversas áreas para discutir mecanismos que facilitassem a descrição, a organização, a recuperação e o acesso a recursos informacionais disponíveis na Internet.

O principal resultado foi a criação de um conjunto de 13 elementos de metadados (Subject, Title, Author, Publisher, OtherAgent, Date, ObjectType, Form, Identifier, Relation, Source, Language e Coverage), chamado de “Dublin Core Metadata Element Set” ou simplesmente Dublin Core (DC), para a descrição de recursos disponíveis na Internet.

As informações completas sobre o workshop estão disponíveis em:

- <http://dublincore.org/workshops/dc1>

### **2º: “OCLC/UKOLN Metadata Workshop”**

Realizado em Warwick, Reino Unido, entre 1 e 3 de abril de 1996, teve como objetivo estudar a representação do padrão DC para recursos da Internet.

Um dos principais resultados foi uma proposta para representação dos elementos do DC definidos no 1º Workshop em HTML e sua representação usando a sintaxe SGML, definindo uma possível DTD para o conjunto de elementos.

Um consenso entre os participantes foi a necessidade de uma arquitetura que pudesse agregar a diversidade de padrões existentes na Web, possibilitando a integração e o intercâmbio entre eles. A proposta

que surgiu dessa discussão foi a Arquitetura Warwick, que tem como proposta possibilitar a interoperabilidade entre os padrões de metadados distintos existentes na Web.

As informações completas sobre o workshop estão disponíveis em:

- <http://dublincore.org/workshops/dc2>
- <http://www.dlib.org/dlib/july96/07weibel.html>

### **3<sup>o</sup>: “CNI/OCLC Workshop on Metadata for Networked Images”**

Realizado em Dublin, Ohio EUA, em 24 e 25 de setembro 1996, teve como objetivo dar continuidade à discussão sobre o padrão DC e suas propostas de mudanças.

O principal resultado foi o consenso na proposta de mudança de alguns elementos de descrição do padrão DC e o aumento de dois novos elementos (Description e Rights Management), totalizando 15 elementos (Title, Author or Creator, Subject and Keywords, Description, Publisher, Other Contributor, Date, Resource Type, Format, Resource Identifier, Source, Language, Relation, Coverage e Rights Management), resultando na versão 1.0 do padrão DC (<http://dublincore.org/documents/1998/09/dces/>).

As informações completas sobre o workshop estão disponíveis em:

- <http://www.dlib.org/dlib/january97/oclc/01weibel.html>
- <http://www.dlib.org/dlib/january97/oclc/01weibel.html>

#### **4 °: “NLA/DSTC/OCLC Dublin Core Down Under”**

Realizado em Canberra, Austrália, entre 3 e 5 de março de 1997, teve como objetivos: discutir as questões de extensibilidade, possibilitando acrescentar novos elementos ao padrão DC, minimizando a duplicação de esforços na criação de novos padrões de acordo com a necessidade; facilitar a interoperabilidade, o entendimento da estrutura dos elementos DC e a clareza na definição de certos elementos.

O principal resultado foi a verificação de dois grupos de pensamento: os minimalistas, que defendiam a simplicidade do padrão DC como característica importante para o seu uso e para a troca de informações entre as comunidade que utilizam o padrão DC, e os estruturalistas, que defendiam uma maior estruturação do padrão para tornar seus elementos mais usuais.

As informações completas sobre o workshop estão disponíveis em:

- <http://www.dstc.edu.au/cgi-bin/redirect/rd.cgi?http://archive.dstc.edu.au/DC4>
- <http://www.dlib.org/dlib/june97/metadata/06weibel.html>

#### **5 °: “The 5th Dublin Core Metadata Workshop”**

Realizado em Helsinki, Finlândia, entre 6 e 8 de outubro de 1997, teve como temas: a utilização do padrão DC em RDF e com o protocolo Z39.50; a definição do elemento data que depois de diversas discussões foi definida como a data de criação ou da disponibilização do recurso; a discussão dos elementos cobertura e relação, e a utilização de sub-elementos no DC.

Os principais resultados foram: a criação de mais grupos de discussão sobre os elementos data e cobertura; a utilização de sub-

elementos; a relação entre os elementos fonte e relação; progressos na utilização da sintaxe RDF.

As informações completas sobre o workshop estão disponíveis em:

- <http://www.lib.helsinki.fi/meta/DC5.html>
- <http://www.ariadne.ac.uk/issue12/metadata>
- <http://www.dlib.org/dlib/february98/02weibel.html>

### **6 °: “The 6th Dublin Core Metadata Workshop”**

Realizado em Washington, EUA, entre 2 e 4 de novembro de 1998, teve como objetivos a consolidação do desenvolvimento de diversos grupos de trabalho DC, a troca de experiências entre implementações piloto e o encaminhamento de discussões com o propósito de promover a interoperabilidade entre implementações DC.

Os principais resultados foram o levantamento da necessidade de discussões com outras comunidades e o avanço nas áreas de padronização, formalização de sintaxes alternativas e um refinamento na semântica dos elementos e seus qualificadores.

As informações completas sobre o workshop estão disponíveis em:

- <http://dublincore.org/workshops/dc6>
- <http://www.dlib.org/dlib/april99/04weibel.html>

### **7 °: “The 7th Dublin Core Metadata Workshop”**

Realizado em Frankfurt, Alemanha, entre 25 e 7 de outubro de 1999, teve como objetivo principal consolidar o desenvolvimento de vários grupos de trabalho DC, através de trocas de experiências entre os

mesmos, proporcionando ao DC maior interoperabilidade entre sistemas de metadados heterogêneos.

O principal resultado foi a integração entre os grupos de trabalho DC para troca de informações.

As informações completas sobre o workshop estão disponíveis em:

- <http://dublincore.org/archives/1999/dc7/index.htm>

### **8º: “The 8th International Dublin Core Metadata Workshop”**

Realizado em Ottawa, Canadá, entre 4 e 6 de outubro de 2000, teve como objetivo consolidar o desenvolvimento dos vários grupos de trabalho, provendo ao padrão DC o suporte a interoperabilidade entre sistemas de metadados heterogêneos. Teve como principais temas a questão lingüística em metadados, políticas e requisitos funcionais para registros de metadados, a representação estruturada de alguns elementos como autor, colaborador e editor, e estender o padrão DC para outras entidades de metadados com domínios específicos.

As informações completas sobre o workshop estão disponíveis em:

- <http://www.ifla.org/udt/dc8/index.htm>
- <http://www.dlib.org/dlib/december00/weibel/12weibel.html>

### **9º: “International Conference on Dublin Core and Metadata Applications”**

Realizado em Tokyo, Japão, entre 22 e 26 de outubro de 2001, foi o primeiro na Ásia e também o primeiro a incluir conferências, tendo como objetivos promover um fórum para discutir mais

profundamente o desenvolvimento do padrão DC e de padrões relacionados, além de discutir novas idéias sobre metadados e aplicações, não limitadas ao DC e, finalmente, prover tutoriais para criação, gerenciamento e uso de aplicações de metadados.

As informações completas sobre o workshop estão disponíveis em:

- <http://www.nii.ac.jp/dc2001>

Uma iniciativa importante do DCMI é a ligação que existe com outras iniciativas e instituições que estudam metadados, como:

- CEN (European Committee for Standardization):  
<http://www.cenorm.be/>
- IEEE/LOM (The Institute of Electrical and Electronics Engineers, Inc.): <http://www.ieee.org/>
- IETF (Internet Engineering Task Force):  
<http://www.ietf.org/>
- NDLTD (Networked Digital Library of Theses and Dissertations): <http://www.ndltd.org/>
- NISO (North American Information Standardization Organization): <http://www.niso.org/>
- Open Archives Initiative: <http://www.openarchives.org/>
- W3C (World Wide Web Consortium): <http://www.w3.org/>

A interação com desenvolvedores de aplicação e de infraestrutura técnica para a Web é um braço importante do DCMI, pois o suporte a uma arquitetura é essencial para a adoção de padrões.

Nesse sentido, o DCMI possui uma forte ligação com o W3C e com comunidades de desenvolvedores em RDF e XML, possibilitando a essas ferramentas possuir codificação para o padrão DC, como por exemplo HTML e XML.

Os elementos do padrão DC estão normalmente embutidos dentro do próprio documento descrito em HTML, XML e outros, mas podem estar separados do recurso, dependendo da situação e da forma como o recurso está distribuído.

### **3.3 - O padrão de metadados Dublin Core (DC)**

O padrão de metadados DC é um conjunto de quinze elementos, com o objetivo de descrever um recurso eletrônico.

Desenvolvido na língua inglesa, pois teve sua origem nos EUA, a versão 1.1 já foi traduzida para vários idiomas, havendo um grupo de trabalho no DCMI estudando, juntamente com o “World Wide Web Consortium” (W3C), formas de juntar as versões traduzidas utilizando tecnologia RDF.

O RDF é base para o processamento de metadados e tem como principal objetivo proporcionar interoperabilidade entre aplicações que trocam informações eletrônicas da Web. Possui um modelo de representação e uma sintaxe para codificação e transporte dos metadados. Normalmente essa sintaxe é representada utilizando-se XML (<http://www.w3.org/TR/1999/REC-rdf-syntax-19990222>).

O padrão de metadados DC se caracteriza por:

- Simplicidade: como a maioria dos elementos tem um entendimento semântico simples, o padrão DC pode ser facilmente gerado pelo responsável do documento sem a necessidade de extensos treinamentos;
- Interoperabilidade semântica: diferentes modelos de descrição interferem na habilidade das pesquisas entre áreas. A existência de um modelo comum aumenta a possibilidade de interoperabilidade entre essas áreas;

- Consenso internacional: a participação de mais de vinte países no DCMI para a busca de escopo internacional na WEB e de uma infra-estrutura adequada contribui para um consenso internacional;
- Extensibilidade: o padrão DC é um modelo simplificado de descrição, que possui flexibilidade e extensibilidade na elaboração de modelos, ou seja, permite que novos elementos possam ser acrescentados para atender a uma necessidade de descrição de um determinado recurso. Esses novos elementos, juntamente com os elementos do DC, permitem que várias comunidades em diversas áreas possam utilizar o padrão DC, trocar informações e ter acesso a elas;  
(<http://www.dublincore.org/documents/2001/04/12/usageguide>)
- Flexibilidade: seus elementos são opcionais, podem ser repetidos se necessário, e modificáveis utilizando-se de qualificadores, que mostraremos nesse capítulo.

Um exemplo de extensibilidade pode ser visto no Projeto "Gateway to Educacional Material" (GEM), descrito por Sutton (1998), no qual os recursos são descritos por elementos do DC juntamente com elementos do GEM: DC.contributor, DC.coverage, DC.creator, DC.date, DC.description, DC.format, DC.identifier, DC.language, DC.publisher, DC.relation, DC.rights, DC.source, DC.subject, DC.title, DC.type, GEM.audience, GEM.cataloging, GEM.duration, GEM.essentialresources, GEM.grade, GEM.pedagogy, GEM.quality e GEM.standars.

Essas características são fatores que levaram à adoção do padrão de metadados DC para descrever as imagens contidas no Banco de Dados do Rural Mídia (SOUZA; CATARINO; SANTOS, 1997).

No site do DCMI podemos encontrar uma lista de projetos submetidos, totalizando cerca de 61 (sessenta e um) em fevereiro de 2002, nas áreas de Artes e Humanidades, Bibliografia, Negócios, Educação, Meio Ambiente, Matemática, Medicina, Ciência e Tecnologia e outras (<http://www.dublincore.org/projects>).

O padrão DC também tem sido adotado por vários países, como Austrália, Canadá, Dinamarca, Finlândia, Irlanda e Reino Unido (<http://www.dublincore.org/news/adoption>).

Para entendermos o funcionamento do padrão de metadados DC, vamos analisar a versão 1.1, composta de 15 elementos. Essa versão é uma iniciativa de recomendação de metadados, e o DCMI entende por recomendação que as especificações são fixas e suportadas para adoção da comunidade DC.

### 3.3.1 - Atributos

Cada elemento DC é definido usando um conjunto de 10 atributos padrões da ISSO/IEC 11179 para a descrição dos elementos de dados (<http://www.dublincore.org/documents/1999/07/02/dces>).

Os atributos são os seguintes:

- **Nome** - O nome atribuído para o elemento de dado;
- **Identificador** - O identificador único atribuído para o elemento de dado;
- **Versão** – A versão do elemento de dado;
- **Registro de Autoridade** – A entidade autorizada a registrar o elemento de dado;
- **Língua** – A língua na qual o elemento de dado está especificado;
- **Definição** – A declaração que claramente representa o conceito e a natureza essencial do elemento de dados;

- **Obrigação** – Indica se o elemento de dados é obrigatório ou não;
- **Tipo do dado** – Indica o tipo do dado que pode ser representado no conteúdo do elemento de dado;
- **Máxima Ocorrência** – Indica algum limite para a repetição dos elementos de dados;
- **Comentário** – Uma observação relativa à aplicação dos elementos de dados;

Seis desses atributos são comuns para todos os elementos do DC, possuindo um valor único. São eles:

- **Versão:** 1.1;
- **Registro de Autoridade:** Dublin Core Metadata Initiative;
- **Língua:** Inglês;
- **Obrigação:** Opcional;
- **Tipo do dado:** *String* de caracteres;
- **Máxima Ocorrência:** ilimitada;

Analisaremos, portanto, os atributos Nome, Identificador, Definição e Comentário, que não são comuns nos 15 elementos do padrão DC, para entendermos como é realizada a descrição de recursos.

Cada definição DC refere-se a um recurso existente descrito. A “Internet Engineering Task Force” (IETF), que é uma comunidade internacional aberta preocupada com a evolução da arquitetura da Internet e de seu uso facilitado para os usuários, define na Request for Comments (RFC) 2396 um recurso como alguma coisa que tem identidade (<http://www.ietf.org/rfc/rfc2396.txt>). Para os propósitos do DC, um recurso será tipicamente uma informação, mas pode ser aplicado mais amplamente.

### 3.3.2 - Elementos do padrão DC

As definições dos elementos do padrão de metadados DC estão definidas na RFC 2413 (<http://www.ietf.org/rfc/rfc2413.txt>).

O atributo identificador permanece na língua original, ou seja, na língua inglesa, pois é dessa forma que o mesmo deve aparecer na descrição de um recurso, por exemplo, em uma página Web.

Segue abaixo a descrição detalhada de cada um dos quinze elementos que compõem o padrão DC, com os atributos que não são comuns (<http://www.dublincore.org/documents/1999/07/02/dces>). Utilizaremos como exemplos principais o site do DCMI, um texto da Revista Ciência da Informação on-line (CIONline) e um quadro de Vincent Van Gogh digitalizado:

#### 1- Título

Nome: Título

Identificador: **Title**

Definição: Um nome dado para o recurso.

Comentário: Tipicamente, um Título será o nome pelo qual o recurso é formalmente conhecido, podendo ser o próprio título, o nome ou uma descrição curta.

Exemplos:

- ❑ Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core (texto)
- ❑ Dublin Core Metadata Initiative –DCMI (site)
- ❑ Girassol (imagem)

## 2- Autor

Nome: Autor

Identificador: **Creator**

Definição: Um entidade primeiramente responsável pelo conteúdo do recursos.

Comentário: Exemplos de um Autor incluem uma pessoa, uma organização ou um serviço. Tipicamente, o nome de um autor deve ser usado para indicar uma entidade.

Exemplos:

- ❑ Souza, Márcia Izabel Fujisawa.; Vendrusculo, Laurimar Gonçalves; Melo, Geane Cristina (autores)
- ❑ DCMI (site)
- ❑ Gogh, Vincent Van (autor)

## 3- Assunto

Nome: Assunto e palavras-chave

Identificador: **Subject**

Definição: O tema (objeto ou ponto principal) do conteúdo do recurso.

Comentário: Tipicamente, um Assunto será expresso com palavras-chave, frases-chave, descritores ou códigos de classificação que descrevem o tema do recurso (indica o conteúdo informativo). A prática mais recomendada é selecionar os valores de um vocabulário controlado ou de um esquema de classificação formal.

Exemplos:

- Metadados, Dublin Core, Informação eletrônica
- Metadados, Dublin Core
- pintura, quadro

#### **4- Descrição**

Nome: Descrição

Identificador: **Description**

Definição: Um relato do conteúdo do recurso.

Comentário: Descrição pode incluir (mas não é limitado a): um "abstract", tabelas de conteúdos, referências para uma representação de conteúdo ou um texto livre de relato do conteúdo.

Exemplos:

- "Este artigo aborda a necessidade de adoção de padrões de descrição de recursos de informação eletrônica, particularmente, no âmbito da Embrapa Informática Agropecuária. O Rural Mídia foi desenvolvido utilizando o modelo DC para descrição de seu acervo, acrescido de pequenas adaptações introduzidas diante da necessidade de adequar-se a especificidades meramente institucionais. Este modelo de metadados baseado no Dublin Core, adaptado para o Banco de Imagem, possui características que endossam a sua adoção, como 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)."  
(Resumo)

- “The Dublin Core Metadata Initiative is an open forum engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models. DCMI's activities include consensus-driven working groups, global workshops, conferences, standards liaison, and educational efforts to promote widespread acceptance of metadata standards and practices” (descrição do site)
- “Óleo sobre tela, Girassóis é uma das obras mais famosas de Van Gogh, e retrata o símbolo do poder e da beneficência da vida” (descrição da imagem)

## 5- Editor

Nome: Editor

Identificador: **Publisher**

Definição: Uma entidade responsável por tornar o recurso disponível

Comentário: Exemplos de editor incluem uma pessoa, uma organização ou um serviço. Tipicamente, o nome de um editor deve ser usado para indicar a entidade.

Exemplos:

- IBICT
- DCMI (site)
- Ediouro S.A. (editora)

## 6- Colaborador

Nome: Colaborador

Identificador: **Contributor**

Definição: Uma entidade responsável por fazer contribuições para o conteúdo do recurso.

Comentário: Exemplos de um Colaborador incluem uma pessoa, uma organização ou um serviço. Tipicamente, o nome do colaborador deve ser usado para indicar a entidade.

Exemplos:

- Gracio, Jose Carlos Abbud (uma pessoa)
- UNESP (entidade)

## 7- Data

Nome: Data

Identificador: **Date**

Definição: Uma data associada com um evento no ciclo de vida do recurso.

Comentário: Tipicamente, data será associada com a criação ou a disponibilização do recurso. Recomendação para melhor uso de codificação dos valores de data é definido na norma ISO 8601 (<http://www.w3.org/TR/1998/NOTE-datetime-19980827>) e segue o formato YYYY-MM-DD, onde YYYY é o ano, MM é o mês e DD o dia.

Exemplos:

- 2002-03-28
- 2000
- 1995

## 8- Tipo

Nome: Tipo do recurso

**Identificador: Type**

Definição: A natureza ou a espécie do conteúdo do recurso.

Comentário: Tipo inclui termos descrevendo categorias gerais, funções, espécies ou níveis de agregação para conteúdo. Recomendação para melhor uso é selecionar valores de vocabulário controlado. Para descrever manifestação física ou digital dos recursos, deve-se usar o elemento FORMATO.

Segue abaixo a lista de termos, sugerida pelo DCMI, que pode ser utilizada para identificar o elemento tipo do recurso (<http://dublincore.org/documents/2000/07/11/dcmi-type-vocabulary>):

- 1- "Collection": a coleção é um agregação de itens;
- 2- "Dataset": o conjunto de dados é uma informação codificada em uma estrutura definida. Exemplos: listas, tabelas e banco de dados;
- 3- "Event": evento é uma ocorrência baseada em tempo. Exemplos: Conferência, Workshop etc.
- 4- "Image": imagem é uma representação visual do recurso: exemplos: fotografia, pintura, desenho, animação, filme, mapa etc.
- 5- "Interactive Resource": é um recurso que requer interação com o usuário. Exemplos: página Web, objetos de aprendizagem multimídia, serviços de chat, realidade virtual etc.
- 6- "Service": é um sistema com funções de valor para o usuário final. Exemplos: serviço de banco, serviço de autenticação, servidor de Web etc.
- 7- "Software": é um programa de computador compilado ou fonte.

8- "Sound": é um recurso cujo conteúdo é composto de áudio: Exemplos: sons, CD de áudio etc.

9- "Text": é um recurso que contém palavras para leitura. Exemplos: livros, artigos, dissertações, poemas etc.

Exemplos:

- ❑ Interactive Resource
- ❑ Text
- ❑ Image

## 9- Formato

Nome: Formato

Identificador: **Format**

Definição: A manifestação física ou digital do recurso.

Comentário: Tipicamente, formato pode incluir o tipo da mídia ou as dimensões do recurso. O Formato pode ser usado para determinar o software, hardware ou outro equipamento necessário para mostrar ou operar o recurso. Exemplos de dimensões incluem tamanho e duração. A recomendação para melhor uso é selecionar valores de vocabulário controlado, como a lista de tipos e subtipos de mídias da Internet (<http://www.isi.edu/in-notes/iana/assignments/media-types/media-types>) ("MIME types") definindo formatos de mídia para computador.

Exemplos:

- ❑ text/html
- ❑ text/pdf
- ❑ image/jpeg

## 10- Identificador

Nome: Identificador de recurso

Identificador: **Identifier**

Definição: Uma referência não ambígua para o recurso dentro de um dado contexto.

Comentário: A recomendação para melhor uso é identificar o recurso pelo significado de uma *string* ou número conforme um sistema de identificação formal. Exemplo de sistemas de identificação formal incluem o Identificador de Recursos Uniforme (Uniform Resource Identificador - URI), o Localizador de Recursos Uniforme (Uniform Resource Locator - URL), o Identificador de Objetos Digitais (Digital Object Identificador - DOI) e o Número Internacional Normalizado para Livros (International Standard Book Number - ISBN).

Exemplo:

- ❑ <http://www.ibict.br/cionline/290100/29010010.pdf>
- ❑ <http://dublincore.org/>
- ❑ image012-Gogh.jpeg (nome do arquivo)

## 11- Fonte

Nome: Fonte

Identificador: **Source**

Definição: Uma referência para o recurso do qual o presente recurso é derivado.

Comentário: O presente recurso pode ser derivado de uma fonte de recurso inteira ou em parte. A recomendação para melhor uso é

identificar o recurso pelo significado da *string* ou do número conforme o sistema de identificação formal.

Exemplo:

- ❑ <http://a.b.org/> (um site)
- ❑ Vida e obra de Vincent Van Gogh (de onde foi feita a cópia digital da imagem)

## 12- Língua

Nome: Língua

Identificador: **Language**

Definição: Uma língua do conteúdo intelectual do recurso.

Comentário: A recomendação para melhor uso dos valores do elemento língua é definida pela RFC 1766 (<http://www.ietf.org/rfc/rfc1766.txt>) que inclui um código de língua em 2 letras (do padrão ISO 639), seguido opcionalmente pelo código de país em 2 letras também (do padrão ISO 3166).

Exemplos:

- ❑ pt-br (Português - Brasil)
- ❑ en-us (Inglês – Estados Unidos)
- ❑ fr (Francês)

## 13- Relação

Nome: Relação

Identificador: **Relation**

Definição: Uma referência para o recurso relacionado, como versão de um trabalho, tradução de um trabalho ou parte de um trabalho.

Comentário: A recomendação para melhor uso é referenciar o recurso pelo significado da *string* ou do número conforme um sistema de identificação formal.

Exemplos:

- ❑ é versão da canção de Elton John
- ❑ é baseado no filme de 1920

#### **14- Cobertura**

Nome: Cobertura

Identificador: **Coverage**

Definição: O âmbito do conteúdo do recurso.

Comentário: Cobertura tipicamente inclui localização espacial (o nome de um lugar ou suas coordenadas geográficas), período temporal (um rótulo, uma data ou intervalo de datas do período) ou jurisdição (como o nome de uma entidade administrativa).

A recomendação para melhor uso é selecionar valores de vocabulário controlado, como do Thesaurus de Nomes Geográficos (TGN) (<http://www.getty.edu/research/tools/vocabulary/tgn/index.html>) e, quando for apropriado, nomes de lugares ou períodos de tempo são usados em preferência a identificadores numéricos como conjunto de coordenadas ou intervalo de tempo.

Exemplos:

- ❑ 1995-2002
- ❑ Estados Unidos
- ❑ UNESP, Marília, São Paulo, Brasil

## 15- Direitos

Nome: Gerenciamento de Direitos

Identificador: **Rights**

Definição: Informações sobre direitos do recurso.

Comentário: Tipicamente, um elemento Direitos conterá uma declaração de gerenciamento de direitos para o recurso. Informações de Direitos freqüentemente abrangem Direitos de Propriedade Intelectual (Intellectual Property Rights - IPR), Copyright, e várias propriedades de Direitos. Se o elemento Direitos é ausente, nenhuma suposição pode ser feita sobre o "status" desse ou de outro direito com respeito a esse recurso.

Exemplos:

- Copyright 1995-2002 DCMI All Rights Reserved
- <http://dublincore.org/about/copyright/#copyright> (site)

Seguem três exemplos de descrição de um recurso utilizando o padrão DC:

**Exemplo 1** – Texto disponível na Revista Ciência da Informação on-line.

<b>Elemento</b>	<b>Conteúdo</b>
<b>DC.Title</b>	Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core
<b>DC.Creator</b>	Souza, Márcia Izabel Fujisawa.
<b>DC.Creator</b>	Vendrusculo, Laurimar Gonçalves
<b>DC.Creator</b>	Melo, Geane Cristina
<b>DC.Subject</b>	Metadados
<b>DC.Subject</b>	Dublin Core
<b>DC.Subject</b>	Informação eletrônica
<b>DC.Description</b>	Este artigo aborda a necessidade de adoção de padrões de descrição de recursos de informação eletrônica, particularmente, no âmbito da Embrapa Informática Agropecuária. O Rural Mídia foi desenvolvido utilizando o modelo Dublin Core (DC) para descrição de seu acervo, acrescido de pequenas adaptações introduzidas diante da necessidade de adequar-se a especificidades meramente institucionais. Este modelo de metadados baseado no Dublin Core, adaptado para o Banco de Imagem, possui características que endossam a sua adoção, como 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)
<b>DC.Publisher</b>	IBICT
<b>DC.Date</b>	2002-03-28
<b>DC.Type</b>	Interactive Resource
<b>DC.Format</b>	text/html
<b>DC.Identifier</b>	<a href="http://www.ibict.br/cionline/290100/29010010.pdf">http://www.ibict.br/cionline/290100/29010010.pdf</a>
<b>DC.Language</b>	pt-br
<b>DC.Coverage</b>	Brasil
<b>DC.Rights</b>	IBICT

## Exemplo 2 - Página Web do DCMI

<b>Elemento</b>	<b>Conteúdo</b>
<b>DC.Title</b>	Dublin Core Metadata Initiative –DCMI
<b>DC.Creator</b>	DCMI
<b>DC.Subject</b>	Metadados
<b>DC.Subject</b>	Dublin Core
<b>DC.Description</b>	The Dublin Core Metadata Initiative is an open forum engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models. DCMI's activities include consensus-driven working groups, global workshops, conferences, standards liaison, and educational efforts to promote widespread acceptance of metadata standards and practices
<b>DC.Publisher</b>	DCMI
<b>DC.Date</b>	2000
<b>DC.Type</b>	Text
<b>DC.Format</b>	text/pdf
<b>DC.Identifier</b>	<a href="http://dublincore.org/">http://dublincore.org/</a>
<b>DC.Language</b>	en-us
<b>DC.Coverage</b>	1995-2002
<b>DC.Rights</b>	Copyright 1995-2002 DCMI All Rights Reserved

## Exemplo 3 - Imagem digitalizada de uma tela de Vincent Van Gogh

Elemento	Conteúdo
<b>DC.Title</b>	Girassol
<b>DC.Creator</b>	Gogh, Vincent Van
<b>DC.Subject</b>	Tela
<b>DC.Subject</b>	Pintura
<b>DC.Description</b>	Óleo sobre tela, Girassóis é uma das obras mais famosas de Van Gogh, e retrata o símbolo do poder e da beneficência da vida
<b>DC.Publisher</b>	Ediouro S.A.
<b>DC.Date</b>	1995
<b>DC.Type</b>	Image
<b>DC.Format</b>	image/jpeg
<b>DC.Identifier</b>	image012-Gogh.jpeg

Podemos notar nos exemplos acima que alguns elementos, como Creator e Subject, aparecem mais de uma vez, mostrando a característica do DC de poder contar com elementos repetidos. Também podemos verificar que alguns elementos não aparecem na descrição, pois não são necessários, mostrando a característica dos seus elementos de serem opcionais.

Algumas vezes necessitamos de um maior detalhamento dos elementos que compõem o DC, como o Date, que poderia conter diversas datas: data da criação, data da disponibilização do recurso na Web e data da última atualização, para dar maior especificidade às informações contidas nos metadados.

### 3.3.3 – Qualificadores para os elementos DC

Sentindo a necessidade de representar de maneira mais detalhada os recursos expressos pelos elementos do padrão DC, o DCMI

desenvolveu, através de seus grupos de trabalho e com aprovação de seu comitê, qualificadores para o conjunto de elementos do padrão DC.

Esses qualificadores foram estudados sobre vocabulários controlados e sistemas de classificação existentes e não pretendem ser um conjunto rígido. Ao contrário disso, o DCMI está aberto a sugestões e melhorias que possam ser implementadas no futuro.

Os usuários podem livremente utilizar ou não esses qualificadores, podendo até desenvolver qualificadores adicionais para sua aplicação, mas tendo em mente que esses qualificadores não serão entendidos por outras aplicações e não poderão, portanto, ser reutilizados por outras comunidades.

O DCMI definiu duas classes de qualificadores:

- 1- Elemento de refinamento: esses qualificadores dão mais especificidade a um elemento, detalhando-o melhor.
- 2- Esquema de codificação: esses qualificadores identificam esquemas para o valor do elemento. Incluem vocabulário controlado e notações formais de representação (sistemas de classificação).

Os qualificadores propostos são mostrados resumidamente no quadro abaixo:

Conjunto de elementos DC	Elementos de refinamento	Elementos do esquema de codificação
<b>1. Title</b>	Alternative	-
<b>2. Creator</b>	-	-
<b>3. Subject</b>	-	LCSH MeSH DDC LCC UDC
<b>4. Description</b>	Table Of Contents Abstract	-
<b>5. Publisher</b>	-	-
<b>6. Contributor</b>	-	-
<b>7. Date</b>	Created Valid Available Issued Modified	DCMI Period W3C-DTF
<b>8. Type</b>	-	DCMI Type Vocabulary
<b>9. Format</b>	Extent	-
	Medium	IMT
<b>10. Identifier</b>	-	URI
<b>11. Source</b>	-	URI
<b>12. Language</b>	-	ISO 639-2 RFC 1766
<b>13. Relation</b>	Is Version Of Has Version Is Replaced By Replaces Is Required By Requires Is Part Of Has Part Is Referenced By References Is Format Of Has Format	URI
<b>14. Coverage</b>	Spatial	DCMI Point ISO 3166 DCMI Box TGN
	Temporal	DCMI Period W3C-DTF
<b>15. Rights</b>	-	-

Quadro 2 – Qualificadores do padrão DC

Como podemos observar no quadro acima, os elementos Creator, Publisher, Contributor e Rights não possuem qualificadores, sejam eles de refinamento ou de esquema de codificação.

Os qualificadores DC possuem os seguintes atributos:

- **Nome:** símbolo único atribuído ao qualificador e utilizado na representação em HTML por exemplo.
- **Rótulo:** Um rótulo legível dado ao qualificador para leitura humana.
- **Definição:** Uma declaração que representa o conceito e a natureza do qualificador.
- **Comentário:** Informações adicionais associadas ao qualificador (se houver).
- **Veja também:** Um link para mais informações sobre o qualificador (se houver).

O valor de todos os atributos foram traduzidos, com exceção do atributo Nome, que deve ser utilizado no seu formato original, ou seja, na língua inglesa, como o atributo Identificador do conjunto de elementos do padrão DC.

Seguem abaixo os qualificadores detalhados, sejam eles de refinamento ou de esquema de codificação, para os elementos do padrão DC que os possuem (<http://dublincore.org/documents/dcmes-qualifiers/>):

#### 1- Elemento: **Title**

Elemento de refinamento:

Nome: **alternative**

Rótulo: Alternativo

Definição: outra forma usada para título como alternativa ou para substituir o original.

Comentário: Este qualificador pode incluir abreviações ou tradução.

Esquema de codificação: **não possui**

## 2- Elemento: **Creator**

Elementos de refinamento: **não possui**

Esquemas de codificação: **não possui**

## 3- Elemento: **Subject**

Elemento de refinamento: **não possui**

Esquemas de codificação:

Nome: **LCSH**

Rótulo: LCSH

Definição: Library of Congress Subject Headings

Nome: **MESH**

Rótulo: MeSH

Definição: Medical Subject Headings

Veja também:

<http://www.nlm.nih.gov/mesh/meshhome.html>

Nome: **DDC**

Rótulo: DDC

Definição: Dewey Decimal Classification

Veja também: <http://www.oclc.org/dewey/index.htm>

Nome: **LCC**

Rótulo: LCC

Definição: Library of Congress Classification

Veja também:

<http://lcweb.loc.gov/catdir/cpsolcco/lcco.html>

Nome: **UDC**

Rótulo: UDC

Definição: Universal Decimal Classification

Veja também: <http://www.udcc.org/>

#### 4- Elemento: **Description**

Elementos de refinamento:

Nome: **tableOfContents**

Rótulo: Tabela de Conteúdo

Definição: Uma lista de subunidades do conteúdo do recurso.

Nome: **abstract**

Rótulo: Abstract ou resumo

Definição: Um sumário do conteúdo do recurso.

Esquema de codificação: **não possui**

#### 5- Elemento: **Publisher**

Elementos de refinamento: **não possui**

Esquemas de codificação: **não possui**

#### 6- Elemento: **Contributor**

Elementos de refinamento: **não possui**

Esquemas de codificação: **não possui**

#### 7- Elemento: **Date**

Elementos de refinamento:

Nome: **created**

Rótulo: Criação

Definição: Data da criação do recurso.

Nome: **valid**

Rótulo: Validade

Definição: Data (freqüentemente um intervalo) da validade do recurso.

Nome: **available**

Rótulo: Disponível

Definição: Data (freqüentemente um intervalo) em que o recurso se tornará ou tornou-se disponível.

Nome: **issued**

Rótulo: Distribuição

Definição: Data da distribuição formal (igual a publicação) do recurso.

Nome: **modified**

Rótulo: Modificado

Definição: Data na qual o recurso foi mudado.

Esquemas de codificação para todos os elementos de refinamento:

Nome: **Period**

Rótulo: Esquema DCMI para período

Definição: Uma especificação de limite para intervalo de tempo.

Veja também: <http://dublincore.org/documents/dcmi-period/>

Nome: **W3CDTF**

Rótulo: W3C-DTF

Definição: Regras de codificação do W3C para data e hora baseado na norma ISO 8601

Veja também: <http://www.w3.org/TR/NOTE-datetime>

#### 8- Elemento: **Type**

Elemento de refinamento: **não possui**

Esquema de codificação:

Nome: **DCMIType**

Rótulo: Esquema DCMI para tipo de vocabulário

Definição: Uma lista de tipos usada para categorizar a natureza ou o gênero do conteúdo do recurso. Essa lista é composta pelos seguintes valores: Collection, Dataset, Event, Image, Interactive Resource, Service, Software, Sound ou Text.

Veja também: <http://dublincore.org/documents/dcmi-type-vocabulary>

#### 9- Elemento: **Format**

Elementos de refinamento:

Nome: **extent**

Rótulo: Medida

Definição: O tamanho ou a duração do recurso.

Nome: **medium**

Rótulo: Meio

Definição: O material ou o transporte físico do recurso.

Veja também: <http://www.isi.edu/in-notes/iana/assignments/media-types/media-types>

Esquema de codificação para medium:

Nome: **IMT**

Rótulo: IMT

Definição: O Esquema de tipos de meio de armazenamento da Internet (Internet Media Types), como: text/html, application/pdf, image/jpegvideo/mpeg etc.

#### 10- Elemento: **Identifier**

Elemento de refinamento: **não possui**

Esquema de codificação:

Nome: **URI**

Rótulo: URI

Definição: Uniform Resource Identifier

Veja também: <http://www.ietf.org/rfc/rfc2396.txt>

#### 11- Elemento: **Source**

Elemento de refinamento: **não possui**

Esquema de codificação:

Nome: **URI**

Rótulo: URI

Definição: Uniform Resource Identifier

Veja também: <http://www.ietf.org/rfc/rfc2396.txt>

#### 12- Elemento: **Language**

Elemento de refinamento: **não possui**

Esquemas de codificação:

Nome: **ISO639-2**

Rótulo: ISO 639-2

Definição: ISO 639-2: Código para representação de nomes de linguagens.

Veja também: <http://lcweb.loc.gov/standards/iso639-2/langhome.html>

Nome: **RFC1766**

Rótulo: RFC 1766

Definição: RFC 1766: Código para identificação da linguagem que especifica duas letras da norma ISO 639, seguidas opcionalmente de duas letras para o país da norma ISO 3166.

Veja também: <http://www.ietf.org/rfc/rfc1766.txt>

### 13- Elemento: **Relation**

Elementos de refinamento:

Nome: **isVersionOf**

Rótulo: é uma versão de

Definição: O recurso descrito é uma versão, edição ou adaptação do recurso citado.

Nome: **hasVersion**

Rótulo: tem uma versão

Definição: O recurso descrito tem uma versão, edição ou adaptação do recurso citado.

Nome: **isReplacedBy**

Rótulo: é substituído por

Definição: O recurso descrito é substituído pelo recurso citado.

Nome: **replaces**

Rótulo: Substitui

Definição: O recurso descrito substitui o recurso citado.

Nome: **isRequiredBy**

Rótulo: É requerido por

Definição: O recurso descrito é requerido pelo recurso citado, física ou logicamente.

Nome: **requires**

Rótulo: Requer

Definição: O recurso descrito requer o recurso citado para suportar sua função, distribuição ou coerência de conteúdo.

Nome: **isPartOf**

Rótulo: É parte de

Definição: O recurso descrito é uma parte física ou lógica do recurso citado.

Nome: **hasPart**

Rótulo: Tem parte

Definição: O recurso descrito inclui o recurso citado física ou logicamente.

Nome: **isReferencedBy**

Rótulo: É referenciado por

Definição: O recurso descrito é referenciado, citado ou apontado pelo recurso citado.

Nome: **references**

Rótulo: Faz referência

Definição: O recurso descrito faz referência, cita ou aponta para o recurso citado.

Nome: **isFormatOf**

Rótulo: É formatado do

Definição: O recurso descrito tem o mesmo conteúdo intelectual do recurso citado, mas apresentado em outro formato.

Nome: **hasFormat**

Rótulo: Has Format

Definição: O recurso descrito pré-existe ao recurso citado, o qual essencialmente possui o mesmo conteúdo intelectual apresentado em outro formato.

Esquemas de codificação para todos os elementos de refinamento:

Nome: **URI**

Rótulo: URI

Definição: Uniform Resource Identifier

Veja também: <http://www.ietf.org/rfc/rfc2396.txt>

#### 14- Elemento: **Coverage**

Elemento de refinamento:

Nome: **spatial**

Rótulo: Espacial

Definição: Contém características espaciais do conteúdo intelectual do recurso.

Esquemas de codificação para **spatial**:

Nome: **Point**

Rótulo: DCMI Point

Definição: O DCMI Point identifica um ponto no espaço usando suas coordenadas geográficas.

Veja também:

<http://dublincore.org/documents/dcmi-point/>

Nome: **ISO3166**

Rótulo: ISO 3166

Definição: ISO 3166 é o código para a representação de nomes de países.

Veja também:

<http://www.din.de/gremien/nas/nabd/iso3166ma/codlstp1/index.html>

Nome: **Box**

Rótulo: DCMI Box

Definição: O DCMI Box identifica uma região do espaço usando seus limites geográficos.

Veja também:

<http://dublincore.org/documents/dcmi-box/>

Nome: **TGN**

Rótulo: TGN

Definição: Getty Thesaurus of Geographic Names

Veja também:

[http://shiva.pub.getty.edu/tgn\\_browser/](http://shiva.pub.getty.edu/tgn_browser/)

Elemento de refinamento:

Nome: **temporal**

Rótulo: Temporal

Definição: Contém características temporais do conteúdo intelectual do recurso.

Esquemas de codificação para **temporal**:

Nome: **Period**

Rótulo: Período DCMI

Definição: Uma especificação de limites de um intervalo de tempo.

Veja também:

<http://dublincore.org/documents/dcmi-period/>

Nome: **W3CDTF**

Rótulo: W3C-DTF

Definição: Regras de codificação para data e hora da W3C baseadas na norma ISO 8601

Veja também: <http://www.w3.org/TR/NOTE-datetime>

#### 15- Elemento: **Rights**

Elementos de refinamento: **não possui**

Esquemas de codificação: **não possui**

Esses qualificadores são representados juntamente com os elementos do padrão DC separados do recurso ou dentro dele, como em uma página Web descrita por exemplo em HTML.

### **3.4 – Representação do DC em HTML**

Ainda hoje, a linguagem de marca HTML é a mais utilizada na construção de páginas e na disponibilização de recursos na Internet. Apesar de sintaticamente limitada para a descrição de recursos complexos, o uso de qualificadores DC dá à linguagem HTML a possibilidade de suprir essa deficiência.

Seguem abaixo as instruções de como representar os elementos do padrão DC e seus qualificadores em HTML, segundo Cox, Miller e Powell (2000).

Dois códigos de marcação do HTML, <link> e <meta>, da marca <head>, são usados para registrar os metadados.

A marca <link> permite estabelecer a relação com outro documento, e no caso do DC indica a localização dos esquemas e as definições dos termos usados para a descrição do recurso e estabelece um prefixo para a descrição dos elementos, como:

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/">
```

Neste caso, define o prefixo DC para o padrão DC e faz a ligação para o esquema e para as definições de elementos localizado em <http://purl.org/dc/elements/1.1/>.

A marca <meta> faz o registro dos metadados, através dos seguintes atributos:

- name: Nome do elemento de metadados
- content: Valor do elemento
- schema: Indica o esquema usado para o valor
- lang: Indica a linguagem natural
- dir: Indica o sentido do texto para o valor do elemento

Os principais atributos da marca <meta> para a descrição de um recurso são "name" e "content" e seguem a seguinte sintaxe:

```
<meta name="DC.elemento" content="valor">
```

onde *elemento* é um dos 15 elementos do padrão DC; *valor* é, como o próprio nome diz, o valor desse elemento, e DC é o prefixo definido na marca <link>.

Apesar da linguagem HTML não suportar diretamente os qualificadores do padrão DC, é possível incorporá-los acrescentando um ponto e o elemento de refinamento (ER) após o elemento:

```
<meta name="DC.elemento.ER" content="valor">
```

Outro atributo importante é o "scheme", que indica o esquema de codificação do valor do elemento, ou seja, se ele está contido

em uma lista de valores (vocabulário controlado) ou segue um esquema pré-definido.

Uma sintaxe complexa para representar os elementos DC em HTML pode ser expressa como:

```
<link
  rel="schema.DC"
  href="http://purl.org/dc/elements/1.1/"
  title="DCMES plus DCMI recommended qualifiers">
<meta name="DC.Elemento" content="valor sem padrão">
<meta name="DC.Elemento.ER" scheme="esquemaA"
  content="Valor codificado de acordo com o esquema
  esquemaA">
<meta name="DC.Elemento.ER" scheme="listaB"
  content="Valor selecionado da lista listaB">
<meta name="DC.Elemento.ER" lang="linguaC"
  content="Valor expresso na linguagem linguaC">
```

onde:

- Elemento é um dos 15 elementos do padrão DC,
- ER representa um elemento de refinamento do DC,
- esquemaA é um esquema de codificação de valores pré-definido,
- listaB é um vocabulário controlado,
- linguaC é um código da linguagem utilizada.

Podemos através dessa sintaxe representar outros padrões de metadados também, bastando para isso defini-lo na marca <link>.

Segue abaixo um exemplo da descrição de metadados de uma página Web em HTML (COX; MILLER; POWELL, 2000):

```
<link
  rel="schema.DC"
  href="http://purl.org/dc/elements/1.1/"
  title="DCMES plus DCMI recommended qualifiers">
<meta name="DC.Identifier"
  scheme="URI"
```

```

        content = "http://www.ukoln.ac.uk/metadata/resources/
        dc/datamodel/WD-dc-rdf/figure1.gif" >
<meta name="DC.Title"
    lang="en"
    content = "A simple RDF assertion">
<meta name="DC.Type"
    scheme="DCMIType"
    content = "image">
<meta name="DC.Date.created"
    scheme="W3CDTF"
    content = "1999-04-27">
<meta name="DC.Coverage.temporal"
    scheme="DCMIPeriod"
    content = "start=1999-04-27">
<meta name="DC.Creator"
    content = "Miller, Paul">
<meta name="DC.Creator"
    content = "Brickley, Dan">
<meta name="DC.Format.extent"
    content = "4033 bytes">
<meta name="DC.Format.media"
    scheme="IMT"
    content = "image/gif">
<meta name="DC.Relation.isVersionOf"
    lang="en"
    content = "Figure 1 from RDF Model and Syntax">
<meta name="DC.Relation.isVersionOf"
    scheme="URI"
    content = "http://www.w3.org/TR/REC-rdf-
    syntax/fig1.gif">

```

O exemplo acima mostra a utilização de elementos sem o uso de qualificadores, como Identifier, Title, Type e Creator, juntamente com elementos e qualificadores, como Date, Coverage, Format e Relation.

Pelo que foi mostrado neste capítulo, pode-se observar que o DCMI é uma iniciativa importante na busca de um padrão de metadados que possa facilitar a catalogação e a recuperação de recursos de informação da Web, principalmente pela dimensão que tomou, atingindo vários países do mundo e agregando profissionais de diversas áreas.

Com isso, o padrão de metadados DC torna-se um forte candidato a modelo na recuperação de informação nos domínios da Internet, como Bibliotecas, Museus, Área governamental, Meio Ambiente,

Publicação, Agricultura e outras. Essa amplitude de atuação pode ser observada pelo fato de o padrão DC ter sido traduzido para 25 línguas e formalmente adotado por sete governos (<http://www.dublincore.org/about/overview>).

O padrão DC pode ser representado por diferentes sintaxes, sendo comum a representação no formato HTML, RDF usando XML e em formatos genéricos do tipo Elemento = "valor", podendo seus elementos estar dentro do recurso ou separado deles.

Com relação a seus elementos, apesar de nenhum deles ser de uso obrigatório, é natural que se utilize um conjunto mínimo de elementos capazes de descrever de maneira adequada um recurso. O fato de esses elementos serem opcionais dá uma liberdade a quem utiliza o DC para poder determinar quais elementos são necessários para a descrição do recurso.

A simplicidade dos elementos é outro fator importante, pois permite uma descrição rápida e ao mesmo tempo eficiente do recurso, gerando uma característica que é um dos pontos fortes do padrão DC, a possibilidade da descrição por não especialistas em catalogação. Isso permite maior rapidez na descrição dos elementos, e bem como um acompanhamento do crescimento da Internet.

A extensibilidade que o padrão oferece possibilita que, em determinado domínio de utilização, novos elementos possam ser acrescentados para a descrição do recurso, sem interferir na descrição através dos elementos DC. Isso dá uma liberdade grande para que o recurso possa conter elementos não definidos no DC, mas vitais para a sua descrição e conseqüentemente para sua busca e recuperação.

No próximo capítulo, descrevemos projetos que utilizam metadados como ferramenta de catalogação e recuperação de informação, com o objetivo de identificar os padrões utilizados, como estão sendo

utilizados e verificar em quais projetos o padrão DC está sendo utilizado e como especificamente.

## **Capítulo 4**

# **Projetos que utilizam Metadados**

A utilização de metadados, juntamente com as novas tecnologias aplicadas ao tratamento da informação, tem sido uma solução para a documentação, a pesquisa e a recuperação de informações, principalmente aquelas armazenadas eletronicamente, como as disponíveis na Internet.

A diversidade das informações armazenadas gera uma variedade de conteúdos disponíveis eletronicamente, cada um com suas características específicas. Essa diversidade também leva à existência de vários padrões de metadados para descreverem essas informações, cada um com sua especificidade, mas também com características comuns.

Mas como os padrões de metadados estão sendo utilizados na prática? Que padrões estão sendo adotados? Onde estão sendo utilizados? Como o padrão DC está sendo utilizado? Estas são perguntas que surgiram no desenvolvimento da pesquisa.

Com o objetivo de respondê-las, realizamos neste capítulo uma análise de projetos que utilizam metadados para a documentação, a busca e a recuperação de informações armazenadas em meio eletrônico.

#### **4.1 - Metodologia para seleção dos projetos**

Realizamos uma busca exaustiva em fontes de pesquisa até dezembro de 2001, quando selecionamos, entre os resultados obtidos, projetos e aplicações, na língua portuguesa ou inglesa, que utilizam metadados como ferramenta para documentação, pesquisa e recuperação de informação . Utilizamos como palavra-chave metadados, sem especificar um padrão, um tipo de informação específica ou uma área de interesse.

As fontes de pesquisa incluíram publicações científicas nas áreas de Ciência da Informação e Exatas. Com o objetivo que verificar o

que estava disponível na Web, também foi realizada uma pesquisa utilizando um site de busca comercial, disponível na Internet.

As fontes pesquisadas foram as seguintes:

- Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) / Revista Ciência da Informação: [www.ibict.br](http://www.ibict.br) e link para a Revista. A Revista Ciência da Informação é uma das revistas nacionais mais conceituadas na área de Ciência da Informação.
- Programa Biblioteca Eletrônica (ProBE): [www.probe.br](http://www.probe.br). O ProBE oferece acesso a diversas revistas científicas internacionais, e para Metadados, que possui pouco material no Brasil, isso é muito importante (<http://www.probe.br/probe-hist.htm>).
- “Scientific Electronic Library Online” (SciELO): [www.scielo.br](http://www.scielo.br). O Modelo SciELO é o produto da cooperação entre a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), o Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (BIREME) e instituições nacionais e internacionais relacionadas com a comunicação científica e editores científicos, que disponibiliza periódicos científicos na Internet ([http://www.scielo.org/model\\_p.htm](http://www.scielo.org/model_p.htm)).
- “Institute of Electrical and Electronics Engineers” (IEEE): [www.ieee.org](http://www.ieee.org). O IEEE é um instituto conceituado e que possui publicações importantes da área engenharia elétrica, computação e tecnologia.
- Internet, através do site de busca Google: [www.google.com.br](http://www.google.com.br). O Google é o site busca mais completo e mais utilizado atualmente para busca de informações na Internet.

## **4.2 - Projetos selecionados**

Segue abaixo uma descrição resumida dos projetos selecionados e a seguir um quadro comparativo de cada aplicação com suas respectivas variáveis em ordem cronológica:

### **1 - Sistema de Gerenciamento de Informação**

#### **“Stennis” - SIMS**

O SIMS, descrito por Saack-Giguette e Lopez (1993), é um projeto que tem como objetivo tratar (processar, catalogar, arquivar e acessar) os dados coletados pelo Programa Aéreo do Centro Espacial Stennis, que coleta esses dados através de dois Scanners, o TIMS e o CAMS, para um acesso rápido e eficiente às informações armazenadas.

Devido ao grande volume de dados que devem ser tratados, o SIMS utiliza metadados e Inteligência Artificial para o acesso eficiente aos arquivos de dados armazenados. A linguagem utilizada no projeto é o Prolog.

O padrão de metadados utilizado foi desenvolvido especificamente para o projeto.

### **2 - Modelo para composição de pratos culinários**

O Modelo de metadados para composição de pratos culinários, descrito por Klensin (1995), é proposto para tratamento de composição de pratos de comida, possibilitando a troca de informações em nível mundial.

Esses dados são utilizados para estudos de dieta, planejamento de menu, estudos epidemiológicos comparativos e uma série de outros propósitos relacionados com a promoção da saúde pública.

Para tratar as complicações com a descrição de pratos culinários, como terminologia, identificação de componentes, unidades de medidas etc., a representação das informações utilizando metadados é essencial para resolver esses problemas, e um padrão próprio foi escolhido, para sua representação através da linguagem de marca Standard Generalized Markup Language (SGML).

### **3 - “Leicester University Metadata Project” (LUMP)**

O Projeto, apresentado por Madsen, Foog e Ruggles (1994), apresenta um modelo de metadados, que tem como objetivo armazenar informações sobre Banco de Dados autônomos distribuídos e heterogêneos, que são acessados por protocolos como WWW, Gopher, X.500, Wide Area Information Service (WAIS) e outros.

Utiliza a linguagem de inteligência artificial Prolog para a construção de um Banco de Dados de questões, o que facilita futuras pesquisas, juntamente com uma linguagem própria para a realização das pesquisas de usuários no sistema, o “Metadata Query Language” (MQL).

Utiliza um padrão de metadados próprio, que proporciona a integração de diferentes Bancos de Dados heterogêneos e distribuídos pela Internet, de forma transparente para o usuário, facilitando suas pesquisas e dando ao sistema características de simplicidade, flexibilidade, extensibilidade (integração de vários tipos de recursos) e integração (de sistemas heterogêneos).

### **4 - Netskills**

O Netskills, descrito por Ingham, Caughey e Little (1997), é um projeto que oferece treinamento em serviços de rede para acadêmicos e que incorporou o padrão DC com o objetivo de facilitar a indexação e a pesquisa das informações.

Utiliza uma ferramenta chamada W3Object, que é específica para a administração de serviços avançados na Web. O W3Object utiliza a tecnologia distribuída de objetos, ou seja, os recursos disponíveis na Web são tratados como objetos.

Os elementos acrescentados ao padrão DC, para atender ao projeto, são: endereço eletrônico, fone, endereço, URL e número do Fax. As pesquisas ao Bando de Dados são realizadas através de uma interface Web.

## **5 - Sistema baseado no padrão "Semantic Header"**

O Sistema, apresentado por Desai (1997), foi desenvolvido com o objetivo de atender à crescente demanda de informações disponíveis na Internet e à necessidade de colocar toda essa informação à disposição das pessoas, visto que os mecanismos atuais de busca não satisfazem plenamente as necessidades dos usuários, pois os resultados são geralmente parciais.

A utilização de metadados para a indexação, juntamente com interfaces gráficas para a pesquisa e recuperação das informações, é a solução apresentada por Desai (1997).

Entre as vantagens para a utilização de metadados estão o baixo custo da indexação (pois não há necessidade de ser realizada por um profissional), a descrição semântica do recurso, a recuperação pelo conteúdo, a possibilidade de expressar dependência semântica (diminuir redundâncias) e o acesso aos metadados antes de acessar o recurso propriamente dito.

Foram estudados os padrões DC e o Semantic Header para a definição do padrão a ser utilizado.

A opção foi pelo Semantic Header, pelo fato de ele conter elementos que melhor descrevem os recursos, como "abstract" e anotações, que não estão descritos especificamente no padrão DC (DESAI,

1997). Deve-se observar, entretanto, que o padrão DC possui o elemento Descrição, que descreve o conteúdo do recurso e pode conter o “abstract”, contrariando a observação feita por Desai (1997).

A partir dessa escolha, foram desenvolvidas as interfaces para a entrada e a pesquisa das informações, utilizando um sistema especialista que realiza todas as consistências e usa termos padronizados para alguns elementos. A alteração dos dados é realizada somente pelo autor/criador do recurso.

As informações são armazenadas em um Sistema de Bancos de Dados Distribuídos e replicados de acordo com o assunto ou proximidade do recurso, possibilitando um melhor gerenciamento e acesso dos usuários.

## **6 - “Gateway to Educacional Material” – GEM**

O GEM, descrito por Sutton (1998), é um projeto desenvolvido pela Universidade Syracuse dos EUA a pedido do então Presidente Bill Clinton, e tem como objetivo organizar materiais de pesquisa na área de educação encontrados em sites na Internet, ou seja, catalogar endereços da Web que contenham informações da área de Educação para o acesso de estudantes e professores.

Para atingir esse objetivo, foi definido que seria necessário utilizar um padrão de metadados rico semanticamente. O padrão DC serviu de base para essa definição, sendo acrescentado mais nove elementos para se chegar ao padrão adequado: audience, cataloging, duration, essential, resources, grade, pedagogy, quality e standards.

Cada membro do consórcio GEM é responsável por criar e armazenar os metadados de seu próprio recurso, utilizando um editor de textos comum ou uma aplicação desenvolvida em JAVA, o GEM Cat. Os metadados podem ser descritos no próprio recurso HTML do site ou armazenados separadamente do recurso. Essas informações geram um

índice nos sites locais. Periodicamente esses índices locais são agregados em um banco de dados único, na “ERIC Clearinghouse on Information and Technology”. Dois mecanismos de busca a essas informações agregadas estão disponíveis: um acesso público através de uma interface GEM e outro acesso através do protocolo Z39.50.

## **7 - Nordic Web Index – NWI**

O NWI, descrito por Ardo e Lundberg (1998), é um sistema aberto para identificação e indexação de toda informação disponível na Web nos domínios Nórdicos (Dinamarca, Finlândia, Islândia, Noruega e Suécia). Está distribuído em computadores espalhados por esses países e ligados na Internet, o que o leva a ter a característica de escalabilidade.

Com o objetivo de utilizar um padrão de metadados que fosse aberto e de acesso livre, optou-se pelo padrão DC, acrescido de alguns elementos, para atender ao NWI, juntamente com o protocolo Z39.50 para a troca de informações. Os elementos acrescidos ao padrão DC não foram citados por Ardo e Lundberg (1998).

Para a implementação do projeto, adotaram uma política de utilização de software livre.

O sistema possui três componentes:

- 1- um robô de busca na Internet (“harvester”), que faz a coleta da informações;
- 2- um banco de dados, para facilitar a pesquisa das informações;
- 3- uma interface de usuário, para tornar a informação disponível.

## **8 - "Alexandria Digital Library" (ADL)**

O ADL, descrito por Hill et al (1999), é um projeto para pesquisa em banco de dados, focado em informações georeferenciadas/geoespaciais (mapas, fotos aéreas, imagens e dados de lugares em particular), e que possui dados de várias coleções distribuídas em Bibliotecas Digitais.

O conceito de coleções em uma Biblioteca Digital refere-se a um grupo de objetos que não necessariamente precisam estar presentes fisicamente na Biblioteca.

A chave para a integração de uma variedade de coleções em uma Biblioteca Digital são os metadados, pois possibilitam a representação das características inerentes e contextuais da coleção.

As características inerentes são aquelas que podem ser obtidas por computador, através da análise do conteúdo (cobertura temporal, tipo da informação, formato etc.). Por sua vez, as características contextuais são aquelas fornecidas pelo responsável da coleção (título, responsável, escopo, propósito, esquema de metadados, contato etc.).

O Projeto ADL tem vários propósitos no desenvolvimento das coleções de metadados:

- em um formulário XML, os metadados registram a coleção com um interface cliente para o usuário, desenvolvida em JAVA,
- em um formulário HTML, é usado para a documentação do usuário e eventualmente será utilizado para descrever a coleção para os agentes de busca da Internet,
- usado para o gerenciamento interno das coleções.

O padrão de metadados utilizado para descrever as coleções é próprio, o Esquema de metadados ADL, baseado no Federal Geographic Data Committee's (FGCD) com extensões do MARC.

### **9 - Coleção de Moda Histórica Digitalizada (CMHD)**

O Projeto, descrito por Zeng (1999), tem como objetivo desenvolver um catálogo para a coleção digitalizada de moda histórica do século XVIII até XX do Museu da "Kent State University", nos Estados Unidos. Para atingir essa meta, analisaram-se formatos de metadados para serem aplicados à coleção, que pudessem proporcionar acesso "on-line" para as descrições e para as imagens digitalizadas.

O museu conta com cerca de 20.000 objetos dessa época, sendo a maior parte constituída de vestidos regionais e de moda.

O acesso digital tem as vantagens de preservar o material original e permitir que pesquisadores, estudantes e o público em geral possam ter acesso visual a essas coleções sem o perigo de danificá-las. Essas vantagens fizeram o número de sites de museus crescerem na WWW.

Foi decidido analisar padrões que pudessem descrever os objetos em três dimensões, como os do museu, e que também já tivessem sido estudados por profissionais da informação.

Foram selecionados três padrões para a análise, o Anglo-american Cataloguing Rule (AACR) em uso com o formato USMARC (hoje conhecido como MARC21), o DC e o Visual Resources Association (VRA). Foram considerados para a seleção os critérios de compatibilidade (com outras instituições) e simplicidade (para um fácil acesso de usuários, sem a necessidade de "browsers" específicos ou "plug-ins" adicionais).

O padrão escolhido foi o VRA modificado, sendo acrescentado cinco novos elementos para a descrição das informações digitalizadas. O VRA modificado atende melhor à descrição de dados em

três dimensões, como as do museu, e possibilitará também a exportação de dados para outros museus e também para os padrões DC e USMARC.

### **10 - Modelo do IME**

O modelo, apresentado por Garcia (1999), foi desenvolvido no Instituto Militar de Engenharia (IME) com o objetivo de possibilitar a representação e a recuperação de imagens, como fotografia, pintura ou uma gravura qualquer.

Propõe utilizar, para o armazenamento das informações, um Banco de Dados Orientado a Objetos e, para representá-las, descritores associados aos dados contidos nas imagens, ou seja, metadados.

A utilização de metadados tem como objetivo descrever não só as informações técnicas, mas também aquelas relacionadas ao conteúdo semântico das imagens.

Foram definidos descritores para representar as imagens considerando dois padrões de metadados existentes, o DC e o SAIF (Spatial archive and Interchange Format), por serem largamente utilizados e por poderem representar imagens.

Foi desenvolvido um protótipo, no qual foram utilizados o SGBDOO Jasmine, isto é, um Sistema de Banco de Dados Orientado a Objetos e multimídia para o armazenamento das informações, e ferramentas de implementação do próprio Jasmine. As páginas de consulta e resposta foram desenvolvidas em HTML.

### **11 - "Reserch, Education, Economic Information System" – REEIS**

O REEIS, descrito por Cortez (1999), foi desenvolvido a pedido do Departamento de Agricultura dos EUA para proporcionar um

mecanismo de localização de programas, projetos e pesquisa com foco em comida, agricultura, recursos naturais e desenvolvimento rural.

O projeto usa metadados com o objetivo de documentar e guiar os usuários na seleção de um banco de dados de interesse. Utiliza vocabulário controlado padronizado na documentação e na pesquisa do banco de dados desejado.

O padrão utilizado é o DC, acrescido de alguns elementos que não são citados no texto de Cortez (1999). As principais justificativas para a utilização do DC são de que esse padrão possui um conjunto de elementos projetados para recursos da WWW e um vocabulário uniforme, possibilitando uma eficiente recuperação das informações.

## **12 - Rural Mídia**

O Rural Mídia de 1997, descrito por Souza, Catarino e Santos (2000), é um Banco de Imagens (fotos, ícones e gravuras) desenvolvido pela Embrapa Informática Agropecuária, que utiliza metadados para a documentação e recuperação dessas imagens.

Para descrever adequadamente as informações técnicas e semânticas contidas em uma imagem, a Embrapa optou pela utilização do padrão de metadados DC acrescido de 3 elementos: categoria, acesso e contato.

O acesso às informações, que são armazenadas em um banco de dados desenvolvido em MS-Access 97, é realizado através de páginas Web. São utilizados o banco de dados SQL para formular a consulta e o programa Apache 1.3 para as pesquisas. A entrada de dados é feita também através da Web.

### **13- Biblioteca Digital Brasileira (BDB)**

A BDB, descrita por Marcondes e Sayão (2001), é um projeto em Ciência e Tecnologia, desenvolvido pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), que tem como objetivo fomentar mecanismos de publicação eletrônica na Internet pela comunidade acadêmica brasileira de textos completos, teses e artigos de periódicos na área de Ciência e Tecnologia, além de possibilitar a interoperabilidade desses recursos, heterogêneos e distribuídos, através do acesso via um portal único, sem a necessidade de navegar e consultar site a site.

Para atingir esses objetivos, é importante atentar para a questão da interoperabilidade entre as bibliotecas digitais, o que possibilita o acesso rápido e integrado a esses recursos, armazenados em diferentes servidores de rede espalhados pela Internet.

O modelo de interoperabilidade adotado pela BDB aproxima-se bastante dos modelos do portal da Networked Digital Library of Theses and Dissertations -NDLTD (<http://rocky.dlib.vt.edu/~etdunion/cgi-bin/index.pl>) e do Cross Archive Searching Service – ARC (<http://arc.cs.odu.edu/>). Os sistemas realizam “harvesting” de metadados de provedores de dados, realizando consultas diretamente nesses provedores e alimentando uma base de dados centralizada de metadados, onde também são realizadas as pesquisas.

O “harvesting” consiste em uma coleta de dados nos sites que contêm as informações, utilizando o protocolo Z39.50, tendo a BDB em seu servidor um programa cliente Z39.50 para acesso aos servidores de informações que possuem um programa servidor Z39.50.

A BDB utiliza o padrão DC, por ser uma referência emergente nesta área, por ser resultado de um intenso trabalho de discussão e padronização em nível internacional do DCMI e por ser utilizado em diferentes sistemas.

Utilizam os padrão DC acrescido de alguns elementos para suportar características especiais em alguns documentos, e que não foram citados.

### **4.3 - Variáveis analisadas**

Para podermos analisar os projetos e poder responder às perguntas levantadas no início do capítulo, escolhemos seis variáveis para análise, levando-se em consideração o que aparece de comum entre os projetos e os objetivos propostos neste capítulo.

As variáveis selecionadas foram:

1. Local: descreve o nome da instituição onde o projeto foi desenvolvido ou está sendo utilizado e tem o objetivo de identificar o tipo da instituição, se privada ou pública;
2. País: descreve o país onde o projeto foi desenvolvido ou onde está sendo utilizado e tem por objetivo identificar onde metadados estão sendo utilizados no contexto mundial;
3. Ano: descreve o ano de publicação do artigo, com a proposta de identificar quando foi disponibilizado para a comunidade;
4. Área: descreve em que área de atuação está sendo aplicado o projeto, com o objetivo de identificar as áreas de utilização de metadados;
5. Tipo: descreve a espécie da informação que está sendo tratada, permitindo analisar o tipo do recurso descrito e armazenado em meio eletrônico;
6. Padrão: descreve qual padrão de metadados está sendo utilizado, e nos permite analisar quais padrões estão

sendo utilizados para descrever os recursos armazenados e, mais especificamente, onde o padrão DC está sendo utilizado e como.

Optamos por não utilizar nenhum vocabulário controlado para as variáveis local, área e tipo, e utilizamos as descrições apontadas nos projeto para termos com mais clareza o tipo de informação tratada.

#### **4.3.1 - Quadro comparativo das variáveis selecionadas**

O quadro a seguir mostra as informações levantadas de cada aplicação, com relação às variáveis definidas para a análise.

Projetos	Local	País	Ano	Área	Tipo	Padrão
<b>SIMS</b>	Nasa	EUA	1993	Pesquisa Espacial	Imagem (fotos da Terra)	Próprio
<b>Modelo para composição de pratos culinários</b>	"United Nations University"	EUA	1995	Alimentos	Composição de nutrientes	Próprio
<b>LUMP</b>	"Leicester University"	Reino Unido	1994	Web	Documentos e arquivos	Próprio
<b>Netskills</b>	"Newcastle University"	EUA	1997	Web	Endereços da Web	DC modificado
<b>Sistema baseado no padrão "Semantic Header"</b>	"Concordia University"	Canadá	1997	Biblioteca Digital	Documentos na Web	Semantic Header modificado
<b>GEM</b>	"Syracuse University"	EUA	1998	Educação	Endereços da Web (sites na área de Educação)	DC modificado
<b>NWI</b>	"Lund University Library"	Países Nórdicos	1998	Web	Endereços da Web (Domínios Nórdicos)	DC modificado
<b>ADL</b>	"University of California"	EUA	1999	Biblioteca Digital	Informações georeferenciadas (mapas, imagens, fotos etc.)	Próprio
<b>CMHD</b>	Museu da "Kent State University"	EUA	1999	Coleção de moda histórica	Texto e imagem	VRA Core modificado
<b>Modelo para representação e recuperação de imagem</b>	Instituto Militar de Engenharia	Brasil	1999	Biblioteca Digital	Imagem	Próprio
<b>REEIS</b>	"U.S. Department of Agriculture" (USDA)	EUA	1999	Agricultura, alimentos, recursos naturais e desenvolvimento rural	Documentos (contidos em Banco de Dados)	DC modificado
<b>Rural Mídia</b>	Embrapa	Brasil	2000	Rural	Imagem (foto, ícones e gravuras)	DC modificado
<b>BDB</b>	IBICT	Brasil	2001	Biblioteca Digital	Documentos eletrônicos (teses, artigos de periódicos, trabalhos etc)	DC modificado

Quadro 3 – Variáveis de análise dos projetos selecionados

#### **4.4 - Análise dos projetos selecionados**

Todos os projetos analisados buscam, com a utilização de metadados, tornar a descrição dos recursos eletrônicos adequada para que a pesquisa e a recuperação das informações armazenadas nestes recursos possam ser rápidas e principalmente eficientes.

Apesar dos vários motivos apontados para a utilização de metadados, tais como controle do grande volume de dados, disponibilização, catalogação, organização, uso de padrões abertos, indexação, intercâmbio, gerenciamento e descrição de informações, o objetivo principal é sempre tornar a informação disponível e de fácil acesso para o usuário final, de forma que o mesmo possa ter acesso aos dados desejados.

Quando se define o que são metadados e quais são os objetivos com sua utilização, vamos ao encontro das necessidades apontadas nos projetos analisados.

Com exceção do NWI (ARDO; LUNDBERG, 1998), que utiliza um robô de busca automático nos sites da Internet dos países Nórdicos ("harvester") para realizar a alimentação do Bando de Dados que contém a descrição dos metadados, os outros projetos, que especificaram como é feita a alimentação das informações de metadados, utilizam a Web com suas ferramentas e linguagens (HTML, JAVA, XML e W3Object) para criar os mecanismos (interfaces com o usuário) com os quais o próprio usuário, "dono" da informação, possa fazer a descrição do recurso.

Essa característica de simplicidade na descrição de recursos eletrônicos, que torna possível ao responsável pelo recurso descrevê-lo, é um fator importante apontado em várias aplicações para a utilização de padrões de metadados.

Em todos os projetos analisados, o mecanismo de acesso às informações disponíveis é a Web. As ferramentas e as linguagens utilizadas para a construção dos mecanismos de acesso à Web são as mesmas utilizadas comumente para a construção de páginas, ou seja, HTML, JAVA e XML.

No caso de XML, a vantagem que a linguagem permite é a criação de “tags” que representem os elementos definidos, dando flexibilidade na escolha do padrão de metadados a ser utilizado.

Abaixo segue a análise individual de cada variável selecionada:

**- Local**

<b>Projetos</b>	<b>Local</b>
<b>SIMS</b>	Nasa
<b>Modelo para composição de pratos culinários</b>	“United Nations University”
<b>LUMP</b>	“Leicester University”
<b>Netskills</b>	“Newcastle University”
<b>Sistema baseado no padrão “Semantic Header”</b>	“Concordia University”
<b>GEM</b>	“Syracuse University”
<b>NWI</b>	“Lund University Library”
<b>ADL</b>	“University of California”
<b>CMHD</b>	Museu da “Kent State University”
<b>Modelo para representação e recuperação de imagem</b>	Instituto Militar de Engenharia
<b>REES</b>	“U.S. Department of Agriculture” (USDA)
<b>Rural Mídia</b>	Embrapa
<b>BDB</b>	IBICT

Quadro 4 – Variável local dos projetos selecionados

Os projetos analisados mostram a utilização de metadados pelo setor público. São todas instituições públicas, cabendo às Universidades o maior número de projetos.

Quando realizamos a pesquisa pelo site de busca Google ([www.google.com](http://www.google.com)), notamos que poucas empresas do setor privado utilizam metadados. Uma delas é a Microsoft, que utilizará XML para intercâmbio de metadados ([http://www.microsoft.com/brasil/pr/md\\_coalition.htm](http://www.microsoft.com/brasil/pr/md_coalition.htm)).

### - País

Projetos	País
SIMS	EUA
Modelo para composição de pratos culinários	EUA
LUMP	Reino Unido
Netskills	EUA
Sistema baseado no padrão "Semantic Header"	Canadá
GEM	EUA
NWI	Países Nórdicos
ADL	EUA
CMHD	EUA
Modelo para representação e recuperação de imagem	Brasil
REEIS	EUA
Rural Mídia	Brasil
BDB	Brasil

Quadro 5 – Variável país dos projetos selecionados

A utilização de metadados é maior no exterior que no Brasil, mostrando que o uso de padrões de metadados no Brasil ainda está em processo inicial. Os dois projetos brasileiros analisados referem-se à utilização de metadados com imagens.

Em uma pesquisa pelo site de busca Google ([www.google.com](http://www.google.com)), utilizando metadados como termo de busca, tivemos como resultado 1.950 entradas nos sites do Brasil. Quando realizamos a mesma pesquisa em sites da Web, utilizando o termo "metadata", tivemos 1.350.000 entradas como resultado.

**- Ano**

<b>Projetos</b>	<b>Ano</b>
<b>SIMS</b>	1993
<b>Modelo para composição de pratos culinários</b>	1995
<b>LUMP</b>	1994
<b>Netskills</b>	1997
<b>Sistema baseado no padrão "Semantic Header"</b>	1997
<b>GEM</b>	1998
<b>NWI</b>	1998
<b>ADL</b>	1999
<b>CMHD</b>	1999
<b>Modelo para representação e recuperação de imagem</b>	1999
<b>REEIS</b>	1999
<b>Rural Mídia</b>	2000
<b>BDB</b>	2001

Quadro 6 – Variável ano dos projetos selecionados

É interessante notar que, já em 1993, a NASA se preocupava em criar um padrão para descrição de dados eletrônicos obtidos de seus satélites, possibilitando assim melhor pesquisa e recuperação das informações.

No Brasil, os projetos publicados são mais recentes: 2000 para o Rural Mídia e 1999 para o Modelo do Instituto Militar de Engenharia.

**- Área**

<b>Projetos</b>	<b>Área</b>
<b>SIMS</b>	Pesquisa Espacial
<b>Modelo para composição de pratos culinários</b>	Alimentos

Quadro 7 – Variável área dos projetos selecionados

Projetos	Área
LUMP	Web
Netskills	Web
Sistema baseado no padrão "Semantic Header"	Biblioteca Digital
GEM	Educação
NWI	Web
ADL	Biblioteca Digital
CMHD	Coleção de moda histórica
Modelo para representação e recuperação de imagem	Biblioteca Digital
REEIS	Agricultura, alimentos, recursos naturais e desenvolvimento rural
Rural Mídia	Rural
BDB	Biblioteca Digital

Quadro 7 – Variável área dos projetos selecionados (cont.)

As áreas onde são aplicados metadados para a documentação e a recuperação de informação são muito distintas, mostrando assim a necessidade existente, em todas as áreas que se preocupam em tratar a informação, de organizar essas informações para a sua melhor recuperação e utilização das mesmas.

As áreas de maior destaque foram as de Banco de Dados, Web e Agricultura. Outras que apareceram na pesquisa foram: Pesquisa Espacial, Educação, Alimentos e Coleção de Moda Histórica.

#### - Tipo

Projetos	Tipo
SIMS	Imagem (fotos da Terra)
Modelo para composição de pratos culinários	Composição de nutrientes
LUMP	Documentos e arquivos
Netskills	Endereços da Web

Quadro 8 – Variável tipo dos projetos selecionados

Projetos	Tipo
Sistema baseado no padrão "Semantic Header"	Documentos na Web
GEM	Endereços da Web (sites na área de Educação)
NWI	Endereços da Web (Dominós Nórdicos)
ADL	Informações georeferenciadas (mapas, imagens, fotos, etc.)
CMHD	Texto e imagem
Modelo para representação e recuperação de imagem	Imagem
REEIS	Documentos (contidos em Banco de Dados)
Rural Mídia	Imagem (foto, ícones e gravuras)
BDB	Documentos eletrônicos (teses, artigos de periódicos, trabalhos etc)

Quadro 8 – Variável tipo dos projetos selecionados (cont.)

Os tipos de dados a serem tratados utilizando metadados concentram-se em informações de documentos, imagem, endereços Web e informações multimídia (som, imagem e texto).

Nos projetos do Brasil, a utilização de metadados foi para informações do tipo imagem.

#### - Padrão

Projetos	Padrão
SIMS	Próprio
Modelo para composição de pratos culinários	Próprio
LUMP	Próprio
Netskills	DC modificado
Sistema baseado no padrão "Semantic Header"	Semantic Header modificado
GEM	DC modificado
NWI	DC modificado
ADL	Próprio
CMHD	VRA Core modificado

Quadro 9 – Variável padrão dos projetos selecionados

<b>Projetos</b>	<b>Padrão</b>
<b>Modelo para representação e recuperação de imagem</b>	Próprio
<b>REEIS</b>	DC modificado
<b>Rural Mídia</b>	DC modificado
<b>BDB</b>	DC modificado

Quadro 9 – Variável padrão dos projetos selecionados (cont.)

Dos projetos analisados, seis utilizam o padrão DC, acrescido de alguns elementos para a descrição das informações a serem tratadas. Isso mostra a aceitação que existe em cima do padrão DC, pois além desses projetos, outros três analisaram o padrão DC para decisão sobre o padrão que deveria ser utilizado. Desses três, o Modelo do IME (GARCIA, 1999) optou pela utilização de um padrão próprio; a Coleção de Moda Histórica Digitalizada (ZENG, 1999) optou pelo padrão VRA Core modificado; e o Sistema baseado no padrão "Semantic Header" (DESAI, 1997), pelo padrão Semantic Header modificado.

A utilização de padrões próprios também é bastante significativa, totalizando cinco projetos. O que se observa é que em três desses projetos, o Modelo para composição de pratos culinários, o SIMS (SAACK-GIGUETTE; LOPEZ JR., 1993) e o LUMP (MADSEN; FOOG; RUGGLES, 1994), publicados antes de 1996, a definição do padrão não se baseia em um padrão existente, como ocorre com ADL (HILL et al, 1999), que se baseou no FGDC content standard com MARC e com o Modelo do IME (GARCIA, 1999), que utilizou com base os padrões DC e o "Spatial Archieve and Interchange Format" (SAIF).

No Projetos CMHD (ZENG, 1999), sistema baseado no padrão "Semantic Header" (DESAI, 1997) e Modelo do IME (GARCIA, 1999), o DC não foi utilizado como padrão, mas foi citado como fonte de pesquisa e comparação com outros padrões para se definir a melhor solução.

Dos projetos apresentados após o surgimento do padrão DC, apenas no ADL (HILL et al, 1999), que é uma publicação de 1999, o padrão DC não é citado como referência para a definição do padrão a ser utilizado no projeto.

A característica do DC, de ser um padrão aberto, ou seja, que permite o acréscimo de novos elementos para a descrição dos recursos, é um dos fatores que levam ao seu uso freqüente como padrão escolhido ou como referência. Essa abertura é apontada em CMHD (ZENG, 1999) como um problema, pois dá ao DC uma característica de alta generalidade. Essa generalidade é mostrada nos projetos que utilizam o DC como padrão, pois todos eles acrescentaram novos elementos ao DC para representar melhor as informações desejadas, como é mostrado no quadro 2. Nesse caso, a generalidade é colocada como uma vantagem para a adoção do padrão DC.

Zeng (1999) cita uma característica oposta a essa do DC no padrão MARC, colocando-o como um formato com vários elementos. O problema gerado por essa especificidade é a difícil catalogação dos recursos por usuários não-especialistas, sendo necessário para a atividade de definição dos elementos um especialista em catalogação. Por essas características do DC e do MARC, Zeng (1999) optou pela utilização do padrão VRA Core modificado para a descrição de dados que contém imagem e texto.

Existe também no artigo de Zeng (1999) uma preocupação em analisar cuidadosamente os padrões MARC e DC, com o objetivo de criar mecanismos que possibilitem a migração dos dados do padrão VRA modificado para esses dois padrões, possibilitando assim o intercâmbio de informações entre instituições que utilizam esses padrões.

Essa preocupação aparece também em outros projetos, mostrando que na utilização de metadados deve-se ter a preocupação com o intercâmbio de informações, não somente em relação aos mesmos

padrões utilizados, mas também em relação àqueles com que possa ter equivalências.

Para Kerhervé (1997), citado em Barreto (1999), pesquisas desenvolvidas na área de metadados devem se dirigir à integração e à interoperação de padrões diferentes, levando a projetos e a pesquisas de modelos de metadados que possam ser adaptados a aplicações específicas, ao desenvolvimento de ferramentas para integração de padrões de metadados diferentes, e a gerenciadores de metadados extensíveis que permitam funções de acesso, transferência, descoberta ou análise para o desenvolvimento de aplicações específicas.

Os projetos que utilizam o padrão DC, também fazem referência à necessidade do uso de vocabulário controlado para a descrição de vários elementos, permitindo assim maior consistência na fase de descrição dos elementos e na realização das pesquisas.

Pode-se observar, nos projetos analisados, que os padrões de metadados estão sendo utilizados com sucesso, mostrando que, quando apresentados juntamente com os recursos computacionais existentes, são uma ferramenta bastante eficiente na documentação, na pesquisa e na recuperação de informações armazenadas em recursos eletrônicos.

Os padrões adotados são diversos, mas nota-se que o padrão DC é o mais utilizado, independente da área ou do tipo de informação de que se deseja tratar.

O padrão DC não tem sido utilizado no seu formato básico, ou seja, com seus 15 elementos. Por seu próprio conceito inicial, que é o de apresentar um conjunto mínimo de descritores, sempre é acrescido de alguns elementos, dependendo da aplicação desejada. Pode ser apontado como uma referência para a descrição de recursos eletrônicos, mas não na sua forma pura, ou seja, com os seus 15 elementos.

## **Capítulo 5**

### **Considerações finais**

Vivemos em um mundo cercado de informações por todos os lados, onde tudo pode ser traduzido em informação, tudo é portador de uma informação e tudo por ser transformado em conhecimento, desde o choro de um bebê que informa algum problema até um jornal que informa o que acontece no mundo.

Necessitamos organizar essas informações para torná-las disponíveis, possibilitando que as mesmas possam gerar conhecimento nos indivíduos, para que estes possam usar esse conhecimento para finalidades pessoais, decisões públicas e aplicações em diversas áreas, como política, cultura, de pesquisa, comercial e social.

A preocupação com o tratamento da informação é um dos principais pilares da Ciência da Informação. Os avanços e o desenvolvimento da Ciência e Tecnologia, em especial da informática e das telecomunicações, que criou a possibilidade do armazenamento de grande quantidade de informações em meio eletrônico, permeou o surgimento de uma das maiores fontes de informação da atualidade, a Internet, uma rede interligada de computadores espalhados pelo mundo.

A Internet colocou à disposição das pessoas diversas informações e o acesso quase instantâneo a elas. Isso possibilita transformar essa informação em conhecimento para a humanidade e com isso gerar novas informações.

Mas, como toda nova tecnologia, o armazenamento digital de informações e a Internet também têm seus problemas a serem resolvidos, e um deles é como descrever essa informação armazenada em meio eletrônico, para sua busca e recuperação.

Como é de consenso entre diversos profissionais da área de Ciência de Informação, a utilização de metadados juntamente com ferramentas adequadas é uma solução para a descrição de informações armazenadas em meio eletrônico, principalmente aquelas disponíveis na

Internet, atualmente a maior rede de informações distribuída e heterogênea do mundo.

A descrição através de metadados proporciona, entre outras coisas, qualidade tanto para a representação de um recurso, como para o resultado de uma pesquisa.

Apesar de o conceito de metadados ser anterior ao surgimento dos documentos digitais, sua utilização para a descrição de recursos armazenados em meio eletrônico, como aqueles disponíveis na Internet, é recente.

O DCMI, com sua origem em Ohio em 1995, é uma das primeiras iniciativas e hoje uma das mais importantes para o estudo de metadados, pois congrega profissionais de diversas áreas e de diversos países, preocupados com a questão do tratamento de recursos disponíveis na Web. Existem diversos braços do DCMI espalhados por muitos países do mundo.

No Brasil, a Sociedade da Informação (TAKAHASHI, 2000) e a proposta da Biblioteca Digital Brasileira (MARCONDES; SAYÃO, 2001) são as principais iniciativas em nível nacional, levantadas na pesquisa, na indicação e na utilização de metadados para a busca e na recuperação de informações armazenadas em meio eletrônico e disponíveis na Internet.

O número de publicações sobre metadados disponíveis no Brasil, seja em revistas científicas, livros ou sites, ainda é pequeno se compararmos com o que já existe internacionalmente.

A abordagem sobre metadados descrita neste trabalho busca preencher um pedaço desse vazio, na medida em que procura:

- mostrar seu histórico;
- os conceitos relacionados ao assunto;
- uma descrição detalhada do padrão DC com os objetivos de possibilitar sua utilização na descrição de

recursos disponíveis na Internet e viabilizar estudos de interoperabilidade envolvendo o padrão;

- mostrar projetos que utilizam metadados na prática.

Podemos verificar também que o número de definições para metadados é grande e dependente do domínio de sua utilização. A definição mais comum, "*dados sobre dados*", é muito restrita para o significado de metadados, pois este representa um conceito mais amplo na descrição de informações para busca e recuperação.

Por esse motivo, defendemos a definição de metadados como **um conjunto de elementos que descrevem as informações contidas em um recurso, com o objetivo de possibilitar sua busca e recuperação.**

Dos padrões de metadados analisados nos capítulos 3 e 4, verificamos que o padrão DC é o mais apropriado para a descrição de recursos disponíveis na Web, pois:

- possui uma comunidade internacional de pesquisadores envolvidos na pesquisa contínua de soluções, o DCMI;
- esse escopo dá ao DC um consenso internacional, ou seja, diversos países e comunidades na Internet o utilizam;
- esse consenso internacional permite ao DC possuir uma capacidade maior de interoperabilidade;
- sua simplicidade dá ao DC um entendimento semântico simples de seus elementos, facilitando sua utilização por não catalogadores ou pelo próprio dono do recurso;
- caracteriza-se pela extensibilidade, ou seja, elementos podem ser acrescentados para atender a uma necessidade específica;

- a divisão do DC em elementos e qualificadores (refinamento e esquema de qualificação) dá uma alta cobertura na sua utilização por diferentes comunidades na Internet;
- os diversos grupos de trabalho do DCMI, os Workshop realizados anualmente desde 1995 e um site bem estruturado e com várias informações disponíveis, possibilitam que o DC possa estar em contínua evolução e atender às iniciativas que o utilizam, acompanhando a constante mutação da Internet.

Mas qual seria a melhor solução, o melhor padrão de metadados para a descrição dos recursos disponíveis na Internet?

Apontamos três soluções e propomos qual a melhor resposta a essa pergunta:

#### 1- Uso de um padrão de metadados único

É pouco provável imaginar que a Internet possa ter um padrão único de metadados para descrever seus recursos, pois atualmente diversos padrões são utilizados, cada um com sua especificidade.

Ficaria quase impossível que esses recursos, com seus metadados já descritos, migrassem para um padrão único.

Essa solução também é contrária a um dos principais motivos que levou ao crescimento da Internet, o fato de não haver controle.

#### 2- Flexibilidade total

Seguindo o que rege a Internet, essa seria uma possibilidade, mas correríamos o risco de a descrição dos recursos por metadados tornar-se algo de difícil compreensão devido à amplitude que poderia alcançar.

Seria quase impossível que mecanismos de busca pudessem buscar essas informações e organizá-las de forma que os usuários pudessem realizar uma pesquisa.

### 3- Aproximação dos padrões: interoperabilidade

Como a adoção de um padrão único e a flexibilidade total para a descrição de recursos da Web parecem bastante improváveis, a solução mais adequada seria a que considera os diversos padrões existentes integrados de forma a atender a comunidades distintas, ou seja, propiciar interoperabilidade aos diversos padrões existentes.

Nesse sentido, estudos sobre interoperabilidade entre padrões distintos têm um papel importante na implementação da solução que nos parece mais viável, juntamente com a adoção de padrões que sejam administrados por instituições e comunidades sérias.

Como vimos em Cromwell-Kessler(1998) e em Barreto (1999), estudos sobre a interoperabilidade entre padrões distintos de metadados estão sendo realizados, mostrando ser este um caminho a ser seguido. Nessas pesquisas são estabelecidas as correspondências entre os elementos que compõem cada padrão.

Com o objetivo de prover a interoperabilidade entre os distintos padrões de metadados existentes, estudos sobre a arquitetura de metadados também têm tido papel importante, como a arquitetura Warwick, proposta em 1996 no 2º Workshop promovido pelo DCMI, e o RDF, uma recomendação da W3C de 1999. O enfoque de ambas as arquiteturas é a definição de uma estrutura para propiciar interoperabilidade entre aplicações, facilitando a troca de informações contidas nos recursos disponíveis na Web.

O uso dessas arquiteturas para o transporte de metadados, oriundos de padrões diferentes, maximiza a interoperabilidade entre eles, pois permite um melhor processamento das informações descritas por seus elementos (BARRETO, 1999).

Ferramentas, como os programas que realizam busca automática de metadados na Internet (“havester”) e protocolos de comunicação de dados, como o Z39.50, também contribuem para a interoperabilidade entre padrões de metadados.

Como vimos em alguns projetos estudados, mesmo aqueles que não adotaram o padrão DC para a descrição de seus recursos, realizaram uma análise do DC. Isso mostra que o DC pode ser apontado como um parâmetro inicial de análise para a descrição de recursos na Web, mas não único, por ter sido desenvolvido para atender às necessidades da Internet e ter uma comunidade forte, internacional e com profissionais de diversas áreas envolvidos no desenvolvimento do padrão, o DCMI.

Outro ponto de partida para uma comunidade definir um padrão de metadados para a descrição de seus recursos deve ser o estudo dos padrões e das experiências existentes em seu domínio.

Para futuros trabalhos na área de metadados, que poderiam ajudar no seu desenvolvimento e dariam um entendimento maior das questões que envolvem o assunto, podemos sugerir:

- Estudos mais aprofundados sobre a questão da interoperabilidade entre padrões distintos, pois esse parece ser o melhor caminho para o problema da descrição de recursos na Web;
- Como utilizar as arquiteturas de metadados para ajudar a resolver o problema da interoperabilidade;
- Aplicação de metadados como solução para a interoperabilidade em Bibliotecas Digitais, distribuídas pela Internet ou em servidores de dados distintos;
- Utilização da linguagem XML juntamente com a arquitetura RDF, para descrição de recursos por metadados em projetos e aplicações;

- Utilização específica do padrão Dublin Core em aplicações, projetos e pesquisas diretamente ligados à Web;
- Desenvolvimento de aplicações na área da computação para implementar ferramentas que utilizem metadados (padrões, arquiteturas, interfaces com o usuário etc.) na descrição, na busca e na recuperação de informações.
- Criação de uma linha de pesquisa que possibilite o desenvolvimento de pesquisas na área;
- Discussão dos conceitos relacionados a metadados nos cursos de graduação e pós-graduação em Ciência de Informação;
- Participação de pesquisadores brasileiros no DCMI, pois este é atualmente uma das iniciativas internacionais mais sérias na área de metadados. Isso possibilitaria a participação em grupos de trabalhos e um acompanhamento dos avanços que estão acontecendo, não só do DC, como também sobre metadados.

Dessa maneira, esperamos que este trabalho possa contribuir na área da Ciência da Informação como uma referência histórica-conceitual, abrindo novas discussões para a continuidade de um tema importante para a descrição de recursos disponíveis em meio eletrônico: metadados.

## Referências

- ALVESTRAND, H. Tags for the identification of languages. In: INTERNET ENGINEERING TASK FORCE, 1995. Disponível em: <<http://www.ietf.org/rfc/rfc1766.txt>>. Acesso em: 6 fev. 2002.
- AN OVERVIEW of the Dublin Core Metadata Initiative. In: DUBLIN CORE METADATA INITIATIVE. Disponível em: <<http://dublincore.org/about/overview>>. Acesso em: 5 fev. 2002.
- ARDO, A.; LUNDBERG, S. A regional distributed WWW search and indexing service: the desire way. *Computer Networks and ISDN Systems*, v. 30, p. 173-183, 1998.
- ATAÍDE, M. E. M. O lado perverso da globalização na sociedade da informação. *Ciência da Informação*, Brasília, v. 26, n. 3, p. 268-270, 1997.
- BACA, M. (Ed.). *Introducción a los metadatos vías a la información digital*. Traducido al español por Marisol Jacas-Santoll. Los Angeles, CA: J. Paul Getty Trust, 1998. 43 p.
- BARITE, M. Organización del conocimiento: un nuevo marco teórico-conceptual en bibliotecología y documentacion. In: CARRARA, K. (Org.). *Educação, universidade e pesquisa*. São Paulo: FAPESP, 2001. p. 35-46.
- BARRETO, A. A. A questão da Informação. *São Paulo em Perspectiva*, v. 8, n. 4, p. 3-8, 1994.
- BARRETO, A. A. A oferta e a demanda da informação: condições técnicas, econômicas e políticas. *Ciência da Informação*, Brasília, v. 28, n. 2, p. 168-173, 1999a.

BARRETO, A. A . Os destinos da ciência da informação: entre o cristal e a chama. *Informação e Sociedade: Estudos, João Pessoa*, v. 9, n. 2, p. 371-382, 1999b.

BARRETO, C. M. Modelo de metadados para a descrição de documentos eletrônicos na Web. In: DEPARTAMENTO DE ENGENHARIA DE SISTEMAS DO INSTITUTO MILITAR DE ENGENHARIA, 1999. Disponível em: <<http://ipanema.ime.eb.br/~de9/teses/1999/cassia.zip>>. Acesso em: 08 jul. 2002.

BAX, M. P. Introdução às linguagens de marca. *Ciência da Informação on-line*, Brasília, v. 30, n. 1, p. 32-38, 2001. Disponível em: <<http://www.ibict.br/>>. Acesso em: 17 maio 2002.

BECKETT, D.; MILLER, E.; BRICKLEY, D. An XML encoding of simple Dublin Core Metadata. In: DUBLIN CORE METADATA INITIATIVE, 2001. Disponível em: <<http://dublincore.org/documents/2001/04/11/dcmes-xml>>. Acesso em: 17 maio 2002.

BERNERS-LEE, T.; FIELDING, R.; MASINTER, L. Uniform Resource Identifiers (URI): generic syntax. In: INTERNET ENGINEERING TASK FORCE, 1998. Disponível em: <<http://www.ietf.org/rfc/rfc2396.txt>>. Acesso em: 14 fev. 2002.

BOUCHE, R. Ciência da informação: ciência da forma. *Ciência da Informação*, Brasília, v. 17, n. 2, p. 99-104, 1988.

BRYAN, M. An introduction to the Extensible Markup Language (XML) In: SGML CENTRE, 1997. Disponível em: <[www.sgml.u-net.com/xmlintro.htm](http://www.sgml.u-net.com/xmlintro.htm)>. Acesso em: 17 maio 2002.

BURNETT, K.; NG K. B.; PARK, S. A comparison of the two traditions of metadata development. *Journal of the American Society for Information Science*, v. 50, n. 13, p. 1209-1217, 1999.

CENDÓN, B. V. Ferramentas de busca na Web. *Ciência da Informação*, Brasília, v. 30, n. 1, p. 39-49, 2001.

CORTEZ, E. M. Use of metadata vocabularies in data retrieval. *Journal of the American Society for Information Science*, v. 50, n. 13, p. 1218-1223, 1999.

COX, S.; MILLER, E.; POWELL, A. Recordinf qualified Dublin Core metadata in HTML meta elements. In: DUBLIN CORE METADATA INITIATIVE, 2000. Disponível em: <<http://dublincore.org/documents/dcq-html>>. Acesso em: 26 jun. 2002.

CROMWELL-KESSLER W. Correspondencias entre metadatos e interoperabilidad: qué significa todo esto?. In: BACA, M. (Ed.). *Introducción a los metadatos vias a la información digital*. Traducido al español por Marisol Jacas-Santoll. Los Angeles, CA: J. Paul Getty Trust, 1998. p. 21-24.

DAHLBERG, I. Current trends in knowledge organization. In: ORGANIZACIÓN DEL CONOCIMIENTO EN SISTEMAS DE INFORMACIÓN: ACTAS DEL I ENCUESTRO DE ISKO. Madrid, v. 1, p. 7-25, 1995.

DCMI type vocabulary. In: DUBLIN CORE METADATA INITIATIVE, 2000. Disponível em: <<http://dublincore.org/documents/2000/07/11/dcmi-type-vocabulary>>. Acesso em: 6 fev. 2002.

DESAI, B. C. Supporting discovery in virtual libraries. *Journal of the American Society for Information Science*, v. 48, n. 3, p. 190-204, 1997.

DUBLIN Core metadata element set, version 1.1: reference description. In: DUBLIN CORE METADATA INITIATIVE, 2000. Disponível em: <<http://dublincore.org/documents/1999/07/02/dces>>. Acesso em: 24 fev. 2000.

DUBLIN Core Metadata Initiative workshops. In: DUBLIN CORE METADATA INITIATIVE, 2000. Disponível em: <<http://dublincore.org/workshops>>. Acesso em: 05 fev. 2002.

DUBLIN Core Metadata Initiative. Disponível em: <<http://dublincore.org>>. Acesso em: 24 fev. 2000.

DUBLIN Core projects. In: DUBLIN CORE METADATA INITIATIVE. Disponível em: <<http://dublincore.org/documents/1999/07/02/dces>>. Acesso em: 24 fev. 2000.

EVANS, T. *HTML simples e rápido*. Tradução de Mario Moror Fecchio. Revisão técnica de Edison Raymundi Junior. São Paulo: Makron Books, 1996. 185 p.

FOSKETT, D. J. Ciência da Informação como disciplina emergente: implicações educacionais. In: GOMES, H. E. *Ciência da informação ou informática?*. Rio de Janeiro:Ed. Calunga, 1980. p. 53-70.

GARCIA, S. S. Metadados para documentação e recuperação de imagens. Dissertação (Mestrado) – Instituto Militar de engenharia (IME), Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1999, 138 p.

GETTY thesaurus of geographic names. In: GETTY. Disponível em: <<http://www.getty.edu/research/tools/vocabulary/tgn/index.html>>. Acesso em: 14 fev. 2002.

GILL, T. Los metadatos y la World Wide Web. In: BACA, M. (Ed.). *Introducción a los metadatos vías a la información digital*. Traducido al español por Marisol Jacas-Santoll. Los Angeles, CA: J. Paul Getty Trust, 1998. p. 10-20.

GILLILAND-SWETLAND, A. J. La definición de los metadatos. In: BACA, M. (Ed.). *Introducción a los metadatos vías a la información digital*. Traducido

al español por Marisol Jacas-Santoll. Los Angeles, CA: J. Paul Getty Trust, 1998. p. 1-9.

GUIMARÃES, J. A. C. Perspectivas de ensino e pesquisa em organização do conhecimento em cursos de Biblioteconomia: uma reflexão. In: CARRARA, K. (Org.). *Educação, universidade e pesquisa*. São Paulo:FAPESP, 2001. p. 61-72.

HILL, L. L. et al. Collection metadata solutions for digital library applications. *Journal of the American Society for Information Science*, v. 50, n. 13, p. 1169-1181, 1999.

HILLMANN, D. Using Dublin Core. In: DUBLIN CORE METADATA INITIATIVE, 2001. Disponível em: <<http://dublincore.org/documents/2001/04/12/usageguide>>. Acesso em: 6 fev. 2002.

INGHAM, D. B.; CAUGHEY, S. J.; LITTLE, M. C. Supporting highly manageable Web services. *Computer Networks and ISDN Systems*, v. 29, p. 1405-1416, 1997.

KLENSIN, J. C. When the metadata exceed the data: data management with uncertain data. *Statistics and Computing*, v. 5, p. 73-84, 1995.

LASSILA, O; SWICK, R. R. Resource Description Framework (RDF) Model and Syntax Specification. In: WORD WIDE WEB CONSORTIUM, 1999. Disponível em: < <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222>>. Acesso em: 10 nov. 2001.

LE COADIC, Y.-F. *A ciência da informação*. Tradução de Maria Yêda F. S. de Filgueiras Gomes. Brasília, DF: Briquet de Lemos/Livros, 1996. 119 p.

MADSEN, M. S.; FOOG, I.; RUGGLES, C. Metadata systems: integrative information technologies. *Libri*, v. 44, n. 3, p. 237-257, 1994.

MARCONDES, C. H.; SAYÃO L. F. Integração e interoperabilidade no acesso a recursos informacionais eletrônicos em C&T: a proposta da Biblioteca Digital Brasileira. *Ciência da Informação on-line*, Brasília, v. 30, n. 3, p. 24-33, 2001. Disponível em: <<http://www.ibict.br/>>. Acesso em: 17 maio 2002.

MARINO, M. T. Suporte de metadados à interoperabilidade em ambientes de aplicações científicas. IN: \_\_\_\_\_. Integração de informações em ambientes científicos na Web: uma abordagem baseada na arquitetura RDF. 2001. Dissertação (Mestrado) – Instituto de Matemática e Núcleo de Computação Eletrônica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, cap. 2, p. 5-31. Disponível em: <<http://genesis.nce.ufrj.br/dataware/Metadados/Teses/Teresa/capitulo2.pdf>>. Acesso em: 20 jul 2002.

MASON, R. O. What is an information professional?. *Journal of Education for Library and Information Science*, v.31, n.2, p.122-138, 1990.

MEDIA Types. In: INFORMATION SCIENCE INSTITUTE. Disponível em: <<http://www.isi.edu/in-notes/iana/assignments/media-types/media-types>>. Acesso em: 14 fev. 2002.

MOREIRA, W.; MOSTAFA, S. P. As garantias no texto de Froehlich. *Transinformação*, Campinas, v. 10,n. 2, 1998. Disponível em: <<http://www.puccamp.br/~biblio/transinformacao/old/vol10n2/pag38.html>>. Acesso em: 27 jun. 2002.

ORGANIZATION of the Dublin Core Metadata Initiative. In: DUBLIN CORE METADATA INITIATIVE. Disponível em: <<http://dublincore.org/about/organization>>. Acesso em: 05 fev. 2002.

ORTIZ-REPISO JIMENEZ, V. Que enseñamos despues del MARC?. *Organizacion Conoscimiento Sistema Informacion Documentacion*, v. 3, p. 217-225, 1999.

PEOPLE involved in the Dublin Core Metadata Initiative. In: DUBLIN CORE METADATA INITIATIVE. Disponível em:

<<http://dublincore.org/about/participants>>. Acesso em: 05 fev. 2002.

PIMENTEL, M. G. C.; OLIVEIRA, M. C. F. Conceitos fundamentais das estruturas de arquivos. In: INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO – USP SÃO CARLOS. Disponível em:

<<http://www.icmsc.sc.usp.br/manuals/Disciplinas/sce183/fsc6.html>>.

Acesso em: 8 fev. 2000.

PINHEIRO, L. V. R.; LOUREIRO, J. M. M. Traçados e limites da ciência da informação. *Ciência da Informação*, Brasília, v. 24, n. 1, p. 42-53, 1995.

ROSETTO, M. Metadados: novos modelos para descrever recursos de informação digital. In: INTEGRAR: CONGRESSO INTERNACIONAL DE ARQUIVOS, BIBLIOTECAS, CENTROS DE DOCUMENTAÇÃO E MUSEUS, 1., São Paulo. *Textos...* São Paulo: Imprensa Oficial do Estado, 2002, p. 485-498.

SAACK-GIGUETTE, M.; LOPEZ JR., A. M. A frame-based design for the TIMS and CAMS metadata for a Stennis information management system. *Journal of Systems and Software*, v. 20, n. 1, p. 87-92, 1993.

SANTOS, P. L. V. A. C.; SANT´ANA, R. C. G. Transferência da informação: análise para valoração de unidades de conhecimento. *DataGramaZero – Revista de Ciência da Informação*, v. 3, n. 2, 2002. Disponível em: <[http://www.dgzero.org/abr02/F\\_I\\_aut.htm](http://www.dgzero.org/abr02/F_I_aut.htm)>. Acesso em: 20 jul 2002.

SAYÃO, L.F. Bases de dados: a metáfora da memória científica. *Ciência da Informação*, Brasília, v. 25, n. 3, p. 314-318, 1996.

SMIT, J. W. Introdução à arquivologia. In: \_\_\_\_\_. Curso de Especialização Organização de Arquivos, 14. São Paulo : USP:ECA, 2000. Cap.1, p. 19-31.

SOUZA, M. I. F.; VENDRUSCULO L. G.; MELO G. C. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. *Ciência da Informação on-line*, Brasília, v. 29, n. 1, p. 93-102, 2000. Disponível em:  
<<http://www.ibict.br/cionline/artigos/2910010.htm>>. Acesso em: 13 jul. 2000.

SOUZA, T. B.; CATARINO, M. E.; SANTOS, P. C. Metadados: catalogando dados na Internet. *Transinformação*, Campinas, v. 9, n. 2, 1997. Disponível em: <<http://www.puccamp.br/~biblio/tbsouza92.html>>. Acesso em: 9 fev. 2000.

SUTTON, S. A. Gateway to educational materials (GEM): metadata for networked information discovery and retrieval . *Computer Networks and ISDN Systems*, v. 30, p. 691-693, 1998.

TAKAHASHI, T. (Org.). Conteúdo e identidade cultural. In: \_\_\_\_\_. *Sociedade da Informação no Brasil: livro verde*. Brasília, DF: Ministério da Ciência e Tecnologia, 2000. Cap. 5, p. 59-66. Disponível em:  
<[http://www.socinfo.org.br/livro\\_verde/download.htm](http://www.socinfo.org.br/livro_verde/download.htm)>. Acesso em: 3 fev. 2002.

USING Dublin Core. In: DUBLIN CORE METADATA INITIATIVE, 2001. Disponível em: <<http://dublincore.org/projects/subjects.html>>. Acesso em: 6 fev. 2002.

WEIBEL, S. et al. Dublin Core Metadata for Resource Discovery. In: INTERNET ENGINEERING TASK FORCE, 1998. Disponível em:  
<<http://www.ietf.org/rfc/rfc2413.txt>>. Acesso em: 14 fev. 2002.

WEIBEL, S. *A proposed convention for embedding metadata in HTML*. In: WORLD WIDE WEB CONSORTIUM, 1996. Disponível em:  
<<http://www.w3.org/Search/9605-Indexing-Workshop/ReportOutcomes/S6Group2.html>>. Acesso em: 1 fev. 2002.

WOLF, M.; WICKSTEED, C. Date and Time Formats. In: WORLD WIDE WEB CONSORTIUM, 1997. Disponível em: <<http://www.w3.org/TR/1998/NOTE-datetime-19980827>>. Acesso em: 14 fev. 2002.

ZENG, M. L. Metadata elements for object description and representation: a case report from a Digitized Historical Fashion Collection Project. *Journal of the American Society for Information Science*, v. 50, n. 13, p. 1193-1208, 1999.