

# UNIDADE 4

## LINGUAGENS DE MARCAÇÃO E METADADOS

---

### 4.1 OBJETIVO GERAL

Expor algumas linguagens de marcação que possuem relação com a Ciência da Informação, apresentando suas características e sua importância na recuperação de informação e interoperabilidade de dados.

Mostrar a importância da utilização de metadados para catalogação em ambientes digitais, discutindo sobre alguns tipos e padrões de metadados existentes.

### 4.2 OBJETIVOS ESPECÍFICOS

Esperamos que, ao final desta unidade, você seja capaz de:

- a) entender claramente o que são linguagens de marcação;
- b) compreender de forma clara a importância das linguagens de marcação na representação descritiva em ambientes eletrônicos;
- c) entender claramente o que são metadados e sua importância na atividade de representação descritiva da informação;
- d) entender de forma clara o conceito de padrões de metadados.

### 4.3 PRÉ-REQUISITOS

Antes de você iniciar o estudo desta unidade, reveja seu material da disciplina Instrumentos de Representação Descritiva da Informação. Você vai precisar do que aprendeu nela para entender o conteúdo da disciplina atual, pois uma complementa a outra.

---



## 4.4 INTRODUÇÃO

Em todos os setores de nossa sociedade ocorrem transformações culturais que constituem etapas essenciais ao desenvolvimento das civilizações – processos cíclicos semelhantes de gênese, crescimento, colapso e desintegração. Estes ciclos também ocorrem na Ciência da Informação.

Se pararmos para pensar, desde as primeiras regras de tratamento da informação até as atuais, a Biblioteconomia viu a gênese e a desintegração de vários métodos e técnicas de organização e tratamento da informação (não vemos mais tantos tabletes de argila em uso por aí) ocasionados, principalmente, pela mudança dos formatos e suportes desta informação.

Além, também, das alterações nas necessidades dos usuários desta informação, os novos ciclos surgem de fragmentos consolidados dos antigos ciclos, ou seja, os novos padrões de descrição bibliográfica sempre surgiram a partir do que os antigos padrões tinham de melhor.

Acompanhando esse processo cíclico do tratamento e organização da informação, estamos presenciando uma nova gênese – a do tratamento e organização da informação virtual, digital. Contudo, isto não representa a desintegração de toda metodologia desenvolvida até aqui pelos profissionais da informação na organização de bibliotecas, arquivos, centros de documentação, onde o suporte da informação é físico e palpável. Apenas surge um novo formato e suporte informacional neste novo milênio, com características distintas que terão de receber um novo tipo de tratamento, pois são recursos informacionais abstratos, livres e etéreos como a comunicação oral.

E é neste novo ambiente informacional que surgem as bibliotecas digitais, que resgatam um conceito antigo, mas que atualmente constitui uma ferramenta imprescindível para o tratamento da informação virtual: o metadado, que pode ser definido literalmente como “dado sobre o dado”.

No ambiente *web*, o metadado é a etiqueta que define ou identifica um dado ou informação, através das ferramentas de marcação existentes nas linguagens de marcação: *Standard Generalized Markup Language* (SGML), *Hipertext Markup Language* (HTML) e *eXtensible Markup Language* (XML). O metadado tem diversas funções em um documento eletrônico, podendo ser utilizado para estruturar um documento ou recurso informacional, como também para descrever “bibliograficamente” este recurso ou para administrar este recurso quanto a datas, restrição de acesso e validação.

Entretanto, é imprescindível um entendimento mais claro dos conceitos e definições mais relevantes desta nova área de atuação do profissional da informação, introduzindo este profissional no assunto, bem como apresentando os principais padrões de metadados que estão sendo desenvolvidos e utilizados no ambiente *web*.

## 4.5 ARQUITETURAS DE METADADOS

---

Na organização da *web*, além dos metadados, das linguagens de marcação e dos padrões de metadados, existem também inúmeras pesquisas sobre arquiteturas de metadados, que têm por objetivo dar suporte à codificação e ao transporte de uma grande variedade de metadados criados de forma independente, provendo entre eles interoperabilidade:

- a) **semântica:** quando possibilita a compreensão de seus elementos oriundos de padrões distintos;
- b) **sintática:** quando provê uma maneira padrão para a representação e intercâmbio de metadados;
- c) **estrutural:** quando provê representação para modelo de dados distintos.

Dentre as arquiteturas de metadados existentes, Barreto (1999) cita cinco estruturas que têm sido desenvolvidas e pensadas atualmente.

- a) **arquitetura Warwick:** seu objetivo é prover o padrão *Dublin Core* de mecanismos de extensibilidade e possibilitar o compartilhamento de abordagens descritivas distintas, através do modelo de dados baseado em: estrutura pacote e recipiente;
- b) **Kahn & Wilensky Framework:** é uma arquitetura conceitual para implementação de bibliotecas digitais no contexto *web* que utiliza como modelo de dados uma versão orientada ao objeto similar ao da arquitetura *Warwick*;
- c) **Meta Content Framework (MCF):** essa arquitetura provê um modelo conceitual e um sistema de tipos para a representação da estrutura de organização da informação em diversos ambientes. Utiliza como modelo de dados o tipo estruturado segundo um *Directed Labeled Graph* (DLG) com NÓS, ARCOS e RÓTULOS;
- d) **arquitetura de modelagem em quatro camadas:** nessa arquitetura, tenta-se prover um modelo conceitual para a integração, federação e interoperação de propostas e padrões de metadados existentes, de forma a possibilitar a manipulação homogênea de conceitos nos diversos níveis de modelagem, além de mecanismos de extensibilidade. Esses níveis são: dado (“João, 30”), modelo (empregado – nome, idade), meta-modelo (entidade, atributo, relacionamentos), meta-meta-modelo (metaentidades; metaclasses, metaatributos);
- e) **Resource Description Framework (RDF):** a arquitetura RDF foi estruturada para ser a base do processamento de metadados, provendo interoperabilidade entre as aplicações, sendo especificamente utilizada pelo padrão RDF. Seu modelo de dados baseia-se no tripé: recurso, propriedade e valor.

Contudo, vale lembrar que as mais citadas e interessantes para a área da Ciência da Informação são a arquitetura *Warwick* e a arquitetura RDF.

Em uma comparação entre essas arquiteturas, observa-se que a arquitetura de *Kahn & Wilensky* não define como dados e metadados podem ser associados dentro da estrutura do objeto digital. Já a arquitetura *Warwick* não define a estrutura de dados das propriedades dentro dos pacotes e recipientes, nem possui uma sintaxe para o transporte desses dados.

A arquitetura MCF, por sua vez, não descreve como associar dados e metadados dentro de uma mesma estrutura. Em relação à arquitetura de quatro níveis, pode-se observar que esta necessita de um formalismo de representação para a informação nas diversas camadas e modelagem.

Finalmente, a arquitetura RDF utiliza como modelo básico RDF e RDFS, que permitem modelar esquemas de aplicação, preservando a semântica dos recursos em qualquer padrão de metadados.

#### 4.5.1 Nome, identificação única e permanência

A identificação única é necessária tanto para o controle inventarial do acervo de uma biblioteca quanto para seu acesso específico para empréstimos e/ou consultas. Ela é um número ou qualquer conjunto alfanumérico que identifica cada item de um acervo especificamente. Poderíamos pensar nela como os documentos das pessoas que nos identificam de maneira única, como o Registro Geral (RG), o Cadastro de Pessoa Física (CPF), entre outros.

Na biblioteca tradicional, isso é feito com o número de registro ou de tomo, utilizado principalmente para controle de inventário e de empréstimos dos itens físicos do acervo, além do *International Standard Book Number* (ISBN), que é a identificação única de cada livro no mundo todo. No caso dos acervos digitais, e aqui se incluem as bibliotecas digitais, os repositórios digitais e os portais de informação, o nome e a identificação única são tão importantes quanto os números de tomo e de ISBN são importantes em uma biblioteca tradicional.

Essa identificação única é representada na *web* através do *Uniform Resource Locator* (URL), ou seja, através do endereço eletrônico de cada item em um acervo digital. Um URL, que em português é chamado de Localizador-Padrão de Recursos, é o endereço de um recurso (como um arquivo, uma impressora, etc.), disponível em uma rede; seja a internet, ou mesmo uma rede corporativa como uma intranet.

Os URLs não são adequados para serem utilizados como identificadores únicos, pois confundem, em um único lugar, vários itens que deveriam estar separados. Dessa forma, sempre que um arquivo é movido, o documento geralmente se perde completamente:

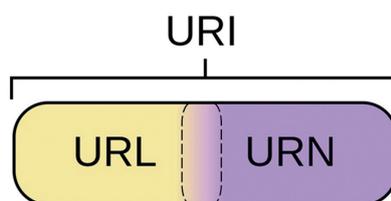
- a) o tipo de método de acesso (http);
- b) o nome da máquina;
- c) a localização do documento;
- d) o nome do arquivo do documento que pode ser ou não único.

Além dos URLs, existem mais dois recursos importantes para a identificação única dos recursos digitais:

- a) o *Uniform Resource Identifier* (URI) (Identificador Uniforme de Recursos): consiste em uma cadeia de caracteres compacta usada para identificar ou denominar um recurso na internet. Seu principal propósito é permitir a interação com representações do recurso através de uma rede, tipicamente a rede mundial, usando protocolos específicos. URIs são identificados em grupos definidos por uma sintaxe específica e protocolos associados;
- b) o *Uniform Resource Name* (URN) (Nome Uniforme de Recurso): é um tipo de URI que tem por objetivo a identificação única do recurso, de forma persistente e independente da sua localização.

Assim, a identificação única e o nome de cada recurso digital em um acervo *on-line* são localizados com a utilização combinada desses três recursos. A Figura 30 mostra como eles interagem na arquitetura de informação de um acervo digital:

Figura 30 - Interação entre os recursos de identificação



Fonte: Stack Overflow.<sup>11</sup>

Como se observa na Figura 30, o URI irá conter o URL e o URN, que se integraram para fornecer: um nome, uma localização e uma identificação únicas a um recurso informacional digital, de maneira que esse recurso não se perca com a alteração do endereço URL.

No universo dos acervos digitais, existem três esquemas propostos como uma tentativa de solucionar o problema de nome permanente:

- a) **Persistent Uniform Resource Locator** (PURL): URL permanente, desenvolvido pela OCLC. Separa o nome de um documento de sua localização, aumentando a probabilidade de ele ser localizado. Esse recurso permite a continuidade de referências a recursos de rede que podem migrar de máquina para máquina para negócios, uso social ou razões técnicas;
- b) **Uniforme Resource Name (URN)**: desenvolvido pela *Internet Engineering Task Force* (IETF). Contém um nome identificador autorizado (uma central autorizada com a tarefa de nomear identificadores) e um objeto identificador (designado pela central de autoridade);
- c) **Digital Object Identifier (DOI)**: é uma iniciativa da *Association of American Corporation for National Research Initiatives*, projetada para prover um método através do qual objetos digitais podem ser realmente identificados e acessados. É um padrão para identificação de documentos em redes de computadores, como a internet.

<sup>11</sup> Disponível em: <http://pt.stackoverflow.com/questions/43224/qual-a-diferen%C3%A7a-entre-url-e-uri>.

Atualmente, cresce a preocupação com a segurança de objetos digitais na internet. Por isso, foi criado o DOI, um sistema para localizar e acessar materiais na *web* – especialmente, publicações em periódicos e obras protegidas por *copyright*, muitas das quais localizadas em acervos digitais

Além de ser um mecanismo utilizado para garantir o pagamento de direitos autorais através de um sistema de distribuição de textos digitais, o DOI também é útil para auxiliar a localização e o acesso de materiais na *web*, facilitando a autenticação de documentos.

## 4.6 LINGUAGENS DE MARCAÇÃO

Os metadados utilizados pelas bibliotecas digitais são criados com os recursos das linguagens de marcação, base dos *sites* da internet. Desta forma, é de vital importância que o profissional bibliotecário conheça esta ferramenta.

Primeiramente, é importante definir o termo **marcação**. De acordo com Barreto (1999, p. 25), marcação “é qualquer informação adicional acrescentada ao texto de um documento.” Assim, as linguagens de marcação acrescentam marcas, etiquetas, “metadados” aos recursos informacionais da *web*, descrevendo-os, formatando-os e administrando-os.

Assim, de acordo com Bax (1998, 2001), linguagens de marcação são padrões públicos criados para minimizar o problema de transferência de um formato de representação da informação para outro. Elas identificam cada “entidade informacional” relevante de um documento, como parágrafos, títulos, tabelas ou gráficos. São elas, principalmente: SGML, HTML e XML.

Existem várias linguagens de marcação, mas todas elas são originárias do SGML, que foi a primeira grande linguagem de marcação da *web*. Contudo, no contexto das bibliotecas digitais, as mais importantes são o SGML, o HTML e o XML, que são utilizadas pelos padrões de metadados que apresentaremos mais adiante. Destas três linguagens, Barreto (1999) destaca as seguintes características principais:

a) **Standard Generalized Markup Language (SGML):**

- fornece um esquema de marcação simples, independente de plataforma e extremamente flexível;
- a formatação do documento tem que ser anexada por uma aplicação específica;
- não impõe qualquer conjunto específico de tipos de elementos: seus usuários podem criar seus próprios tipos de elementos;
- é uma metalinguagem, ou seja, uma linguagem usada para definir linguagens de marcação.

#### b) **Hipertext Markup Language (HTML):**

- é baseada na linguagem SGML;
- foi criada especialmente para codificação dos documentos disponíveis através dos servidores *web* e interpretados pelos navegadores;
- os documentos HTML são transportados pelo protocolo *HyperText Tranfer Protocol* (HTTP);
- oferece um conjunto de rótulos (elementos) padrão, não permitindo que os usuários criem seus próprios rótulos;
- a informação de apresentação (*layout*) está integrada à marcação da página.

#### c) **eXtensible Markup Language (XML):**

- surgiu em novembro de 1996 com o objetivo de definir uma linguagem de marcação que tivesse o poder da SGML, mas fosse fácil de implementar na *web*;
- permite que os usuários criem seus próprios rótulos;
- a aparência do documento é determinada por folhas de estilo XSL que podem ser anexadas ao documento;
- fornece hiperligações compatíveis com HTML;
- tem como propósito fundamental a descrição de informações sendo a base dos atuais estudos sobre padrões de metadados.

Os padrões de metadados mais antigos foram baseados em SGML. Conforme foram surgindo novas necessidades de acesso à informação em tecnologias mais modernas e diferenciadas, alguns padrões de metadados passaram a utilizar o HTML. Atualmente, com a necessidade de acesso de internet em sistemas com pouca memória, como aparelhos celulares, a necessidade de uma linguagem mais simples levou ao desenvolvimento do XML, que tem como principais vantagens: retomar a liberdade de criação de elementos (*tags*) do antigo SGML e, como no SGML, a formatação volta a ser desvinculada da formatação da página *web*, o que facilita o trabalho de descrição dos recursos informacionais da *web*. No Quadro 4, podemos verificar, além das principais diferenças entre o HTML e o XML, apontadas por Décio (2000), também exemplos destas linguagens para uma comparação visual. Hoje novas versões do HTML também são encontradas em uso, como o XHTML, entre outras. No quadro a seguir, são listadas as principais diferenças entre as linguagens de marcação mais utilizadas:

Quadro 4 - XML versus HTML

XML	HTML
<ul style="list-style-type: none"> <li>- Separa o conteúdo da apresentação.</li> <li>- As tags são definidas pelo próprio usuário.</li> <li>- Objetiva a descrição do objeto digital.</li> </ul>	<ul style="list-style-type: none"> <li>- Mistura conteúdo e apresentação.</li> <li>- As tags são padronizadas, e o usuário não pode criar tags.</li> <li>- Objetiva a formatação e estruturação do documento digital.</li> </ul>
<p style="text-align: center;"><b>Exemplo de XML</b></p> <hr/> <pre> &lt;menu&gt;   &lt;estabelecimento&gt;Restaurante Selete&lt;/estabelecimento&gt;   &lt;grupos&gt;     &lt;grupo&gt;       &lt;titulo&gt;Aperitivos&lt;/titulo&gt;       &lt;itens&gt;         &lt;item&gt;           &lt;descricao&gt;Polenta Frita&lt;/descricao&gt;           &lt;preço&gt;R\$2,00&lt;/preço&gt;         &lt;/item&gt;         &lt;item&gt;           &lt;descricao&gt;Queijo Prato porção&lt;/descricao&gt;           &lt;preço&gt;R\$3,00&lt;/preço&gt;         &lt;/item&gt;       &lt;/itens&gt;     &lt;/grupo&gt;     &lt;grupo&gt;       &lt;titulo&gt;Especialidades da Casa&lt;/titulo&gt;       &lt;itens&gt;         &lt;item&gt;           &lt;descricao&gt;Feijão com Arroz&lt;/descricao&gt;           &lt;preço&gt;R\$1,50&lt;/preço&gt;         &lt;/item&gt;         &lt;item&gt;           &lt;descricao&gt;Bife Acebolado&lt;/descricao&gt;           &lt;preço&gt;R\$3,00&lt;/preço&gt;         &lt;/item&gt;       &lt;/itens&gt;     &lt;/grupo&gt;   &lt;/grupos&gt; &lt;/menu&gt; </pre>	<p style="text-align: center;"><b>Exemplo de Documento HTML</b></p> <hr/> <pre> &lt;H2&gt;Restaurante Selete&lt;/H2&gt; &lt;TABLE WIDTH="480" CELLSPACING="1"&gt;   &lt;TR&gt;&lt;TD&gt;&lt;H3&gt;Aperitivos&lt;/H3&gt;&lt;/TD&gt;&lt;/TR&gt;   &lt;TR&gt;     &lt;TD&gt;Polenta frita&lt;/TD&gt;     &lt;TD&gt;R\$2,00&lt;/TD&gt;     &lt;TD&gt;Queijo Prato porção&lt;/TD&gt;     &lt;TD&gt;R\$3,00&lt;/TD&gt;   &lt;/TR&gt;   &lt;TR&gt;&lt;TD&gt;&lt;H3&gt;Especialidades da casa&lt;/H3&gt;&lt;/TD&gt;&lt;/TR&gt;   &lt;TR&gt;     &lt;TD&gt;Feijão com Arroz&lt;/TD&gt;     &lt;TD&gt;R\$1,50&lt;/TD&gt;     &lt;TD&gt;Bife acebolado&lt;/TD&gt;     &lt;TD&gt;R\$3,00&lt;/TD&gt;   &lt;/TR&gt; </pre>

Fonte: Produção da própria autora.



## 4.7 METADADOS: DEFINIÇÃO E TIPOS

Neste cenário novo da Ciência da Informação, surge o termo metadado como instrumento de tratamento, estruturação e administração dos recursos de informação da internet. Embora não se trate de um conceito novo (longe disso), o termo vem chamando a atenção dos estudiosos da área sobre sua definição exata. A definição encontrada na literatura da área conceitua metadados, de maneira bem simples, como sendo dados sobre dados.

Podemos explicar melhor se transferirmos para o ambiente das bibliotecas tradicionais: quando preenchemos a planilha de um novo livro em um sistema de automação de bibliotecas, ou mesmo manualmente, cada campo a ser preenchido é precedido por um termo que identifica a informação que deverá ser colocada neste campo específico (autor, título, edi-

ção, data, etc.). Essa “etiqueta”, que identifica o dado que será colocado naquele campo específico, é o metadado.

Desta forma, podemos ampliar o conceito e dizer que o metadado se refere a descrições estruturadas, armazenadas como dados de computador que tentam descrever as propriedades essenciais de outros dados específicos de um banco de dados de computador e dos objetos de dados que compõem a informação na *web*. E esses metadados nas bibliotecas digitais são representados pelas *tags* ou metaetiquetas das linguagens de marcação.

Nesse sentido, metadado não é mesmo um conceito novo, principalmente no contexto biblioteconômico. Em bibliotecas, os metadados estão incluídos nos índices, nos *abstracts* e nos registros do catálogo. Além disso, até meados dos anos 1990, já era um termo utilizado por comunidades envolvidas com administração e interoperabilidade de dados geoespaciais.

De acordo com Gilliland-Swetland (2000), em geral todo objeto de informação, independentemente de sua forma física ou intelectual, tem três características, que podem ser representadas por metadados:

- a) **conteúdo:** relacionado ao que o objeto de informação contém;
- b) **contexto:** localiza o objeto informacional no tempo e no espaço;
- c) **estrutura:** dá a relação de um objeto informacional com outros. Pode ser intrínseco ou extrínseco.

Este autor também nos alerta para o fato de que em um ambiente onde o usuário tem acesso incondicional aos objetos informacionais, o metadado irá também certificar a autenticidade do conteúdo; estabelecer e documentar o contexto do conteúdo; identificar e explorar as relações estruturais dos objetos de informação externas e internas com outros objetos de informação; fornecer pontos de acesso para os usuários recuperarem estes objetos informacionais e recuperar informações analíticas e resumidas dos objetos de informação que possam ter sido tratados por profissionais da informação (GILLILAND-SWETLAND, 2000).

Diante desta realidade conceitual, é imprescindível que o profissional da informação observe alguns aspectos importantes sobre metadados, antes de utilizá-los em suas pesquisas e em suas bibliotecas digitais. De acordo com Gilliland-Swetland (2000, p. 9):

Metadados não são necessariamente digitais;  
Metadados são mais que simplesmente a descrição de um objeto digital;  
Metadados podem ser gerados por várias fontes automáticas e/ou manuais;  
Metadado acompanha o objeto de informação atualizando-se e modificando-se de acordo com este;  
Um metadado de um objeto de informação pode ser ao mesmo tempo o dado de um outro objeto de informação.

Seguindo a linha de raciocínio traçada até aqui, podemos dizer que os metadados podem ser divididos em cinco tipos distintos, de acordo com sua função em um banco de dados de uma biblioteca digital, conforme o Quadro 5:

**Quadro 5 - Tipos de metadados quanto à sua função**

TIPO	DEFINIÇÃO	EXEMPLOS
Administrativo	Metadado utilizado na administração de recursos de informação	- Aquisição de informação - Direitos de reprodução - Critérios de seleção para digitalização, etc.
Descritivo	Metadado para descrição de recursos de informação	- Catalogação de registros - Índices especializados, etc.
De preservação	Metadado utilizado para preservação de recursos de informação	- Documentação das condições físicas dos recursos, etc.
Técnico	Metadado utilizado para conhecer as funções de um sistema ou o comportamento dos metadados	- <i>Hardware</i> e <i>software</i> - Dados de segurança - Documentação, etc.
De uso	Metadado relativo ao nível e tipo de uso de um recurso de informação	- Registros de exibição - Sumário de reuso e de versões, etc.

Fonte: Adaptado de Gilliland-Swetland (2000, p. 5).

De acordo com esta classificação de *Gilliland-Swetland*, podemos visualizar outras funções além da descritiva, como também podemos observar que um banco de dados de uma biblioteca digital irá utilizar todos esses tipos de metadados em um único banco de dados, pois, como em uma biblioteca tradicional, a administração, a preservação, as informações técnicas e a limitação de utilização de um registro de informação estão tão presentes quanto a descrição.

Este mesmo autor também identifica os metadados de acordo com seus atributos e suas características, como podemos observar no Quadro 6:

**Quadro 6 - Atributos e características dos metadados**

(continua)

ATRIBUTO	CARACTERÍSTICAS	EXEMPLOS
Origem do metadado	Metadado interno	- Nome do arquivo e cabeçalho - Estrutura de diretórios
	Metadado externo	- Registro de catalogação - Direitos e informações legais
Método de criação do metadado	Metadado automático	- Indexação de palavras-chave
	Metadado manual	- Registro de descrição ( <i>Dublin Core</i> )

Quadro 6 - Atributos e características dos metadados

(conclusão)

ATRIBUTO	CARACTERÍSTICAS	EXEMPLOS
Natureza do metadado	Metadado leigo	- <i>Metatag</i> criado em páginas pessoais
	Metadado especialista	- Cabeçalhos de assunto especializado - Registros MARC
<i>Status</i>	Metadado estático	- Título, proveniência, data de criação, etc.
	Metadado dinâmico	- Estrutura de diretório - Resolução de imagens, etc.
	Metadado a longo prazo	- Formato técnico - Informação de propriedade - Documentação de administração e preservação
	Metadado a curto prazo	- De caráter transacional
Estrutura	Metadado estruturado	- MARC - TEI a EAD - Formato de base de dados local
	Metadado não estruturado	- Anotações e arquivos não estruturados
Semântica	Metadado controlado	- Vocabulários controlados - Controle de autoridades - AACR2R
	Metadado não controlado	- Notas de texto livre - <i>Metatags</i> de HTML
Nível	Metadados coletivos	- Relativos a coleções de objetos de informação
	Metadados individuais	- Informações individuais dentro das coleções.

Fonte: Adaptado de Gilliland-Swetland (2000, p. 6-7).

Estas características e atributos, transportados para a área da Ciência da Informação, dão a noção exata do papel do profissional da informação na *web* e de como os metadados poderão constituir-se em ferramentas deste profissional para a organização das bibliotecas digitais.

No que diz respeito ao atributo **origem do metadado**, a Ciência da Informação trabalhará efetivamente com o **metadado externo**, ou seja, aquele metadado inserido *a posteriori* pelo bibliotecário para o tratamento da informação contida em um recurso informacional, tomando apenas conhecimento do funcionamento do metadado interno para melhor utilização das máquinas de recuperação da *web*, em busca de informação relevante para seus usuários, externas à coleção de sua biblioteca digital específica.

Quanto ao atributo **método de criação do metadado**, o profissional da informação será responsável e utilizará efetivamente em uma biblioteca digital o **metadado manual** para a descrição de um recurso informacional, utilizando-se de padrões de metadados como o *Dublin Core* para este fim. O metadado automático é utilizado pelos diretórios e pelas máquinas de busca e se mostra excessivamente revocatório, prejudicando a relevância destes sistemas de recuperação da informação.

Já em relação ao atributo **natureza do metadado**, os metadados utilizados pelos profissionais de Ciência da Informação em geral serão os **metadados especialistas** baseados em vocabulário controlado e formatos de entrada de dados padrão, como o MARC e o *Dublin Core*. Os metadados leigos, por sua vez, são naturais das páginas da *web* de responsabilidade das mais diversas pessoas leigas.

Outro atributo dos metadados interessante é o *status*. Aqui, o profissional da informação irá trabalhar especificamente com os **metadados estáticos** e com os **metadados a longo prazo**. Contudo, poderão surgir recursos informacionais dos quais o profissional da informação necessitará manusear também metadados dinâmicos e a curto prazo.

Em relação ao atributo **estrutura**, uma biblioteca digital irá utilizar preferencialmente **metadados estruturados** como os padrões de metadados. Entretanto, uma biblioteca digital desenvolvida por um leigo, provavelmente, lançará mão de metadados não estruturados.

No que tange ao atributo **semântica**, os profissionais da informação desenvolvem em suas bibliotecas digitais catálogos que utilizam **metadados controlados** de acordo com as normas de tratamento da informação vigentes. Os metadados não controlados são utilizados, geralmente, por leigos.

E, finalmente, no atributo **nível**, uma biblioteca digital poderá trabalhar tanto com **metadados individuais** como com **coletivos**, como nas bibliotecas tradicionais em que existem obras individuais e coleções.

## 4.8 PADRÕES DE METADADOS

---

Nos dias atuais, constantemente nos deparamos com algum tipo de padronização e, na área de tratamento da informação, isso é mais frequente, até mesmo porque a padronização é o fator que garante a total interoperabilidade entre os sistemas de recuperação da informação atuais,



principalmente os automatizados. Padrões podem ser entendidos como sendo uma forma de se fazer algo da mesma maneira, da mesma forma que todos os outros profissionais fazem. Esta padronização é garantida por códigos, normas, patentes, tabelas, enfim, qualquer documentação que torne os procedimentos em uma determinada área normalizados.

Estes procedimentos normalizados na área de Ciência da Informação e, em especial, no tratamento de informação da *web*, tem por objetivos, segundo Moura (2001, p. 4):

- a) fornecer um conjunto de termos para descrição de uma ou mais categorias de recursos;
- b) obter resultados mais precisos no processo de busca de recursos de informação na *web*;
- c) estabelecer padrões de metadados, modelos e protocolos e sua integração a mecanismos de busca na *web*.

Mas de onde surgem estes padrões, quem os define e como são aceitos e incorporados pelas áreas profissionais? Brickley e Guha (1999, p. 2) afirmam que padrões

[...] podem ser totalmente definidos pelo mercado; [...] também são o resultado da pressão de grupos de peritos bem-intencionados; [...] ou a consequência de uma legislação de interesse público. [...] podem ocorrer como uma expressão de um consenso dentro de uma comunidade grande. Quando eles nascem por este último motivo tendem a ser mais duradouros, mas este consenso é difícil de ser alcançado.

Se pararmos para pensar, na Ciência da Informação, os padrões sempre foram o resultado dos esforços de grupos de especialistas bem-intencionados, preocupados com a organização do conhecimento humano como um todo. Entretanto, quando estes esforços se concretizam, o tratamento da informação adquire padrões com o consenso de toda a comunidade de profissionais da informação, tornando-se padrões consolidados e duradouros.

Assim, podemos dizer que a Ciência da Informação e, em especial, a Biblioteconomia, após anos de estudos e esforços de seus especialistas, podem se orgulhar em dizer que alcançaram o mais difícil dos níveis de padronização de seus procedimentos profissionais: estes padrões são a expressão de um consenso dentro da Ciência da Informação. Mas é importante ressaltar que essa realidade alcança os sistemas de recuperação da informação manuais e automatizados.

Já o tratamento da informação no ambiente *web* e, mais particularmente, nas bibliotecas digitais ainda não alcançou este nível, até mesmo porque é uma realidade muito recente. O que já existe no caso das bibliotecas digitais ainda é a pressão de grupos peritos bem-intencionados, na tentativa de se chegar a um consenso de padrão de metadados. E ainda padrões que se estabeleceram através de uma legislação de imposição do padrão, como é o caso do *Government Information Location Service* (GILS), que é utilizado em todos os órgãos do governo norte-americano.

Dentre os padrões de metadados existentes, hoje na internet e em especial no ambiente de bibliotecas digitais, podemos observar dois grupos

de estudos na área distintos: os grupos ligados às tradições bibliotecônicas e os grupos não ligados a estas tradições. Cada um desses grupos possui vários padrões de metadados desenvolvidos por instituições diversas. Mas somente entre os ligados à tradição biblioteconômica de tratamento da informação é que percebemos um trabalho com direcionamentos interligados, pois todos eles se basearam no formato MARC para serem desenvolvidos. O Quadro 7 apresenta alguns padrões de metadados, divididos pela sua origem ou não na Biblioteconomia moderna.

Quadro 7 - Origens dos padrões de metadados

LIGADOS À TRADIÇÃO BIBLIOTECÔNOMICA	NÃO LIGADOS À TRADIÇÃO BIBLIOTECÔNOMICA
<ul style="list-style-type: none"> <li>- Formato MARC</li> <li>- Padrão DUBLIN CORE – OCLC</li> <li>- Padrão GILS – National Archives dos EUA.</li> <li>- Padrão EAD – Universidade da Califórnia.</li> <li>- Padrão RDF – W3C, baseado nos resultados dos estudos de Dublin Core.</li> </ul>	<ul style="list-style-type: none"> <li>- Padrão TEI – Associação de Computadores e Humanidades</li> <li>- Padrão AIFA – Grupo IETF (<i>Internet Engineering Task Force</i>)</li> <li>- Padrão SOIF – Universidade do Colorado</li> <li>- FDGC e UDK – Agências de controle e proteção ambiental</li> <li>- <i>Metatags</i> em HTML</li> <li>- Entre outros.</li> </ul>

Fonte: Produção da própria autora.

De acordo com Gill (2001), dentre estes esforços de padrões de metadados, três são particularmente pertinentes no contexto das bibliotecas digitais, sendo os mais importantes atualmente no estudo da organização de recuperação dos recursos informacionais da *web*:

- a) **Metatags de HTML (Keywords e Description):** utilizando o *tag* <META>, o HTML permite que sejam especificadas informações adicionais a respeito do documento, como a inclusão de um texto descritivo sobre o documento (*Meta Description*), ou de uma lista de palavras-chaves (*Meta Keywords*). Veja exemplo na figura a seguir.

Figura 31 - Metatags de HTML (Keywords e Description)

```
<META NAME="Description" CONTENT="Gateway to Africa on the Web: art, literature, business, trade, culture, politics, music, news, events, sports, travel, food and discussions">
<META NAME="Keywords" CONTENT="AFRICAN, DIRECTORY, TRAVEL, AFRICA, BUSINESS, CUISINE">
```

Fonte: Produção da própria autora.

- b) **PADRÃO DUBLIN CORE:** baseado no Formato MARC e desenvolvido pela OCLC e pela *National Center for Supercomputer Applications* (NCSA), utiliza-se das *tags* do HTML para a identificação de seus 15 elementos de descrição. Veja exemplo na figura a seguir.

Figura 32 - Padrão Dublin Core

```
<link rel="schema.DC" href="http://purl.org/dc">
<meta name="DC.Title" content="XML bible">
<meta name="DC.Creator" content="Harold, Elliotte Rusty">
<meta name="DC.Subject" content="XML (Document Markup Language">
<meta name="DC.Description" content="System Timed Out (Library of Congress
Online Catalog).">
<meta name="DC.Publisher" content="IDG Books Woldwide">
<meta name="DC.Date" scheme="W3CDTF" content="1999">
<meta name="DC.Type" scheme="DCMIType" content="Dataset">
<meta name="DC.Format" content="text/html">
<meta name="DC.Format" content="2085 bytes">
<meta name="DC.Identifier" content="http://catalog.loc.gov/cgi-bin/Pwebrecon.
cgi?v4=25&ti=1,25&SEQ=20020806104233&Search_Arg=xml&Search_Code=TAL-
L&PID=3127&CNT=25&SID=3">
<meta name="DC.Language" content="English">
<meta name="DC.Relation" content="Book">
<meta name="DC.Coverage" content="Foster City, California">
<meta name="DC.Rights" content="Harold, Elliotte Rusty IDG Books Woldwide">
```

Fonte: Produção da própria autora.

- c) **ESQUEMA RDF:** desenvolvido pelo consórcio *World Wide Web Consortium* (W3C), baseado nos resultados das pesquisas do padrão *Dublin Core*, utiliza as *tags* do XML para a definição dos elementos descritivos de um recurso informacional. No esquema RDF, o uso do XML permite que os elementos descritivos sejam criados de acordo com a necessidade de descrição do recurso. Veja exemplo na figura a seguir.

Figura 33 - Esquema RDF

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
<rdf:Description about="http://catalog.loc.gov/cgi-bin/Pwebrecon.
cgi?v4=25&ti=1,25&SEQ=20020806104233&Search_Arg=xml&Search_
Code=TALL&PID=3127&CNT=25&SID=3">
  <s:título>
    XML bible
  </s:título>
  <s:autor>
    Harold, Elliotte Rusty
  </s:autor>
  <s:assunto>
    XML (Document Markup Language
  </s:assunto>
</rdf:Description>
</rdf:RDF>
```

Fonte: Produção da própria autora.

## 4.8.1 Padrões de metadados ligados às tradições biblioteconômicas

São eles:

- a) **Machine Readable Cataloging record (MARC):** o MARC é o mais antigo dos padrões de metadados, elaborado em 1960 pela LC para atender à necessidade de se ter um padrão para entrada de dados bibliográficos nos primeiros sistemas de automação de biblioteca que despontavam na época.

Criado com o objetivo de catalogar automaticamente o acervo das bibliotecas para permitir o empréstimo entre bibliotecas em redes interligadas, atualmente é o formato padrão adotado para entrada de dados bibliográficos em sistemas de automação bibliotecária e para a catalogação cooperativa.

Este formato, inclusive atualmente, possui versões adaptadas para cada país, possuindo uma família de formatos específicos para cada necessidade local: USMARC (norte-americano), CAN/MARC (canadense), etc. Hoje, no Brasil, adotamos o MARC21, que se constitui em uma tentativa de unificação dos padrões norte-americanos, ingleses e canadenses, publicada em 1999.

Até 1980, o MARC era totalmente adequado para a catalogação de arquivos de computador, mas não possuía recursos suficientes para catalogação em redes de bibliotecas. A catalogação cooperativa era realizada através de cópias dos bancos de dados em CD-ROMs, que eram distribuídos às bibliotecas participantes de uma rede.

Esta situação foi solucionada em 1995, quando um novo campo, o 856, onde seriam armazenadas informações para localização de recursos da rede, fez com que os registros catalográficos em MARC pudessem ser acessados e importados via internet.

O MARC nada mais é do que um formato padrão para um “banco de dados” bibliográfico. Ele não é um padrão de descrição bibliográfica ou de catalogação. A entrada de dados em um registro MARC e a definição dos pontos de acesso deste registro são baseadas nas normas de catalogação do AACR2R. A relação do MARC com o AACR2R, a ISDB ou outro código de catalogação é simplesmente a padronização de entrada de dados.

Com a utilização do formato MARC, um sistema de automação de bibliotecas irá fornecer ao catalogador os campos de descrição necessários para o registro de um item bibliográfico ou documental. Mas a maneira correta e normalizada de inserir as informações nos campos definidos pelo MARC é toda definida pelo código de catalogação (AACR2R). Desta forma, o MARC fornecerá os insumos necessários para que um *software* bibliográfico administre adequadamente o seu banco de dados.

Um registro MARC é formado pelas seguintes partes:

- **líder:** é o primeiro registro do MARC e não aparece na tela do *software*. Possui 24 posições de 00 a 23, com informações administrativas do registro como: *status* do registro, tipo de registro, nível bibliográfico, tipo de controle, endereço do registro na base, nível de catalogação, forma de catalogação descritiva entre outras;

- **diretório:** é criado automaticamente pelo *software*. Fornece a estrutura dos campos do registro. O diretório irá informar ao sistema as etiquetas (metadados) de cada campo, o tamanho dos campos em número de dígitos e a posição inicial de cada campo no banco de dados;
- **campos fixos ou de controle:** identificado pelas etiquetas de 00X a 008, contém informações sobre: o número de controle (como um número de tombo), data e hora da última alteração do registro, informações específicas adicionais sobre o material registrado (público-alvo, forma, natureza do conteúdo, etc.) e informações gerais sobre qualquer tipo de material (lugar, idioma, fonte de catalogação, etc.);
- **campos variáveis:** que se dividem em campos de controle de 01X a 09X, preenchidos com os números de controle atribuído pelas agências catalogadoras como a LC, e campos de dados de X00 a 9XX, com as informações descritivas dos itens catalogados e os pontos de acesso para criação dos índices invertidos do sistema, preenchidos de acordo com o AACR2R, LC e *Thesaurus*.

b) **Dublin Core:** o *Dublin Core* é um padrão de metadados criado para facilitar a descoberta de recursos informacionais na *web*.

Conceito: é um padrão de metadados criado para facilitar a descoberta de recursos eletrônicos na rede. Foi criado pela OCLC e pelo NCSA, em 1995, seus elementos descritivos foram baseados nos campos variáveis mais importantes do MARC para descrição de itens informacionais. Permite o uso de qualificadores como o MARC para uma especificação maior na descrição de um objeto digital.

O *Dublin Core* possui 15 elementos descritivos, que são:

- *title*: título do objeto;
- *creator*: responsáveis pelo conteúdo intelectual do objeto;
- *subject*: tópico relacionado ao objeto descrito;
- *description*: contém uma descrição textual do objeto;
- *publisher*: agente responsável por tornar o objeto disponível;
- *contributor*: outros "autores" do conteúdo intelectual do objeto;
- *date*: data de publicação;
- *type*: tipo do objeto;
- *format*: formato de dado do objeto;
- *identifier*: identifica o recurso de forma única;
- *source*: objetos dos quais o objeto descrito é derivado;
- *language*: idioma relativo ao conteúdo intelectual do objeto;
- *relation*: indica um tipo de relacionamento com outros objetos;
- *coverage*: localização espacial e duração temporal do objeto;
- *rights*: contém referência ou direitos de propriedade.

O *Dublin Core* é especificado em HTML, contudo, com a utilização da arquitetura RDF, seus elementos são identificados utilizando-se o XML.

c) **Encoded Archival Description (EAD):** o desenvolvimento do EAD foi uma iniciativa da *Universidade de Califórnia, Berkeley*, em 1993. A meta do projeto era investigar a viabilidade de desenvolver um padrão não proprietário de codificação, para descrição dos recursos informacionais da rede da universidade. O EAD é baseado no SGML, não sendo muito difundido na *web*. Este padrão agrupa seus elementos descritivos em dois segmentos, sendo que cada um deles, respectivamente:

- provê informações sobre a "autoria" do arquivo;
- provê informações sobre o "corpo" do arquivo.

d) **Government Information Location Service (GILS):** este padrão de metadados teve seus estudos iniciados em 1992, fruto de uma parceria que englobou o *Office of Management and Budget*, o *National Archives* e a *Records Administration*, órgãos governamentais dos Estados Unidos. Tendo sido aprovado em 1994 pelo Departamento de Comércio, foi adotado como padrão federal para processamento de informações do governo norte-americano e, em 1995, tornou-se o padrão de metadados obrigatório por lei em todos os departamentos governamentais dos Estados Unidos.

O propósito do GILS é auxiliar o público em geral a localizar e acessar tanto informações como fontes de informações. Para isso, define cerca de 70 atributos, denominados elementos centrais (*core elements*, além de aproximadamente mais 100 elementos herdados do protocolo Z39.50).

Contudo, cada elemento de GILS pode ser:

- obrigatório ou opcional;
- repetível ou não repetível;
- controlado ou não controlado;
- elemento de agrupamento ou não agrupamento.

Desta forma, GILS pode ser entendido como: um serviço de coleta e disseminação de metadados; uma arquitetura para projeto físico e implementação e um esquema para descrição de metadados;

e) **Resource Description Framework Schema (RDF):** o RDF foi desenvolvido pelo W3C, com base nas pesquisas do *Dublin Core* e da arquitetura *Warwick*. Assim, o RDF é dividido em duas aplicações: o *RDF Schema* e o RDF propriamente dito, que será discutido quando falarmos das arquiteturas de metadados.

A parte descritiva deste padrão de metadados é chamada de esquema RDF e define as propriedades dos recursos (título, autor, assunto, tamanho, etc.), os tipos de recursos e suas semânticas. Este padrão providencia informação sobre a interpretação das declarações dadas em um modelo de dados RDF e é baseado na linguagem XML. O esquema RDF constitui a base descritiva da arquitetura RDF, ou modelo de dados.

O modelo conceitual é composto por dois elementos: *sem:Conceito* e *sem:Dependência*. O elemento *sem:Conceito*, definido como uma classe através das primitivas *rdfs:subClassOf* e *rdf:Class*, denota as entidades do mundo real, de forma similar ao conceito de classe encontrado em ontologias. O elemento *sem:Dependência*, definido como uma propriedade através das primitivas *rdfs:subClassof* e *rdf:Property*, denota relacionamentos direcionados entre dois elementos *sem:Conceito*, indicando que um influencia o outro. (MARINO, 2001, p. 65).

#### 4.8.2 Padrões de metadados não ligados às tradições biblioteconômicas

São eles:

- a) **Text Encoding for Information Interchange (TEI):** o TEI é um esforço cooperativo desenvolvido com a meta de definir um jogo de diretrizes genéricas para a representação de documentos eletrônicos. Seu projeto é patrocinado por três associações profissionais: *Association for Computational Linguistics (ACL)*, *Association for Literary and Linguistics Computing (ALLC)* e *Association for Computing and the Humanities (ACH)*. Devido a esta raiz linguística, o padrão TEI está principalmente interessado em texto e é baseado em SGML.

Os jogos de etiquetas são o princípio básico da organização do TEI e estão divididos em quatro grupos:

- **jogos de etiquetas centrais:** contêm os elementos necessários a todos os tipos de documentos;
- **jogos de etiquetas básicos:** são conjuntos de elementos específicos para documentos de uma classe em particular, tais como verso, prosa, drama, etc.;
- **jogos de etiquetas adicionais:** são elementos apropriados para o tratamento especializado ou detalhado do texto em classes diferentes de documentos;
- **jogos de etiquetas auxiliares:** são elementos com funções mais especializadas.

- b) **Internet Anonymous FTP Archive (IAFA):** o padrão IAFA foi desenvolvido para aplicação em ambientes que utilizam o protocolo *File Transfer Protocol (FTP)*, um dos componentes do conjunto de protocolos TCP/IP, anterior ao HTTP.

Os modelos IAFA foram desenvolvidos pelo grupo IETF, com o objetivo de prover aos administradores do espaço FTP um conjunto de descritores para a documentação dos recursos de informação disponíveis em seus arquivos, tornando possível a sua utilização pelos mecanismos de indexação de recursos.

Não é um padrão muito difundido na *web*, até porque o protocolo FTP não é muito utilizado atualmente;

- c) **Summary Object Interchange Format (SOIF):** este padrão de metadados foi concebido como parte da arquitetura do sistema *Harvest* (um conjunto de ferramentas integradas para coletar, extrair, organizar, pesquisar, “cachear” e replicar informações relevantes na internet) na *Universidade do Colorado*, em 1994. O SOIF especifica vários elementos descritores, dos quais 17 são comuns a todos os tipos de dados coletados, tais como *abstract*, *author*, *description* e *title*, além de uma sintaxe padrão para o intercâmbio de informações estruturadas entre os diversos subsistemas *Harvest*.

As etiquetas do SOIF podem ser geradas tanto manual como automaticamente pelos coletores *Harvest*.

- d) **Padrões de metadados para dados georreferenciados:** os padrões de metadados para informações georreferenciais são os mais antigos do ambiente eletrônico e são utilizados por instituições de pesquisas na área de controle ambiental, ecologia, astronomia, etc.

Dentre os padrões ligados às tradições biblioteconômicas, o GILS também tem sido utilizado pelo governo norte-americano para descrição, organização e disseminação de informações georreferenciais. Além do GILS, existem outros que são muito utilizados:

- **Federal Geographic Data Commite (FGDC):**

Objetivo: Fornecer um conjunto de terminologias e definições comuns para documentação de dados espaciais e digitais.

- **Catologue of Data Source (CDS):**

Objetivo: Armazenar a descrição do conjunto dos dados que já estão disponíveis.

- **Umwelt Datenkatalog (UDK):**

Objetivo: Coletar e recuperar metadados ambientais, ou seja, o metadado deve informar “quem” possui os dados, “onde” eles estão e em qual formato estão armazenados.



## Multimídia

### Conheça mais sobre os padrões de metadados:

- MARC: <https://www.loc.gov/marc/>;
- Dublin Core*: <http://dublincore.org/documents/2005/11/07/usageguide/qualifiers.shtml>;
- EAD: <http://www2.archivists.org/groups/technical-subcommittee-on-encoded-archival-description-ead/encoded-archival-description-ead/>;
- GILS: <http://www.archives.gov/records-mgmt/policy/gils.html#intro>;
- RDF: <http://www.w3.org/TR/rdf-schema/>;

- f) TEI: <http://www.tei-c.org/Vault/ISC/J31>;
- g) AIFA: <https://www.ietf.org/wg/concluded/iafa.html>;
- h) SOIF: <http://harvest.sourceforge.net/harvest/doc/html/manual-7.html>;
- i) FGDC: <https://www.fgdc.gov/>;
- j) CDS: <http://archive.rec.org/REC/Programs/Telematics/DETERMINE/EuroInfoSession/SJensen.html>;
- k) UDK: [http://www.ncgia.ucsb.edu/conf/SANTA\\_FE\\_CD-ROM/sf\\_papers/guenther\\_oliver/my\\_paper.html](http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/guenther_oliver/my_paper.html).

---

## 4.9 CONCLUSÃO

---

Podemos observar que tanto as linguagens de marcação quanto os metadados representam as tecnologias de ponta atuais para a implementação dos produtos da representação descritiva em ambiente eletrônico e digital.

É importante observar também que, apesar de estar envolto em uma “nuvem” de dúvidas, o conceito de metadados é mais antigo do que os textos científicos parecem nos informar. O que seria novo talvez fosse somente sua aplicação em ambientes não convencionais ou eletrônicos.

Podemos perceber que eles já existiam embutidos nos instrumentos de representação não só descritiva, mas também temática, utilizados na construção dos catálogos e das fontes de informação manuais, mas sem que os profissionais bibliotecários se dessem conta disso.

A diferença é que, no ambiente eletrônico, a importância dos metadados se intensifica, e este começa a ser percebido no contexto da representação da informação. Para sua utilização, as linguagens de marcação se mostraram mais eficientes que as antigas linguagens de programação na implementação de bases de dados em ambiente *web*.



### 4.9.1 Atividade

---

**Atende ao objetivo 1 a 4**

#### **UTILIZAÇÃO DO PADRÃO DUBLIN CORE**

Faça o que se pede a seguir:

1. Entre nos links abaixo:

[http://www.dgz.org.br/ago06/F\\_I\\_art.htm](http://www.dgz.org.br/ago06/F_I_art.htm);

<http://www.informacaoesociedade.ufpb.br/ojs2/index.php/ies/issue/view/43/showToc>.

2. Escolha um artigo de cada periódico *on-line*.
3. Acesse um dos editores de Dublin Core abaixo:  
<http://www.ukoln.ac.uk/metadata/dcdot/>:  
Extrator de metadados  
  
<http://www.library.kr.ua/dc/dcreditunie.html>:  
Extrator de metadados e conversor de metadados DC para USMARC
4. Preencha os campos do *Dublin Core* de cada artigo selecionado.
5. Gere o *Dublin Core* em HTML, XML e qualquer outra linguagem disponível nesses editores.
6. Salve no Word.

### Resposta comentada

Nesta última atividade, o objetivo é apresentar aos alunos a existência de extratores de *Dublin Core* gratuitos na internet, que podem ser utilizados também para a edição desse padrão de metadados, gerando-o automaticamente nas mais diversas linguagens de marcação e/ou nos mais diversos padrões de metadados.

O objetivo é demonstrar que, mesmo que o bibliotecário não entenda muito de linguagens de marcação, existem ferramentas para auxiliá-lo na construção de bibliotecas e repositórios digitais.

## RESUMO

Nessa última unidade, foram abordadas as ferramentas disponíveis para a representação descritiva gerar seus produtos em ambiente eletrônico e digitais.

No caso dos catálogos eletrônicos e também dos manuais, os metadados são indispensáveis e se materializam na forma de padrões de metadados, no primeiro, e na forma de campos subentendidos nas fichas catalográficas, estabelecidos pelos códigos de catalogação. Isso também se aplica no caso das fontes de informação manuais e das bases de dados.

Nesse contexto, as linguagens de marcação permitem a implementação dessas regras e desses padrões em ambiente *web*, indo além das bases de dados geradas em SGBD, utilizadas inicialmente na informatização dos produtos da representação descritiva.



## Sugestão de Leitura

ALMEIDA, Maurício Barcellos. Uma introdução ao XML, sua utilização na internet e alguns conceitos complementares. **Ciência da Informação**, Brasília, v. 31, n. 2, p. 5-13, maio/ago. 2002.

BURNARD, Lou. **Text Encoding for Information Interchange: an Introduction to the Encoding Initiative**. [S.l.]: TEI, 1995. Disponível em: <http://www.tei-c.org/Vault/ISC/J31>. Acesso em: 11 fev. 2015.

CATHRO, Warwick. **Metadata: Overview**. Sydney: National Library of Australia, 1997. Disponível em: <http://nla.gov.au/nla/staffpaper/cathro3.html>. Acesso em: 18 nov. 1999.

ELEMENTOS do núcleo de metadata "Dublin Core": descrição de referência. [S.l.: s.n., 20--?]. Disponível em: [http://jlbnd.bn.pt/DCed/elementos\\_dublin\\_core/dcmes/dcmes11-20000518.htm](http://jlbnd.bn.pt/DCed/elementos_dublin_core/dcmes/dcmes11-20000518.htm). Acesso em: 4 ago. 2001.

GÜNTHER, Oliver; LESSING, Helmut; SWOBODA, Walter. **UDK: a European Environmental Data Catalogue**. Santa Barbara: NCGIA, 1996. Disponível em: [http://www.ncgia.ucsb.edu/conf/SANTA\\_FE\\_CD-ROM/sf\\_papers/guenther\\_oliver/my\\_paper.html](http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/guenther_oliver/my_paper.html). Acesso em: 12 fev. 2015.

HARDY, Darren R. et al. **Harvest User's Manual**. [S.l.: s.n.], 2002. Disponível em: <http://harvest.sourceforge.net/harvest/doc/html/manual.html#toc7>. Acesso em: 12 fev. 2015.

SOCIETY OF AMERICAN ARCHIVISTS. **Encoded Archival Description (EAD)**. [S.l.: s.n.], c2017. Disponível em: <http://www2.archivists.org/groups/technical-subcommittee-on-encoded-archival-description-ead/encoded-archival-description-ead>. Acesso em: 11 fev. 2015.

THE FEDERAL Geographic Data Committee. Reston, [201-?]. Disponível em: <https://www.fgdc.gov/>. Acesso em: 12 fev. 2015.

THE GOVERNMENT Information Locator Service. **National Archives**, [S.l., 20--?]. Disponível em: <http://www.archives.gov/records-mgmt/policy/gils.html#intro>. Acesso em: 11 fev. 2015.

USING Dublin Core: Dublin Core Qualifiers. **Dublin Core Metadata Initiative**, [S.l.], 2005. Disponível em: <http://dublincore.org/documents/2005/11/07/usaguide/qualifiers.shtml>. Acesso em: 11 fev. 2015.

VELLUCCI, Sherry L. Bibliographic Relationships. In: INTERNATIONAL CONFERENCE ON THE PRINCIPLES AND FUTURE DEVELOPMENT OF AACR, 1997, Toronto. **Proceedings...** Toronto: American Library Association Publishing, 1998, p. 105-147.

## REFERÊNCIAS

---

BARRETO, Cássia Maria. **Modelo de metadados para a descrição de documentos eletrônicos na web**. 1999. 190 f. Dissertação (Mestrado em Ciências em Sistemas de Computação) – Instituto Militar de Engenharia, Rio de Janeiro, 1999.

BAX, Marcello Peixoto. Introdução às linguagens de marcas. **Ciência da Informação**, Brasília, v. 30, n. 1, p. 32-38, jan./abr. 2001.

BRICKLEY, Dan; GUHA, R. V. **Resource Description Framework (RDF) Schema**: especificação do esquema da estrutura de descrição de recursos. [S.l.: s.n., 1999?]. Disponível em: <http://www.geocities.com/SiliconValley/Orchard/5320/Rdf.html>. Acesso em: 7 fev. 2002.

GILLILAND-SWETLAND, Anne J. **Introduction to Metadata: Setting the Stage**. [S.l.: s.n., 2000?]. Disponível em: <http://www.getty.edu/research/institute/standards/intrometadata/>. Acesso em: 7 jun. 2002.

INTERNET Anonymous FTP Archives. **IETF**, [S.l.], 1994. Disponível em: <https://www.ietf.org/wg/concluded/iafa.html>. Acesso em: 12 fev. 2015.

MOURA, Ana Maria de Carvalho; FERNANDES, Cezar S. A metadata approach to represent and visualize sites on the web. In: INTERNATIONAL WORKSHOP ON INFORMATION INTEGRATION TECHNOLOGIES ON THE WEB, 2001, Itaipava. **Papers...** Itaipava: [s.n.], 2001. Disponível em: <http://www.ipanema.ime.br/~anamoura/METADATA.html>. Acesso em: 19 dez. 2001.











UNIVERSIDADE FEDERAL  
DO RIO DE JANEIRO



MINISTÉRIO DA  
EDUCAÇÃO



Agência Brasileira do ISBN  
ISBN 978-85-85229-40-5



9 788585 229405

Agência Brasileira do ISBN  
ISBN 978-85-85229-41-2



9 788585 229412