

Metadata – Conceito e sua relevância para as bibliotecas

José Borbinha

Biblioteca Nacional

Direcção de Serviços de Inovação e Desenvolvimento

1749-081 Lisboa

Tel.: 217982083

E-mail: jose.borbinha@bn.pt

RESUMO

O conceito de “metadata” emergiu como uma nova realidade associada à Internet e aos recursos digitais em geral. A sua importância tem sido reconhecida a vários níveis, de desde os serviços emergentes de comércio electrónico até às bibliotecas e arquivos.

Esta comunicação foca este fenómeno, segundo a perspectiva das bibliotecas e arquivos. Para tal apresentam-se e discutem-se definições em torno do conceito de metadata e de topologias relacionadas. São apresentadas ainda algumas das iniciativas internacionais mais relevantes para esta temática, assim como pistas para uma agenda de trabalho com realce para os temas emergentes ou mais relevantes em discussão.

Esta apresentação baseia-se na experiência do envolvimento do autor em projectos e iniciativas nacionais e internacionais, que se têm verificado quer como representante da Biblioteca Nacional quer como investigador na área da Engenharia Informática.

PALAVRAS-CHAVE: Metadata, Publicação Digital, Bibliotecas Digitais.

INTRODUÇÃO

A «biblioteca tradicional» tem estendido até à exaustão os formatos MARC – *M*ACHINE *R*EADABLE *C*ATALOGUE, com a redefinição ou definição de novos campos e sub-campos, visando abarcar tudo o que se mostre potencialmente útil ou necessário. A proliferação no entanto dos géneros de recursos, tanto em diversidade técnica como na especialização dos conteúdos, modelos e estruturas, com relevância para os recursos digitais, tem levado à necessidade de outras soluções, talvez mais adequadas a cada caso. Esta é uma realidade estranha à «biblioteca tradicional», habituada a ver no catálogo, o seu recurso de informação principal, uma estrutura uniforme e cuidadosamente normalizada.

Como parte da resposta a este problema nasceu o conceito de metadata. No contexto da «biblioteca tradicional», este conceito tem sido identificado quase exclusivamente com os formatos bibliográficos tradicionais da família MARC, mas na realidade ele pretende responder a um problema mais vasto. Nesta perspectiva pode-se definir genericamente metadata como **informação estruturada sobre ou representativa de um recurso** (documento ou obra em geral). Tal pode consistir por exemplo em informação de indexação, de descrição bibliográfica, de descrição estrutural, de termos e condições de acesso, etc.

PARADIGMA DIGITAL

O problema dos requisitos de metadata para um paradigma *on-line* foi abordado de forma alargada durante 1998 e 1999 por um grupo de trabalho internacional [1]. Aí se promove uma visão de comparação entre a «Biblioteca Tradicional» e a «Biblioteca Digital», que podemos aqui resumir nos seguintes pontos:

Dinamismo e persistência: As bibliotecas tradicionais são sistemas estáveis e de mutação lenta. As bibliotecas digitais tenderão a ser altamente dinâmicas, adaptáveis, e nalguns casos mesmo, efémeras.

Diversidade dos recursos: As bibliotecas tradicionais possuem objectos atómicos, fisicamente definidos, e na sua grande maioria impressos. As bibliotecas digitais tenderão a possuir, ou apenas a providenciar acesso, a recursos de informação interligados e de conteúdos fractais, consistindo em objectos multimédia com estruturas e dimensões variáveis e limites nem sempre definidos.

Auto-descrição dos recursos: As bibliotecas tradicionais possuem objectos com estruturas lineares, pobres em metadata de descrição formal do seu contexto temático e conteúdo intelectual. As bibliotecas digitais tenderão a suportar artefactos estruturados, com metadata e informação de contexto significativa, que pode ser extraída e processada automaticamente.

Paradigmas de publicação: As bibliotecas tradicionais, e especialmente as bibliotecas especializadas e académicas, possuem objectos criados pelas suas próprias comunidades de utilizadores e avaliados previamente por esquemas formais de publicação. O novo paradigma irá permitir a qualquer um publicar os seus trabalhos de uma forma informal, tendendo a avaliação a emergir posteriormente, por métodos formais ou por novos métodos de avaliação.

Distribuição e ubiquidade: As bibliotecas tradicionais caracterizam-se como sistemas de controlo centralizado e com relativamente poucos pontos de acesso. As bibliotecas digitais tenderão a ser distribuídas mas ao mesmo tempo ubíquas.

Recursos virtuais e personalizados: Nas bibliotecas tradicionais o controlo físico e lógico dos objectos tende a ser comum, com os objectos regra geral fisicamente organizados em colecções. Nas bibliotecas digitais o controlo físico e o controlo lógico tenderão a ser separados, permitindo a oferta

de colecções e objectos virtuais, de forma dinâmica e ainda à medida das necessidades ou perfil do utilizador.

Acesso e propriedade intelectual: A tradição da biblioteca pública é a do acesso livre e universal. As bibliotecas digitais tenderão a seguir este princípio, mas poderão ou terão de concretizar também novos métodos de controlo para responder a novos requisitos relacionados com a propriedade intelectual e termos e condições de acesso intrínsecas aos novos géneros de recursos com que terão de lidar.

Interactividade: As bibliotecas tradicionais suportam a comunicação com os utilizadores num sentido, com fraca interacção. As «Bibliotecas Digitais» tenderão a ser sistemas bidireccionais, com ciclos de interacção e capacidade de resposta rápidos, e mesmo com capacidade de encetar reacções pro-activas em relação às necessidades dos utilizadores.

Relacionamento social: As bibliotecas tradicionais são sistemas baseados num modelo com uma via de acesso, onde um consumidor acede na procura de um recurso. As bibliotecas digitais tenderão a ser sistemas com duas vias de acesso, onde os produtores dos recursos também terão a possibilidade de procurar os tradicionais ou potenciais novos consumidores, podendo a partir daí estabelecer novos tipos de relacionamentos sociais.

METADATA: CARACTERIZAÇÃO

Da perspectiva apresentada no ponto anterior resultam os requisitos segundo a Figura 1, relacionados com os cenários da Figura 2.

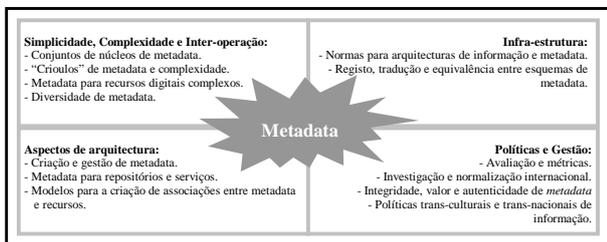


Figura 1: Contexto e requisitos do problema da metadata.

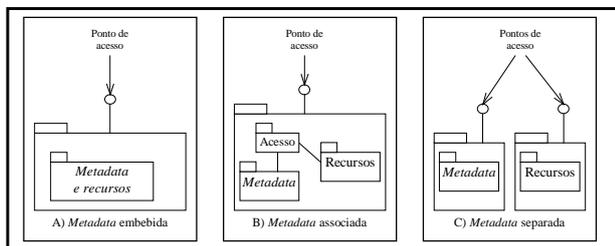


Figura 2: Cenários possíveis de organização e de pontos de acesso a estruturas de informação e às estruturas descritivas de metadata relacionadas.

O cenário da metadata embutada é aquele que se encontra, ou que pelo menos se tem tentado promover, para os recursos digitais em formato HTML – *Hypertext Modeling Language*. A perspectiva é que os recursos de informação incluam essa metadata embutada nos cabeçalhos dos

documentos (o formato HTML permite, através de uma "tag" <META>, a especificação de qualquer elemento de metadata pretendido, o qual será ignorado pelas máquinas de processamento de HTML mas poderá ser aproveitado por outros sistemas que o reconheça).

A metadata associada é a perspectiva que se pretende promover com o modelo emergente baseado em XML – *Extensible Markup Language*.

Finalmente, a perspectiva da metadata separada não anda longe do modelo da biblioteca tradicional, com bases de dados bibliográficas existentes em sistemas próprios e os recursos, isto é, os livros, existindo nas prateleiras.

A oferta de simples mas poderosos sistemas de criação de recursos digitais (editores de texto, ambientes de edição multimédia, etc.), permite hoje em dia concretizar facilmente qualquer projecto individual de publicação digital. Isto faz prever um futuro caótico, pulverizado de grandes e pequenos editores, com diferentes níveis de formalização. É pois de esperar que a quantidade de itens com que as bibliotecas venham a ter que lidar aumente de forma substancial, ao mesmo tempo que aumentará a sua especificidade, trazendo enormes problemas ao seu registo "manual" de acordo com as regras de catalogação tradicionais. Nesse mundo será difícil fazer vigorar regras normalizadas na criação de estruturas de metadata, especialmente se as vantagens no curto prazo não forem evidentes para quem for suposto dever segui-las.

Tudo isto perspectiva um cenário inevitável mas nada cómodo para as bibliotecas, obrigadas a encarar agora o facto de terem de reutilizar metadata criada no exterior e em contextos não normalizados. O problema complica-se ainda mais devido ao facto de a incerteza da dinâmica actual não permitir a definição segura de uma topologia para essa metadata. Tal deve-se não só à instabilidade no exterior da biblioteca (emergência de novos agentes, modelos e géneros de conteúdos), mas também à incerteza dos papéis que a «Biblioteca Digital» será chamada a desempenhar. Assim, não se julga possível estabelecer e sustentar uma topologia óptima para essa metadata, restando-nos apenas o campo da análise da situação actual e a tentativa da sua extrapolação.

Requisitos para metadata

Segundo as mais recentes iniciativas e projectos, podemos identificar requisitos para metadata na «Biblioteca Digital» para as necessidades apresentadas na Tabela 1.

Uma reacção imediata à identificação destas classes de requisitos pode levar aparentemente à conclusão da necessidade de normalização de estruturas próprias para cada uma dessas classes. A prática tem no entanto demonstrado que nem sempre terá de ser esse o caminho, pois o problema pertence também ao reino dos processos e dos contextos de utilização.

O projecto BIBLINK [2], por exemplo, foi um projecto Europeu com o objectivo de estudar até que ponto metadata produzida pelos editores tradicionais poderia vir a ser reutilizada nas bibliotecas. O ponto de partida foi a definição de uma estrutura de metadata própria, reflectindo essencialmente as necessidades das bibliotecas. No final tal

veio a mostrar-se ineficaz, devido essencialmente ao facto de os editores não fornecerem a informação nem no formato nem com os controlos de autoridades acordados. A justificação foi que tal implicaria mais custos para a sua estrutura, os quais não se mostraram dispostos a assumir. Temos assim aqui um claro erro de estratégia, já que teria talvez sido mais eficaz para as bibliotecas terem investido por um lado no conhecimento dos sistemas nativos utilizados por esses editores, e por outro na redefinição dos seus requisitos e sistemas próprios, de forma a poderem tirar partido dessa informação independentemente de como ela fosse produzida.

Objectivo	Descrição (com referência a contextos e projectos onde os requisitos têm sido abordados)
Descrição bibliográfica dos recursos	Descrição bibliográfica e identificação dos recursos, tais como títulos, autores, informação de indexação ou classificação, resumos, etc. (MALVINE [3], BIBLINK, NDLTD [4], DiTeD [5]).
Gestão administrativa dos recursos	Informação para administração do recurso, tais como elementos sobre a sua forma de aquisição, custo, condições de acesso e reprodução, etc. (BIBLINK).
Preservação dos recursos	Requisitos e condições, técnicas ou formais, para preservação a longo prazo (NEDLIB [6], PANDORA [7]).
Descrição estrutural e técnica dos recursos	Requisitos técnicos para manipulação do recurso (equipamento e aplicações), incluindo informação estrutural de autenticação, chaves de codificação ou de descodificação, etc. (NDLTD, DiTeD, SMIL [8]).
Acesso, uso e reprodução dos recursos	Estabelece termos e garantias de acesso e uso, registo de acessos (MALVINE, NDLTD, DiTeD).
Gestão administrativa da própria metadata	Informação sobre a própria metadata do recurso, nomeadamente a sua data e forma de criação, proveniência, autenticidade, etc. (DCMI [9], DiTeD).

Tabela 1: Classes de requisitos para metadata.

Atributos da Metadata

Segundo os seus atributos, identificam-se ainda quatro classes de atributos a levar em consideração para a metadata (os exemplos quando se referem são contextos onde esses atributos são suportados ou onde se podem verificar):

Origem e modo de criação: processo e agente criador, mecânico ou humano (exemplo: catalogador humano, palavras-chave extraídas por uma aplicação informática, informação fornecida pelo editor, etc.).

Estrutura e semântica: metadata organizada segundo estruturas semânticas normalizadas ou não, e ainda seguindo controlo de autoridades e vocabulários ou não (exemplo: Dublin Core, MARC, EAD – *Eencoding Archival Description*, utilização das AACR – *Anglo-American Cataloguing Rules*, etc.).

Estatuto: metadata estática, dinâmica, persistente ou temporária (exemplo: elementos estáticos como título ou autor, ou elementos dinâmicos como preço, condições de acesso e uso, etc.).

Nível: nível de descrição detalhado ou básico, descrição aplicada a itens ou a colecções, etc. (exemplo: Dublin Core, MARC).

INTER-OPERAÇÃO

Um outro requisito genérico que se impõe ainda no problema da metadata num paradigma de redes e recursos digitais é o de garantir a chamada inter-operação entre sistemas. O objectivo é o de tentar perceber que suportes

serão necessários para a oferta de um ambiente digital cooperante, com os utilizadores a beneficiar de acesso a mais que um sistema, sendo estes talvez mesmo heterogéneos. No fundo o problema é o de tentar perceber como poderá a «Biblioteca Digital» integrar o projecto da «Semantic Web» [10], isto é, a evolução da World Wide Web para um universo de acesso ontológico, onde os recursos passem a ser reconhecidas como recursos de informação em vez de simples dados. Duas peças fundamentais desse projecto, relevantes para a perspectiva da «Biblioteca Digital», são a DCMI – Dublin Core Metadata Initiative, e a linguagem RDF – Resource Description Framework.

Iniciativa Dublin Core

Na perspectiva de inter-operação da DCMI, não se pretende que a biblioteca importe metadata, mas sim que os sistemas onde ela se aloja e a forma como com eles se pode dialogar sejam conhecidos dos serviços que deles queiram tirar partido (por inter-operação).

A estratégia delineada pela DCMI tem sido a de abordar o problema de “baixo para cima”, numa perspectiva de criar uma solução baseada num modelo que seja extensível e adaptável às diferentes circunstâncias e contextos de aplicação. Assim, e partindo do princípio de que não só formatos de metadata especializados iriam proliferar no futuro mas também novas áreas e novos géneros de recursos teriam de vir a ser considerados, começou por se definir uma estrutura com um núcleo de elementos que pudessem ser usados como plataforma de inter-operação. Esta opção foi tomada em detrimento de uma atitude rígida ou global, que tentasse ver o problema de “cima para baixo” ou de uma forma excessivamente abrangente (segundo a filosofia MARC).

O resultado foi a definição de um grupo de quinze elementos, conforme se apresentam na Tabela 2 e se descrevem em apêndice. A primeira lista de elementos e as suas definições formais foram finalizadas na versão 1.0, posteriormente revistas na versão 1.1, de 11 de Julho de 1999, levando em conta as regras de definições de atributos ISO 11179.

Conteúdo	Propriedade intelectual	Instância
Assunto	Contribuinte	Data
Cobertura	Criador	Formato
Descrição	Direitos	Identificador
Fonte	Editor	Tipo
Língua		
Relação		
Título		

Tabela 2: Âmbito dos elementos do formato Dublin Core.

Os elementos deste núcleo de metadata podem-se apresentar em três grupos, de acordo com o seu âmbito. Cada elemento possui um nome descritivo que pretende transmitir o significado semântico do mesmo. A presença de cada um dos elementos num esquema de metadata é opcional ou susceptível de ser repetido quando necessário. Para promover uma inter-operação global, a descrição do valor de alguns dos elementos poderá ser associada ainda a vocabulários controlados, assumindo-se que esses

vocabulários controlados serão desenvolvidos para assegurar inter-operação em domínios específicos.

Este é assim um grupo simples de elementos de metadata mas que se espera funcional para descrever as características essenciais de documentos e artefactos digitais. A sua utilização principal deverá ser em cenários de inter-operação, para estabelecer correspondências entre diferentes sistemas especializados, conforme ilustrado na Figura 3. Seria assim possível assegurar por exemplo uma pesquisa cruzada de recursos entre os espaços descritos pelas estruturas A, B e C, sendo possível ainda tirar-se o máximo de partido do núcleo Dublin Core para o cruzamento entre os espaços B e C. A importância da definição deste núcleo é assim a de permitir o desenvolvimento de ferramentas e sistemas normalizados e que o reconheçam automaticamente, permitindo uma utilização generalizada e a baixo custo.

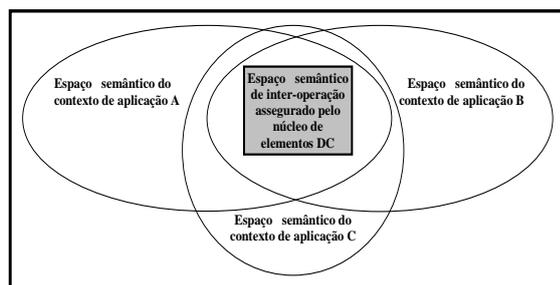


Figura 3: O núcleo Dublin Core como espaço de inter-operação (para descoberta de recursos) entre diferentes contextos de metadata.

A definição do Dublin Core permite ainda a especialização de cada um dos seus elementos de acordo com os contextos de aplicação. Assim, será possível por exemplo definir campos como “*creator.first*” e “*creator.second*” para recursos em contextos que exijam a identificação de um primeiro e segundo autor. Esse tipo de informação poderá ser usado de forma otimizada por ferramentas e sistemas que reconheçam esse “dialecto” definido a partir do Dublin Core, podendo ser ainda explorada de forma útil por outros sistemas e contextos que reconheçam apenas a núcleo ou outras especializações, caso em que considerariam apenas o campo “*creator*” como duplicado (e daí a importância de se definir a possibilidade de repetição de cada campo logo desde o início).

Muito se tem especulado sobre o valor de todo este trabalho, nomeadamente sobre o nível de penetração e qualidade dos registos Dublin Core criados e acessíveis na Internet. É um facto que não se atingiu o objectivo inicial de criar um formato e uma infra-estrutura técnica que pudesse gerir eficazmente a metadata associada a “cada página na Web”, mas talvez seja igualmente altura de reconhecer que esse objectivo acaba por não ser assim tão relevante face a outros resultados emergentes. São assim já vários os projectos e iniciativas que usam com sucesso o núcleo Dublin Core para representação primeira de metadata, com especial destaque para os portais e redes de portais temáticos usando a tecnologia do projecto ROADS [11]. De facto esta situação até acaba por ser preferível, pois foca a atenção do problema em meta-serviços especializados nessa área, e assim mais fáceis e

susceptíveis de serem reconhecidos e certificados por outros serviços que os usem do que os serviços isolados. Um outro modelo importante e relacionado com os mesmos objectivos da DCMI é o modelo GILS – *Governmental Information On-Line*. Criado nos Estados Unidos, este modelo destina-se a descrever, para efeito de descoberta, recursos governamentais [12]. Um recente entendimento entre esta iniciativa e a DCMI pretende promover a coordenação entre as mesmas, especialmente aos níveis de inter-operação.

RDF

O objectivo da definição da linguagem RDF – *Resource Description Framework*, é o de permitir a descrição de recursos por estruturas formais de metadata que possam vir a ser criadas e processadas automaticamente [13]. A sua definição tem sido promovida pelo W3C – *World Wide Web Consortium*, num esforço paralelo ao do desenvolvimento da linguagem XML. Esta definição consiste de duas partes, uma definindo um modelo e sintaxe e outra a forma de declaração de esquemas, ou contextos, baseados nessa sintaxe.

Um objectivo é que os sistemas que venham a ser criados para a manipulação de metadata possam reconhecer automaticamente e de forma otimizada descrições RDF. Os ambientes de criação de recursos (sistemas de “*authoring*”), assim como os clientes de pesquisa e acesso (“*browsers*”) e as máquinas e *robots* de pesquisa e indexação, passarão deste modo a reconhecer automaticamente não só objectos XML mas também as suas contrapartidas RDF.

Uma declaração RDF consiste num tríplice de conteúdo [recurso, propriedade, valor]. O recurso representa a entidade a descrever, tais como documentos, “*páginas web*”, serviços, etc. (tudo o que puder ser referenciado por um URI). A propriedade é a característica do recurso que se pretende declarar, cujos valores podem ser intrínsecos ao recurso (por exemplo, *author*=“Vitor”), ou consistir ainda em referências qualificadas para outros recursos, permitindo assim a construção de estruturas descritivas em árvore ou mesmo em rede (por exemplo, *author*=<http://authors.org/vitor>).

A linguagem RDF será assim o mecanismo ideal para a representação de qualquer estrutura de metadata, nomeadamente as baseadas em Dublin Core e suas possíveis extensões. Um obstáculo inicial a este propósito, detectado na definição original da linguagem RDF pelo grupo de trabalho DC-Multilingual, foi o da necessidade de dotar as propriedades de uma declaração RDF com um nome destinado a leitura humana. O modelo original requeria para cada propriedade apenas a declaração de um identificador, na forma de uma sequência de caracteres única no espaço de aplicação do momento, o qual se destinava a ser usado pelas máquinas processadores de RDF mas que poderia ser assim também usado para visualização humana. Tal mostrou-se inadequado para ambientes em que seja requerida a utilização de mais do que uma língua, o que tornou necessária a recomendação, posteriormente aceite, para a criação de um novo campo de descrição do nó, separado do seu identificador (permitindo deste modo passar a definir *author.value*=“Vitor” e *author.name*=“autor”).

Apesar de definida independentemente, a linguagem RDF pode ser declarada como uma aplicação de XML. Isto coloca-a em paralelo com a HTML, promovendo a perspectiva de que afinal metadata não será mais do que outro nome a dar a determinadas estruturas de dados ou informação, apenas dependendo talvez da perspectiva e do seu contexto de utilização. O facto de o valor de uma propriedade poder ser um recurso em si leva mesmo a que se possa redefinir um cenário com as duas especializações que se apresentam na Figura 4, em que o ponto de acesso a um recurso pode ser a sua página HTML, que contém então a sua metadata descritiva, ou poderá ser a própria declaração RDF, a qual contém por referência o recurso em si.

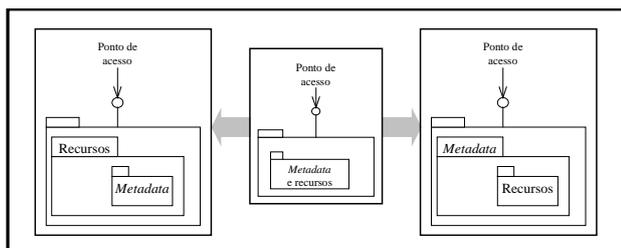


Figura 4: Cenários possíveis de metadata embebida descrita em RDF.

Estamos assim perante uma ferramenta de impacto fundamental na Biblioteca Digital, embora a sua aplicabilidade prática ainda seja alvo de muitas interrogações. O esforço posto nos últimos anos na definição de estruturas normalizadas como XML, RDF e Dublin Core tem sido tremendo a nível mundial, mas ainda sem resultados concretos no ponto de vista da área da metadata. Se é verdade que se tem demonstrado ser possível hoje em dia mudar radical e rapidamente certos paradigmas, também tem é verdade que muitos caminhos tiveram que ser arreados por falta dessa componente essencial que é a capacidade de transformar modelos teóricos em realidades concretas capazes de ser operadas a uma escala "normal", e não em contextos de excepção.

CONCLUSÕES

Como conclusão, será assim de esperar uma realidade em que a «Biblioteca Digital» terá de vir a lidar com múltiplas estruturas de metadata. Tal deve-se quer aos diferentes requisitos dos conteúdos e processos, quer às diferentes origens dessa informação. É mesmo de esperar que um mesmo recurso possa vir a ter, numa mesma biblioteca, múltiplos registos de metadata, nalguns casos sobrepondo-se mesmo. Para estes casos tem sido aventada a hipótese ou necessidade de cruzamentos entre essas múltiplas estruturas, o que levou à identificação da necessidade de registos formais para as mesmas.

Este é um problema novo, sendo a iniciativa relacionada mais relevante a esse nível actualmente em curso o projecto SCHEMAS, uma iniciativa de curto prazo, destinada a lançar luz sobre o problema e a servir de plataforma de discussão sobre os requisitos para esse tipo de sistemas.

CITAÇÕES E REFERÊNCIAS

1. SCHAUBLE, Peter; SMEATON, Alan (1998) - An International Research agenda for Digital Libraries: Summary Report of the Series of Joint NSF-EU

Working Groups on Future Directions for Digital Libraries Research. October 12, 1998.
<http://www.iei.pi.cnr.it/DELOS/NSF/Brussrep.htm>

2. <<http://www.echo.lu/libraries/en/projects/biblink.html>>
3. <<http://www.malvine.org>>
4. <<http://www.ndltd.org>>
5. <<http://dited.bn.pt>>
6. <<http://www.kb.nl/nedlib>>
7. <<http://www.nla.gov.au/pandora>>
8. HOSCHKA, P. (1998). Synchronized Multimedia Integration Language (SMIL) 1.0 Specification. World Wide Web Consortium, W3C Proposal Recommendation PR-smil-19980409, April 1998.
9. <<http://purl.org/dc>>
10. BERNERS-LEE, Tim; FISCHETTI, Mark (2000). Weaving the Web – The Past, Present and Future of the World Wide Web by its Inventor. Texere, 2000.
11. <<http://www.ilt.bris.ac.uk/roads>>
12. <<http://www.gils.net>>
13. MILLER, Eric (1998) - An Introduction to the Resource Description Framework. D-Lib Magazine, May 1998.
<<http://www.delib.org/dlib/may98/miller/05miller.html>>

APÊNDICE

Elementos do Núcleo de Metadata "Dublin Core", Versão 1.1: Descrição de Referência (versão em língua portuguesa)

Estado do documento: Este documento é uma Recomendação da Iniciativa de Metadata "Dublin Core" (Dublin Core Metadata Initiative). Uma recomendação significa que as especificações estão estáveis, e que são apoiadas para adopção pela comunidade "Dublin Core".

Este documento é a descrição de referência dos Elementos do Núcleo de Metadata "Dublin Core", versão 1.1. Este documento substitui a versão 1.0 do Conjunto de Elementos de Metadata "Dublin Core" (Dublin Core Metadata Element Set, version 1.0). Ver a página da Iniciativa "Dublin Core" (<http://purl.org/dc/>) para mais informação sobre os encontros de trabalho, relatórios, documentos dos grupos de trabalho, projectos, e novos desenvolvimentos relacionados com o Conjunto de Elementos de Metadata "Dublin Core".

Este documento resume as definições actualizadas para os elementos de metadata "Dublin Core", tal como foram definidos originalmente em [RFC2413]. Estas novas definições serão oficialmente conhecidas como Versão 1.1.

As descrições dos elementos de metadata são normalizadas. Esta normalização ajuda a melhorar a consistência com outras comunidades de metadata, permitindo ainda uma melhor clarificação, definição do alcance e consistência interna nas definições dos elementos de metadata "Dublin Core".

Cada elemento "Dublin Core" é definido usando um conjunto de dez atributos da norma ISO/IEC 11179 [ISO11179], para a descrição de elementos de dados. Estes consistem em:

- Nome: A etiqueta atribuída ao elemento de dado.
- Identificador: O identificador único atribuído ao elemento de dado.
- Versão: A versão do elemento de dado.
- Autoridade de Registo: A entidade autorizada a registar o elemento de dado.
- Língua: A linguagem na qual o elemento de dado é definido.
- Definição: Uma afirmação que representa claramente o conceito e a natureza do elemento de dado.
- Obrigação: Indica se o elemento de dados é sempre obrigatório ou não (contém um valor).
- Tipo dos Dados: Indica o tipo de dados que podem ser representados no valor do elemento de dado.
- Máxima Ocorrência: Indica qualquer limite à repetição do elemento de dado.
- Comentário: Uma nota relativa à aplicação do elemento de dado.

Seis dos referidos dez atributos são comuns a todos os elementos "Dublin Core". Esses são, com os seus respectivos valores:

- Versão: 1.1
- Autoridade de Registo: Dublin Core Metadata Initiative
- Língua: pt ("en" para a versão original)
- Obrigação: Opcional
- Tipo dos Dados: Cadeia de caracteres
- Máxima Ocorrência: Ilimitada

Estes atributos não serão repetidos nas definições seguintes, mas é importante ter presente que eles fazem parte da definição formal dos elementos.

As definições aqui fornecidas incluem ambas as formas conceptuais e de representação dos elementos "Dublin Core". O atributo Definição captura o conceito semântico, enquanto os atributos Tipo dos Dados e Comentário capturam a representação dos dados. Por sua vez o atributo Identificador deverá ser entendido não como um nome mas apenas como uma etiqueta ou código único identificador do elemento, o qual será sempre representado pela palavra em língua Inglesa equivalente ao valor do atributo Nome (é esta etiqueta que deverá ser identificada pelas aplicações e sistemas que interpretem e processem registos de metadata "Dublin Core", por isso a necessidade de a manter numa forma única e compatível para qualquer aplicação, independentemente da língua do contexto).

Cada elemento "Dublin Core" refere-se ao recurso a ser descrito. Um recurso é definido em [RFC2396] como "tudo o que tiver uma identidade" ("anything that has identity"). Para o presente efeito, um recurso será tipicamente um serviço ou um recurso de informação, mas o conceito poderá ser aplicado de forma mais alargada.

Elemento: Título
 Nome: Título
 Identificador: title
 Definição: O nome dado ao recurso.
 Comentário: Tipicamente, um Título será o nome pelo qual o recurso é formalmente conhecido.

Elemento: Criador
 Nome: Criador
 Identificador: creator
 Definição: A entidade responsável em primeira instância pela existência do recurso.
 Comentário: Exemplos de Criador incluem uma pessoa, uma organização, ou um serviço. Tipicamente, o nome de um Criador deve ser usado para indicar uma entidade.

Elemento: Assunto
 Nome: Assunto e Palavras-chave
 Identificador: subject
 Definição: O tópico do conteúdo do recurso.
 Comentário: Tipicamente, um Assunto deverá ser expresso por palavras-chave, frases, ou códigos de classificação que descrevem o conteúdo do recurso. Como boa prática recomenda-se a selecção de termos de vocabulários controlados, ou de sistemas de classificação formais.

Elemento: Descrição
 Nome: Descrição
 Identificador: description
 Definição: Uma descrição do conteúdo do recurso.
 Comentário: Descrições podem incluir, sem estarem limitadas a tal: um resumo, um índice, uma referência a uma representação gráfica do conteúdo, ou uma descrição textual.

Elemento: Editor
 Nome: Editor
 Identificador: publisher
 Definição: Uma entidade responsável por tornar o recurso acessível.
 Comentário: Exemplos de um Editor incluem uma pessoa, uma organização ou um serviço. Tipicamente, o nome de um Editor deve ser usado para indicar a entidade.

Elemento: Outro Contribuinte
 Nome: Outro Contribuinte
 Identificador: contributor
 Definição: Uma entidade responsável por qualquer contribuição para o conteúdo do recurso.
 Comentário: Exemplos de Outro Contribuinte incluem uma pessoa, organização ou serviço. Tipicamente, o nome de um Outro Contribuinte deve ser usado para indicar a entidade.

Elemento: Data
 Nome: Data
 Identificador: date
 Definição: Uma data associada a um evento do ciclo de vida do recurso.
 Comentário: Tipicamente, uma Data deve ser associada à criação ou disponibilidade do recurso. Como boa prática recomenda-se para codificação de valores de datas um perfil da norma ISO 8601 [W3CDTF], segundo o formato AAAA-MM-DD.

Elemento: Tipo
 Nome: Tipo do Recurso
 Identificador: type
 Definição: A natureza ou género do conteúdo do recurso.
 Comentário: Tipos incluem termos descrevendo categorias genéricas, funções, géneros, ou níveis de agregação para o conteúdo. Recomenda-se como boa prática a selecção de valores a partir de vocabulários controlados (por exemplo, a lista do documento de trabalho "Dublin Core Types" [DCT1]). Para descrever a manifestação física ou digital do recurso, deve ser usado o elemento Formato.

Elemento: Formato
 Nome: Formato
 Identificador: format
 Definição: A manifestação física ou digital do recurso.
 Comentário: Tipicamente, o Formato deve incluir o tipo de meio do

recurso, ou as suas dimensões. Este elemento deve ser usado para determinar as aplicações informáticas ou qualquer tipo de equipamento necessário para reproduzir ou operar com o recurso. Exemplos de dimensões incluem tamanho e duração. Como boa prática recomenda-se a selecção de valores a partir de vocabulários controlados (como por exemplo a lista de "Internet Media Types" [MIME] definindo formatos e meios).

Elemento: Identificador
Nome: Identificador do Recurso
Identificador: identifier
Definição: Uma referência não ambígua ao recurso, definida num determinado contexto.
Comentário: Como boa prática recomenda-se a identificação do recurso por meio de uma cadeia de caracteres ou por um número de acordo com um sistema de identificação formal. Exemplos de sistemas de identificação formais incluem o "Uniform Resource Identifier" (URI) (incluindo o "Uniform Resource Locator" (URL)), o "Digital Object Identifier" (DOI) e o "International Standard Book Number" (ISBN).

Elemento: Fonte
Nome: Fonte
Identificador: source
Definição: Uma referência a um recurso de onde o presente recurso possa ter derivado.
Comentário: O presente recurso pode ter derivado do recurso Fonte na sua totalidade ou apenas em parte. Como boa prática recomenda-se a referência ao recurso fonte através de um identificador em conformidade com um sistema de identificação formal.

Elemento: Língua
Nome: Língua
Identificador: language
Definição: A língua do conteúdo intelectual do recurso.
Comentário: Como boa prática recomenda-se para valores do elemento Língua a utilização do RFC 1766 [RFC1766], o qual inclui um código de língua de duas letras (retirado da norma ISO 639 [ISO639]), seguido opcionalmente por um código de duas letras para o país (retirado da norma ISO 3166 [ISO3166]). Por exemplo, 'en' para Inglês, 'fr' Francês, ou 'en-uk' para o Inglês do Reino Unido.

Elemento: Relação
Nome: Relação
Identificador: relation
Definição: Uma referência a um recurso relacionado.
Comentário: Como boa prática recomenda-se referir o recurso através de uma cadeia de caracteres ou número em conformidade com um sistema de identificação formal.

Elemento: Cobertura
Nome: Cobertura
Identificador: coverage
Definição: A extensão ou alcance do recurso.
Comentário: Cobertura inclui tipicamente uma localização espacial (o nome de um lugar ou coordenadas geográficas), um período no tempo (a sua designação, data, ou intervalo de tempo), ou jurisdição (o nome de uma entidade administrativa). Como boa prática recomenda-se a selecção de valores de vocabulários controlados (como

por exemplo o "Thesaurus of Geographic Names" [TGN]), devendo ainda ser usados, quando apropriado, preferencialmente nomes de lugares e designações de períodos no tempo, em vez de identificadores numéricos tais como coordenadas ou intervalos de datas.

Elemento: Direitos
Nome: Gestão de Direitos
Identificador: rights
Definição: Informação de direitos sobre o recurso ou relativos ao mesmo.
Comentário: Tipicamente, este elemento deverá conter uma declaração de gestão de direitos sobre o recurso, ou uma referência a um serviço que fornecerá essa informação. Tal poderá compreender informação sobre direitos de propriedade intelectual, direitos de autor, ou outros. A ausência deste elemento não permite formular qualquer hipótese válida sobre quaisquer direitos que possam incidir sobre o recurso.

Referências:

[DCT1] List of Resource Types. Dublin Core Draft Working Group Report. <<http://purl.org/DC/documents/wd-typelist.htm>>

[ISO11179] ISO 11179 – Specification and Standardization of Data Elements, Parts 1-6. <<ftp://sdct-sunsv1.ncsl.nist.gov/x318/11179/>>

[ISO639] ISO 639 – Codes for the representation of names of languages. <<http://www.oasis-open.org/cover/iso639a.html>>

[ISO3166] ISO 3166 – Codes for the representation of names of countries. <<http://www.oasis-open.org/cover/country3166.html>>

[MIME] Internet Media Types. <<http://www.isi.edu/in-notes/iana/assignments/media-types/media-types>>

[RFC1766] Tags for the Identification of Languages, Internet RFC 1766. <<http://www.ietf.org/rfc/rfc1766.txt>>

[RFC2396] Uniform Resource Identifiers (URI): Generic Syntax, Internet RFC 2396. <<http://www.ietf.org/rfc/rfc2396.txt>>

[RFC2413] Dublin Core Metadata for Resource Discovery. Internet RFC 2413. <<http://www.ietf.org/rfc/rfc2413.txt>>

[TGN] Getty Thesaurus of Geographic Names. <http://shiva.pub.getty.edu/tgn_browser/>

[W3CDTF] Date and Time Formats, W3C Note. <<http://www.w3.org/TR/NOTE-datetime>>

