

Um gato preto num quarto escuro – falando sobre metadados

Ana Alice Baptista

Altamiro Barbosa Machado¹

Metadados significa, basicamente, dados sobre os dados. Num ambiente caótico como a Internet, os dados já não são suficientes: são precisos metadados para a descrição semântica dos recursos. No entanto, os metadados per se são também insuficientes. Como acontece noutras áreas e, como acontece também em relação à descrição dos recursos físicos, a padronização é um elemento chave na sua utilização em grande escala. O Dublin Core (DC) e o Resource Description Framework (RDF) são duas recomendações de dois organismos diferentes: a DCMI (Dublin Core Metadata Initiative) e o W3C (World Wide Web Consortium). De forma a poder ser utilizado em larga escala, a DCMI optou por definir de forma ampla a semântica do DC, deixando as questões ligadas à sintaxe abertas e indefinidas. Esta é a razão pela qual o DC e o RDF combinam tão bem: o RDF traz as regras sintácticas nas quais os DC pode ser embebido. Neste artigo, faremos uma descrição geral do DC, do RDF e do RDF Schema, mostrando também alguns exemplos da duas aplicação. Posteriormente serão retiradas algumas conclusões sobre a sua aplicabilidade a falaremos sobre o seu futuro.

Palavras-Chave: *Publicação em linha, Metadados, XML, Dublin Core, RDF, Esquema RDF.*

1 INTRODUÇÃO

Num ambiente caótico como a Internet, os dados já não são suficientes. Embora os sistemas de indexação e de recuperação da informação estejam cada vez mais sofisticados, a relevância e a precisão nas respostas que os sistemas devolvem aos utilizadores ainda não alcançaram níveis satisfatórios.

De modo a melhorar a eficácia e a eficiência dos serviços de informação, torna-se necessária a utilização de metadados semânticos. No entanto, com a utilização de diferentes metadados e estruturas de metadados pelos diversos tipos de serviços e *software*, o problema mantém-se. Conforme acontece noutras áreas, a padronização é um elemento chave para a utilização eficaz dos metadados.

O *Dublin Core Metadata Element Set* (DCMES), geralmente denominado apenas de Dublin Core (DC), uma recomendação do *Dublin Core Metadata Initiative* (DCMI)², é um conjunto de 15 elementos para catalogação/descrição de objectos, suficientemente amplo e flexível para ser usado nas mais diversas situações.

De maneira a que tal flexibilidade seja possível, estes elementos estão definidos de uma forma muito imprecisa quer sintáctica, quer semanticamente. Sendo esta uma das suas grandes vantagens, permitindo a sua facilidade de utilização e rápida disseminação, também acabou por levar a variadas interpretações, utilização descoordenada e, por vezes, conflitual entre diferentes

projectos, especialmente no que se refere aos mecanismos de qualificação. Estes servem para fornecer mais precisão semântica aos elementos do DC e para indicar esquemas de codificação que ajudam na interpretação dos valores dos elementos do DC.

O RDF é uma recomendação do *World Wide Web Consortium*³ (W3C) para a descrição semântica de recursos (Lassila e Swick, 1999). A especificação do modelo e sintaxe do RDF (Lassila e Swick, op. cit.) fornece uma sintaxe e um modelo para a codificação em XML de vocabulários como é o caso do Dublin Core. O RDF permite a interoperabilidade entre aplicações, e a serialização em XML é uma das mais utilizadas devido, não só a promover essa interoperabilidade, mas também ao facto de o XML ser cada vez mais utilizado para “transportar” informação.

Nas duas secções seguintes explicamos mais profundamente o Dublin Core e o Dublin Core Qualificado, o RDF e o RDF Schema, recorrendo a alguns exemplos que demonstram a sua aplicação. Depois, retiraremos algumas conclusões sobre a utilização destas tecnologias agora e no futuro.

2 DUBLIN CORE E DUBLIN CORE QUALIFICADO

«O Dublin Core é um conjunto de elementos de metadados cujo objectivo é facilitar a descoberta de recursos electrónicos. (...) A construção de um consenso internacional e interdisciplinar à volta de um conjunto nuclear de elementos é a característica central do Dublin Core. O progresso representa o conhecimento emergente e a experiência colectiva de vários actores na arena da descrição de recursos.»⁴

O DCMI, na sua página *Web*⁵, reclama algumas características para *Dublin Core Metadata Element Set* (DCMES), que passamos a citar:

- Simplicidade;
- Interoperabilidade Semântica;
- Consenso Internacional;
- Extensibilidade;
- Modularidade de Metadados na Web;

A quase todos os elementos do DCES estão associados qualificadores de refinamento de elemento e/ou de esquema de codificação. A tabela 1 (Apêndice I) mostra os qualificadores até agora aprovados pelo *Dublin Core Usage Committee* (DCUC).

Os qualificadores de refinamento de elemento são, como o próprio nome sugere, outros termos associados ao elemento que especificam com mais precisão o significado desse mesmo elemento. «Um elemento refinado partilha o significado do elemento não qualificado, mas com um âmbito mais restrito» [3].

Por exemplo, o elemento *Date* tem como qualificadores de refinamento de elemento os termos *Issued*, *Created*, *Available*, *Modified* e *Valid*.

Os qualificadores de esquema de codificação, por seu turno, «identificam esquemas que ajudam na interpretação do valor de um elemento» (Dublin Core Qualifiers, 2000). Por exemplo, o mesmo elemento *Date* tem como qualificadores de esquema de codificação os esquema W3C-DTF⁶ e DCMI⁷ *Period*.

Os grupos de trabalho (*Working Groups* – WGs) do DCMI responsáveis pelo desenvolvimento dos qualificadores do DC⁸ realizaram diversos rascunhos de documentos (*Working Drafts* – WDs) quer sobre a utilização dos qualificadores nos projectos mais relevantes a nível mundial (Title: DC Agents..., 1999; Kunze; DC Relation..., 1999), quer sobre propostas de qualificadores para quase todos os elementos do DCMES (Kunze; Bearman, 1997; Iannella, 1999; Final Draft, 1999; Powell, 1999; Guenther, 1999; DC Working Draft, 1999; DC Date Qualifiers, 1999 e Rusch-Feja, 2000). A última entrega de rascunhos de propostas de qualificadores foi em Dezembro de 1999. O resultado da votação do *Dublin Core Usage Committee* (DCUC) foi tornada pública nas listas de correio electrónico em 17 de Abril de 2000. Em 11 de Julho de 2000 o DCMI fez uma *press release* (Khudairi, 2000) na qual estes qualificadores (Dublin Core Qualifiers, 2000) foram apresentados ao público em geral.

Existem muitos projectos em todo o mundo a usar o Dublin Core, e alguns a usar o Dublin Core Qualificado há já algum tempo (eLib Collection⁹; vCard¹⁰; OCLC CORC Project¹¹; BIBLINK Project¹²; The SuperJournal Project¹³; Nordic Metadata/BIBSYS¹⁴; Australian Government Locator Service (AGLS)¹⁵; National Library of Australia¹⁶; Gateway to Educational Materials (GEM)¹⁷; The Arts & Humanities Data Service (AHDS)¹⁸; John Wiley & Sons¹⁹; entre outros) (Title: DC Agents..., 1999 e DC Relation, 1999). No entanto, estes qualificadores são específicos a cada projecto e, desta forma, diferem numa extensão maior ou menor entre eles. O esforço, experiência e *feedback* destes pioneiros foram a base de trabalho para os Grupos de Trabalho em Qualificadores do DCMI poderem propor um conjunto de Qualificadores do DC ao DCUC.

3 RESOURCE DESCRIPTION FRAMEWORK (RDF)

O *Resource Description Framework* (RDF) é «uma aplicação XML recomendada pelo W3C para codificar, fazer o intercâmbio e reutilizar metadados normalizados (...) O RDF é desenhado para facilitar ao *software* perceber o suficiente sobre um *Web site*, de modo a que possa descobrir recursos, catalogar o conteúdo do *site*, escalonar esse conteúdo, perceber que possui o conteúdo e sob que termos e a que preço este pode ser usado, e fazer outras coisas que um *Web spider* ou um agente inteligente possa querer fazer» (Harold, 1999).

Um *statement* RDF faz declarações sobre recursos, usando um propriedade e tendo com resultado da aplicação dessa propriedade ao recurso, um valor. Um *statement* pode ser visto como um triplo composto por três elementos: recurso (sujeito), propriedade (predicado) e valor (objecto). Um recurso pode ser qualquer

coisa identificável por um *Uniform Resource Identifier* (URI) (Harold, Op. Cit.). O modelo RDF é simplesmente um modelo de triplos, o que o torna muito poderoso, mas difícil de implementar. O *parser* RDF é responsável por ler, verificar a sintaxe RDF transformar o código escrito na sintaxe RDF num conjunto de triplos e, eventualmente, num grafo RDF. Este não passa de uma representação gráfica desse conjunto de triplos: um grafo em que cada propriedade, ou predicado, é representado por um arco. Depois, só nos restam os sujeitos e os objectos que, sendo recursos, podem também ser sujeitos de uma outra propriedade. E assim por diante.

A Figura 1 (apêndice I) apresenta um exemplo de uma descrição RDF de um documento que é um comentário a um artigo científico. O código aí exemplificado corresponde à tabela de triplos apresentada a seguir (ver tabela 2) e ao grafo apresentado na Figura 2.

O RDF está dividido em duas partes, compreendendo duas especificações diferentes: (1), O *RDF Model and Syntax Specification* (RDFMSS) que é uma recomendação do W3C e que «apresenta um modelo para representar metadados RDF, assim como uma sintaxe para codificar e transportar metadados de uma forma que maximize a interoperabilidade de servidores e clientes *Web* desenvolvidos independentemente» (Lassila e Swick, Op. Cit.); e (2) O *RDF Schema Specification* (Brickley e Guha, 2000) que é uma especificação candidata (Berners-Lee, 1995) do W3C desde 27 de Março de 2000 e que «define uma linguagem de especificação de esquemas²⁰.

Com o RDF Schema, podem-se desenhar e implementar de uma forma consistente, vocabulários de metadados específicos. Estes podem, ainda, ser mais desenvolvidos no seio de outros projectos gerando, assim, uma rede de esquemas de metadados. Por exemplo, determinados termos de um vocabulário a ser desenhado podem perfeitamente ser definidos como refinamentos de elementos do DC ou de outro qualquer vocabulário anteriormente definido. Na Figura 3 (apêndice III) pode-se ver um exemplo de parte de um esquema RDF em que se definem duas propriedades, sendo uma delas uma subpropriedade do elemento Type do DC.

4 CONCLUSÕES

Os metadados semânticos são, neste momento, incontornáveis à utilização eficiente e eficaz da Internet. De facto, assim, como aconteceu com os recursos físicos, também são necessárias formas de descrever capazmente os recursos electrónicos. De qualquer modo, dada a disparidade de sistemas a todos os níveis, a padronização é essencial, sob pena de os problemas relacionados com a descrição de recursos se manterem ou, até, agravarem.

Tanto o *Dublin Core Element Set*, como o *Resource Description Framework*, são recomendações de organismos amplamente reconhecidos a nível mundial, tanto pela comunidade científica como pela comunidade empresarial. Por outro lado, embora a especificação do RDF Schema ainda seja “apenas” uma

recomendação candidata, a sua utilização já está a ser feita, principalmente no seio da comunidade científica.

O aparecimento de novos vocabulários, sejam os seus termos derivação de outros ou não, e o seu registo em espaços próprios para o efeito (chamados os metadata registries), poderá levar à constituição de uma rede de metadados em cima da rede de dados ou de documentos, propriamente dita. Através do recurso a esta estrutura, abre-se caminho aos mecanismos de pesquisa e recuperação da informação para “entenderem” os metadados semânticos e, assim, poderem fornecer ao utilizador resultados mais fiáveis, relevantes e precisos.

É exactamente esse o objectivo de todo o trabalho em desenvolvimento à volta dos metadados semânticos.

5 FUTURO

O *Dublin Core Element Set* é já um *standard de facto* na Internet. É já de alguns anos a esta parte largamente utilizado quer em projectos científicos, quer em projectos e aplicações comerciais. Projectos conjuntos entre a DCMI e outras organizações para desenvolvimentos futuros aplicados às mais variadas áreas são uma constante. Quer estes vinguem, quer não, o núcleo, o conjunto central de elementos para a descrição de recursos electrónicos é já o DCES.

Não fornecendo uma sintaxe, existem várias maneiras de embeber o DCES em documentos web e em bases de dados. Apesar de a sintaxe RDF/XML ser uma das mais recomendadas, não é possível neste momento, no entanto, afirmar que será esta efectivamente a ganhar mais adeptos.

De facto, o RDF/XML não é fácil de aprender, constituindo esta, neste momento, uma das maiores barreiras à sua disseminação em larga escala. Existem alternativas, sendo a maior os *Topic Maps*, um padrão ISO a suscitar bastante curiosidade na comunidade científica. Embora concorrentes em determinados aspectos, o RDF e os *Topic Maps* são, por outro lado, complementares, o que pode levar a alguma aproximação futura.

Esta área é, neste momento, palco de uma grande convulsão resultante quer de diferentes interesses, quer do facto de haver ainda muitas coisas por clarificar e muita para aprender. Os primeiros passos estão a ser dados. As direcções não são claras, e é difícil prevêr, mesmo a curto prazo, o que vai acontecer. De qualquer modo, parece-nos que não estaremos longe da verdade se dissermos que, seja qual forem os padrões escolhidos, a utilização de metadados na Internet não tem retorno e será cada vez mais necessária e fundamental.

6 REFERÊNCIAS BIBLIOGRÁFICAS

[LASSILA, O.; SWICK, R. R. \(Ed.\). *Resource framework: model and syntax specification*. \[S. l.\] : W3C Consortium, 1999. Disponível em: <<http://www.w3.org/TR/REC-rdf-syntax/>>](http://www.w3.org/TR/REC-rdf-syntax/)

DUBLIN core: a simple content description model for electronic resources. [S. l.] : DCMI, [2000?].

[DUBLIN](http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/) core qualifiers. [S. l.] : DCMI, 2000. Disponível em: <<http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/>>

DC AGENTS WORKING GROUP REVIEW OF CREATOR/CONTRIBUTOR/PUBLISHER. *Qualifier usage*. Dublin Core : Metadata initiative, 1999. Disponível em: <<http://archive.dstc.edu.au/RDU/DC-Agent/qualifier.html>>

KUNZE , A. J. (Ed.). *Date element working draf.*, Dublin Core : Metadata Initiative [1999?]. Disponível em: <<http://mirrored.ukoln.ac.uk/dc/documents/wd-date-current.htm>>

DC RELATION/SOURCE WORKING GROUP. *Review of relation qualifier usage*. Dublin Core : Metadata Initiative, 1999. Disponível em: <<http://mirrored.ukoln.ac.uk/groups/relation-qualifierreview.htm>>

BEARMAN, D. (Ed.). *Relation element working draft*. Dublin Core : Metadata Initiative, 1997. Disponível em: < <http://purl.org/dc/documents/wd-relation-current.htm>>

IANNELLA, R. (Ed.). *DC agent qualifier*. Dublin Core : Metadata Initiative, 1999. Disponível em: <<http://www.mailbase.ac.uk/lists/dc-agents/files/wd-agent-qual.html>>

FINAL draft. Dublin Core : Metadata Initiative, 1999. Disponível em <<http://www.mailbase.ac.uk/lists/dc-coverage/files/wd-coverage-qual.htm>>

POWELL, A. (Ed.). *DC format working group: proposed format qualifiers*. Dublin Core : Metadata Initiative, 1999. Disponível em: <<http://www.ukoln.ac.uk/metadata/resources/dc/dc-format-wg/proposals-all.html>>

GUENTHER, R. (Ed.). *Type working group: list of resource types*. Dublin Core : Metadata Initiative, 1999. Disponível em: <<http://mirrored.ukoln.ac.uk/dc/documents/wd-typelist.htm>>

DC WORKING DRAFT. Dublin Core : Metadata Initiative, 1999. Disponível em: <<http://www.mailbase.ac.uk/lists/dc-subdesc/files/wd-subdesc-qual.htm>>

CHILDRESS, E. (Ed.). *DC date qualifier*. Dublin Core : Metadata Initiative, 1999. Disponível em: <<http://www.mailbase.ac.uk/lists/dc-date/files/prop-19991214.html>>

RUSCH-FEJA, D. (Ed.). *DC title working group proposal for title qualifier*. Dublin Core : Metadata Initiative, 2000. Disponível em: <<http://mirrored.ukoln.ac.uk/dc/groups/qualifierproposal-title.htm>>

KHUDAIRI, V. S. (Ed.). *Press release: releases recommended qualifiers*. Dublin Core : DCMI, 2000. Disponível em: <<http://www.zotgroup.com/development/dcmi/dcqualifiers>>

HAROLD, E. R. *XML bible*. Foster City : IDG Books Worldwide, 1999.
BRICKLEY, D.; GUHA, R. V. (Ed.). *Resource description framework schema specification 1.0: W3C candidate recommendation*. [S. l.] : W3C Consortium, 2000. Disponível em: <<http://www.w3.org/TR/2000/CR-rdf-schema-20000327/>>

BERNERS-LEE, T., *W3C technical reports and publication*. [S. l. : s. n.], 1995. Disponível em: <<http://www.w3.org/TR/>>

7 NOTAS

¹ O prof. doutor Altamiro Barbosa Machado, meu orientador de doutoramento e director do Departamento de Sistemas de Informação da Universidade do Minho, faleceu no dia 23 de Março do corrente ano. Por uma questão de fidelidade e honra à sua memória, mantenho o nome dele em todas as publicações relacionadas com o meu doutoramento que estava a escrever aquando do seu falecimento. (Ana Alice)

² <http://dublincore.org>

³ <http://www.w3.org>

⁴ The Dublin Core: a simple content description model for electronic resources, DCMI.

⁵ <http://dublincore.org>

⁶ <http://www.w3.org/TR/NOTE-datetime>

⁷ <http://dublincore.org/documents/dcmi-period/>

⁸ *Working Groups* directamente envolvidos nas propostas de Qualificadores: Agents, Coverage, Date, Format, Relation, Title, Type. Alguns destes WG's tratam mais que um qualificador.

⁹ <http://eee.ukoln.ac.uk/metadata/cld/simple/>

¹⁰ <http://www.imc.org/rfc2426>

¹¹ <http://www.oclc.org/oclc/research/projects/corc/>

¹² <http://hosted.ukoln.ac.uk/biblink/wp8/fs/bc-semantic.htm>

¹³ <http://www.superjournal.ac.uk/sj/>

¹⁴ <http://www.bibsys.no>

¹⁵ http://www.naa.gov.au/govserv/agls/user_manual.htm

¹⁶ <http://www.nla.gov.au/meta>

¹⁷ http://www.geminfo.org/Workbench/Metadata/GEM_Element_List.html

¹⁸ <http://ahds.ac.uk/public/metadata/discovery.html>

¹⁹ <http://www.wiley.co.uk>

²⁰ *Schema Specification Language no original em Inglês.*

