

RESENHA

MÉNDEZ RODRÍGUEZ, Eva. **Metadados y recuperación de información**: estándares, problemas y aplicabilidad en bibliotecas digitales. Gijón: Trea, 2002. 429 p.

O mundo depara-se com o tempo de uma nova explosão, a explosão digital, conseqüência das dinâmicas e revolucionárias formas de registro de conhecimentos que, na obra de Eva Méndez Rodríguez, estão relacionadas ao fenômeno da recuperação da informação e centradas, tanto na questão fundamental da normalização, quanto nas possibilidades metodológicas de aplicação da meta-informação para organizar os acervos digitais.

Não são poucos os desafios a enfrentar por quem, no contexto do atual milênio, se propõe a escrever uma obra que envolve, por um lado, aspectos relacionados aos problemas de recuperação da informação, como padrões e métodos para estruturar registros do conhecimento e, por outro, questões conceituais que a eles dizem respeito, no sentido, tanto de evitar imprecisões terminológicas ou descontextualizações, quanto de definir categorias em suas relações recíprocas. Enfocando o que se denomina metadados, fenômeno protagonista de um modelo que vem se impondo e despertando interesse no mundo da informação, o livro *Metadados y recuperación de información*: estándares, problemas y aplicabilidad en bibliotecas digitales (*Metadados e recuperação da informação*: padrões, problemas e aplicação em bibliotecas digitais) volta-se ao estudo dos aspectos relacionados com as iniciativas de descrição de recursos digitais a partir de suas estruturas.

Ao enfrentar o desafio, a autora, cuja história acadêmica localiza-se na Espanha, graduada em Biblioteconomia e Documentação na Universidad de León, licenciada e doutorada na Universidad Carlos III de Madrid, onde leciona desde 1997, revelou, ao defender a tese da qual resultou esta obra, uma requintada fórmula para integrar abrangência temática com profundidade analítica. Muito jovem ainda, a autora já produziu relevantes trabalhos, entre muitas participações como convidada em eventos internacionais, co-autoria do livro *Sociedad de la Información: política, tecnología e industria*, autoria e co-autoria de diversos trabalhos em periódicos e anais de eventos científicos, além de atividades como membro de entidades e grupos de trabalho.

O tema vem despertando a curiosidade e o debate científico, tanto no campo acadêmico, quanto nos setores profissionais e profissionalizantes, especialmente devido à posição revolucionária da Internet como fonte, depósito e recurso da informação e às

alternativas metodológicas para solucionar os problemas de recuperação, decorrentes das contradições entre as práticas tradicionais e o contexto em transformação.

Com cautela cartesiana, a autora desenvolve o texto em três partes, procurando primeiramente encadear as questões relacionadas ao processo de transição da informação à metainformação, com destaque para os metadados, em seguida aos aspectos da normalização a eles relacionados, incluindo-se os formatos, os esquemas e os padrões e, finalmente, ao acesso global à informação, incluindo a recuperação, os metadados e as bibliotecas digitais.

A primeira parte desenvolve-se em dois capítulos, sendo o primeiro destinado a questões teóricas sobre o conceito, a justificativa, o fundamento, a caracterização, o uso e a tipologia dos metadados, passando por aspectos relativos às precisões terminológicas e ao uso histórico do termo. Reúne um quadro referencial evidenciando ao leitor a necessidade de contar com metadados para otimizar a busca e a recuperação de informação na Internet e, mais precisamente, relacionando-os ao contexto metodológico das bibliotecas virtuais. Salienta que como termo e conceito principal da obra, a palavra *metadados*, em virtude de sua excessiva utilização e diferentes significados e do fato de ser considerada “de moda”, no mundo da informação, determinou um classificador para precisar seu alcance, conteúdo ou campo de aplicação. Ao estudar a evolução semântica do termo na literatura sobre sistemas de gestão de bases de dados, a partir da década de 1980 e até meados da seguinte, a autora não se furta a analisar criticamente todos os significados que o termo teve, tanto na sua trajetória, quanto nos diferentes setores do conhecimento atual. Encontra-o mais especificamente em comunidades envolvidas na gestão e interoperabilidade de dados geoespaciais, e em geral no planejamento e na manutenção dos sistemas de gestão de dados. Para estas primeiras comunidades de gestores de informação, que utilizavam metainformação de forma semelhante a como é atualmente entendida, os metadados referiam-se a um conjunto de normas industriais, junto a uma documentação adicional interna e externa, e outros dados necessários para a identificação, representação, interoperabilidade, gestão técnica, funcionamento e utilização dos dados abrigados em um sistema de informação. Esse componente normativo de estruturação e definição da informação é o que persiste no significado que foi conferido no início do século XXI, e aceito na obra.

Outro aspecto importante na definição dos metadados é que podem requerer enfoques diferentes, como, por exemplo, em relação aos recursos de informação local, estatal ou federal, tais como publicações governamentais, ou endereço Web de uma agência federal ou de uma administração pública; informação relacionada a museologia: apresentações do patrimônio cultural compostas de um catálogo de exposição, imagens, murais de textos e arte

folclórica; livros, revistas, periódicos, ou coleções de materiais relacionados pertencentes a bibliotecas digitais. Este último é privilegiado no trabalho, já que a edição eletrônica e as bibliotecas digitais são uma realidade que o profissional da informação tem de enfrentar com o esforço de normalizar e de estruturar, o que tem caracterizado seu trabalho durante séculos, e que agora deve adequar-se a este novo contexto informativo.

Além destas áreas de informação, com uma forte e já consolidada trajetória no âmbito dos metadados, outros coletivos trabalham de maneira semelhante, defendendo modelos estruturais aplicáveis a seus tipos de informação ou desenvolvendo formatos de metainformação *ad hoc*, como, por exemplo, os sistemas de informação geográfica e geoespacial, ou o mundo da informação de arquivo, do comércio eletrônico, do coletivo da informação vinculada à educação, assim como outros menos comuns ou mais específicos, como a aplicação da informação oceanográfica. Via de regra, para se garantir a utilidade de qualquer conjunto de dados, como por exemplo, um conjunto de dados estatísticos, é preciso incluir informação descritiva sobre os mesmos, tanto em relação aos sistemas de informação geográficos (SIG), que necessitam passar informação (metadados) sobre as aplicações dos dados, quanto aos sistemas de gestão de bases de dados (SGBD).

Todos os exemplos apresentados pela autora têm uma relação direta com a visão bipartida que ela apresenta nos capítulos que se sucedem, sobre a extensão e diferenciação do termo *metadados*, a partir da definição genérica, e ambígua, de “dados sobre os dados” e aos diferentes domínios suscetíveis de aplicá-los, com distintas interpretações do termo, que ela divide em duas tendências fundamentais: a tendência do *tudo metadados* e, por outro lado, a interpretação exaustiva que restringe o conceito de metadados ao contexto da informação eletrônica distribuída através da Web que entende os documentos como objetos, quer dizer, *metadados no sentido estrito*, a ela aplicados. Longe de querer construir uma classificação categórica do conceito de metadados, já que, em muitos casos, os mesmos autores reconhecem ambos sentidos, a autora admite que a divisão de posturas e sua apresentação em forma de dicotomia simplificam seu objetivo de passar uma visão profunda do conceito em questão.

Já, no segundo capítulo da primeira parte, a autora discorre sobre o problema da atribuição de metadados, refletindo sobre o processo técnico dos recursos eletrônicos que permitiram a evolução dos processos de catalogação à atribuição de metadados. Após discutir a quem se deve atribuir a sua criação, ela nos apresenta respostas às questões relativas ao *onde* e ao *como* armazená-los, assim como às conseqüências dessas decisões para a criação e funcionamento de um sistema de informações baseado em metadados. Finaliza este capítulo

tratando da metodologia para associa-los aos recursos de informação eletrônica, incluindo ferramentas e aplicações, entre as quais distingue as que se destinam à criação de metaetiquetas (*metatags*) e à criação/gestão de metadados, apresentando ainda considerações sobre os critérios de escolha de software para sua criação.

A segunda parte da obra, Normalização de metadados: formatos, esquemas e padrões é a parte mais técnica do livro e desenvolve-se em dois capítulos, o primeiro referente aos modelos de metadados e o segundo à normalização, propriamente dita. Inicia a sessão descrevendo os modelos de metadados, desde linguagens de marca até as propostas existentes. Do mesmo modo que na primeira parte, nessa apresentação a autora revela seu amplo referencial e conhecimento de causa, expondo as duas faces da normalização e as iniciativas, projetos e normas mais destacadas.

Correndo o risco da redundância, por já ter tratado nos capítulos anteriores dos modelos, a autora opta pelo aprofundamento mais descritivo e crítico sobre os principais esquemas de meta-informação, a partir das linguagens sobre as quais se constitui a WWW. Desse modo, justifica a não-inclusão do formato MARC, embora haja quem o defenda como o primeiro modelo de metadados. Analisa, portanto, apenas os esquemas e soluções surgidos à raiz da Internet, partindo das demandas concretas das comunidades no sentido de organizar, intercambiar e recuperar sua informação. Entre as linguagens para estruturação da informação, destaca inicialmente a *Standard Generalized Markup Language* (SGML), que permite a definição de metalinguagens e facilita o intercâmbio de informação de recursos eletrônicos amplamente estruturados. Analisa a transição da linguagem HTML (*HyperText Markup Language*), para a *eXtensible Markup Language* (XML). Este fenômeno ampliou a importância de uma forma normalizada para definir a estrutura dos documentos, mas apesar de todas as vantagens que oferece, não substitui o HTML, razão pela qual foi criada uma norma de transição (*eXtensible HyperText Markup Language*). O XHTML proporciona a sintaxe XML a documentos com elementos HTML, acrescentando vantagens a esta linguagem.

Atores da cadeia informativa estão sendo incentivados, portanto, a mudar suas metodologias de trabalho diante das transformações nas tecnologias Web e suas linguagens de marca. Ao analisar o intercâmbio de informação de recursos eletrônicos amplamente estruturados, a autora acena com modelos que possibilitam a representação da informação de maneira estruturada e flexível, mapeável e intercambiável entre diversos formatos. Entre as propostas e iniciativas de metadados, são contemplados primeiramente aqueles de propósito geral, ou seja, os destinados à representação e descrição de documentos eletrônicos, categoria

à qual se subordinam os metadados HTML (metaetiquetas), modelo sintático de definição para armazenar a expressão do conteúdo de um documento e como meta-informação mais básica e o Dublin Core, modelo semântico concreto para expressar o conteúdo de uma maneira homogênea. Quanto aos metadados de propósito específico, pretendem ajustar-se à temática e às particularidades da informação eletrônica para a qual são formulados. Para ilustra-los, a autora destaca como exemplo paradigmático a *Text Encoding Initiative (TEI Header)*, mencionando também outros esquemas, alguns de estrutura sintática prescrita e outros delineados independentemente da sintaxe que utilizem. Apresenta o RDF, como um “meta-modelo” de metadados, seus princípios e alcance, já que o *Resource Description Framework* é o padrão do W3C que possibilitará a Web Semântica e uma maior interoperabilidade e integração entre distintos sistemas de informação digital baseados em metadatos.

Finalizando a segunda parte da obra, a autora trata, no capítulo 4, da normalização, analisando criticamente a validade dos modelos, esquemas e formatos estudados no capítulo anterior e dos modos de enfrentar o problema da adoção de normas no entorno informativo da WWW. Parte do pressuposto de que a apreciação da Rede como um sistema global de informação fundamenta-se no estabelecimento de padrões e métodos consistentes para definir a informação. Discute o problema da padronização distinguindo tipos de padrões utilizados na gestão da informação: as normas relativas à estrutura, que definem os campos e suas relações; os padrões sobre o conteúdo, que abrangem tanto as regras de catalogação quanto a sintaxe dos dados e as normas destinadas à valorização semântica dos dados, ou seja, vocabulários específicos que definem a forma e a importância do conteúdo.

Ao tratar dos padrões de metadados, em relação à casuística e à formalização, a primeira circunstância destacada pela autora é a capacidade normalizadora de instituições como o Consórcio Web, distinguindo-se na Internet os padrões *de fato*, os *formais* ou de direito e os tipos mistos, que mesclam os anteriores, chamados padrões *formais de fato*. Em relação ao nível de implantação dos padrões utilizados, a autora cita os padrões *reconhecidos e estáveis*, os *dependentes* ou utilizados em pequeno número de provedores ou empresas informáticas e os *proprietários*, cujo uso não se pode garantir. Nesse contexto, o destaque dado ao projeto SCHEMAS deve-se ao fato de que ele é o fórum das instituições e organizações que desenvolvem iniciativas de metadados na Europa, emprestando especial empenho na normalização e procurando ratificar iniciativas no sentido de padronizá-las. Desse esforço resultou um relatório de 22 iniciativas de normalização implantadas no continente europeu. Partem dos modelos, de padrões básicos, padrões interdomínios e padrões

de um domínio, os quais, por sua vez, aplicam-se a seis domínios concretos e apresentam sete fases de evolução em seu ciclo de vida, desde a identificação das necessidades, passando pela programação coletiva, as redações preliminares e os momentos de discussão, até chegar à aprovação do texto e recomendação. Para conhecer o nível de utilização de um formato de metadados e seu desenvolvimento no tempo, a autora seleciona o caso paradigmático do Dublin Core Metadata Initiative (DCMI), em cuja origem há uma orientação para a organização da informação eletrônica e que veio a se converter em um padrão americano.

Para fechar a última parte do capítulo, a autora refere-se à problemática relacionada à normalização do vocabulário que, malgrado os avanços tecnológicos e as decorrentes metodologias para um acesso mais rápido à informação, parece advertir que o entendimento da natureza do processo de indexação é mais crítico agora do que nunca. Assim, ela esboça outro aspecto da normalização como os padrões de valor semântico (tesauros, ontologias, classificações etc), cuja descrição corresponde a uma estrutura de metadados, discutindo suas características e peculiaridades, para concluir pela necessidade de incentivar a interoperabilidade, tanto do ponto de vista das estruturas semânticas dos dados, como dos valores dos metadados.

A terceira parte da obra, sobre o acesso global à informação, inclui três capítulos densos, relacionando, no primeiro, os metadados à recuperação de informação na Internet; no segundo, às bibliotecas digitais e, no terceiro e último, enfocando o trinômio globalização, metadados e acesso à informação.

Iniciando o capítulo cinco, onde discute a recuperação na rede, a autora tece considerações gerais sobre os aspectos distintivos do *browsing* e da navegação e o paradigma de interrogação, apontando para o problema da relevância. Discute a seguir as tendências dos sistemas de busca Web e a busca em texto completo frente à busca em metadados, analisando vantagens e inconvenientes de cada um dos tipos.

As bibliotecas digitais relacionam-se aos metadados, no capítulo seguinte, em que é possível acompanhar a discussão sobre a biblioteca como conceito e sua diversidade designativa, especialmente ao se apresentar a questão sobre o que é de fato uma biblioteca digital, virtual ou eletrônica. A autora discute sobre o oxímoro do termo bibliotecas virtuais e sua realidade prática, apreciando essa realidade sob dois aspectos, enquanto produtora ou possuidora de sua coleção digital e em sua relação com o fenômeno *subject gatewa*, e o conceito de coleção distribuída.

No último capítulo do livro a autora posiciona-se ao tratar mais especificamente do acesso universal à informação, defendendo a integração metadados e interoperabilidade e conclui com as perspectivas de futuro, direcionadas a uma Web Semântica.

Pode-se inferir que os conceitos principais da obra, os metadados e as bibliotecas digitais, apresentam-se num contexto teórico fartamente analisado e discutido, convergindo para soluções metodológicas extensamente analisadas, equilibrando-se com a profundidade esperada de um texto resultante de estudos de doutorado. Além de apresentar o estado da questão e as posturas teóricas diferenciadas em cada momento, a autora apresenta com convicção e sistematicamente seus próprios argumentos. Assim juntamente com o Doutor José A. Moreira González, professor catedrático da Universidad Carlos III de Madrid, que apresenta o livro, pode-se afirmar que uma obra como esta deve ser bem-vinda, pois “reúne tantos e tão representativos significantes, que seu término causa enorme satisfação pelos aspectos de qualidade e profunda análise que o leitor poderá comprovar em suas páginas”.

Essa competência manifesta-se, por outro lado, na erudição da autora e na atenção que dá ao produto de suas reflexões, contemplando o leitor com um glossário explicativo da terminologia utilizada, um índice de siglas e acrônimos, que remete aos endereços eletrônicos de cada termo e uma extensa bibliografia, de indiscutível qualidade estrutural e gráfica.

É importante observar que, além de ser a primeira obra sobre o tema, publicada na Espanha, sua autora é um dos nomes que mais tem se destacado no campo específico dos metadados, transcendendo as fronteiras europeias do mundo acadêmico e profissional.

Quanto à seriedade e a forma equilibrada com que a autora tratou a documentação, pode-se observar no decorrer da leitura, especialmente no que se refere à abrangência temática, à qualidade, atualidade e uso crítico do quadro teórico, à pertinência das citações, à seqüência lógica e organizada das informações, que ela é proporcional aos cuidados editoriais com o livro. Este faz parte de uma coleção de títulos voltados à área de Ciência da Informação, cuja qualidade gráfica corresponde às expectativas do leitor mais exigente, o que vem se tornando uma tradição da Ediciones TREA, de Gijón (Astúrias).

Seria recomendável a edição traduzida da obra, mas enquanto isso não acontece, para facilitar a obtenção da obra pelos leitores, seu representante pode ser encontrado no endereço jmpujol@pujolamado.com, de Pujol y Amado, coordenador para a América. No Brasil também se pode contatar diretamente a importadora LETRAVIVA, responsável pelos pedidos. Sempre se encontrará informação atualizada sobre o fundo editorial, novos títulos, representantes e distribuidores no *site* da editora, em www.trea.es.

Em suma, vale a leitura do livro, pois acompanhar a trajetória protagonizada pelos metadados e as bibliotecas digitais, animada pela dialética das metamorfoses tecnológicas, é um exercício que poderá servir de contraponto às análises mais ligeiras que se vêm apresentando sobre o tema.

[Leilah Santiago Bufrem](#)

Professora Titular do Departamento de Ciência e Gestão da Informação

Universidade Federal do Paraná

Av. Prefeito Lothário Meissner, 3400 – CEP 80210-170 Campus III – Jardim Botânico

Tel. (41) 360-4422 – Fax: (41) 360-4420 Curitiba – Paraná

Originais recebidos em 02/08/2004