

A VISIBILIDADE DOS DADOS DE BIBLIOTECA NA WEB [Julho/2015]

Normas e padrões da área são os suportes da atividade bibliotecária. A informação estruturada e codificada permite implementar sistemas de recuperação que otimizam a circulação e o uso dos recursos de biblioteca.

A biblioteca e seus procedimentos têm características particulares que possibilitam que a informação tratada, possa ser intercambiada com outras bibliotecas aderentes aos mesmos padrões.

Apesar da eficiência testada, estes procedimentos estão restritos ao ambiente da biblioteca. Torna-se uma limitação significativa se o fluxo da informação normatizada for solicitada por distintas áreas do conhecimento. A comunicação ou o intercâmbio se não impossível são, na melhor das hipóteses, pouco aproveitáveis, quando não sujeitos a alterações.

Ao visualizar o ambiente atual da Web, e o contínuo fluxo de informações ajustáveis às mais diversas configurações de sistemas e serviços digitais, verifica-se que os recursos de biblioteca não são compatíveis com um cenário de interação de dados.

Os recursos, no padrão bibliotecário, tornam-se invisíveis aos usuários deste ambiente digital. E, se a informação não é usada, ela é desnecessária. A biblioteca enquanto organismo responsável pela disponibilização de conteúdos registrados, torna-se esvaziada do interesse público.

Assim, este preâmbulo é para introduzir algumas das tendências conceituais que permeiam a área da Biblioteconomia. Sinalizar aos bibliotecários, em especial os catalogadores, sobre a amplitude de demanda para a atividade biblioteconômica em produzir dados bibliográficos que sejam aproveitados na esfera da Rede.

Nesse propósito, toma-se como base o artigo "[Ending the Invisible Library: Linked Data](#)", de Matt Enis, que aborda as tendências de mudanças dos recursos bibliográficos ajustados ao conceito de [linked data](#) (dados ligados ou vinculados), e explorados por ferramentas de busca na web. Uma indagação diante da situação atual: Qual a utilidade de uma busca semântica baseada no conceito dos dados vinculados?

A explicação é comentada por Jeff Penka, diretor de desenvolvimento de produtos da empresa [Zepheira](#), por meio de um exemplo de pesquisa realizada na ferramenta de busca do Google, com o termo "[Chevy Chase](#)". Além da lista de ligações retornada, um painel aparece na tela do computador exibindo fotos do ator acompanhada de uma biografia resumida, data de nascimento, altura física, nome completo, dados familiares: cônjuges e filhos, e uma lista de filmes e de programas de televisão por ele estrelado. Ao acrescentar à consulta as letras "MD", o painel se altera ao mostrar imagens, dados de clima e outras informações sobre a cidade de "Chave Chase, MD", nos Estados Unidos. Podemos repetir a pesquisa com personalidades ou regiões brasileiras, e o painel trará informações destacadas do termo consultado.

Este recurso do Google é denominado [Knowledge Graph](#), uma base de conhecimento lançada em maio de 2012, contendo mais de 500 milhões de dados coletados em fontes que incluem [Freebase](#), [Wikipedia](#), e a [CIA World Factbook](#), além de outros 3.5 bilhões de fatos sobre e relacionados entre esses e outros dados. Há, ainda, informações baseadas em assuntos que as pessoas procuram na web. Visa melhorar os resultados das ferramentas de busca com informações de pesquisa semântica.

Com o crescimento do Knowledge Graph, o recurso do painel tornou-se um diagrama popular que antecipa o que os usuários buscam, exhibe fatos específicos sobre pessoas, organizações, lugares, eventos e fatos históricos, e outros tópicos relacionados em formato de apresentação conveniente e acessível. Este aspecto demonstra que a tecnologia galga um uso generalizado. Recorde-se que a web teve início como uma coleção de páginas estáticas, navegáveis por links. Na atualidade, a web é um ambiente cada vez mais de dados e das relações entre os objetos destes dados.

Se o bibliotecário desejar saber mais sobre o significado desta teia de dados terá, certamente, que familiarizar sua linguagem e saberes com os conceitos da tecnologia semântica, linked data, BIBFRAME e Schema.org. Terminologias que encerram a maneira como o bibliotecário irá se comunicar, e dispor os seus recursos bibliográficos de forma significativa na web. E para isso, requer promover alterações em seus protocolos de normalização de dados.

Textos e eventos, sobre a questão, ressaltam que os bibliotecários (de países anglo-saxão) têm consciência das mudanças, em especial, no que se refere ao formato de catalogação legível por máquina ([MARC](#)), tornado praticamente anacrônico em um mundo interconectado em rede.

A percepção é reforçada pelo fato de nenhuma outra profissão adotar o formato MARC. Segundo Roy Tennant, autor de "[MARC Must Die](#)", quando a biblioteca adquire um software para lidar com os registros bibliográficos estruturados no formato, está limitada ao nicho de mercado dos fornecedores de sistemas. De seu lado, os fornecedores projetam sistemas que basicamente recebem e fornecem registros em MARC.

A situação aponta um problema delicado para as bibliotecas, com a obsolescência do Formato MARC. No Brasil, muitas bibliotecas só agora, no século 21, vivem o rejuvenescimento com a substituição de seus catálogos bibliográficos em fichas, ou migrando a base catalográfica construída em WinISIS para sistemas baseados no formato MARC 21. Nem saíram de uma situação e se deparam com outra.

Em 2011, as empresas responsáveis pelas principais ferramentas de busca na web, como Google, Yahoo e Microsoft (Bing), anunciaram parceria denominada [Schema.org](#) para desenvolver esquemas de dados estruturados. A finalidade é permitir aos projetistas web incluir informações em sites que identificam entidades e os relacionamentos existentes entre elas.

Para o internauta, o resultado da parceria pode ser exemplificado da seguinte forma: o site oficial de um artista poderia usar a marcação estruturada dos dados para indicar a programação do artista (local, endereço e data da performance). Quando a ferramenta de busca rastreia a web, lê as marcações, o que pode melhorar os resultados das pesquisas. No caso Google, permite que o painel do Knowledge Graph exiba os resultados relacionados. É, neste ambiente das ferramentas de busca, que os registros formatados em MARC têm pouca relevância.

Em outro exemplo, utiliza-se o serviço de busca do Google para pesquisar bibliotecas norte-americanas, como a [New York Public Library](#) (NYPL). Um painel da Knowledge Graph, caracteriza a biblioteca como modelo de negócio, com logotipo e endereços. Se acrescentar à consulta do nome da NYPL, o título de algum best-seller, a ferramenta de busca irá retornar informações relacionadas com o serviço [Overdrive](#) e [Bibliocommons](#) uma vez que estes sistemas apresentam dados em formato rastreáveis por motores de busca.

É provável que as bibliotecas possam se beneficiar com estes dados relacionados do OverDrive, se o nome de uma instituição for procurado, em conjunto com o título de um livro. Em geral, os usuários não utilizam as ferramentas de busca desta forma. Porém, mesmo que a utilizem, certamente, iriam visualizar uma relação de links para sites de bibliotecas ou suas direções no Facebook, contendo alguma menção sobre a obra, eventos já passados, e outras informações relacionadas aos serviços bibliotecários.

Para Kevin Ford, coordenador de projetos digitais no [Network Development and MARC Standards Office](#), da Library of Congress (LC), é preciso dar início a um intercâmbio e formatação dos dados bibliográficos que possibilite torná-los visíveis e vistos (utilizáveis) por outras organizações do universo digital (Facebook, Yahoo, Microsoft, Google, etc.). Baseado neste princípio, se esboça alguns dos objetivos do Bibliographic Framework for the Digital Age([BIBFRAME](#)), até pelo fato de os usuários de bibliotecas iniciarem suas buscas a partir destas ferramentas da web.

O projeto do BIBFRAME teve sua gênese no ano de 2011, como decorrência dos testes de aplicação da RDA (Recursos: Descrição e Acesso), conduzida pela National Agricultural Library, National Library of Medicine, LC e outras duas dezenas de instituições. Os testes suscitaram pouca confiança entre várias das bibliotecas participantes, quanto as mudanças preconizadas pela nova norma descritiva (em relação ao AACR2), se haveria benefícios significativos sem o acompanhamento de mudanças complementares na concepção do velho formato MARC.

Em relação a este sentimento da comunidade bibliotecária norte-americana, Deanna Marcun, bibliotecária associada para serviços de biblioteca na Library of Congress, no documento "[A Bibliographic Framework for the Digital Age](#)", explica os planos da LC para o desenvolvimento de uma estrutura substituta ao Formato MARC. Destaca que a preocupação das agências bibliográficas, envolvidas no teste, teve como foco principal a estrutura do MARC e a dificuldade no relacionamento dos elementos bibliográficos, além da capacidade restrita de utilizar [Uniform Resource Locator](#) (URL) em um ambiente de linked data.

No Brasil, a adoção da RDA parece ocorrer de forma isolada e silenciosamente particular. O problema emanado nos testes norte-americanos não causa maior preocupação por parte da comunidade bibliotecária no país. Observações empíricas sugerem que para algumas bibliotecas brasileiras, o AACR2 e o formato MARC atendem satisfatoriamente as necessidades descritivas. De outro lado, como anteriormente mencionado, há bibliotecas que a pouco tempo experimentam a estrutura do MARC na codificação de seus registros. Entretanto, as implicações do conceito da web semântica e do linked data está distante do fazer de muitas agências bibliográficas nacionais, apesar dos fatos e acontecimentos se desenvolverem em tempo real na web.

A outra ação envolvida é o [schema.org](#), que segue em paralelo. Neste projeto se agrega o W3C Schema Bib Extend Community Group (também conhecido como [Schemabibex](#)). Trata-se de grupo de trabalho composto por bibliotecários, fornecedores de sistemas e instituições bibliográficas. O grupo desenvolve esforços para expandir a proposta do [schema.org](#) com vista a melhor representar as informações produzidas pelas bibliotecas para coleta e indexação pelas ferramentas de busca.

Debate em relação ao projeto [schema.org](#) salienta a desconsideração, com as peculiaridades dos recursos de biblioteca, em sua concepção. Segundo Carol Jean Godby, pesquisadora sênior da OCLC, o projeto carece da representação das entidades do Grupo I dos FRBR (Requisitos Funcionais para Registros Bibliográficos), os conceitos de obra, expressão, manifestação e item. Nem há distinção clara entre conteúdo e suporte. Definiu-se poucos relacionamentos entre obras. Também inexistente o conceito de coleção e série. Não há modelos de intercâmbio envolvendo os recursos de bibliotecas e as instituições que tanto fornecem ou recebem tais recursos, como as próprias bibliotecas, editoras, agregadores de conteúdo e provedores de dados.

Entretanto, representantes do [schemabibex](#), em respostas as críticas, observam que não há, no trabalho desenvolvido, o interesse de substituir o formato MARC pelo [schema.org](#). A intenção é encontrar uma forma melhor de representar o registro bibliográfico tendo como base qualquer formato descritivo (MARC, [ONIX](#), [Dublin Core](#), [EAD](#), etc.). Com o projeto [schema.org](#) a finalidade é efetuar melhorias gerais no processo de descoberta realizado pelas ferramentas de busca; e desenvolver acréscimos ao vocabulário do projeto que estaria habilitado para representar os recursos bibliográficos. Na aplicação do esquema houve a implementação em softwares bibliográficos de código aberto como: [Evergreen](#), [Koha](#) e [VuFind](#). Versões atualizadas do Evergreen já expressam o conceito linked data intercambiando dados com a LC e [WorldCat](#) da [OCLC](#).

Apesar das discussões sobre os distintos padrões, os trabalhos prosseguem. Em relação ao [schema.org](#), indicações dos textos: [SWIB 2014: RDFa with schema.org for web applications](#); e [White hat SEO:Structured Web Data](#).

Sobre o BIBFRAME, a LC optou por desenvolvê-lo a partir do zero. Em maio de 2012 formalizou parceria com a empresa Zepheira, visando impulsionar o desenvolvimento do novo formato. O foco principal, do projeto, foi migrar o formato MARC 21 para um modelo de dados vinculados preservando, tanto quanto possível, os aspectos benéficos e as características robustas do histórico formato, ao invés de simplesmente adaptar o [schema.org](#) para uso em biblioteca.

Para acelerar a substituição do MARC pelo BIBFRAME, consta dos planos da LC oferecer uma ferramenta gratuita para a conversão de registros MARC. Entretanto, a padronização de vocabulários ainda está em processo, bem como, a transição de padrões é estimado entre dois a cinco anos. A crença entre os profissionais envolvidos é que o BIBFRAME irá levar as bibliotecas para o mundo do linked data.

Basicamente, o que o formato BIBFRAME irá possibilitar é tornar a estrutura do registro bibliográfico coletável pelas ferramentas de busca de maneira que elas extraiam informações significativas. Por exemplo, o Google pode informar ao usuário por meio da análise dos dados bibliográficos, em qual biblioteca o livro está depositado, até mesmo, informar a localização física do material na coleção, e a sua disponibilidade de empréstimo.

Na observação de Jeremy Nelson, bibliotecário de metadados e sistemas do [Colorado College](#), em Colorado Spring (EUA), a abordagem sobre estruturação de dados que funciona com as ferramentas de busca tem o potencial de transformar a maneira como é descoberta o recurso da biblioteca, além de transformar a forma como a biblioteca investe na gestão dos dados.

O tratamento dos dados bibliográficos em formato BIBFRAME colabora para o desenvolvimento de novas plataformas de armazenamento de dados, não delimitado aos sistemas integrados de bibliotecas. Exemplo de plataforma é encontrado no [HTML5 Django Mobile Aristotle Library Apps](#), ferramenta desenvolvida para descoberta, acesso e gerenciamento de recursos de biblioteca.

O entendimento na área da Biblioteconomia é que o linked data pode produzir melhores resultados de pesquisa para os usuários. Outro impacto é sobre os sistemas integrados de biblioteca, que também se tornam arcaicos como o formato MARC. Certamente, devem ter uma nova concepção no processamento dos dados bibliográficos, ajustados às novas tendências, e dentro de um custo financeiro mais acessível, face a provável concorrência das ferramentas de busca como opção para a recuperação de registros bibliográficos.

A empresa Zepheira, parceira da LC, anunciou em 2014 a [Iniciativa LibHub](#), um projeto, ainda em fase de demonstração, com a finalidade de estabelecer redes de bibliotecas aderentes ao padrão BIBFRAME na vinculação de dados e de outros recursos. A intenção é demonstrar como o novo padrão pode aumentar a visibilidade dos recursos de biblioteca na web aberta.

Dados estruturados sob tais processos de marcação, e as interligações entre as instituições bibliográficas por meio dos seus dados, devem ajudar as ferramentas de busca a confirmarem com precisão a autoridade de referência das publicações, pessoas, organizações, lugares, eventos, e recursos eletrônicos.

Ademais, há a elevação da classificação da biblioteca no índice de resultados das pesquisas em geral e, potencialmente, dos seus recursos apresentados em painéis gerados pelo Knowledge Graph, com base nos dados estruturados.

A expectativa em torno do BIBFRAME não se resume a simples alternativa de substituição ao formato MARC, mas a de servir como processo de capacitação das bibliotecas em se comunicarem, unificadamente, de maneira a serem entendidas na web.

As análises precoces do BIBFRAME e LibHub indicam que estes formatos de ligação de dados expõem os recursos de biblioteca de forma mais ampla para o público. Também é destacado que os padrões podem ajudar as bibliotecas a garantir melhor controle sobre os seus próprios registros.

Uma das primeiras experiências no uso da Iniciativa LibHub é apresentada pela [Denver Public Library](#) (EUA), por meio da sua gerente de serviços de coleções Rachel Fewell, ao comentar que:

"Nós queremos ser capazes de avançar e atender os usuários da biblioteca onde eles estejam [na web]. Nós não queremos que os

usuários potenciais da biblioteca tenham que vir ao nosso site para entender o que estamos oferecendo [...]. Nós trabalhamos muito duro na criação de excelentes registros de catálogos, e queremos expô-los a tantas pessoas quanto possível, por isso olhamos a isto como um próximo passo lógico [...]. Por que esconder o nosso trabalho no formato MARC e que não está disponível à grande maioria das pessoas na internet?”

Outro produto promovido pela empresa Zepheira é o [Linked Data Readiness Assessment](#). Trata-se de um serviço para auxiliar as bibliotecas na análise do estado atual de suas coleções e sistemas, de forma a determinar quais os procedimentos a serem adotados para tornar visíveis os seus recursos. Em realidade é uma espécie de auditoria na qual as bibliotecas são submetidas à diagnóstico dos seus sistemas, relacionamentos, dados e objetivos institucionais. Um programa estruturado em etapas é definido, baseado em cronograma de eventos que deve permitir melhor compreensão dos impactos e de ganho de confiança em explorar as oportunidades que o Linked Data e BIBFRAME oportunizam.

Nota-se que qualquer mudança, enfrentada pela biblioteca, deve ser precedida de alguma metodologia de gestão que subsidie sua capacidade de enfrentamento, adequação e assimilação gerencial. Algo pouco usual nas bibliotecas brasileiras.

A empresa também oferece neste seu pacote, uma plataforma de educação online denominada “[Practical Practitioner](#)” para ajudar os bibliotecários gestores a disseminarem na equipe, o aprendizado sobre o novo padrão bibliográfico, e a compreensão dos conceitos de forma detalhada. O curso não é gratuito, seu custo é de \$1.400 dólares, com duração de cinco semanas. Inclui vídeo-aulas, exercícios práticos, fóruns de discussão e contato com especialistas, além de participação na comunidade de ex-alunos do programa.

A inserção dos recursos de biblioteca nas plataformas e mídias sociais da web não se dá por iniciativas isoladas ou voluntárias de dispor os dados bibliográficos na rede. Para tal finalidade, métodos e adequação de protocolos uniformes são projetados de maneira a preservar os princípios da Biblioteconomia, no que se refere a organização e representação da informação registrada. A finalidade é dar acesso, uso e reaproveitamento dos dados e recursos bibliográficos para ambientes distintos das bibliotecas. E que, assim, façam circular as informações por outras aplicações e serviços digitais.

A sobrevivência das bibliotecas e do trabalho bibliotecário, no mundo digital atual, não está na conservação radical de suas raízes, mas na maneira como elas poderão germinar, proveitosamente, em novos terrenos de cultivos e de manejos abertos da informação.

Sobre Fernando Modesto

Bibliotecário e Mestre pela PUC-Campinas, Doutor em Comunicações pela ECA/USP e Professor do departamento de Biblioteconomia e Documentação da ECA/USP.