

# Os benefícios do Linked Data para catálogos de bibliotecas: reflexões preliminares

**Paula Regina Ventura Amorim Gonzalez** (UFES) - paulaventuramorim@gmail.com

**Felipe Augusto Arakaki** (UNB) - fe.arakaki@gmail.com

**Morgana Carneiro Andrade** (Ufes) - morganaandrade@hotmail.com

**Ana Carolina Simionato** (UFSCar) - acsimionato@ufscar.br

## **Resumo:**

*Atualmente os catálogos possuem um estrutura delineada por uma proposta de intercambiar fichas catalográficas. Entretanto, na década de 90, com o desenvolvimento e propagação da Web, novas tecnologias surgiram para agilizar e ampliar as possibilidades de representação, recuperação e acesso à informações. Nesse contexto, vislumbrando cenários de possíveis aplicações dessas tecnologias em catálogos, o objetivo deste trabalho é discutir os benefícios do Linked Data, para multidimensionar as possibilidades e benefícios de acesso em catálogos. Para compor os procedimentos metodológicos desta pesquisa, optou por uma pesquisa exploratória e de caráter qualitativo. Para tanto, as bibliotecas precisam estruturar os metadados existentes de forma a se adequarem às diretrizes Linked Data. Como discorrido no trabalho, o enriquecimento dos metadados pelo conceito Linked Data e conjunto o uso das tecnologias semânticas disponíveis, permitem multidimensionar as possibilidades e benefícios de acesso em catálogos para a descoberta dos dados e recursos por motores de busca na Web. Nesse sentido, a aplicação do conceito Linked Data é delineada a partir da reestruturação dos dados de cada registro no catálogo, pensando em uma modelagem de dados que possa conduzir para os relacionamentos entre os mesmos, e não, para uma nova forma de visualização dos dados.*

**Palavras-chave:** *Linked Data, Catalogação,*

**Eixo temático:** *Eixo 12: V EEPC Encontro de Estudos e Pesquisas em Catalogação*



# XXVIII Congresso Brasileiro de Biblioteconomia e Documentação

Vitória, 01 a 04 de outubro de 2019.

## Introdução

Desde a década de 60, os catálogos possuem uma estrutura delineada por uma proposta de intercambiar fichas catalográficas. Essa estrutura faz com que os dados dos catálogos e registros bibliográficos tenham característica monolítica, isto é, os dados estão armazenadas em um único bloco, de forma sedimentar, que determina que a máquina compreenda o registro de forma linear, do início do registro pelos campos líder e pelos campos de controle, para assim, chegar na descrição do recurso informacional.

Entretanto, na década de 90, com o desenvolvimento e propagação da *Web*, novas tecnologias surgiram para agilizar e ampliar as possibilidades de representação, recuperação e acesso à informações. Nesse contexto, as linguagens de marcação e uma estrutura mais flexível tornam-se triviais para a interoperabilidade dos dados (ZENG, 2010).

Nessa perspectiva, a abertura dos dados é o princípio para que os ambientes digitais se adequem às novas tecnologias propostas. Ressalta-se que a estrutura atual dos catálogos, baseada ainda nas tecnologias da década de 60, impede a abertura dos dados das bibliotecas para que robôs de busca possam processar rapidamente os dados das bibliotecas e apresentar em um buscador (COYLE, 2010). Mesmo com a possibilidade de publicação do *Machine Readable Cataloging* (MARC) em *Extensible Markup Language* (XML), a estrutura de campos, subcampos e indicadores com diversos significados, dificulta a análise e o tempo de resposta dos robôs de buscas eficiente, que precisam fazer consultas mais rápidas em maiores quantidades de dados no menor tempo possível.

Ainda com a proposta dos modelos conceituais, *Functional Requirements for Bibliographic Records* (FRBR), *Functional Requirements for Authority Data* (FRAD), *Functional Requirements for Subject Authority Data* (FRASAD) e IFLA *Library Reference Model* (LRM), atrelada às tecnologias da *Web Semântica*, fomentem a fragmentação dos registros monolíticos para uma estrutura de registro modular, contendo diversos registros.

Vislumbrando cenários de possíveis aplicações dessas tecnologias em catálogos, o objetivo deste trabalho é discutir os benefícios do *Linked Data*, para multidimensionar as possibilidades e benefícios de acesso em catálogos.

Para compor os procedimentos metodológicos desta pesquisa, optou por uma pesquisa exploratória e de caráter qualitativo. Assim, foram buscados documentos na BRAPCI, *Web of Science*, *Scopus* e *Library and Information Science Abstract*. O período pesquisado abrange desde a proposta do *Linked Data* (2006) até o ano de 2018. Entre os trabalhos pesquisados, foram selecionados para análise os que abordam e apresentam os benefícios e possíveis aplicações do *Linked Data* em catálogos.

### **Benefícios da publicação de dados em *Linked Data***

A *Web Semântica* projetada por Berners-Lee, Hendler e Lassila (2001) foi idealizada como uma extensão da *Web*, viabilizando um ambiente onde esse conteúdo é inteligível por programas especiais ou agentes inteligentes de *software*, com a possibilidade de analisar e fazer inferências sobre os conteúdos dos documentos. A informação é fornecida com um significado bem definido, possibilitando o trabalho cooperativo com o uso dos computadores (BERNERS-LEE; HENDLER; LASSILA, 2001).

Berners-Lee (2006) descreve que a *Web Semântica* não deve ser caracterizada apenas com a publicação de dados na *Web*, mas sim, com a construção de relacionamentos entre esses dados e ligados por meio de estruturas semânticas propiciadas pelo uso do *Resource Description Framework* (RDF), *Resource Description Framework Schema* (RDFS), *Uniform Resource Identifier* (URIs), linguagem SPARQL, ontologias *Web Ontology Language* (OWL), *Simple Knowledge Organization Schemes* (SKOS), com base nas diretrizes do *Linked Data*.

Conceitua-se o termo *Linked Data* como um conjunto de melhores práticas para a publicação e ligação de dados estruturados na *Web* (BERNERS-LEE, 2006). Van Hooland e Verborgh (2014) reforçam esse conceito e esclarecem que o *Linked Data* não configura-se como uma ferramenta.

Em sua definição, Berners-Lee (2006) definiu quatro princípios que servem como sugestões de estruturação de dados. Cunha, Lóscio e Souza (2011) esclarecem que, o primeiro princípio justifica o uso de referências URI para identificação, documentos *Web* e conteúdos digitais, objetos do mundo real e conceitos abstratos. O segundo princípio argumenta o uso de URIs HTTP para identificar objetos e conceitos abstratos que são definidos pelo primeiro princípio, viabilizando essas URIs serem diferenciáveis sobre o protocolo HTTP, isto é, no processo de recuperação de uma representação, identificado por uma URI, em que este recurso pode possuir várias representações como documentos HTML, RDF, XML entre outros. O terceiro princípio defende o uso do RDF como modelo para a publicação dos dados estruturados na *Web*. A partir do RDF, oportuniza a

descrição de significados sobre recursos, qualificando agentes de software a explorar os dados de maneira automática, seja agregando, interpretando ou mesclando dados. O quarto princípio diz respeito de um URI a outro conjunto de dados, o que permitirá a interação de informações como por exemplo a ligação de um recurso a outro dentro na rede.

Para melhor direcionar e fomentar que mais instituições utilizem os princípios do *Linked Data*, o W3C publicou em 2014, um conjunto de melhores práticas para publicação de dados em *Linked Data* (W3C, 2014). Outro grupo de trabalho do W3C discutiu os diversos benefícios para pesquisadores, estudantes, organizações, bibliotecários, arquivistas, curadores, desenvolvedores e fornecedores, os problemas de conciliação dos princípios do *Linked Data* para dados de bibliotecas, além de apresentar recomendações para publicação de dados em *Linked Data* (W3C, 2011). De acordo com Coyle (2010), as bibliotecas possuem grande número de títulos e materiais raros, entretanto há muito sobreposição de conteúdo entre as informações dos catálogos e a Web. Outro ponto destacado pela autora, é que existem dados não bibliográficos que podem ser relacionados a dados bibliográficos.

Dentre os benefícios decorrentes da adoção do *Linked Data*, destacados por W3C (2011), Saorin (2012), Mitchell (2016) e Smith-Yoshimura (2018), pelas bibliotecas estão: necessidade de publicar *Linked Data* para reutilização em projetos futuros; potencialização da interoperabilidade e reuso de dados; fornecimento de dados estáveis, dados integrados e normalizados produzidos pelas instituições; otimização da descoberta e o reuso dos recursos e dos seus metadados em acesso aberto; criação de um conjunto de dados abertos e global compartilhados que podem ser usados e reutilizados para descrever recursos, com uma quantidade limitada de esforço redundante em comparação com os processos de catalogação atuais uso dos identificadores baseados na Web e fará com que as descrições de recursos atualizadas sejam citadas diretamente pelos catalogadores; uso de identificadores compartilhados permitirá reunir descrições de recursos fora de seu ambiente de domínio, em todos os conjuntos de dados de herança cultural e até mesmo na Web em geral; catalogadores poderão concentrar seus esforços em seu domínio de conhecimento local, em vez de precisar recriar descrições existentes que já foram elaboradas por outras pessoas; o *Linked Data* é mais durável e robusto do que os formatos de metadados que dependem de uma estrutura de dados específica.

Os benefícios decorrentes da adoção dos *Linked Data* pelas bibliotecas pode ser extensivo a toda Web. Coyle (2010) exemplifica que os verbetes da *Wikipédia* podem ser vinculados aos acervos de biblioteca e os dados das bibliotecas podem enriquecer informações bibliográficas na Web, pois, as

bibliotecas possuem controle sobre nomes e informações sobre os autores (data de nascimento e falecimento) que eram usados para diferenciar nomes semelhantes, agora são essenciais para análise do status dos direitos autorais, por exemplo.

Para tanto, essas bibliotecas precisaram preparar os metadados existentes de forma a se adequarem às diretrizes *Linked Data*. Passado alguns anos, nota-se que as bibliotecas em geral ainda não avançaram de maneira oportuna. Para alguns autores, isso pode ser reflexo da falta da compreensão generalizada pelos profissionais de informação (BOWEN, 2010). Os movimentos observados nesse sentido estão relacionados em parte pela adoção do *Dublin Core* em repositórios e dos novos instrumentos como a RDA e o BIBFRAME.

### **Considerações Finais**

Como discorrido no trabalho, o enriquecimento dos metadados pelo conceito *Linked Data* e conjunto o uso das tecnologias semânticas disponíveis, permitem multidimensionar as possibilidades e benefícios de acesso em catálogos para a descoberta dos dados e recursos por motores de busca na *Web*.

Nesse sentido, a aplicação do conceito *Linked Data* é delineada a partir da reestruturação dos dados de cada registro no catálogo, isto é, uma atualização das tecnologias utilizadas na construção do catálogo, e ainda pensando em uma modelagem de dados que possa conduzir para os relacionamentos entre os mesmos, e não, para uma nova forma de visualização dos dados.

De forma conclusiva, no uso do conceito *Linked Data*, a nova estrutura irá trazer uma maior dinamicidade, visibilidade e comunicação entre os dados já produzidos pela biblioteca, e o impacto direto nas ações dos bibliotecários estará atrelado ao reuso dos dados, diminuindo o retrabalho, as inconsistências e duplicações do mesmo valor no registro.

Considera-se emergente estas ações nas instituições e centros de informação, visto a busca incessante para a disseminação e provimento de acesso aos usuários, para um conteúdo que não está devidamente representado ou ainda que não multidimensione as possibilidades e benefícios de acesso em catálogos.

### **Referências**

BERNERS-LEE, T. **Linked Data**: design issues. [S.l.]: W3C, 2006. Disponível em: <http://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 20 mar. 2019.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. **Scientific American**, p. 29-37, 2001. Disponível em: [https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American\\_%20Feature%20Article\\_%20The%20Semantic%20Web\\_%20May%202001.pdf](https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf). Acesso em: 20 mar. 2019.

BOWEN, J. F. Moving library metadata toward linked data: opportunities provided by the eXtensible catalog. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, Pittsburgh, 20-22 Oct. 2010. **Proceedings...** Pittsburg, 2010. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/1010/979>. Acesso em: 20 mar. 2019.

COYLE, K. Library data in a modern context. **ALA TechSource**, n. 1, p. 5-13, 2010. Disponível em: <https://journals.ala.org/index.php/ltr/article/view/4630>. Acesso em: 19 abr. 2019.

CUNHA, R. B. C.; LÓSCIO, B. F.; SOUZA D. **Linked data**: da web de documentos para a web de dados. 2011. Disponível em: [http://www.cin.ufpe.br/~daise/arquivos/publications/2011/2011\\_Cap%204%20Linked%20Data%20-%20MC2.pdf](http://www.cin.ufpe.br/~daise/arquivos/publications/2011/2011_Cap%204%20Linked%20Data%20-%20MC2.pdf). Acesso em: 20 jun. 2018.

MITCHELL, E. T. The current state of linked data in libraries, archives, and museums. In: MITCHELL, E. T. (Org.). **Library Linked Data**: early activity and development. **Library Technology Reports**, v. 52, n. 1, p. 5-33, 2016.

SAORÍN, Tomás. Cómo linked open data impactará en las bibliotecas a través de la innovación abierta. **Anuario ThinkEPI**, v. 6, 2012. Disponível em: <http://www.thinkepi.net/como-linked-open-data-impactara-en-las-bibliotecas-a-traves-de-la-innovacion-abierta>. Acesso em: 20 fev. 2019.

SMITH-YOSHIMURA, K. Analysis of 2018 International Linked Data Survey for Implementers. **Code4lib Journal**, n. 42, Nov. 2018. Disponível em: <https://journal.code4lib.org/articles/13867>. Acesso em: 30 jan. 2019.

VAN HOOLAND, S.; VERBORGH, R. **Linked Data for Libraries, Archives and Museums**: how to clean, link and publish your metadata. UK: Facet Publishing, 2014.

W3C. **Best Practices for Publishing Linked Data**. W3C Working Group Note 09 January 2014, 2014. Disponível em: <https://www.w3.org/TR/ld-bp/>. Acesso em: 27 abr. 2019.

W3C. **Library Linked Data Incubator Group Final Report**. W3C Incubator Group Report, 2011. Disponível em: <http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>. Acesso em: 27 abr. 2019.

ZENG, M. L. Domain-specific markup languages and descriptive metadata: their functions in scientific resource discovery. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, n. esp., p. 166-176, 2º sem. 2010. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2010v15nesp2p164>. Acesso em: 27 abr. 2019.