



***Linked data* em bibliotecas: discussões iniciais de implementação no Sistema Integrado de Bibliotecas do Município de São Carlos (SIBISC)**

Linked data in libraries: initial discussions for implementation at Integrated Municipal Library System of the of São Carlos in Brazil

Felipe Augusto Arakaki, Universidade de Brasília – fe.arakaki@gmail.com

Ana Carolina Simionato Arakaki, Universidade Federal de São Carlos – acsimionato@ufscar.br

Eixo 12 - I Fórum de Catalogação

1 INTRODUÇÃO

O *Linked Data*, traduzido como dados conectados, foi apresentado em 2006 por Tim Berners-Lee e é considerado como as boas práticas para a abertura e conexão entre os dados abertos, apresentando uma construção aplicada ao “sentido” dos recursos digitais às máquinas.

O *Linked data* origina-se da materialização das tecnologias propostas para a *Web Semântica*. Sendo que os propostos da *Web Semântica* orientam ao desenvolvimento implementável de tecnologias para a estruturação dos dados, utilizando-se de uma série de tecnologias semânticas (BERNERS-LEE; HENDLER; LASSILA, 2001). Tais tecnologias abarcavam desde linguagens para a estruturação e a representação dos conteúdos como o *eXtensible Markup Language* (XML) e o *Resource Description Framework* (RDF), passando por protocolos para recuperação de dados como o *SPARQL Protocol and RDF Query Language* (SPARQL) e linguagens para a construção de ontologias como a *Web Ontology Language* (OWL), chegando até a estruturas básicas para a identificação única de conteúdos como o *Uniform Resource Identifier* (URI).

A partir dessas tecnologias semânticas, qualquer uma instituição, como exemplo as bibliotecas, podem disponibilizar os dados de seus acervos, a fim de cooperar com as redes de relacionamento e com a crescente interoperabilidade semântica e sintática entre esses dados. Desse modo, pensando no desafio e



dificuldade das bibliotecas brasileiras em realizar tais atividades, foi proposta uma parceria com uma biblioteca pública como estudo de caso para o desenvolvimento de procedimentos e metodologia para publicação de dados abertos conectados.

Ressalta-se também que o comportamento de busca de informações dos usuários foi alterado ao longo dos anos e com isso, há uma exigência para que as bibliotecas possuam catálogos adaptados a essa nova realidade. O problema de pesquisa se contextualiza pela questão de como fomentar a disponibilização e o acesso a dados de bibliotecas no contexto dos dados abertos conectados?

Desse modo, o objetivo desse trabalho consiste em apresentar as reflexões iniciais, como o processo de curadoria e higienização dos dados do Sistema Integrado de Bibliotecas do Município de São Carlos (SIBISC) para adequação dos procedimentos necessários para publicação de dados abertos conectados.

2 FUNDAMENTAÇÃO TEÓRICA

O uso do *Linked data* em bibliotecas ainda é incipiente, principalmente se considerar a realidade brasileira. Jesus (2021) fez uma revisão sistemática de literatura sobre o uso do *Linked data* em bibliotecas e analisou 20 instituições no que tinham indícios de uso de tecnologias semânticas e/ou que seguiam os princípios do *Linked data*.

Para que uma biblioteca possa disponibilizar seu conjunto de dados em *Linked data* precisa estar de acordo com alguns princípios e boas práticas. As primeiras orientações e princípios do *Linked data* foram estabelecidas por Tim Berners-Lee, criador do *World Wide Web* (W3C) e idealizador da *Web Semântica*. O autor estabeleceu quatro princípios para que os dados pudessem ser considerados conectados:

1. Use URIs como nomes para as coisas;
 2. Use HTTP URIs para que as pessoas possam procurar esses nomes;
 3. Quando alguém procurar um URI, fornecer informações úteis, usando os padrões (RDF, SPARQL);
 4. Incluir links para outros URIs, para que eles possam descobrir mais coisas.
- Berners-Lee (2006, não paginado, tradução nossa)



Posteriormente, no ano de 2014, foi proposta as boas práticas para publicação de dados conectados. No documento apresenta 10 boas práticas para publicação de dados conectados.

A primeira boa prática corresponde à “Preparar as Partes Interessadas”, nela fala da importância do planejamento e definição dos procedimentos que serão adotados para publicação de dados com a equipe envolvida.

Após definido os processos e planejamento que serão realizados, deve-se selecionar e definir um conjunto de dados (*dataset*) para publicação dos dados.

Uma vez definido esse conjunto de dados, deve-se adotar ou criar um modelo de dados. É destacado pelo documento, envolver os participantes que criam o *dataset* para auxiliar na definição da modelagem dos dados.

Depois de estabelecida a modelagem dos dados, deve-se adotar uma licença apropriada para disponibilização dos dados.

A quinta boa prática está relacionada ao uso de bons *Uniform Resource Identifier* (URI). Conforme apontado por Berner-Lee (2006) a URI é um dos pilares para possibilitar que os dados fiquem conectados. Há uma preocupação com a persistência das URIs para que a informação que está sendo conectada, sempre fique acessível e não deixe URIs “quebrados”.

A sexta boa prática de acordo com o documento do W3C (2014) refere-se ao uso de vocabulários para descrição dos objetos do *dataset*. Há uma preocupação e recomendação para utilização de vocabulários já criados e na possibilidade de estender os metadados para facilitar o processo de interoperabilidade dos dados. Observa-se ainda, a importância da documentação de uso desses vocabulários de dados.

Por conseguinte, há orientações para conversão dos dados para as instruções do RDF. O modelo RDF é caracterizado pela tríade recurso, propriedade e valor. O recurso é a “coisa” a ser descrita, pode ser um livro, uma pessoa, um mapa. A propriedade é o atributo desta coisa, pode ser o título, nome, tipo. Já o valor é a informação que deve ser preenchida referente ao recurso, por exemplo, um livro (recurso), tem um título (propriedade) chamado de “Vidas secas” (valor). Então, todo o *dataset* deverá ser convertido para essa estrutura (recurso-propriedade-valor).



A oitava boa prática refere-se ao acesso dos dados pelas máquinas. Recomenda-se que agentes computacionais possam acessar o *dataset* e que esses dados sejam legíveis por máquinas.

De acordo com as boas práticas, uma vez que os dados estiverem estruturados deve-se realizar a divulgação desse *dataset* ao público.

Por fim, a última boa prática, refere-se à responsabilidade em manter os dados disponíveis a longo prazo.

Apesar das boas práticas, trazer orientações para publicação de dados conectados, em 2017, foi publicado de forma complementar as boas práticas para publicação de dados abertos. As boas práticas de publicação de dados abertos possuem 35 boas práticas divididas em 12 categorias, além de destacar os benefícios de cada boa prática (compreensão, processabilidade, descoberta, reuso, veracidade, capacidade de ligação, acesso e interoperabilidade). (WORLD WIDE WEB CONSORTIUM, 2017).

A questão da publicação de dados abertos conectados nas bibliotecas, foi tema de grupo de trabalho no W3C e que apresentou um relatório em 2011 apontando diversas questões referentes aos benefícios, desafios e recomendações para que as bibliotecas possam publicar seus dados. (BAKER et al., 2011). Os autores Baker et al. (2011) afirmam que as bibliotecas precisam disponibilizar seus dados de forma conectada com outros recursos na *Web*, isto é, as “[...] as bibliotecas podem assumir um papel de liderança fundamentado em suas atividades tradicionais: gerenciamento de recursos para uso atual e preservação de longo prazo; descrição dos recursos com base em regras acordadas; e respondendo às necessidades dos buscadores de informações.”

3 PROCEDIMENTOS METODOLÓGICOS

Para realização do estudo, propõe-se uma discussão e levantamento dos procedimentos e tecnologias utilizadas para publicação de dados conectados. A aplicação desses procedimentos está sendo realizada no Sistema Integrado de Bibliotecas do Município de São Carlos (SIBISC), localizada em São Carlos, estado de São Paulo. O SIBISC é um departamento da Secretaria Municipal de Educação,



vinculado ao município de São Carlos. A escolha foi motivada pelos desafios que uma biblioteca pública possui em relação a recursos, pessoal e infraestrutura. Além de representar um ecossistema de informações complexas e similares a muitas outras bibliotecas brasileiras.

O SIBISC é composto por: oito Bibliotecas Públicas Municipais Escolares; duas Bibliotecas Públicas Municipais; uma Biblioteca Pública Distrital; uma Biblioteca Especializada Braille (voltado ao atendimento de deficientes visuais e/ou pessoas com baixa visão); e uma sala de leitura.

Possui, um quadro de dezessete bibliotecários alocados em cada uma das unidades vinculadas diretamente ao sistema. O catálogo é baseado no *software* PHL, tendo a base de usuários compartilhada por toda a rede. O sistema de classificação utiliza a CDD (18ª edição) e classificação de autor, majoritariamente em Cutter, com algumas unidades utilizando PHA. Ao todo, os acervos do SIBISC possuem aproximadamente 100 mil itens, sendo cerca de 20 mil obras bibliográficas.

3 RESULTADOS

Seguindo as boas práticas de publicação de dados conectados, o projeto encontra-se em fase inicial. Para comunicação dos envolvidos, foram propostas até o momento duas ações. A primeira foi a realização de uma *live* aberta para divulgação da proposta e divulgação do trabalho. Posteriormente, uma equipe interdisciplinar composta por bibliotecários, docentes de Biblioteconomia e Ciência da Computação e estudantes de Biblioteconomia, mestrado e doutorado em Ciência da Informação optaram por realizar reuniões de estudos com periodicidade a cada 15 dias. Na primeira reunião, foi traçada uma estratégia para condução dos trabalhos e definição das temáticas que seriam abordadas. Dentre as temáticas estudadas foram: *Web Semântica*, *linked data*, curadoria, sintaxe de codificação, RDF.

Paralelamente um grupo ficou responsável pela extração e análise dos dados para verificar as possibilidades de exportação dos dados e identificação dos metadados utilizados. Devido o sistema de biblioteca utilizar o *software* PHL e que o sistema foi customizado para adaptações e uso pelas bibliotecas, alguns plugins que



eram utilizados para exportação dos dados em XML não estavam funcionando. Isso fez com que os relatórios fossem gerados em planilhas *x/s*.

Muitas discussões estão sendo debatidas, mas espera-se que em breve possa-se publicar os dados em RDF do SIBISC para acesso e consulta de toda a população, além de potencializar a busca e recuperação da informação de seus dados. Acredita-se ainda que os processos estabelecidos poderão contribuir com diretrizes teóricas-metodológicas para bibliotecas brasileiras a publicarem seus dados abertos conectados.

4 CONSIDERAÇÕES FINAIS

As atividades para publicação de dados abertos conectados do Sistema Integrado de Bibliotecas do Município de São Carlos (SIBISC) ainda estão no início, entretanto, é possível identificar alguns desafios que é a adaptação de um conjunto de dados de uma biblioteca para o contexto do *Linked data*.

Como trabalhos futuros, será seguido as etapas das boas práticas para publicação de dados conectados e posteriormente, a análise e adequação do *dataset* às boas práticas para publicação de dados abertos conectados.

REFERÊNCIAS

BAKER, T. et al. 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: 8 set. 2019.

BERNERS-LEE, T. Semantic Web Road Map. World-Wide Web Consortium (W3C). 2001. Disponível em: . Acesso em: 17 fev. 2022.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. Scientific American, May 2001, p. 29-37.

JESUS, Ananda Fernanda de. Recomendações teórico-metodológicas para a publicação de dados bibliográficos abertos e conectados. 2021. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de São Carlos, São Carlos, 2021. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/14228>. Acesso em: 15 jul. 2022.



WORLD WIDE WEB CONSORTIUM. Semantic Web. [S.l.], 2013. Disponível em: <https://www.w3.org/standards/semanticweb/>. Acesso em: 15 jul. 2022.

WORLD WIDE WEB CONSORTIUM, Best Practices for Publishing Linked Data, 2014. Disponível em: <https://www.w3.org/TR/ld-bp/>. Acesso em: 01 out. 2020.

WORLD WIDE WEB CONSORTIUM. Boas Práticas para Dados na Web. São Paulo: W3C, 2017. Disponível em: <https://www.w3.org/Translations/DWBP-ptBR/>. Acesso em: 01 out. 2020.

Apoio financeiro: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).