



XXII Seminário Nacional de
Bibliotecas Universitárias
28 de novembro a 01 de dezembro
Florianópolis - SC

Eixo 4 – Inovação e Tecnologias

As tecnologias *Linked Data* e *Big Data* aplicadas nos serviços de informação em bibliotecas

Linked Data and Big Data technologies applied to library information services

Beatriz Scopel Borges – Universidade do Estado de Santa Catarina (UDESC)
beatriz.borges28@edu.udesc.br

Jéssica Carraro – Universidade do Estado de Santa Catarina (UDESC)
jessica.carraro@edu.udesc.br

Júlia Mattos Fernandes Pedroso – Universidade do Estado de Santa Catarina (UDESC)
julia.pedroso@edu.udesc.br

Ana Maria Pereira – Universidade do Estado de Santa Catarina (UDESC)
ana.pereira@udesc.br

Vanessa Marie Salm – Universidade do Estado de Santa Catarina (UDESC)
vanessa.salm@udesc.br

Resumo: Este trabalho objetiva apresentar resultados parciais do projeto de pesquisa “O uso do *Linked Data* e da *Big Data* pela RDA – *Resource Description and Access* na representação, recuperação e acesso dos dados e informações”. Quanto a metodologia, foi usada abordagem qualiquantitativa, descritiva e exploratória, expondo análise dos dados coletados por amostragem, com questionário enviado para bibliotecas universitárias e nacionais. Como resultados, pretende-se compreender como as práticas de *Linked data* e *Big data* podem melhorar a organização e recuperação dos dados e informação na web, além de disponibilizar a pesquisa concluída posteriormente para a comunidade bibliotecária e acadêmica.

Palavras-Chave: Organização do conhecimento. Recuperação da informação. *Linked data*. *Big data*.

Abstract: This work contains partial results of the research project “The use of *Linked Data* and *Big Data* by RDA – *Resource Description and Access* in the representation, recovery and access of data and information”. Methodology used a quali-quantitative, descriptive, and exploratory approach, exposing analysis of data collected by sampling, with a questionnaire sent to university and national libraries. As a result, it is intended



to understand how Linked data and Big data practices can improve the organization and retrieval of data and information on the web, in addition to making the research completed later available to the library and academic community.

Keywords: Knowledge organization. Information recovery. Linked data. Big data.

1 INTRODUÇÃO

Este relato é um recorte do projeto de pesquisa intitulado “O uso do *Linked Data* e da *Big Data* pela RDA – *Resource Description and Access* – na representação, recuperação e acesso dos dados e informações” vinculado ao curso de Biblioteconomia da Universidade do Estado de Santa Catarina (UDESC), estimado para término em agosto de 2023. A etapa em que se encontra é a análise dos dados das respostas recebidas através do questionário enviado para todos os continentes, denominado “Questionário sobre o uso de ferramentas tecnológicas no processo de catalogação de recursos informacionais” com foco em analisar o conhecimento e uso das duas ferramentas estudadas na pesquisa, por parte dos bibliotecários atuantes em bibliotecas nacionais e universitárias.

Para o presente recorte se faz o questionamento sobre as formas com que as práticas de *Linked Data* e *Big Data* podem auxiliar ou proporcionar melhorias na organização e recuperação de informações na web no contexto das unidades de informação, para isso utilizou-se as respostas de uma das perguntas realizadas no questionário já mencionado, sendo assim possível averiguar o nível de conhecimento dos profissionais da informação a respeito dessas ferramentas e como os mesmos podem utilizá-las a favor da melhor recuperação por parte dos usuários.

O trabalho se justifica, pois, a área das ciências de dados vem sendo muito discutida interdisciplinarmente com a biblioteconomia por se tratar da “[...] extração de informação útil a partir de imensas bases de dados complexas, dinâmicas, heterogêneas e distribuídas” (Bungnion; Manivan-nan; Nicolas, 2017). Além disso auxilia na criação, retenção, transformação e análise de dados advindos dessas bases, sendo a *Big Data* e a *Linked Data* ferramentas relacionadas a essa área, assim, vem se tornando cada vez mais relevante para as unidades de informação estudarem-nas e pensar em estratégias de recuperação e organização da informação utilizando-as.

Não restam dúvidas que as bibliotecas existem para proporcionar possibilidades diversas de informação de modo que o usuário possa encontrar e escolher o que deseja (Mey, 2009). Desse modo, o conhecimento de ferramentas que melhorem os processos de organização da informação na Web por parte dos bibliotecários poderá enriquecer a experiência do usuário no momento de busca e recuperação da informação.

Fato é que a Web apresenta grande volume de dados e informações não estruturadas produzidas por humanos e máquinas, e esse fenômeno recebe o nome de *Big Data* (ORACLE, 2023). O maior desafio da atualidade é desenvolver tecnologias que consigam processar esse grande volume de dados e informações com a finalidade de atender as necessidades dos usuários na Web com eficácia.

Sendo transversal à *Big Data*, a Web semântica, propõe que essa variedade de formatos que foram surgindo ao longo dos anos recebam uma padronização com a finalidade de criação de links semânticos para que um grande volume de dados, faça sentido. Destaca-se o *Resource Description Framework* - RDF como modelo de representação da informação que se liga diretamente a *Linked Data* por fornecer subsídios para a construção de uma semântica na Web que pode ser aplicada em sistemas de recuperação da informação (Sousa; Martins; Ramalho, 2018).

2 METODOLOGIA

Os procedimentos metodológicos utilizados caracterizam de acordo com objetivo, do tipo levantamento bibliográfico, de caráter descritivo e exploratório. Quanto ao procedimento técnico de coleta de dados foi utilizado o questionário semiestruturado online e a análise dos dados foi realizada de acordo com os critérios da análise de conteúdo de Bardin (2016).

No período de 2019 a 2020, para a viabilizar a aplicação dos questionários, foi realizado o mapeamento das unidades de informação de bibliotecas universitárias e bibliotecas nacional por continentes (América, Europa, África, Ásia e Oceania), com levantamento feito nos sites da IFLA e da RDA Toolkit em nível internacional, e em nível nacional na FEBAB e os CRBs, gerando uma lista de 3.679 bibliotecas universitárias e bibliotecas nacionais.

Como ferramenta de coleta de dados, foi enviado questionário semiestruturado com 7 perguntas abertas e 12 perguntas fechadas, nos idiomas português, inglês e espanhol. O encaminhamento do questionário online ocorreu por envio de e-mail aos contatos de diretores, coordenadores e catalogadores das bibliotecas, totalizando 5.433 e-mails coletados e enviados dos quais obtivemos 237 respostas. Por ter sido encaminhado a todos os continentes, verificou-se que o retorno dos questionários não foi representativo estatisticamente, no entanto, enquanto análise qualitativa por amostragem, considera-se importante para compreender como as práticas de *Linked Data* e o contexto *Big Data* podem proporcionar e/ou melhorar a organização e recuperação dos dados na web.

Para a compreensão por amostragem, a Análise de Conteúdo foi realizada em 3 grandes etapas: a 1ª etapa consistiu na organização dos dados e informações do questionário após coleta dos dados. A 2ª etapa estabeleceu a estratégia de critérios de categorização para o tratamento e análise dos dados e informações e a 3ª etapa apresentou a análise dos dados e informações do conteúdo com criticidade, com base em autores da fundamentação teórica da pesquisa.

3 RESULTADOS E DISCUSSÕES

A identidade dos respondentes e suas instituições foram mantidas em anonimato, sendo denominados com letras como A, B, C, D e assim, consecutivamente. De acordo com os respondentes, identificou-se explicações coerentes de como a *Linked Data* e a *Big Data* podem auxiliar na recuperação e organização da informação. Segundo o respondente A, a *Linked Data* e a *Big Data*, permitem “analisar grandes quantidades de dados para a geração de resultados importantes que, em volumes menores, dificilmente seria possível e estruturar e publicar dados conectados na Web por meio de hyperlinks e anotações, de maneira a permitir buscas semânticas”.

A afirmação do respondente, pode ser compreendida de acordo com Paletta e Mucheroni (2014, p. 80), pois segundo os autores:

[...] não só a Web Semântica precisa de acesso aos dados, mas as relações entre os dados devem ser disponibilizadas, para poder criar uma verdadeira Web de Dados (em oposição a uma mera coleção compartilhada de dados). Esta coleção de conjuntos de dados inter-relacionados na Web também pode ser referida como dados vinculados ou *Linked Data*.

A relação entre o *Linked Data* e a Web Semântica, permite maior organização e estruturação dos dados na Web de modo que as máquinas também consigam realizar a leitura destes dados, através de agentes computacionais, bem como sua posterior recuperação e acesso. O principal fruto desta relação é a significação dos dados que são linkados na Web Semântica e que fazem com que ela seja um ambiente rico em informações contextualizadas e complementares entre si (Santarem Segundo; Martins, 2018).

Para descrever a importância do *Linked Data* (LD) e do *Big Data* para a organização, recuperação e acesso à informação e a visibilidade das bibliotecas o respondente B afirma que o “LD augment the visibility and discoverability of library data on Web, where most users now search for information. It also helps to overcome linguistic difference (using URIs) and supports interoperability. LD enables libraries to provide exchangeable, comprehensive and reusable datasets, saving time and money on cataloguing. Big Data and their advanced analysis can increase the efficiency of library information services, offer more customized services to users¹”.

Para o respondente C “Linked Data allows webs of things that may seem to be unrelated to be linked together using URIs and relationship syntax. So starting at a certain point, through linked data you can discover relationships from your sought after data to other things that you may not have suspected have anything to do with what you are looking for. Big Data allows for the mining of large databases to discover underlying relationships that might not be apparent on the face of things²”.

Bizer, Heath, Berners-Lee (2011, tradução nossa) consideram que são necessários 4 passos para publicar *Linked Data* na web, que consistem em:

¹ O LD aumenta a visibilidade e a capacidade de descoberta dos dados da Library Data na Web, onde a maioria dos usuários agora procura informação. Também ajuda a superar diferenças linguísticas (usando URIs) e oferecer suporte à interoperabilidade. O LD permite que as bibliotecas forneçam conjuntos de dados intercambiáveis, abrangentes e reutilizáveis, economizando tempo e dinheiro na catalogação. Big Data e sua análise avançada podem aumentar a eficiência dos serviços de informação da biblioteca, oferecer serviços mais personalizados aos usuários (tradução nossa).

² Linked Data permite a rede das coisas que podem parecer não relacionadas sejam vinculadas usando URIs e sintaxes de relacionamentos. Então, a partir de um certo ponto, por meio do Linked Data você pode descobrir relações entre os dados procurados e outras coisas que você pode não ter suspeitado que tenham algo a ver com o que você está procurando. Big Data permite a mineração de grandes bancos de dados para descobrir relacionamento subjacentes que podem não ser aparentes (tradução nossa).

- a) usar URIs como nomes para coisas;
- b) usar URIs HTTP para que as pessoas possam procurar esses nomes;
- c) fornecer informações úteis, quando alguém procurar um URI, usando padrões (RDF, SPARQL);
- d) incluir links para outros URIs, para que eles possam descobrir mais coisas.

Portanto, é importante que as Unidades de informação sigam as recomendações para publicar na Web, estabelecendo assim, links que permitam disponibilizar seus catálogos online para além dos usuários das bibliotecas. No contexto apresentado pelos respondentes, foi possível observar que esses profissionais têm base conceitual na resposta à questão e teceram comentários sobre alguns usos específicos e amplos sobre a *Linked Data* e a *Big Data*, tais como a análise de grandes quantidades de dados, o acesso aberto de dados vinculados na Web, o enriquecimento de descrições e relações entre dados, pesquisas personalizadas para os usuários e descobertas de oportunidades de melhorias em serviços. Além disso, verificou-se que relacionaram não só os dois conceitos com as unidades de informação, mas também entre a nova diretriz de catalogação, a RDA.

Mesmo com toda positividade nas respostas sobre o conhecimento da *Linked Data* e da *Big Data* para o acesso aberto dos dados e informações dos catálogos das bibliotecas e para recuperação e acesso por parte dos usuários no ambiente Web, a pesquisa revelou também que há muito desconhecimento sobre os termos pesquisados. Há bibliotecários que dizem não conhecer ou conhecer parcialmente sobre o assunto abordado, algumas das justificativas foram a falta de contato prático com as ferramentas, assim como com os conceitos. Com um destaque para um dos respondentes que é diretor da biblioteca e que mesmo sabendo que seu time trabalha com as ferramentas e estuda os conceitos, não sabia o que eram e nem como funcionavam.

O respondente D afirma “Não tenho base teórica suficiente para explorar o assunto de forma satisfatória”. O respondente E afirma que “My team is making use of all these but being an Administrator/Library Director, not directly familiar³.” Já o respondente F afirma que “We have never experienced Linked Data or Big Data and therefore it is difficult to say if the new technology is good or bad. One must

³ Minha equipe está fazendo uso de tudo isso, mas sendo um administrador/diretor da biblioteca, não estou familiarizado diretamente (tradução nossa).

experience it⁴”. O respondente G diz que “Respondendo essas questões percebi que preciso fazer cursos, estou ficando a margem das atualizações da tecnologia da informação na minha atuação como bibliotecário”. Por fim, o respondente H afirma “não sei. Mas fiquei curiosa sobre o assunto e vou pesquisar”.

Destaca-se também que entre os respondentes que não conseguiram identificar por completo o uso da *Linked Data* e da *Big Data* nas unidades de informação, alguns entenderam a importância e as melhorias que as ferramentas podem oferecer e se interessaram por pesquisar mais, fazer treinamentos e se inteirar sobre o assunto. Isso demonstra que a pesquisa despertou o interesse real dos bibliotecários em se atualizar e buscar inovações para as respectivas unidades de informação em que atuam, e, futuramente, usufruir dos benefícios oferecidos pelo *Linked Data* e *Big Data*, tanto à unidade de informação, quanto aos usuários no que diz respeito a organização, recuperação e acesso aberto dos dados e informação dos catálogos das bibliotecas.

4 CONSIDERAÇÕES FINAIS

A presente pesquisa iniciou em meados de 2019, com previsão de término em 2021. No entanto, com a crise sanitária da pandemia, o desenvolvimento foi amplamente prejudicado, fazendo a pesquisa prorrogar e traçar novas estratégias para sua finalização.

Pela larga abrangência de contatos, também se enfrentou barreiras durante a coleta de dados, pois durante a pandemia, muitas bibliotecas fecharam por períodos indeterminados prejudicando o contato com os profissionais. Assim, novamente em 2022, os questionários foram enviados para que os dados fossem coletados, justificando assim a prorrogação da pesquisa. Contudo, muitas unidades de informação retornaram com negativa de participar da pesquisa, não proporcionando a equipe, a interação ou compreensão do porquê não participaram.

Mesmo com as limitações e dificuldades apresentadas, os dados foram coletados e a pesquisa encontra-se em fase de conclusão. Os dados apresentados

⁴ Nunca experimentamos *Linked Data* ou *Big Data* e, portanto, é difícil dizer se a nova tecnologia é boa ou ruim. É preciso experimentá-lo (tradução nossa).

nesse relato são uma prévia do final da pesquisa, e demonstra a importância e a necessidade de conhecer e fazer uso do *Linked Data* e do *Big Data* para a organização, recuperação e acesso de dados e informação dos catálogos das bibliotecas e/ou unidades de informação para acesso aberto na Web, disponível aos usuários.

Como encaminhamentos, pretende-se concluir a pesquisa e divulgar amplamente seus resultados com o objetivo de disponibilizar a comunidade bibliotecária e acadêmica, como as práticas de *Linked Data* e contexto *Big Data* podem proporcionar e/ou melhorar a organização e recuperação dos dados e informação na Web.

REFERÊNCIAS

- BARDIN, L. **Análise de conteúdo**. Ed. rev e ampl. São Paulo: Ed. 70, 2016. 279 p.
- BIZER, C.; HEARTH, T.; BERNERS-LEE, T. Linked Data: the story so far. **International Journal on Semantic Web and Information Systems**, v. 5, n. 3, 2009
- BUGNION, P.; MANIVANNAN, A.; NICOLAS, P. R. **Scala: Guide for Data Science Professionals**. 1st. ed. Birmingham: Packt Publishing, 2017.
- MEY, E. S. A.; SILVEIRA, N. C. **Catálogo no plural**. Brasília, DF: Briquet de Lemos, 2009. 217 p.
- MUCHERONI, Marcos Luiz e PALETTA, Francisco Carlos. O desenvolvimento da Web 3.0: Linked Data e DBPedia. 2014, Anais. São Paulo: Escola de Comunicações e Artes, Universidade de São Paulo, 2014. Disponível em: <http://www.contecsi.fea.usp.br/envio/index.php/contecsi/11contecsi/paper/view/772>
Acesso em: 24 maio 2023.
- SEGUNDO, J. E. S.; MARTINS, D. L. Web semântica e extração de dados na composição de modelo estrutural para dados de resultados de produção científica. **Encontro Brasileiro de Bibliometria e Cientometria**, v. 6, n. 6, p. 124 – 133, 2018. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/117685>. Acesso em: 01 ago. 2023.
- SOUZA, J. L.; MARTINS, P. G. M.; RAMALHO, R. A. S. Modelos de representação semântica na era do big data. **Brazilian Journal of Information Science**, v. 12, n. 3, p. 34–40, 2018. Disponível em: <https://brapci.inf.br/index.php/res/download/99420>
Acesso em: 01 mar. 2023.
- ORACLE. **O que é big data?**. 2023. Disponível em: <https://www.oracle.com/br/big-data/what-is-big-data/>