

LINKED OPEN DATA: conceito, relações e importância na Era da Informação

GUANDALINI, Clara Alcina¹
SANTOS, Amanda Azevedo dos²

GT 4 – Gestão e organização da informação e do conhecimento

Modalidade da apresentação: comunicação oral

Resumo: A velocidade e o volume de compartilhamento de dados crescem exponencialmente, fazendo com que haja diversas perspectivas de como acessar e recuperar esses dados. As áreas de Biblioteconomia e Ciência da Informação buscam estudar as novas ferramentas e conceitos que surgem decorrentes de poder melhorar a forma como os dados são acessados e recuperados. A pesquisa objetiva-se a abordar os conceitos de dados abertos, *Web Semântica*, *Linked Data* e *Linked Open Data*. No referencial teórico apresenta-se primeiramente um contexto desses conceitos, com a intenção de observar como surgiram, como têm se identificado e quais suas relações, com enfoque no conceito de *Linked Open Data*, que vêm sendo cada vez mais pesquisado e posto em prática, tornando-se fundamental e relevante para estruturar e ligar dados na *Web* atual. O conceito de *Web Semântica* também é um ponto de destaque, pois é base fundamental para essas novas ferramentas da tecnologia, sendo considerada uma evolução da *Web* que visa dar significado semântico, oferecendo mais possibilidades aos usuários e *softwares*. A partir desse embasamento tem-se algumas observações em discussão e resultados, e por fim nas conclusões apresentam-se considerações que pretendem identificar o motivo pelo qual o *Linked Open Data* é importante na atualidade e qual a função da Ciência da Informação nesse cenário.

Palavras-chave: *Linked Open Data*. *Web Semântica*. Dados Abertos.

LINKED OPEN DATA: concept, relationships and importance in the Information Age

Abstract: The speed and volume of data sharing grow exponentially, leading to diverse perspectives on how to access and retrieve this data. The areas of Librarianship and Information Science seek to study the new tools and concepts that arise from being able to improve the way data is accessed and retrieved. The research aims to address the concepts of open data, Semantic Web, Linked Data and Linked Open Data. In the theoretical framework, a context of these concepts is presented first, with the intention of observing how they emerged, as they have been identified and which their relations, with focus on the concept of Linked Open Data, that have been increasingly researched and put into practice, making it fundamental and relevant for structuring and connecting data in the current Web. The Semantic Web concept is also a highlight, as it is a fundamental basis for these new technology tools, being considered a Web evolution that aims to give semantic meaning, offering more possibilities to users and software. From this base we have some observations

¹ Graduada em Biblioteconomia e Ciência da Informação, Universidade Federal de São Carlos (UFSCar), clara.alcina@gmail.com;

² Graduada em Biblioteconomia e Ciência da Informação, Universidade Federal de São Carlos (UFSCar), amandahazevedo@outlook.com;

in discussion and results, and finally in the conclusions we present considerations that aim to identify the reason why Linked Open Data is important in the present and what the function of Information Science in this scenario.

Keywords: Linked Open Data. Semantic Web. Open Data.

1 INTRODUÇÃO

Imersos em uma sociedade que consome e cria informações por meio do uso da Tecnologia de Informação e Comunicação (TIC), compartilhar tais informações é feito com grande facilidade, que é viabilizada pela *internet* que permite ao usuário realizar *downloads* e *uploads* que segundo Schons (2007, não paginado) “[...] sua função é permitir o acesso e a partilha de informações.” de diversos recursos informacionais como imagens, vídeos, músicas, textos e *podcasts*, além de dados estruturados ou não estruturados.

Ao que tange os dados compartilhados na *internet*, é necessário buscar informações a respeito desses dados que acrescentem detalhes e deem embasamento, por exemplo: esses dados são confiáveis, se qualquer pessoa pode acessar, como esses dados estão dispostos, quem os inseriu e se eles estão disponibilizados da melhor forma.

Posteriormente observa-se que teses, pesquisas e discussões em áreas da Biblioteconomia e Ciência da Informação passaram a buscar formas de ampliar esses questionamentos e que melhorassem a recuperação e acesso destes dados.

Essa discussão no contexto brasileiro mostra-se em ascensão no que se refere a dados governamentais, visto que o governo brasileiro criou e mantém o Portal Brasileiro de Dados Abertos³ desde 2011 em parceria com a INDA (Infraestrutura Nacional de Dados Abertos), sendo possível encontrar dados de administração pública. Também é importante destacar que no Brasil existe a Lei de Acesso à Informação (Lei 12.527/2011) e o Decreto n.º 8.777, de 2016 sobre Política de Dados Abertos. Santarem Segundo (2015, p. 221) narra que “[...] cultura de publicação de dados abertos no Brasil é muito tímida e quando acontece é carente de melhor infraestrutura, tanto do ponto de vista do armazenamento quanto da real possibilidade de acesso pelos usuários finais [...]”.

No cenário científico, têm-se a preocupação atualmente em como disponibilizar os dados em formato aberto, sendo uma tendência mundial por parte das instituições, agências de fomento e pesquisadores. As instituições têm exigido que os pesquisadores entreguem juntamente com sua pesquisa um plano de gerenciamento de dados, nesse plano deve haver

³ PORTAL BRASILEIRO DE DADOS ABERTOS. Disponível em <<http://dados.gov.br>>. Acesso em: 17 jul. 2018.

um “planejamento” de como os dados serão compartilhados, armazenados e preservados, dessa forma é possível que os dados sejam reutilizados, de acordo com princípios éticos, em outros projetos. “Quanto mais dados forem disponibilizados de forma útil, maior será o nível de transparência e reprodutibilidade e, portanto, mais eficiente o processo científico se tornará, em benefício da sociedade.” (MOLLOY, 2011, p.1, tradução nossa).

1.1 Objetivos

Assim o objetivo da pesquisa é abordar os conceitos de dados abertos, *Web Semântica*, *Linked Data* e *Linked Open Data*, a fim de analisar suas relações e utilidade no aperfeiçoamento de recuperação e compartilhamento de dados.

Os objetivos específicos são:

- Definir o conceito de dados abertos;
- Investigar os princípios *Linked Data*;
- Compreender o movimento *Linked Open Data*.

1.2 Procedimentos Metodológicos

Destarte a pesquisa caracteriza-se como descritiva, de natureza qualitativa e de cunho teórico sobre os temas: Dados Abertos, *Web Semântica* e *Linked Open Data*. Para contextualização teórica e embasamento para os resultados e conclusões realizou-se o levantamento bibliográfico. As fontes para esse artigo foram encontradas nas bases de dados da Revista Informação e Informação e BRAPCI, na ferramenta de busca Google Acadêmico e sites do próprio governo, entre outros.

2 REFERENCIAL TEÓRICO

O conceito de dados abertos pode ser observado a partir de iniciativas que têm como objetivo uma maior transparência e compartilhamento de dados para seu reuso, sendo que o indivíduo possui cada vez mais o poder de buscar a informação que necessita por meio das facilidades que a *internet* possibilita. Conforme Sayão e Sales (2014, p. 77) afirmam:

A ideia de ciência aberta tem muitas faces e muitos significados, porém, o mais eloquente deles é o que reconhece, primordialmente, que o conhecimento científico é um patrimônio da humanidade e, que, portanto, deve estar disponível livremente para que as pessoas – cientistas ou não – possam usá-lo, reusá-lo e distribuí-lo sem constrangimentos tecnológicos, econômicos, sociais ou legais.

A *Open Knowledge Foundation* é uma importante associação que apoia e incentiva iniciativas *Open Data*, no site da Fundação é possível observar sua importância, os projetos que apoiam e algumas definições relevantes. A OKF aponta que as principais características na abertura de dados são: disponibilidade e acesso, reutilização e redistribuição e participação universal. Os dados abertos mostram-se cada vez mais parte importante de uma sociedade participativa, e sua colaboratividade tem trazido benefícios no seu reuso como aplicativos e projetos que auxiliam e facilitam o acesso de informações estruturadas.

O *Linked Data* surge de um conceito em ligar os dados e estruturá-los, em suma: “(...) Assim como a Web provocou uma revolução na publicação e no consumo de documentos, o *Linked Data* tem potencial para permitir uma revolução na forma como os dados são acessados e utilizados.” (BIZER; HEATH; BERNERS-LEE, 2009, tradução nossa).

A maior motivação para o desenvolvimento do *Linked Open Data* está ligada a conexão para compartilhamento não de apenas documentos (o item completo), mas também de dados, de forma que seja compartilhado entre vários e não restrito a um grupo selecionado.

A partir desse pressuposto de dados abertos e *Linked Data*, levanta-se o relacionamento que os dois conceitos apresentam, identificando-se então o *Linked Open Data*. É importante referir-se aos dados abertos, pois possuem papel fundamental no conceito de *Linked Open Data*:

(...) no ambiente digital os dados abertos (Open Data) estão cada vez mais acessíveis a consulta comum e ao mesmo tempo a ligação entre eles. Assim, os dados estarão mais suscetíveis para ligação de dados abertos (Linked Open Data) na Web, facilitando a interoperabilidade e acesso de informação. (ARAKAKI, 2016, p. 22).

Na tese de Arakaki, pode-se observar uma análise interessante a partir do levantamento bibliográfico realizado pelo autor, ele afirma em determinado trecho que: “(...) pode-se dizer que o *Linked Open Data* é a junção dos princípios do *Linked Data* com os princípios do *Open Data* (...)”. (ARAKAKI, 2016, p. 27). Também é interessante observar que o autor aborda anteriormente os princípios do *Linked Data* e do *Linked Open Data*, trazendo destaque do uso deste em catálogos bibliográficos:

O *Linked Data* corresponde em princípios para promover a ligação de dados estruturados. Outro conceito atrelado a esses princípios é a proposta de *Linked Open Data* que consiste na abertura dos dados para promoção do *Linked Data*. Nesse contexto, a proposta do *Linked Open Data* em catálogos bibliográficos torna-se fundamental para construção de catálogos abertos no ambiente da Web, melhor estruturados. (ARAKAKI, 2016, p. 14).

Santarem Segundo também traz uma abordagem notável sobre o *Linked Open Data*, relacionando-o com os conceitos e tecnologias da Web Semântica:

O LOD, que atualmente apresenta-se como a melhor forma de materialização dos conceitos e tecnologias da Web Semântica, é um projeto, com um conjunto de normas a serem seguidas, que usa os mesmos princípios de ligação semântica da Web de Dados, entretanto tem particularidades específicas, indicando um grau de exigência maior na constituição de sua rede de interligações. (SANTAREM SEGUNDO, 2015, p. 225).

Em 2010 foi desenvolvido por Tim Berners-Lee um sistema de classificação de estrelas na utilização do *Linked Open Data* devido à importância da abertura dos dados vinculados, pois para obter essa classificação, sendo de 1 a 5 estrelas, o principal é o que os dados sejam disponibilizados em licença aberta, como demonstrado a seguir.

Classificação	Descrição
★	Disponível na web (qualquer formato), mas com uma licença aberta, para ser Open Data
★★	Disponível como dados estruturados legíveis por máquina (por exemplo, Excel)
★★★	Como em (2), mais formato não-proprietário (por exemplo, CSV em vez de Excel)
★★★★	Todos as especificações acima, mais o uso de padrões abertos do W3C (RDF e SPARQL) para a identificação das coisas
★★★★★	Tudo acima, além de: vincular seus dados aos dados de outras pessoas para fornecer o contexto

Fonte: W3C (BERNERS-LEE, 2010, tradução nossa).

Percebe-se então que o *Linked Open Data* importa-se não somente em ligar os dados com licença aberta, mas também com a qualidade desses dados. Dessa forma essa ferramenta oferece uma recuperação e acesso aos dados com o intuito de relacionar contextos durante a

busca na *Web*, que estejam disponibilizados em acesso aberto, formando assim uma sociedade com poder de conhecimento. Marcondes aponta essa importância propondo um exemplo:

Ao invés dos links para outras páginas como na *Web* convencional, na proposta *Linked Open Data* os links são para recursos como, por exemplo, o registro de um livro num catálogo de biblioteca, ou um verbete numa enciclopédia, e não para meras páginas hipertextuais; além disso os *links* têm significado, são *links* semânticos, expressam a relação entre os recursos linkados. (MARCONDES, 2012, p. 179).

É relevante ressaltar que esses conceitos que foram demonstrados anteriormente possuem como base uma evolução na *Web*, que surge a partir do momento que se começa a pensar em dar significado semântico para conteúdo de páginas da *Web*, fazendo com que haja uma interoperabilidade entre seus conteúdos. Essa evolução é denominada de *Web Semântica* e seu grande propulsor é Tim Berners-Lee, segundo ele: “A *Web Semântica* não consiste apenas em colocar dados na web. Trata-se de fazer links, para que uma pessoa ou máquina possa explorar a rede de dados. Com dados vinculados, quando você possui algum, você pode encontrar outros dados relacionados.” (BERNERS-LEE, 2006, não paginado, tradução nossa).

A *Web Semântica* é então uma evolução da *Web* atual, assim como aponta Marcondes:

A proposta da *Web Semântica* descreve uma *Web* qualitativamente diferente da atual, na qual, ao invés dos conteúdos serem estruturados para serem exibidos e lidos por pessoas, terão uma semântica “inteligível” por programas, os chamados “agentes de software”; ou seja, programas que não sejam criados especificamente para processarem somente um tipo determinado de dado. As tecnologias da *Web Semântica* são a evolução da *Web* atual, em direção a mais amplas possibilidades de interação e de recuperação semântica da informação. (MARCONDES, 2012, p. 176).

3 RESULTADOS E DISCUSSÃO

Visou-se compreender o que o LOD representa atualmente nessa Era da Informação, e como pode beneficiar o desenvolvimento de questões com a sociedade, seja na área científica ou governamental. As iniciativas *Open* tomaram uma grande proporção partindo do conceito primeiramente de acesso aberto, avançou-se no fato de que os dados também precisavam ser compartilhados de forma ética e por meio disso poderiam ser essenciais em conexão com os outros dados. Nessa perspectiva, tem-se que o uso de metadados é de importância fundamental para contextualizar e constituir as informações, de forma a organizar um volume de dados consideráveis, principalmente no meio digital.

Para melhor representação do que vêm sendo discutido o site *The Linked Open Data Cloud* (<https://lod-cloud.net/>) mostra um diagrama de nuvem do LOD, ou seja, conjuntos de dados que foram publicados no formato de dados vinculados contendo 1.224 conjuntos de dados com 16.113 links (até junho de 2018), sendo que pode ser utilizado na Licença de Atribuição *Creative Commons*. Um exemplo de utilização do LOD é a *Europeana Collections*, uma biblioteca virtual que atualmente fornece acesso a mais de 50 milhões de itens (livros, obras de arte, etc.), em sua missão afirma que esse trabalho contribui para uma sociedade aberta, informada e criativa.

A questão é que a Web Semântica torna mais visível os benefícios que todos esses conceitos e suas relações podem oferecer, além dos desafios também para os profissionais da informação.

4 CONCLUSÕES

Desta forma é possível observar que a importância do *Linked Open Data* nos dias de hoje ou na denominada Era da Informação é a demanda em buscar e criar recursos informacionais instantaneamente, o que reflete em um excesso de recursos e informações sendo lançadas na *Web* todos os dias que necessitam de uma estruturação e ligação para possuir contexto e conseguir relacioná-los.

De frente a esta nova rotina adotada por inúmeras pessoas que incluem diversas faixas etárias de crianças a idosos, o campo da Ciência da Informação tem como seu papel principal fazer com que informações providas de qualquer meio ou recurso tenham seu poder de acesso por qualquer pessoa, em qualquer lugar, além de levantar questões e pesquisas que abordem o assunto a fim de desenvolver propostas para ampliações e melhoramentos da representação proporcionando assim uma recuperação com maior precisão por meio de mecanismos de busca. Usar o LOD como um dos meios para esse fim é, sem dúvidas, uma maneira de conseguir que esse objetivo seja alcançado.

Deste modo, o *Linked Open Data* utiliza essa realidade e se transforma justamente em um meio de qualquer pessoa conseguir ter acesso à dados governamentais, culturais, tecnológicos, e utilizá-los para analisar, comparar algo de forma a se construir maior conhecimento sobre um campo do saber, assim como colaborar para alguma pesquisa, e também em gerar inovação; além de oferecer um protocolo acessível e de cunho universal.

Outro ponto importante a se destacar é referente aos dados abertos, que apesar de trazer benefícios e melhorias para a sociedade, maior confiabilidade (por meio de esquemas XML) e facilidade é ainda um tema muito discutido, em decorrência a responsabilidade

intelectual, por dificuldades que as políticas no âmbito empresarial e governamental, incluindo as diretrizes adotadas por editoras de diversas áreas impõem e também por falta de conhecimento sobre o tema, o que reflete em uma tímida apropriação dos dados abertos, sendo que são de relevância na utilização do LOD.

Com questão de tempo e maiores discussões, não apenas pelos campos da Ciência da Informação e da Ciência da Computação, mas de outros campos, é esperado que os dados abertos se popularizem e a sua demanda aumente, influenciando em maior abertura e visibilidade de dados, como por exemplo o investimento de recursos financeiros de organizações políticas, em suma, quanto, onde e em que é investido.

O formato aberto utilizado pelo LOD é, muitas vezes, compreendido de maneira incorreta, as políticas adotadas são rígidas ao ponto de interferir no compartilhamento, ou por ser algo novo, desta maneira a sociedade, incluindo algumas organizações e pesquisadores, ainda não possui propriedade no que concerne ao conhecimento sobre o assunto, ou seja, é importante conscientizar sobre a relevância dos dados abertos e consequentemente os benefícios que a prática do LOD pode trazer em catálogos e buscadores.

REFERÊNCIAS

ARAKAKI, F. A. **Linked Data**: ligação de dados bibliográficos. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista (Unesp), Faculdade de Filosofia e Ciências, Marília, 2016.

BERNERS-LEE, T. **Linked Data**: Design Issues. [S.l.]: W3C, 2006. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 20 dez. 2017.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. **Linked Data** - the story so far International. *Journal on Semantic Web and Information Systems*, 2009.

EUROPEANA COLLECTIONS. **About**. Disponível em: <<https://www.europeana.eu/portal/pt/about.html>>. Acesso em: 15 out. 2018.

EUROPEANA COLLECTIONS. **Our Mission**. Disponível em: <https://pro.europeana.eu/?utm_source=portalmenu&utm_medium=portal&utm_campaign=Portal%2Bmenu>. Acesso em: 15 out. 2018.

MARCONDES, C. H. Linked data - dados interligados - e interoperabilidade entre arquivos, bibliotecas e museus na web. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, v. 17, n. 34, p. 171-192, maio/ago., 2012. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17n34p171>>. Acesso em: 20 dez. 2017

MOLLOY, J. C. The Open Knowledge Foundation: Open Data Means Better Science. **PLoS Biology**, United Kingdom; v.9, december, 2011.



OPEN KNOWLEDGE INTERNATIONAL. **About**. Disponível em:
<<https://okfn.org/about/>>. Acesso em: 17 jul. 2018.

PORTAL BRASILEIRO DE DADOS ABERTOS. **Sobre o dados.gov.br**. Disponível em:
<<http://dados.gov.br/pagina/sobre>>. Acesso em: 17 jul. 2018.

PORTAL BRASILEIRO DE DADOS ABERTOS. **Perguntas frequentes**. Disponível em:
<<http://dados.gov.br/pagina/faq>>. Acesso em: 17 jul. 2018.

SAYÃO, L. F.; SALES, L. F. **Dados abertos de pesquisa: ampliando o conceito de acesso livre**. 2014.

SCHONS, C. H. O volume de informações na Internet e sua desorganização: reflexões e perspectivas.. **Informação & Informação**, [S.l.], v. 12, n. 1, p. 50-65, jul. 2007. ISSN 1981-8920. Disponível em:
<<http://www.uel.br/revistas/uel/index.php/informacao/article/view/1748/1497>>. Acesso em: 17 jul. 2018.

SANTAREM SEGUNDO, J. E. Web semântica, dados ligados e dados abertos: uma visão dos desafios do Brasil frente às iniciativas internacionais. In: XVI ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (XVI ENANCIB), 2015.