

QUALIDADE DE DADOS *LINKED DATA*: ANÁLISE DA TEMÁTICA SOB A PERSPECTIVA DA CIÊNCIA DA INFORMAÇÃO

LINKED DATA QUALITY: ANALYSIS OF THE THEME FROM THE PERSPECTIVE OF INFORMATION SCIENCE

Ananda Fernanda de Jesus^a
José Eduardo Santarem Segundo^b

RESUMO

Objetivo: O presente estudo tem como objetivo identificar como a questão da qualidade de dados *Linked Data* é abordada nos artigos científicos indexados em bases de dados da Ciência da Informação. **Metodologia:** Mapeamento Sistemático da Literatura, pautado em protocolo de pesquisa. **Resultados:** Como resultados foram recuperados 38 artigos, sendo aceitos 9 para compor o *corpus* teórico da pesquisa. **Conclusões:** Conclui-se que a relação entre qualidade de dados *Linked Data* e a Ciência da Informação ainda é pouco explorada através de artigos científicos e que geralmente é realizada sob uma perspectiva aplicada, não sendo identificados estudos com o objetivo de apresentar um referencial teórico aprofundado da temática.

Descritores: Qualidade de dados. *Linked Data*. Ciência da Informação. Mapeamento Sistemático da Literatura.

1 INTRODUÇÃO

Entendido como um conceito multifacetado, a qualidade de dados é tradicionalmente definida pela literatura com base em uma perspectiva contextual, sendo mensurada por meio da capacidade dos dados de atenderem satisfatoriamente às necessidades dos usuários, performando adequadamente as tarefas nas quais serão empregados. (ZAVERI *et al.*, 2012; NOOGHABI; DASTGERDI, 2016; AHMED, 2017).

As preocupações com a qualidade dos dados não são recentes, tendo

^a Doutoranda do Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista (UNESP). Marília, Brasil. E-mail: af.jesus@unesp.br

^b Doutor em Ciência da Informação pela Universidade Estadual Paulista (UNESP). Docente da Universidade de São Paulo (USP). Ribeirão Preto, Brasil. E-mail: santarem@usp.br

como marco a década de 1970. (LANGER *et al.*, 2018). Essas preocupações afetam diferentes tipos de dados, dentre os quais encontram-se os publicados de acordo com os princípios do *Linked Data*.

Esses princípios foram elaborados como uma resposta ao crescimento exponencial de informações disponibilizadas na *Web*, que conduziu a um contexto em que a busca eficiente de informação apenas por usuários humanos tornou-se inviável, sendo necessário a ampliação do uso de agentes computacionais. (ISOTANI; BITTENCOURT, 2015).

Embora os dados publicados como *Linked Data* sejam pautados em uma estrutura pré-estabelecida, eles também são afetados por problemas de qualidade, que se tornam ainda mais complexos devido à própria estrutura desses dados, e a heterogeneidade das suas fontes.

Partindo do contexto apresentado, questiona-se: Como a relação entre qualidade de dados e o *Linked Data* tem sido discutida pela Ciência da Informação a nível nacional e internacional?

Esse artigo tem por objetivo identificar como a questão da qualidade de dados *Linked Data* tem sido abordada nos artigos científicos indexados em bases de dados da Ciência da Informação, para isso foi realizado um mapeamento sistemático nas bases de *dados Information Science & Technology Abstracts* (ISTA); *Library, Information Science & Technology Abstracts with Full Text* (LISTA) e Base de Dados em Ciência da Informação (BRAPCI). Os procedimentos do processo de busca são apresentados na seção 3. A próxima seção apresenta os principais aspectos e conceitos relacionados à qualidade em dados *Linked Data*.

2 BREVES CONSIDERAÇÕES SOBRE QUALIDADE DE DADOS PUBLICADOS COMO *LINKED DATA*

Proposto em 2006, o *Linked Data* é composto essencialmente por quatro princípios que visam permitir a conexão de dados na *Web*, facilitando o uso de agentes computacionais no processo de busca por informações, promovendo a reutilização desses dados para novas tarefas em outros contextos, permitindo resultados de busca mais significativos (que forneçam respostas aos usuários e

não recursos informacionais que contenham essas respostas) possibilitando a serendipidade, ou seja, a descoberta acidental de informações.

De acordo com Berners-Lee (2006, não paginado, tradução nossa) os quatro princípios do *Linked Data* são:

1. Use URIs como nomes para as coisas;
2. Use HTTP URIs, para que as pessoas possam procurar esses nomes;
3. Quando alguém procura um URI, forneça informações úteis, usando os padrões (RDF, SPARQL);
4. Inclua links para outros URIs, para que eles possam descobrir mais coisas.

Os *Uniform Resource Identifier* (URIs) são identificadores que permitem a representação de documentos, locais, endereços da Web, figuras ou até mesmo pessoas. O *Hypertext Transfer Protocol* (HTTP) é, de maneira geral, um protocolo de solicitação/resposta de uso padrão, que permite acesso às informações contidas no ambiente *Web*, realizando a conexão entre dois computadores conectados à internet. (W3C, 2018).

O *Resource Description Framework* (RDF) é um modelo que visa possibilitar a explicitação das relações entre os dados de maneira formal, legível por máquinas. A descrição dessas relações é realizada através de declarações em formato de triplas Recurso + Propriedade + Valor. Nessas declarações, o recurso ou a coisa a ser descrita precisa ser representada utilizando *URIs*. As propriedades devem ser descritas usando vocabulários, criados para atender a domínios específicos e reutilizados sempre que possível. O valor das declarações pode ser tanto um literal (números, nomes, palavras e etc. em linguagem natural) ou um outro *URI*. (W3C, 2014).

Desde da publicação do *Linked Data* em 2006, uma quantidade sem precedente de dados estruturados foi publicada na Web, como afirmam Kontokostas *et al.* (2014).

A *Linked Open Data Cloud* (LOD Cloud), plataforma que reúne dados abertos publicados como *Linked Data*, passou de 12 *datasets* em 2007 para 1301 *datasets* em 2022. (LOD CLOUD DIAGRAM, 2022).

Esses dados seguem as características da publicação do ambiente *Web*, um ambiente aberto e complexo, com grande heterogeneidade de fontes, sem controle de quem podem criar, publicar e recuperar dados, fazendo com que a

qualidade oscile.

Nesse sentido, discutir e avaliar a qualidade de dados é indispensável tanto para os fornecedores de dados, que precisam avaliar a qualidade dos dados que publicam visando ampliar o seu potencial de reuso, quanto para os consumidores de dados, que precisam garantir que os dados atendam satisfatoriamente a sua necessidade.

A qualidade de dados se caracteriza como um conceito multifacetado, podendo ser compreendido e discutido através de diversas perspectivas. Nooghabi e Dastgerdi (2016) apontam como principais vertentes de qualidade a perspectiva intrínseca, a contextual e a representacional. A intrínseca seria uma perspectiva que visa avaliar a qualidade dos dados através de um conjunto de características desses dados, sem levar em consideração o contexto no qual serão aplicados.

Já a perspectiva contextual é a mais amplamente adotada pela literatura e está relacionada ao conceito de “*fitness for use*”, nessa perspectiva a qualidade também é mensurada através de um conjunto de características dos dados, mas as características a serem avaliadas e o peso de cada característica dependem dos objetivos e das tarefas que o usuário pretende executar com esses dados. Nessa perspectiva, um conjunto de dados pode ser perfeitamente adequado para determinada tarefa e não atuarem satisfatoriamente em outro contexto (JURAN, 1988; WANG; STRONG, 1996; ZAVERI *et al.*, 2012).

A perspectiva representacional possui muitas relações com a Ciência da Informação, de acordo com Nelson, Todd e Wixom (2005), seria uma perspectiva mais recente, onde se busca avaliar se os dados apresentam as informações necessárias para sua interpretação, compreensão e aplicação. Ou seja, se esses dados estão corretamente representados.

Melo (2017, p. 48) aponta que “Os problemas de qualidade são classificados em diferentes categorias, chamadas de dimensões, nas quais problemas do mesmo tipo são classificados de acordo com uma categoria específica”. Cada uma dessas dimensões é relacionada a uma ou mais formas de mensurá-las, verificando o quão adequados são os dados em relação a essa dimensão. Esses conjuntos de indicadores são chamados métricas, e são

elaborados com o propósito de calcular a uma pontuação de avaliação para as dimensões, podendo essas pontuações serem quantitativas ou qualitativas (BIZER, 2013; TALLERÁS, 2017).

A avaliação de dados publicados como *Linked Data* demanda o uso de dimensões específicas, tendo em vista que estes se diferem dos demais dados tanto em seu contexto de publicação como em suas estruturas.

Em relação ao contexto de publicação, esses dados são provenientes de domínios distintos, como dados governamentais, dados bibliográficos, elaborados por bibliotecas, dados criados pela comunidade, como os dados provenientes da DBpedia. Tendo sido publicados com objetivos diferentes, esses dados foram criados com diferentes níveis de preocupação com a qualidade, levando a problemas como a falta de confiança nas fontes, desatualização dos dados, informações inconsistentes, incompletas ou mal interpretadas (ASSAF; SENART; TRONCY, 2016; HADHIATMA, 2018).

Já em relação à estrutura, por advirem de domínios distintos, esses dados requerem ontologias e vocabulários diferentes, em muitas situações vocabulários de criação própria, podendo levar a problemas de integração. Monika Rani, Sapna e Mishra (2018) indicam a existência de links inadequados como um problema recorrente em datasets *Linked Data*. Hadhiatma (2018) fala que as relações entre os conjuntos de dados podem estar incorretas ou incompletas, tornando necessário ponderar se esses links são apropriados e úteis.

Portanto, ao avaliar dados publicados de acordo com o *Linked Data*, é necessário criar ou adaptar as dimensões de qualidade existentes para que elas sejam adequadas a esse contexto. Além da preocupação com a adaptação das dimensões e métricas, a estrutura do *Linked Data* faz com que exista uma preocupação com o uso de dados de fontes externas, tendo em vista que esses dados podem passar por atualizações sem aviso prévio, ou até mesmo serem excluídos, comprometendo assim a qualidade dos dados internos, fazendo com que a avaliação seja um processo contínuo e não uma ação única.

Apresentados os principais conceitos relacionados a qualidade de dados *Linked Data*, a próxima seção apresenta os procedimentos metodológicos para

realização de um mapeamento sistemático da literatura, com objetivo de verificar como tem sido discutida a relação entre qualidade de dados e *Linked Data* nas bases de dados temáticas da Ciência da Informação.

3 METODOLOGIA

O presente artigo resulta de uma pesquisa exploratória e descritiva, com resultados quantitativos e qualitativos, realizada através de um Mapeamento Sistemático da Literatura, um método que orienta a condução de levantamentos bibliográficos, pautado em um protocolo de busca e no registro das tomadas de decisão dos pesquisadores, com a finalidade de fornecer um panorama geral da discussão de determinada temática, em busca de assuntos comuns e assuntos pouco explorados. (KITCHENHAM *et al.*, 2010). O Quadro 1, apresentado a seguir, apresenta as informações utilizadas para estruturar o levantamento:

Quadro 1 – protocolo de busca

Protocolo de busca	
Pergunta pesquisa	de Como a relação entre qualidade de dados e Linked Data tem sido discutida pela Ciência da Informação a nível nacional e internacional?
Objetivos	Identificar a forma como têm sido conduzidas as discussões sobre qualidade de dados <i>Linked Data</i> nos artigos indexados em bases temáticas da Ciência da Informação
Estratégia busca	de ("Linked Data" OR "Linked Open Data") AND ("Data Quality")
Bases dados consultada	de ISTA; LISTA; e BRAPCI.
Período abrangido	Sem restrição temporal.
Idiomas	Português, inglês e espanhol.
Critérios Inclusão	de (I) Foco principal é voltado para discutir qualidade de dados publicados de acordo com os princípios do Linked Data
Critérios exclusão	de (E) Não está nos idiomas estabelecidos para a pesquisa; (E) Apenas menciona a temática de interesse; (E) Não aborda a temática de interesse; (E) Não foi possível obter acesso ao documento completo;

Formulário de extração	Enfoque do documento; Domínio através do qual a temática foi discutida; apresenta uma discussão teórica aprofundada sobre qualidade de dados <i>Linked Data</i> (sim; não); traça uma relação explícita entre qualidade de dados <i>Linked Data</i> e Ciência da Informação (sim; não);
Data da coleta	Maio de 2022

Fonte: Autores (2022)

As buscas nas bases *Information Science & Technology Abstracts – (ISTA)*; *Library, Information Science & Technology Abstracts with Full Text (LISTA)* foram realizadas através da estratégia de busca apresentada, sem recorte temporal. As buscas na Base de Dados em Ciência da Informação (BRAPCI) foram realizadas de maneira manual, pelo termo “qualidade de dados”, seguida da aplicação dos critérios de exclusão, apenas foram computados os artigos aceitos nesta pré-seleção. A base foi selecionada visando fornecer um panorama da abordagem da temática através de pesquisas nacionais em Ciência da Informação.

Os documentos recuperados foram sistematizados em quadros resumo, onde foram identificadas as duplicatas e aplicados os critérios de exclusão. Após aceitos os documentos, foi realizada a etapa de extração de informações, onde foram coletados os enfoques dos artigos, por meio de uma análise dos objetivos, métodos e resultados desses trabalhos. Foi coletado o domínio no qual a qualidade de dados *Linked Data* foi discutida nesses artigos, análise feita por meio da identificação do enfoque realizada previamente.

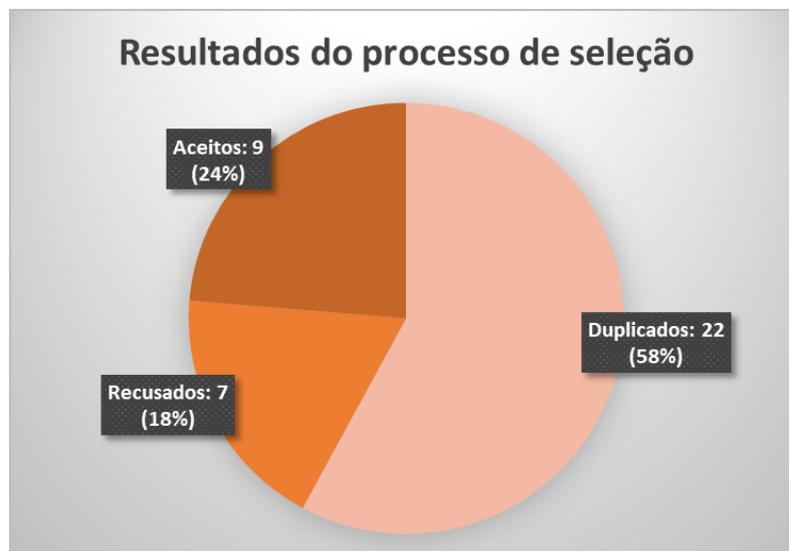
Foi analisado ainda se esses trabalhos apresentaram uma discussão teórica aprofundada sobre qualidade de dados *Linked Data* e se apresentam uma relação explícita entre qualidade de dados *Linked Data* e Ciência da Informação.

4 RESULTADOS: APRESENTAÇÃO E DISCUSSÃO

Essa seção apresenta os resultados obtidos por meio da aplicação do protocolo apresentado no Quadro 1. As buscas foram realizadas em maio de 2022 e retornaram 38 artigos, sendo 22 (57,89%) artigos duplicados, 7 (18,42%)

recusados e 9 (23.68%) documentos aceitos. A Figura 1 representa os resultados desse processo de seleção dos documentos.

Figura 1 - Resultados da seleção dos documentos recuperados



Fonte: Autores (2022)

Todos os documentos recusados foram selecionados com base no critério de exclusão “(E) Apenas menciona a temática de interesse”, sendo, portanto, artigos que mencionaram tanto o contexto do *Linked Data* como alguma questão relacionada à qualidade de dados, porém, o assunto não foi o objetivo central do documento.

Dentre os 9 documentos aceitos para compor o *corpus* teórico da pesquisa, 8 foram recuperados nas bases de dados internacionais (LISTA e ISTA) e apenas 1 na base de dados nacional (BRAPCI). O Quadro 2 apresenta a relação entre os documentos aceitos, seus autores, ano de publicação e uma breve descrição.

Quadro 2 - Relação dos documentos aceitos

TÍTULO	AUTOR(ES)	ANO	DESCRIÇÃO
Linked data practice at different levels of semantic precision: the perspective of libraries, archives and museums.	Isaac; Baker	2015	Discute os diferentes níveis de precisão semântica através da perspectiva dos vocabulários controlados

TÍTULO	AUTOR(ES)	ANO	DESCRIÇÃO
Proposed metrics for data accessibility in the context of linked open data.	Nooghabi; Dastgerdi	2016	Propõe modelos e métricas para avaliação da dimensão “ <i>data accessibility</i> ”
Quality of linked bibliographic data: the models, vocabularies, and links of data sets published by four national libraries.	Tallerås	2017	Aplica o processo de avaliação de qualidade a <i>datasets Linked Data</i> provenientes de bibliotecas
Metodologia de Avaliação de Qualidade para Dados Conectados	Melo; Botega e Santarem Segundo	2017	Propôs uma metodologia para avaliação de qualidade de dados <i>Linked Data</i> com base na identificação de problemas de qualidade, dimensões, métricas e metodologias de avaliação disponíveis na literatura.
How RDA is essential in the reconciliation and conversion processes for quality Linked Data.	Possemato	2018	Discutir como o <i>Resource Description and Access (RDA)</i> pode melhorar a qualidade de dados bibliográficos publicados como <i>Linked Data</i> .
A multi-criteria decision making model for the selection of open data management systems	Máchová; Iněnička	2019	Objetiva propor um modelo de seleção de fontes de dados governamentais abertos publicados como <i>Linked Data</i> , utilizando dimensões de qualidade
Connecting Semantic Situation Descriptions with Data Quality Evaluations—Towards a Framework of Automatic Thematic Map Evaluation.	Homburg	2020	Propõe uma ontologia para avaliar qualidade de dados geoespaciais publicados como <i>Linked Data</i>
Evaluating the quality of linked open data in digital libraries.	Candela <i>et al.</i>	2020	Realizar um estudo de avaliação de qualidade em <i>datasets Linked Data</i> de instituições do patrimônio cultural visando identificar <i>datasets</i> que podem ser potencialmente adequados para esse domínio e ainda propor um padrão de qualidade baseado no <i>RDA</i> .

TÍTULO	AUTOR(ES)	ANO	DESCRIÇÃO
Data Quality Barriers for Transparency in Public Procurement	Soylu <i>et al.</i>	2022	Avaliar um conjunto de dataset governamentais publicados como <i>Linked Data</i> , bem como propor diretrizes para avaliação de fontes em potencial.

Fonte: Autores (2022).

Observa-se no Quadro 2 que o primeiro estudo aceito é relativamente recente, datado de 2015, tendo em vista a publicação dos princípios do *Linked Data* em 2006, e que mesmo antes disso já se discutia a questão da qualidade de dados.

Na Ciência da Informação, a preocupação com a qualidade de dados também não é uma questão recente, ela está presente desde as primeiras tentativas por padronizar os processos de representação da informação.

Dos 9 artigos aceitos, apenas Melo; Botega e Santarém Segundo (2017, p. 17) buscaram traçar uma relação explícita entre a Ciência da Informação e a qualidade de dados *Linked Data*. Os autores apontam que

A Ciência da Informação tem assumido um papel importante nas questões relacionadas à qualidade de dados, em modelos de publicações de dados baseados nas melhores práticas de Dados Conectados, contribuindo para minimizar os problemas de qualidade nos processos de armazenamento e recuperação da informação.

Tem se ampliado no campo da Ciência da Informação os estudos que se preocupam em discutir de maneira profunda tanto a questão da qualidade de dados como estudos que abordam o *Linked Data*, como aponta Jesus (2021, p. 140) ao indicar que existe uma pluralidade temática nas abordagens dos estudos sobre *Linked Data* desenvolvidos no âmbito da CI e que essa pluralidade “[...] demonstra a existência de um amadurecimento das discussões na área, pois passou-se das pesquisas gerais e de caráter teórico para as discussões de temas mais específicos e aplicados”.

Embora as temáticas venham se destacando de maneira individual na Ciência da Informação, ainda é necessário se aprofundar na relação entre ambas, tendo em vista que essa relação entre CI e qualidade de dados *Linked Data* não foi explícita em nenhum dos artigos.

Mesmo não realizando uma relação explícita com a CI, essa relação está presente nos artigos recuperados por meio dos domínios nos quais se discute qualidade de dados. O Quadro 3 apresenta uma relação entre os artigos e o domínio por meio do qual se discute a qualidade de dados *Linked Data*.

Quadro 3 - Relação dos artigos aceitos e respectivos domínios

AUTOR(ES)	ANO	DOMÍNIO
Isaac; Baker	2015	Dados Bibliográficos e dados provenientes de instituições do patrimônio cultural (bibliotecas, arquivos e museus)
Nooghabi; Dastgerdi	2016	<i>Linked Open Data</i> de maneira abrangente, sem foco em um domínio específico
Tallerás	2017	Dados Bibliográficos e dados provenientes de instituições do patrimônio cultural (bibliotecas, arquivos e museus)
Melo; Botega e Santarem Segundo	2017	Dados sobre publicações científicas como prova de conceito, mas o contexto discutido é o do <i>Linked Open Data</i> , de maneira mais abrangente.
Possemato	2018	Dados Bibliográficos e dados provenientes de instituições do patrimônio cultural (bibliotecas, arquivos e museus)
Máchová; Iněnička	2019	Dados Abertos Governamentais
Homburg	2020	Dados Geoespaciais
Candela <i>et al.</i>	2020	Dados Bibliográficos e dados provenientes de instituições do patrimônio cultural (bibliotecas, arquivos e museus)
Soylu <i>et al.</i>	2022	Dados governamentais, focado em transparência de dados públicos

Fonte: Autores (2022)

Destacam-se temas de estudo recorrentes na CI, tais como os dados governamentais abertos. Observa-se ainda um destaque de um objeto clássico da CI, os dados bibliográficos, provenientes de instituições o patrimônio cultural,

tais como bibliotecas, arquivos e museus.

Observa-se que a temática é discutida de forma mais abrangente em apenas um dos artigos, o que ressalta a complexidade de avaliar a qualidade de dados publicados como *Linked Data*. Não basta apenas se preocupar com as particularidades estruturais e contextuais intrínsecas ao *Linked Data*, apresentadas na seção 2, é necessário considerar também as demandas desse domínio específico, a finalidade na qual os dados serão aplicados e as necessidades de seus usuários finais.

Em relação às discussões teóricas aprofundadas sobre qualidade de dados *Linked Data*, observou-se que nenhum dos estudos teve como objetivo promover essa discussão, o foco teórico da maioria dos artigos foi em descrever as particularidades do domínio de aplicação, caracterizando os dados bibliográficos, os dados governamentais e geoespaciais.

Essa questão fica mais evidente na análise dos enfoques temáticos dos artigos, apresentada no Quadro 4.

Quadro 4 - Relação entre enfoques temáticos dos documentos e número de artigos

Enfoque temático	Número de artigos
Propor um artefato para a avaliação de qualidade	5
Realizar a análise de qualidade de um ou mais <i>datasets</i>	2
Apresentar o papel de um instrumento de padronização proveniente da CI na melhoria da qualidade de dados <i>Linked Data</i>	1
Discutir a qualidade da dados através da perspectiva da precisão semântica em vocabulários controlados	1

Fonte: Autores (2022)

Os artigos que buscam propor um artefato para a avaliação de qualidade têm uma proposta mais aplicada, voltada para a apresentação de justificativas que embasam a necessidade desse artefato e ainda para as questões técnicas do funcionamento desse artefato.

Os artigos que buscaram analisar a qualidade de *dataset* também

possuem um foco mais aplicado, focando em justificar a necessidade de análise dos datasets escolhidos e em apresentar as métricas, os artefatos utilizados na avaliação e os resultados do processo.

Somados, 7 dos 9 artigos recuperados tem uma abordagem mais aplicada do assunto, e os 2 restantes discutem a temática a partir de uma perspectiva muito específica, não tendo recuperado nenhum artigo que visa fornecer uma visão geral e aprofundada da temática através de uma abordagem teórica.

Apresentados os resultados do levantamento realizado, a próxima seção apresenta algumas considerações finais.

5 CONSIDERAÇÕES FINAIS

Esse artigo teve seu desenvolvimento pautado no seguinte questionamento: “Como a relação entre qualidade de dados e o *Linked Data* tem sido discutida pela Ciência da Informação a nível nacional e internacional?”

A partir da análise dos resultados apresentados é possível concluir que as discussões sobre a temática a nível internacional ainda são recentes e pouco exploradas pela literatura, mesmo que as temáticas de qualidade de dados e de *Linked Data* sejam frequentemente discutidas pela comunidade científica.

O mesmo ocorre a nível nacional, de maneira ainda mais explícita, tendo em vista que apenas um artigo foi recuperado.

Entre os artigos que abordam a temática existe uma tendência para propostas aplicadas, voltados para a elaboração de artefatos que auxiliem na avaliação de qualidade e para a avaliação de qualidade de *datasets*.

Observou-se que essa tendência voltada para a aplicação tem reflexo no referencial teórico apresentado por esses trabalhos, sendo esse mais voltado para apresentar as especificações técnicas e metodológicas do processo de avaliação de qualidade e em apresentar os resultados obtidos.

Não foi recuperado nenhum artigo cujo objetivo foi discutir a relação através de uma perspectiva abrangente, concentrando-se em fornecer definições, se aprofundar em abordagens, apontando os principais problemas de qualidade que são relativos a dados que seguem os princípios do *Linked Data*.

Também não foi recuperado nenhum artigo que se aprofunde em

explicitar o papel da Ciência da Informação nesse contexto, explicitando em que aspectos a CI pode colaborar para os avanços na qualidade de dados *Linked Data*.

Nesse sentido, a próxima etapa desta pesquisa terá como objetivo realizar um levantamento mais abrangente do assunto, considerando bases temáticas multidisciplinares com a finalidade de apresentar definições e apresentar de maneira explícita os principais aspectos da relação entre Ciência da Informação e qualidade de dados *Linked Data*.

AGRADECIMENTOS

Agradecemos à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo financiamento recebido para o desenvolvimento dessa pesquisa. Processo FAPESP nº2021/03349-0

REFERÊNCIAS

AHMED, H. H. Data quality assessment in the integration process of linked pen data. *In: INTERNATIONAL CONFERENCE ON COMPUTER SYSTEMS AND APPLICATIONS*, 14., 2017. **Anais [...]**. [S. l.]: leee/Acs, 2017. p. 1-6. Disponível em: [10.1109/AICCSA.2017.178](https://doi.org/10.1109/AICCSA.2017.178). Acesso em: 04 jul. 2022.

ASSAF, A.; SENART, A.; TRONCY, R. Towards an objective assessment framework for linked data quality. **International Journal On Semantic Web And Information Systems**, [S. l.], v. 12, n. 3, p. 111-133, jul. 2016. Disponível em: <http://dx.doi.org/10.4018/ijswis.2016070104>. Acesso em: 27 maio 2022.

BERNERS-LEE, T. **Linked data**, 2006. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 08 set. 2020.

BIZER, C. **Quality driven information filtering**: in the context of web based information systems. Berlin: VDM Publishing, 2013. 196 p.

CANDELA, G.; ESCOBAR, P.; CARRASCO, R. C.; MARCO-SUCH, M. Evaluating the quality of linked open data in digital libraries. **Journal Of Information Science**, [S. l.], v. 48, n. 1, p. 21-43, ago. 2020. Disponível em: <http://dx.doi.org/10.1177/0165551520930951>. Acesso em: 26 maio 2022.

HADHIATMA, A. Improving data quality in the linked open data: a survey. **Journal Of Physics: Conference Series**, [S. l.], v. 978, p. 12-26, mar. 2018.

Disponível em: <http://dx.doi.org/10.1088/1742-6596/978/1/012026>. Acesso em: 26 maio 2022.

HOMBURG, T. Connecting semantic situation descriptions with data quality evaluation: towards a framework of automatic thematic map evaluation.

Information, [S. l.], v. 11, n. 11, p. 532, nov. 2020. Disponível em: <http://dx.doi.org/10.3390/info11110532>. Acesso em: 26 maio 2022.

ISAAC, A.; BAKER, T. Linked data practice at different levels of semantic precision: linked data practice at different levels of semantic precision. **Bulletin of the association for information science and technology**, [S. l.], p. 34-39, abr. 2015. Disponível em: [10.1002/bult.2015.1720410411](https://doi.org/10.1002/bult.2015.1720410411). Acesso em: 05 jul. 2022.

ISOTANI, S.; BITTENCOURT, I. **Dados abertos conectados**. São Paulo: Novatec, 2015. 175 p. Disponível em: <http://www.icmc.usp.br/e/b0477>. Acesso em: 19 nov. 2020.

JESUS, A. J. **Recomendações teórico-metodológicas para a publicação de dados bibliográficos abertos e conectados**. 2021. 166 f. 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: 19 nov. 2020.

JURAN, J. M. **Quality Control Handbook**. New York: [s. n.], 1988. 500 p.

KITCHENHAM, B.; PRETORIUS, R.; BUDGEN, D.; BRERETON, O. P.; TURNER, M.; NIAZI, M.; LINKMAN, S. Systematic literature reviews in software engineering: a tertiary study. **Information and Software Technology**. [S. l.], v. 52, p. 792-805, jan. 2010.

KONTOKOSTAS, D.; WESTPHAL, P.; AUER, S.; HELLMANN, S.; LEHMANN, J.; CORNELISSEN, R.; ZAVERI, A. Test-driven evaluation of linked data quality. **International Conference On World Wide Web**, [S. l.], v. 1, n. 23, p. 747-758, 2014. Disponível em: <http://dx.doi.org/10.1145/2566486.2568002>. Acesso em: 26 maio 2022.

LANGER, A.; SIEGERT, V.; GÖPFERT, C.; GAEDKE, M. SemQuire: assessing the data quality of linked open data sources based on dqv. **Current Trends In Web Engineering**, [S. l.], p. 163-175, 2018.

LOD CLOUD DIAGRAM. **About**. [S. l.: s. n.], 2022. Disponível em: <https://lod-cloud.net/#about>. Acesso em: 26 maio 2022.

MÁCHOVÁ, R.; LNěNIČKA, M. A multi-criteria decision making model for the selection of open data management systems. **Electronic Government, An International Journal**, [S. l.], v. 15, n. 1, p. 1-18, 2019. Disponível em: <http://dx.doi.org/10.1504/eg.2019.10017080>. Acesso em: 27 maio 2022.

MELO, J. O. de S. F. **Metodologia de avaliação de qualidade de dados no**

contexto do linked data. 2017. 111 f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências de Marília, Universidade Estadual Paulista, Marília, 2017.

MELO, J. O. S. F.; BOTEGA L. C.; SANTAREM SEGUNDO, J. E. Metodologia de avaliação de qualidade para dados conectados. **Informação & Tecnologia**, João Pessoa, v. 4, n. 2, p. 80-101, 2017. Disponível em: <https://brapci.inf.br/index.php/res/download/105665#:~:text=Acredita%2Dse%20que%20a%20metodologia,no%20diagrama%20do%20projeto%20LOD>. Acesso em: 05 jul. 2022.

MONIKA RANI, H. G.; SAPNA, R.; MISHRA, S. An investigative study on the quality aspects of linked open data. **International Conference On Cloud Computing And Internet Of Things**, [S. l.], p. 33-39, 2018. Disponível em: <http://dx.doi.org/10.1145/3291064.3291074>. Acesso em: 26 maio 2022.

NELSON, R. R.; TODD, P. A.; WIXOM, B.H. Antecedents of information and system quality: an empirical examination within the context of data warehousing. **Journal of Management Information Systems**, [S. l.], v. 21, n. 4, p. 199-235, abr. 2005.

NOOGHABI, M. Z.; DASTGERDI, A. F. Proposed metrics for data accessibility in the context of linked open data. **Program**, [S. l.], v. 50, n. 2, p. 184-194, abr. 2016. Disponível em: <http://dx.doi.org/10.1108/prog-01-2015-0007>. Acesso em: 26 maio 2022.

POSSEMATO, T. How rda is essential in the reconciliation and conversion processes for quality linked data. **Jlis**, [S. l.], v. 1, n. 1, p. 48-60, fev. 2018. Disponível em: <http://dx.doi.org/10.4403/jlis.it-12447>. Acesso em: 26 maio 2022.

SOYLU, A.; CORCHO, Ó.; ELVESÆTER, B.; BADENES-OLMEDO, C.; YEDRO-MARTÍNEZ, F.; KOVACIC, M.; POSINKOVIC, M.; MEDVEŠČEK, M.; MAKGILL, I.; TAGGART, C.; SIMPERL, E.; LECH, T. C.; ROMAN, D. Data quality barriers for transparency in public procurement. **Information**, [S. l.], v. 13, n. 2, p. 99-115, fev. 2022. Disponível em: <http://dx.doi.org/10.3390/info13020099>. Acesso em: 26 maio 2022.

TALLERÁS, K. Quality of linked bibliographic data: the models, vocabularies, and links of data sets published by four national libraries. **Journal Of Library Metadata**, [S. l.], v. 17, n. 2, p. 126-155, abr. 2017. Disponível em: <http://dx.doi.org/10.1080/19386389.2017.1355166>. Acesso em: 26 maio 2022.

W3C. **Best practices for publishing linked data.** 2014. Disponível em: <https://www.w3.org/TR/ld-bp/>. Acesso em: 22 abr. 2021.

W3C. **Links in html documents.** 2018. Disponível em: <https://www.w3.org/TR/html401/cover.html#minitoc>. Acesso em: 08 set. 2020.

WANG, R. Y.; STRONG, D. M. Beyond accuracy: what data quality means to

data consumers. **Journal of Management Information Systems**, [S. l.], v. 12, n. 4, p 5-33, jan. 1996.

ZAVERI, A.; RULA, A.; MAURINO, A.; PIETROBON, R.; LEHMANN, J.; AUER, S.; HITZLER, P. Quality assessment methodologies for linked open data. **SWJ**, [S. l.], v.1, p. 1-5, 2012. Disponível em: <http://www.semantic-web-journal.net/system/files/swj414.pdf>. Acesso em: 22 abr. 2021.

LINKED DATA QUALITY: ANALYSIS OF THE THEME FROM THE PERSPECTIVE OF INFORMATION SCIENCE

ABSTRACT

Objective: The present study aims to identify how the issue of Linked Data data quality is addressed in scientific articles indexed in Information Science databases.

Methodology: As a result, 38 articles were retrieved, 9 being accepted to compose the theoretical corpus of the research. **Conclusions:** It is concluded that the relationship between Linked Data data quality and Information Science is still little explored through scientific articles and that it is usually carried out from a more applied perspective, not being identified studies with the objective of presenting a more in-depth theoretical Reference of the theme.

Descriptors: Data quality. Linked Data. Information Science. Systematic Mapping of Literature.

CALIDAD DE DATOS LINKED DATA: ANÁLISIS DEL TEMA DESDE LA PERSPECTIVA DE LAS CIENCIAS DE LA INFORMACIÓN

RESUMEN

Objetivo: Este estudio tiene como objetivo identificar cómo se aborda el tema de la calidad de los datos de Datos Vinculados en artículos científicos indexados en bases de datos de Ciencias de la Información. **Metodología:** Mapeo Sistemático de la Literatura, basado en un protocolo de investigación. **Resultados:** Como resultado, fueron recuperados 38 artículos, de los cuales 9 fueron aceptados para componer el corpus teórico de la investigación. **Conclusiones:** Se concluye que la relación entre la calidad de los datos de Linked Data y las Ciencias de la Información es aún poco explorada a través de artículos científicos y que generalmente se realiza desde una perspectiva aplicada, no identificándose estudios con el objetivo de presentar un referencial teórico en profundidad del tema.

Descriptores: Calidad de datos. Linked Data. Ciencias de la Información. Mapeo Sistemático de la Literatura.

Recebido em: 08.06.2022

Aceito em: 13.08.2022