



24° ENANCIB
Encontro Nacional de Pesquisa em Ciência da Informação
Perspectivas Contemporâneas na Ciência da Informação
• Vitória - ES • Ancib • PPGCI/UFES



XXIV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XXIV ENANCIB

ISSN 2177-3688

GT 8 – Informação e Tecnologia

O TERMO QUALIDADE DE DADOS NA CIÊNCIA DA INFORMAÇÃO BRASILEIRA: UMA ANÁLISE DE TERMINOLOGIA PONTUAL

THE TERM DATA QUALITY IN BRAZILIAN INFORMATION SCIENCE: AN ANALYSIS OF PUNCTUAL TERMINOLOGY

Ananda Fernanda de Jesus – Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp) Campus de Marília

Mario Barité – Facultad de Información y Comunicación (FIC)

José Eduardo Santarem Segundo – Universidade de São Paulo (USP), Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp) Campus de Marília

Modalidade: Trabalho Completo

Resumo: O presente contexto tecnológico é marcado por um crescimento exponencial no volume de dados, bem como, por uma multiplicação das ferramentas, técnicas e da capacidade de processamento desses dados para a geração de novas informações. Nesse contexto, a qualidade dos dados tem impacto direto na qualidade das informações, tornando ainda mais relevantes os processos de avaliação, seleção e melhoria de qualidade de dados. Compreendendo que a qualidade de dados é um dos objetos de interesse recentes da Ciência da Informação, objetivou-se analisar como é compreendido o termo qualidade de dados na Ciência da Informação Brasileira, partindo da análise terminológica das definições e da discussão dos conceitos correlatos utilizados na produção científica dessa comunidade. Como método para identificação dos estudos, das definições e dos conceitos, foi realizada uma revisão bibliográfica baseada em protocolo de pesquisa, com uma abordagem de terminologia pontual, utilizando o método da grade. Foram identificados 35 documentos, entre artigos, teses e dissertações, dos quais 26 foram aceitos para compor o corpus da pesquisa e 18 apresentaram definições para o conceito estudado. Com base nas definições cunhadas e/ou adotadas por autores da comunidade, foi possível discutir o termo no âmbito da Ciência da Informação e ainda estabelecer uma definição representativa.

Palavras-chave: terminologia pontual; qualidade de dados; Ciência da Informação brasileira.

Abstract: The present technological context is marked by an exponential growth in the volume of data, as well as a multiplication of tools, techniques and the processing capacity of this data to generate new information. In this context, data quality has a direct impact on the quality of information, making the processes of evaluating, selecting and improving data quality even more relevant. Understanding that data quality is one of the recent objects of interest in Information Science, the objective was to analyze

how the term data quality is understood in Brazilian Information Science, starting from the terminological analysis of definitions and the discussion of related concepts used in scientific production of this community. As a method for identifying studies, definitions and concepts, a bibliographic review was carried out based on a research protocol, with a specific terminology approach, using the grid method. 35 documents were identified, including articles, theses and dissertations, of which 26 were accepted to compose the research corpus, 18 presented definitions for the concept studied. Based on the definitions coined and/or adopted by authors from the community, it was possible to discuss the concept within the scope of Information Science and establish a representative definition.

Keywords: punctual terminology; data quality; Brazilian Information Science.

1 INTRODUÇÃO

O contexto tecnológico vigente é marcado por avanços no uso de dados como insumo, seja na tomada estratégica de decisão pelas organizações ou em atividades cotidianas, o uso de dados para alimentar ferramentas, para evidenciar cenários, confirmar ou descartar hipóteses tem se tornado constante, fazendo com que a qualidade desses dados tenha impacto em diferentes setores da sociedade.

Na Ciência da Informação, a discussão da qualidade de dados enquanto objeto de pesquisa ainda é recente, entretanto a preocupação com a qualidade dos dados não. Evidencia-se a preocupação com qualidade das representações de informação criadas no âmbito da Ciência da Informação, que resultam em dados e metadados que precisam ser confiáveis e consistentes. Destaca-se ainda a relação intrínseca entre qualidade de dados e qualidade de informações, como apontada por Santos e Santana (2013, p. 208):

O conceito de dados pode gerar avanços significativos para a área de ciência da informação, pois representa um problema de grande impacto e abrangência para a atividade de pesquisa, reforçando os pressupostos da ciência da informação. Nessa perspectiva, o tratamento da informação, a representação de recursos, a recuperação e a disseminação de informação se tornam áreas vinculadas à descrição, ao armazenamento, à preservação, ao acesso e à gestão de dados.

Nesse sentido, partiu-se do questionamento: como pode ser definido o conceito de qualidade de dados na Ciência da Informação Brasileira? O objetivo da pesquisa é analisar como é compreendido o termo qualidade de dados na Ciência da Informação Brasileira, partindo da análise terminológica das definições e da discussão dos conceitos correlatos utilizados na produção científica dessa comunidade, estabelecendo uma definição para o conceito que seja representativa para a comunidade.

A presente pesquisa é um estudo pontual, já que tem como foco a análise de um só termo e/ou termos próximos ou correlatos. Nesse caso concentra-se no termo “qualidade de dados”, a diferença dos estudos sistemáticos que se concentram em análise terminológica de um domínio ou área específica do conhecimento (Célestin; Godbout; Vachon-L'Heureux, 1984; L'Homme, 2004).

Para a identificação das definições e dos conceitos foi realizada uma revisão bibliográfica, baseada em um protocolo de pesquisa. Para a análise terminológica foi adotado o método de “la grilla”, ou em português a grade, proposto por Barité (2011). O método é utilizado para elaboração de definições em diferentes abordagens terminológicas, e permite, a partir da fragmentação e análise de distintas definições, a proposta de uma definição mais completa e representativa para determinado domínio. A próxima seção apresenta os procedimentos metodológicos empregados na pesquisa.

2 PROCEDIMENTOS METODOLÓGICOS

Caracterizada pelos seus objetivos como exploratória e descritiva, com resultados qualitativos, a presente pesquisa tem como base a condução de um levantamento bibliográfico baseado em protocolo de pesquisa. O quadro 1, apresentado a seguir, contém as informações utilizadas para estruturar o levantamento:

Quadro 1 – Protocolo de pesquisa

| Protocolo de busca | |
|---------------------------------|---|
| Pergunta de pesquisa principal | Como tem sido abordada a relação entre a Ciência da Informação e a qualidade de dados no âmbito nacional? |
| Pergunta de pesquisa secundária | Como os pesquisadores da Ciência da Informação Brasileira definem o conceito de qualidade de dados? |
| Objetivo | Compreender como os pesquisadores da Ciência da Informação Brasileira definem o conceito de qualidade de dados e como eles discutem a relação desse conceito com a Ciência da Informação |
| Estratégia de busca | “Qualidade de dados” |
| Base de dados consultada | BRAPCI, Catálogo de Teses e Dissertações (CAPES), Anais do ENANCIB e do Widat |
| Período abrangido | Sem restrição temporal. |
| Idiomas | Português |
| Coleta | Dezembro de 2023 |
| Crítérios de Inclusão | (I) Discute a criação de artefato de avaliação de qualidade (I) Conduz um processo de avaliação de qualidade (I) Discussão de artefatos de avaliação de qualidade (comparação, avaliação) (I) Abordagem teórica de qualidade de dados |

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

| | |
|------------------------|---|
| Critérios de exclusão | (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 dos documentos, conceito de qualidade de dados; relação entre qualidade de dados e Ciência da Informação; Conceitos correlatos |
| Data da coleta | janeiro 2024 |

Fonte: autores (2024).

Para o registro dos dados correspondentes as definições utilizou-se uma modalidade simples do **método da grade**, proposto por Barité (2011), que permite a análise comparativa das definições, com base em uma segmentação em: enunciados **essenciais** - atributos que necessariamente participam da definição de um objeto; **acidentais** – atributos que somente estão presentes em algumas instancias de um objeto mas que podem ser significativos; e elementos **informativos** – aqueles que não se referem aos atributos de um objeto definido, mas sim a dados ou informações que pretendem ilustrar aos leitores sobre aspectos relacionados com o objeto em seu contexto ou na realidade).

3 RESULTADOS

Com base no protocolo apresentado foram recuperados 35 documentos, entre artigos, monografias e trabalhos apresentados em eventos. Desses documentos, sete foram recusados, um duplicado e vinte e seis aceitos para compor o corpus da pesquisa. Entre os estudos aceitos, 18 apresentam definições para o conceito de qualidade de dados, sejam definições próprias ou definições acrescidas do uso de citações para apresentar e fundamentar o conceito. O quadro 2 apresenta as definições. Os destaques indicam pontos relevantes para o preenchimento da grade, que auxiliaram no entendimento do conceito de qualidade de dados.

Quadro 2 – Definições de qualidade de dados em artigos da Ciência da Informação Brasileira.

| Autores | Definições |
|----------------------------|---|
| BENTANCOURT e ROCHA (2012) | A qualidade pode ser definida em função da conformidade a especificações , denominada qualidade de projeto, que no caso dos metadados refere-se à conformidade com as especificações dos padrões adotados . A norma ISO/IEC 25012 aborda especificamente a qualidade dos dados de forma multidimensional , segundo critérios classificados em duas categorias. A primeira categoria reúne critérios relativos à qualidade inerente dos dados , onde possa atender-se uma demanda informacional a partir dos dados em si. A segunda, por outro lado, dependente do sistema , presume que a qualidade do dado é melhorada ou dependente do sistema computacional em que |

**XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024**

| Autores | Definições |
|--|---|
| | opera. No entanto, também há casos em que ambas as categorias estão presentes. (p. 88) |
| ALMEIDA <i>et al.</i> (2016) | Com foco na menor quantidade possível de anomalias, a qualidade de dados é um fator importante para determinar a precisão da tomada de decisão quando se utiliza informações provenientes de bancos de dados. [...] O nível de qualidade de um banco de dados pode ser medido de forma específica, utilizando os conceitos apresentados no trabalho de Strong, Lee e Wang (1997). Em sua pesquisa, os autores definiram categorias e dimensões para a qualidade de dados. (p. 74). |
| MELO (2017) | Quanto à qualidade de dados, outro fator relevante identificado consiste na forma como as dimensões são definidas, ou seja, são estabelecidas, visando atender às necessidades específicas do domínio ao qual estão sendo aplicadas. Em razão disso, as dimensões variam de acordo com cada metodologia e domínio; assim, algumas dimensões mostraram-se presentes em diversas metodologias, como completude, precisão e timeliness. Ressalte-se o fato de que a maioria das metodologias utiliza a avaliação subjetiva para a definição do que pode ser considerado de qualidade ou não. O domínio do sistema de informação desempenha um papel muito importante para medir a qualidade dos dados, visto que é por meio dele que os requisitos são estabelecidos. Tais requisitos podem ser definidos como itens aos quais o dado deve atender para ser considerado de qualidade. 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. No geral, as dimensões possuem descrições e definições similares, porém a aplicação e como funciona cada dimensão possuem características distintas, de acordo com o domínio no qual estão sendo aplicadas. (p. 48) [...] |
| MELO; BOTEGA; SANTAREM SEGUNDO (2017a) | De acordo com Olson (2003) um dado é de qualidade se satisfaz os requisitos para o seu uso ; conseqüentemente, carece de qualidade quando não os satisfaz. Considerando que cada domínio possui diferentes requisitos de qualidade, a literatura aborda diferentes pontos de vista quanto à aplicação das dimensões de qualidade. |
| MELO, BOTEGA, SANTAREM SEGUNDO (2017b) | Qualidade aplicada a dados indica que se espera que eles atendam à medida de perfeição, precisão e conformidade no domínio no qual estão inseridos. Dados que não descrevem fielmente componentes do mundo real diminuem a efetividade dos sistemas, contribuem de forma negativa para as atividades envolvidas em sua utilização e seus impactos podem ser tanto sociais quanto econômicos. De acordo com Olson (2003) um dado é de qualidade se satisfaz os requisitos para o seu uso; conseqüentemente, carece de qualidade quando não os satisfaz. Considerando que cada domínio possui diferentes requisitos de qualidade , a literatura aborda diferentes pontos de vista quanto à aplicação das dimensões de qualidade. (Wang; Strong, 1996; Veregin, 1999; Lee <i>et al.</i> , 2002; Pipino <i>et al.</i> , 2002; Batini <i>et al.</i> , 2008; Laudon, 1986; Fisher; Kingma, 2001) (p. 83). |
| ESPÍNDOLA <i>et al.</i> (2018) | Entende-se que dados de qualidade também são essenciais para o sucesso de pesquisas científicas, pois os dados científicos requerem a adoção de boas práticas de coleta, armazenamento e recuperação (p. 276). Assim, no contexto da qualidade dos dados em organizações, compreende-se que o dado tem qualidade , quando atende às necessidades dos stakeholders da instituição de onde ele é produzido, armazenado e utilizado. E também quando são empregados recursos (tecnológicos, pessoais e |

**XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024**

| Autores | Definições |
|--|--|
| | <p>financeiros) de forma eficiente e eficaz para sua produção, gerenciamento e atendimento de padrões internos e externos de qualidade. Wang e Strong (1996) afirmam que dados de qualidade são aqueles que estão aptos para o uso de seu consumidor. Mezzananza <i>et al.</i> (2015, p. 148) explicam que isso faz com que a qualidade dos dados seja um conceito que dependente do contexto em que está sendo abordada, pois “[...] um conjunto de dados pode ser considerado apropriado para um uso, embora possa não ser adequado para outro.” (p. 282) Para avaliar a qualidade dos dados é necessária a determinação de atributos, os quais são denominados por Jesiljevska (2017) como dimensões de qualidade de dados. Tais dimensões podem ser utilizadas para mensurar a qualidade de dados em diferentes contextos. Dentre tais dimensões, têm-se, por exemplo: acurácia, temporalidade, consistência e completude dos dados. [...] (p. 283) O controle da qualidade dos dados nas organizações é um dos processos compreendidos pela governança de dados. Por meio da governança de dados são definidas métricas, procedimentos e requisitos que auxiliam a organização a atingir a qualidade de dados necessária para cumprir suas demandas e alcançar seus objetivos (Barata, 2015). (p. 276) A qualidade dos dados pode ser considerada um dos principais focos da governança de dados (Barbieri, 2011). (p. 282).</p> |
| FAGUNDES; MACEDO e FREUND (2018) | <p>Dada a importância das informações disponibilizadas, a qualidade dos dados que geram estas informações tornou-se um dos grandes desafios para as organizações se manterem em um mercado cada vez mais competitivo, sendo que a partir da década de 90 iniciaram-se diversos estudos sobre metodologias e ferramentas para auxiliar no processo de gestão da qualidade dos dados dentro das organizações [...] (196). Para Wang e Strong (1996), o conceito de qualidade dos dados é considerado multidimensional e as suas dimensões são definidas como um conjunto de atributos de qualidade que representam um único aspecto da qualidade de dados. Observa-se na literatura, tanto estudos que propõem que a qualidade dos dados seja considerada de forma genérica, onde a partir de um determinado modelo a qualidade pode ser gerida independente do domínio de aplicação, como também pesquisas que procuraram definir e operacionalizar as dimensões de qualidade de dados específicas para determinados contextos. Wang e Strong (1996), conduziram o que foi considerada a primeira pesquisa empírica com o propósito de identificar as dimensões da qualidade dos dados [...] (p. 197).</p> |
| PICCOLO (2018) | <p>Nas pesquisas, a definição de qualidade dos dados é frequentemente descrita como dados que são “aptos para o uso”, que baseia no julgamento subjetivo da qualidade dos dados por usuários. Para Wang e Strong (1996), um dado com qualidade é aquele que o usuário seja capaz de interpretar, seja acessível, além de relevante, exato e vir de fonte confiável. Todavia, a qualidade dos dados é mais do que simplesmente sua exatidão, pode significar integridade, confiabilidade, consistência e atualidade. Desta forma, autores definem diferentes dimensões para caracterizar tal aspecto qualitativo (p. 13).</p> |
| MOURA JUNIOR; ARAGÃO (2018) | <p>A necessidade de avaliar a qualidade de dados vêm ganhando importância, uma vez que a real contribuição (o valor de negócio) dos dados só pode ser estimada no seu contexto de uso. A metrificação da qualidade de dados constitui-se de um esquema cíclico, iniciando no mapeamento da documentação e do comportamento dos dados, através da observação de trechos das bases.</p> |
| MOREIRA <i>et al.</i> (2020) | <p>Para Todoran <i>et al.</i> (2015), novas questões e problemas relacionados à complexidade, heterogeneidade e qualidade dos dados estão emergindo, tornando-se uma perspectiva cada vez mais relevantes em ambientes informacionais. [...] Todavia, a qualidade dos dados é mais do que simplesmente sua exatidão, pode significar integridade, confiabilidade, consistência e atualidade. Desta forma, autores definem diferentes dimensões para caracterizar tal aspecto qualitativo (p. 13).</p> |

**XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024**

| Autores | Definições |
|----------------------------------|---|
| PICCOLO <i>et al.</i> (2021) | <p>É importante considerar que a qualidade de dados é um conceito de sentido delimitado tradicionalmente na literatura científica, de modo que constitui um campo consolidado do conhecimento, que possui arcabouço teórico próprio. [...] A qualidade de dados, de um modo geral, é compreendida em dimensões. A respeito disso, os autores abaixo mencionados apontam que [...] existem Dimensões mais utilizadas, porém não há um padrão estabelecido de dimensões; cada domínio utiliza dimensões que atendam aos seus requisitos específicos. Dentre as mais utilizadas na literatura constam: completude, precisão, relevância e consistência. (Melo; Botega; Santarem Segundo, 2017, p. 84). Devido ao fato de a qualidade de dados não ser um parâmetro absoluto, mas relativo aos requisitos do domínio e dos usos das atividades pretendidas, não é apropriado tentar estabelecer uma metodologia aplicável, com profundidade satisfatória, irrestritamente a quaisquer situações. Desse modo, dimensões funcionam como aspectos pelos quais a qualidade de dados pode ser avaliada de maneira adequada, dentro de um domínio. Isso pressupõe que, para determinado domínio, sejam definidas essas dimensões, o que pode ser concretizado no âmbito de uma metodologia de avaliação construída com foco nesse contexto específico (p. 6). A partir da consciência de que a qualidade de dados é um parâmetro relativo à implementação de serviços dessa natureza deve passar por preparação, com foco no atendimento a requisitos como os concernentes ao domínio de atuação – que eventualmente pode estar contemplado por literatura específica para qualidade de dados em gestão de dados de pesquisa no referido domínio –, bem como aos usos pretendidos para os dados. (p. 21)</p> |
| GUALDANI <i>et al.</i> (2022) | <p>A qualidade de dados é um tema propício como objeto de estudo de diversos campos da ciência, tornando-se parte de um escopo interdisciplinar que abrange diversas áreas do conhecimento. O tema ganha ainda mais força ao visar a quantidade de dados que é disponibilizada diariamente no mundo todo, muitos deles, podem ser mais bem disponibilizados e apresentados mediante utilização de conceitos e técnicas abordadas nesta pesquisa (p. 485).</p> |
| JESUS; SANTAREM SEGUNDO (2022) | <p>A qualidade de dados é intrínseca à capacidade de atuação satisfatória nas atividades ou aplicações nas quais esses dados vão ser empregados, podendo ser avaliada através de dimensões e métricas específicas para cada domínio. (p. 1) A preocupação com a questão da qualidade dos dados não é algo recente, remetendo à década de 1970, com a busca por estabelecer formas de analisar a qualidade dos dados. 2) Wang e Strong (1996) organizam o processo de avaliação em categorias, dimensões de qualidade e critérios de qualidade. [...]</p> |
| SILVA <i>et al.</i> (2022) | <p>Uma má qualidade de dados pode influenciar negativamente humanos operadores de sistemas e processos computacionais que dependem de parâmetros corretos para produzir melhores informações e contribuir para uma tomada de decisão mais assertiva. “[...] O gerenciamento da qualidade de dados se dá pela “definição de papéis, responsabilidades, políticas e procedimentos relacionados à aquisição, manutenção, representação e disseminação de dados e informações” (Botega <i>et al.</i>, 2017).</p> |
| JESUS e SANTAREM SEGUNDO (2023a) | <p>O processo de avaliação de qualidade de dados é baseado nos conceitos de dimensões e métricas. As dimensões estão relacionadas às características dos dados, ou conjuntos de dados (datasets), que os usuários precisam avaliar para decidir se esses são adequados para a tarefa que pretendem realizar, sendo cada característica geral contemplada por uma dimensão. Para mensurar o quão adequado um dataset está em relação a determinada dimensão, são elaboradas métricas que permitem quantificar e qualificar esses dados. A escolha das dimensões e o peso de cada métrica, nessa perspectiva contextual, precisam ser estabelecidas pelo usuário no início do processo de avaliação, pois estão condicionadas às suas necessidades e às tarefas nas quais os dados serão empregados. Para reutilizar datasets em novas aplicações os usuários precisam realizar o</p> |

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

| Autores | Definições |
|----------------------------------|--|
| | <p>processo de avaliação de qualidade, que pode ser feito tanto de forma manual como utilizando-se ferramentas que tornam o processo automático ou semiautomático. Eles também podem consultar processos de avaliação realizados por outros usuários da comunidade ou por agências de certificação. Nesse sentido a descrição formal da qualidade dos datasets pode ter impacto direto na sua reutilização, auxiliando consumidores de dados no processo de seleção de fontes adequadas para suas aplicações, pois facilita o uso de agentes computacionais e consequentemente adoção de ferramentas automáticas e semiautomáticas de avaliação de qualidade." (p. 3).</p> |
| JESUS e SANTAREM SEGUNDO (2023b) | <p>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, 1995; Zaveri <i>et al.</i>, 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. 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) (p. 4).</p> |
| LEMOS; COELHO JÚNIOR (2023) | <p>Em suma, a obtenção da qualidade de dados em coleções de acervos é consequência da modelagem apropriada de metadados e sua aplicação pelo profissional da informação em processos de organização e representação da informação e do conhecimento (Svenonius, 2000; Taylor, 2004; Hjørland, 2007; Gilliland, 2016) endereçados a soluções de problemas da comunicação efetiva de conhecimento e de seus registros à sociedade (Saracevic, 1996). Gilliland (2016) destaca tipologias de padrão de dados que deveriam ser levadas em consideração na ação de tratamento nos dados, focando padronização, normalização, qualidade e intercâmbio de metadados em ambiente digital, quais sejam i) estrutura de dados: conjunto de elementos de metadados ou esquemas de categorias que formam um registro de informação; ii) valores dos dados: linguagens documentárias, vocabulários controlados, arquivos de autoridade e ontologias de domínio usados para preencher os dados nos elementos de metadados; iii) conteúdo dos dados: regras e códigos de catalogação (p. 3)</p> |

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

| Autores | Definições |
|----------------------|--|
| COELHO JUNIOR (2023) | Existem muitos aspectos sobre qualidade de dados, incluindo modelagem e gerenciamento, controle e garantia de qualidade, análise, armazenamento e acesso (Chapman, 2005). A abordagem usada para lidar com cada um desses aspectos dependerá da aplicação e do nível da qualidade de dados exigida para a utilização desses dados (Usaid, 2009). Assim, um dos principais desafios é determinar qual nível de qualidade de dados é aceitável para o fim almejado. A qualidade não é necessariamente dados isentos de erros, sendo esta, apenas uma das dimensões. Deve-se, portanto, considerar outras dimensões, tais como a precisão, a integridade e a consistência dos dados, que, por fim, corroboraram com a medida da adequação dos dados a um propósito específico (Eckerson, 2002). No entanto, em geral, a qualidade dos dados pode ser vista como um subconjunto da qualidade da informação . Isso ocorre porque a qualidade dos dados se concentra na precisão e na integridade dos dados, enquanto a qualidade da informação também leva em consideração o significado dos dados e como eles são usados (Chapman, 2005). (p. 10-11). |

Fonte: Autores (2024).

O primeiro aspecto analisado foi o potencial gênero do termo. Com base nas definições é possível afirmar que a qualidade de dados na Ciência da Informação é aplicada como um termo poliédrico, representando diferentes conceitos, inclusive em uma mesma definição, conforme o princípio da poliedricidade que encontra Cabré (1999). Segundo a autora a construção dos conceitos especializados, se expressa em: poliedricidade denominativa, cognitiva e funcional (Cabré, 1999, p. 129). Isso implica aceitar que, em determinados contextos terminológicos, o princípio da unicidade promovido pelas escolas clássicas de terminologia, deve resignar-se em benefício da descrição fiel de uma área especializada do saber.

De maneira explícita, a qualidade de dados é definida pelos autores como: um conceito; como um tema de investigação; como um campo consolidado do conhecimento; como um processo; como parte integrante de outros domínios; como um parâmetro para a tomada de decisão, como um nível que pode ser ou não atingido. Pode representar a um campo consolidado de conhecimento, uma área de estudo, um processo, uma medida ou uma ponderação a ser mensurada. Também pode ser entendido como um subcampo relacionado a outros objetos de interesse da Ciência da Informação.

O próximo aspecto analisado nas definições foi em relação aos enunciados essenciais, que permitem especificar o termo em relação ao seu gênero. **Enquanto um campo do conhecimento, um domínio** ou ainda um **tema de investigação**, a qualidade de dados possui um escopo interdisciplinar, de interesse de diversas áreas do conhecimento, no qual são discutidas, investigadas e propostas formas de mensurar e melhorar os níveis de qualidade

dos conjuntos de dados, também são discutidos e formulados artefatos e metodologias que permitam auxiliar na condução de processos de avaliação e melhoria da qualidade dos dados.

Enquanto um processo, a qualidade de dados pode ser abordada pela literatura como um sinônimo de **avaliação de qualidade**. Esse processo é compreendido com base principalmente na estrutura elaborada por Wang e Strong (1996), onde a avaliação é realizada por meio da escolha de dimensões, critérios e métricas que permitem mensurar o nível de qualidade dos dados. As dimensões permitem agrupar os diferentes aspectos que serão avaliados nos conjuntos de dados, sendo compostas por um conjunto de critérios, que descrevem os atributos específicos que serão analisados. As métricas são os indicadores que permitem mensurar a qualidade dos dados em relação a um critério.

Enquanto uma medida ou uma ponderação a qualidade de dados é abordada como um sinônimo dos conceitos de dados de qualidade, qualidade dos dados, ou ainda de qualidade aplicada a dados. (Almeida *et al.*, 2016; Melo; Botega; Santarem Segundo, 2017; Piccolo; 2018). Nessa acepção, a qualidade dos dados é comumente representada na literatura pela expressão “*fitness for use*”, ou adequado ao uso, onde para serem considerados de qualidade os dados devem atender satisfatoriamente às necessidades dos usuários (Juran, 1988).

Em todos esses aspectos o entendimento e aplicação do termo qualidade de dados possui uma **característica multidimensional**, podendo ser abordado por diferentes perspectivas, ou dimensões, que afetam a forma como se realiza o processo de avaliação de qualidade e como se identifica se um conjunto de dados é ou não de qualidade. Em relação a essas dimensões, elas geralmente seguem o processo de categorização apresentado por Wang e Strong (1996), que é a referência mais frequente nas definições, tanto para apresentar as diferentes dimensões, como para descrever o processo de avaliação de qualidade. As principais dimensões, ou ainda perspectivas, do conceito de qualidade de dados são: contextual, intrínseca e a representacional e de acessibilidade.

Na dimensão contextual, discutir, avaliar, mensurar e melhorar a qualidade de dados são ações que dependem das características do domínio e das necessidades do contexto onde serão empregados esses dados. Em uma perspectiva contextual um dado é considerado de qualidade quando atende satisfatoriamente às necessidades do domínio ou de seu contexto de aplicação. Em uma perspectiva intrínseca, se discute aspectos da qualidade dos dados

através de um conjunto de características inerentes a esses dados, com uma abordagem mais genérica, de uma maneira independente do domínio e do contexto de futura aplicação. Nessa perspectiva um conjunto de dados pode ser considerado de qualidade quando está livre de erros e de más formações.

A perspectiva representacional é focada nos processos de descrição, representação e organização dos conjuntos de dados, por meio de vocabulários, metadados, padrões de metadados. Nessa perspectiva se discute e avalia qualidade em relação a ausência ou presença das informações necessárias para que esses conjuntos de dados possam ser encontrados e reutilizados. Nessa perspectiva não se busca identificar o nível de qualidade dos conjuntos de dados, mas sim o quão bem descritos, representados e organizados esses dados estão, e o processo de avaliação de qualidade será voltado para a avaliação dos metadados relacionados ao conjunto de dados.

A perspectiva de acessibilidade está relacionada à capacidade dos usuários de obterem acesso aos dados de maneira eficaz e eficiente. Nessa perspectiva um dado é considerado de qualidade quando pode ser localizado e obtido quando necessário.

Ao analisar as definições do termo, observa-se que esse guarda muitas similaridades e compartilha referências com as definições de qualidade de dados em outras áreas, como ciência e computação, o que parcialmente se justifica por esse ser um objeto recente de interesse da área, que ainda está sendo adaptado e aproximado aos objetos da Ciência da Informação.

Com base nos enunciados informativos e os termos correlatos dos objetos e objetivos dos estudos que compõem o corpus da pesquisa, observa-se a preocupação como representar, organizar, e permitir o uso e o reuso desses conjuntos de dados. Nesse sentido, ressalta-se que na Ciência da Informação, a qualidade de dados pode ser compreendida como um subcampo, ou ainda, uma etapa integrante de processos relacionados à: Curadoria Digital, Curadoria de Dados, Preservação Digital; Gestão de Dados; Governança de Dados; Ciclo de Vida dos Dados. Também se destacam aproximações relacionadas às áreas de Organização e Representação do Conhecimento, sendo frequentemente discutida e aplicada em uma perspectiva representacional, onde se busca avaliar a qualidade das representações criadas em relação a sua capacidade de recuperação, destacando-se a preocupação com os dados e metadados derivados desses processos. (Almeida *et al.*, 2016; Fagundes; Macedo; Freund, 2018; Gualdani *et al.*, 2022; Moreira *et al.*, 2020).

Destacam-se entre os termos correlatos a elaboração e aplicação de instrumentos de padronização, tais como vocabulários, ontologias, códigos de catalogação, padrões de metadados, políticas e boas práticas, que atuam como parâmetros para guiar o processo avaliação de qualidade dos dados na Ciência da Informação. Também é possível traçar correlações com as áreas de informação e Tecnologia, principalmente pelos conceitos correlatos relacionados aos objetos de estudo dos processos avaliação de qualidade de dados na Ciência da Informação, onde se destaca a abordagem interdisciplinar com as áreas tecnológicas, com os processos de elaboração, seleção e aplicação de artefatos que permitam avaliar e melhorar a qualidade dos dados.

Conforme o que sugere a investigação terminológica pontual realizada, a qualidade de dados parece ser um termo que conta com as seguintes definições:

1. Enquanto um campo, ou domínio, a qualidade de dados possui um escopo interdisciplinar, de interesse de diversas áreas do conhecimento no qual são discutidas, investigadas e propostas formas de mensurar e melhorar os níveis de qualidade dos conjuntos de dados, assim como formular artefatos e metodologias que permitam auxiliar na condução de processos de avaliação e melhoria de qualidade de dados. Entre outras possibilidades, pode ser entendido como um subcampo relacionado a objetos de interesse da Ciência da Informação, como a organização e representação da informação e do conhecimento, a gestão da informação e as relações entre informação e tecnologia. **2. Enquanto processo**, a qualidade de dados é abordada pela literatura como um sinônimo do processo de avaliação de qualidade de dados, que é realizado por meio do estabelecimento de critérios, dimensões e métricas que permitem mensurar os níveis de qualidade dos conjuntos de dados. **3. Enquanto uma medida ou uma ponderação**, qualidade de dados é abordada como um sinônimo dos conceitos “dados de qualidade”, “qualidade dos dados” ou “qualidade aplicada a dados”, onde a qualidade dos dados é um aspecto que pode ser mensurado e melhorado com base na perspectiva adotada. O caráter multidimensional da qualidade dos dados perpassa as três acepções identificadas, podendo ser considerar as principais dimensões: a **contextual**, onde discutir, mensurar e melhorar a qualidade de dados são ações que dependem das características do domínio e das necessidades do contexto onde serão empregados esses dados; **intrínseca** onde se discute aspectos da qualidade dos dados por meio de um conjunto de características inerentes a esses dados, avaliando o quão bem estruturados, completos e livres de erros estão; **representacional**, onde se discute e avalia qualidade em relação a ausência ou presença das informações necessárias para que esses conjuntos de dados possam ser recuperados e reutilizados; e a **acessibilidade**, que está relacionada à capacidade dos usuários de obterem acesso aos dados de maneira eficaz e eficiente.

Apresentados discutidos os principais resultados obtidos, a próxima seção apresenta as considerações finais.

4 CONSIDERAÇÕES FINAIS

A presente pesquisa teve como objetivo discutir o termo qualidade de dados na Ciência da Informação Brasileira, com base em uma análise de terminologia pontual, amparada nas definições apresentadas pela comunidade científica.

Para atender o objetivo proposto destaca-se a importância do método da grade (Barité, 2011), um método que tem se consolidado em diferentes abordagens terminológicas. Entende-se que o método apresenta-se como uma importante ferramenta, que ao viabilizar o estabelecimento de definições em um domínio, por meio de abordagens de terminologia pontual, podendo instrumentalizar construções teóricas importantes, inclusive no âmbito do GT 8, núcleo das principais discussões sobre informação e tecnologia na Ciência da Informação Brasileira, perpassado por constantes mudanças, apresentações e proposições de novos termos, marcado pela interdisciplinaridade e pelos constantes câmbios tecnológicos.

Na presente pesquisa, a aplicação da grade permitiu compreender como o termo qualidade de dados é abordado e definido nas pesquisas realizadas pela comunidade da Ciência da Informação Brasileira. Com base nessa análise, conclui-se que o termo qualidade de dados possui uma abordagem poliédrica, com características interdisciplinares e multidimensionais.

Observa-se que as definições do termo guardam muitas similaridades e compartilham referências com as definições de qualidade de dados em outras áreas, como ciência e computação, o que parcialmente se justifica por esse ser um objeto recente de interesse da Ciência da Informação. Embora as primeiras discussões de qualidade de dados na Ciência da Informação tiveram início em 1989, as discussões a nível nacional são mais recentes, com o primeiro estudo identificado em bases de dados em 2012, com publicações constantes apenas a partir de 2016, tendo sua comunidade e discurso ainda em desenvolvimento.

Entretanto, uma análise inicial dos termos correlatos e dos objetos e objetivos dos estudos que compõem o *corpus* da presente pesquisa permite inferir que na Ciência da Informação Brasileira, a qualidade de dados aparece como um subprocesso ou etapa relacionado à Curadoria Digital, Curadoria de dados, Preservação Digital; Gestão de Dados; Governança de Dados; Ciclo de vida dos dados. Podendo ser relacionada aos processos de representação e organização da informação e do conhecimento, tanto em uma abordagem representacional, onde se busca avaliar e discutir o quão bem representados e organizados

estão os conjuntos de dados, como através da utilização de instrumentos de padronização, tais como vocabulário, ontologias, códigos de catalogação e padrões de metadados, como instrumentos para o processo de avaliação de qualidade. Pode se relacionar também com a área de gestão da informação e com os estudos que relacionam informação e tecnologia, especialmente por sua interdisciplinaridade com áreas como a computação.

Com a apresentação da definição, entende-se que o objetivo da presente pesquisa foi atingido e que a pergunta de pesquisa foi respondida. Como estudos futuros, pretende-se realizar uma análise aprofundada dos conceitos correlatos apresentados, categorizando-os e apresentando as devidas relações hierárquicas. Também pretende-se realizar uma análise aprofundada dos objetivos e objetos de estudo dos artigos aceitos e das relações entre Ciência da Informação e qualidade de dados apresentada pelos autores visando compreender como e com que objetivos são conduzidos estudos de qualidade de dados na Ciência da Informação Brasileira.

REFERÊNCIAS

ALMEIDA, C. A. *et al.* Melhoria na qualidade de dados com a aplicação de. **Atoz**: novas práticas em informação e conhecimento, Curitiba, v. 5, n. 2, p. 72, 2016. Disponível em: <http://dx.doi.org/10.5380/atoz.v5i2.47303>. Acesso em: 19 dez. 2023.

BARATA, A. M. **Governança de dados em organizações brasileiras**: uma avaliação comparativa entre os benefícios previstos na literatura e os obtidos pelas organizações. 2015. Dissertação (Mestrado em Ciências) – Programa de Pós-Graduação em Sistemas de Informação, Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2015.

BARITÉ, M. **La garantía literaria como herramienta de revisión de sistemas de organización del conocimiento**: modelo y aplicación. 2011. Tesis (Doctorado) – Universidad de Granada, Granada, 2011. Disponível em: digibug.ugr.es/bitstream/10481/17583/1/19711864.pdf

CABRÉ, M. T. **La Terminología**: representación y comunicación: elementos para una teoría de base comunicativa y otros artículos. Barcelona: Universitat Pompeu Fabra, 1999.

CÉLESTIN, T.; GODBOUT, G.; VACHON-L'HEUREUX, P. **Méthodologie de la recherche terminologique ponctuelle**. Québec: Office de la Langue Française, 1984.

COELHO JÚNIOR, A. **Qualidade de dados em acervos do patrimônio cultural**: uma proposta diagnóstica semiautomática para objetos culturais sob gestão do instituto brasileiro de museus. 2023. Dissertação (Mestrado) – Programa de Pós-Graduação em Ciência da Informação, Universidade Federal do Espírito Santo, Vitória, 2023.

**XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024**

ESPÍNDOLA, P. L. *et al.* Governança de dados aplicada à ciência da informação. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 16, n. 3, p. 274-298, ago. 2018. Disponível em: <http://dx.doi.org/10.20396/rdbci.v16i3.8651080>. Acesso em: 19 jun. 2024.

FAGUNDES, P. B; MACEDO, D. D. J.; FREUND, G. P. A produção científica sobre qualidade de dados em big data: um estudo na base de dados web of science. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 16, n. 1, p. 194, 9 nov. 2017. Disponível em: <https://doi.org/10.20396/rdbci.v16i1.8650412>. Acesso em: 19 dez. 2023.

GUALDANI, F. A. *et al.* Critérios de qualidade de dados em saúde: uma análise quantitativa. **Informação & Informação**, Londrina, v. 27, n. 2, p. 466–490, 2022. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/43782>. Acesso em: 20 dez. 2023.

JESUS, A. F.; SANTAREM SEGUNDO, J. E. A. Questão da qualidade em dados publicados como linked data: um mapeamento sistemático da literatura. In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 22., 2022, Porto Alegre. **Anais [...]**. Porto Alegre: Ancib, 2022. p. 1-16. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxiiencib/paper/viewFile/811/729>. Acesso em: 19 dez. 2022.

JESUS, A. F.; SANTAREM SEGUNDO, J. E. Qualidade de dados Linked Data. **Informação@Profissões**, Londrina, v. 11, n. 2, p. 153-169, set. 2023a. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/infoprof/article/view/48595>. Acesso em: 19 dez. 2023.

JESUS, A. F.; SANTAREM SEGUNDO, J. E. S. A descrição formal da qualidade de dados publicados na web: análise do data quality vocabulary (dqv). **Em Questão**, v. 29, n., 2023b. Disponível em: <https://brapci.inf.br/#/v/229265>.

JURAN, J. M. **Quality Control Handbook**. New York: Mcgraw-Hill. 1988.

LEMOS, D. L. S; COELHO JUNIOR, A. Qualidade de dados em acervos do patrimônio cultural: uma avaliação diagnóstica semiautomática nos objetos culturais sob gestão do instituto brasileiro de museus. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 28, p. 1-22, 7 fev. 2023. Disponível em: <https://doi.org/10.5007/1518-2924.2023.e90510>. Acesso em: 19 dez. 2023.

MELO, J. O. S. F. **Metodologia de avaliação de qualidade de dados no contexto do linked data**. 2017. Tese (Doutorado) – Programa de Pósgraduação em Ciência da Informação, Faculdade de Filosofia e Ciências de Marília, Universidade Estadual Paulista, Marília, 2017. Disponível em: <https://repositorio.unesp.br/server/api/core/bitstreams/a4efef73-6c80-4060-b48b-27dddf311409/content>. Acesso em: 19 dez. 2023.

MELO, J. O. S. F.; BOTECA, Leonardo Castro; SANTAREM SEGUNDO, José Eduardo. Metodologia de Avaliação de Qualidade para Dados Conectados. **Informação & Tecnologia**, [S. l.], v. 4, n. 2, p. 80-101, 20 out. 2017a. Disponível em: <https://doi.org/10.22478/ufpb.2358-3908.2017v4n2.40539>. Acesso em: 19 dez. 2023.

MELO, J. O. S. F.; BOTECA, Leonardo Castro; SANTAREM SEGUNDO, José Eduardo. Metodologia De Avaliação De Qualidade Para Dados Conectados. In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 18., 2017b, Marília. **Anais [...]**. Marília: Ancib, 2021. p. 1-16.

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

L'HOMME, M. C. *La terminologie*: principes et techniques. Montréal: Presses de l'Université de Montréal, 2004.

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**, v. 21 n. 4, p. 199–235.

SANTOS, P. L. V. A. C.; SANTANA, R. C. G. Dado e Granularidade na perspectiva da Informação e Tecnologia: uma interpretação pela ciência da informação. **Ciência da Informação**, Brasília, v. 42, n. 2, p. 199-209, 27 jan. 2015. Disponível em: <http://dx.doi.org/10.18225/ci.inf.v42i2.1382>. Acesso em: 19 jun. 2023.

SILVA, J. N. *et al.* Desenvolvimento de Ontologia Ciente de Qualidade de Informações para a Melhoria de Consciência Situacional no Domínio de Gerenciamento de Emergências. In: WORKSHOP DE INFORMAÇÃO, DADOS E TECNOLOGIA, 1., 2017, Florianópolis. **Anais [...]**. Florianópolis: Ufsc, 2017. p. 1-6. Disponível em: <https://www.widat2017.ufsc.br/wp-content/uploads/2017/09/ST2.2.pdf>. Acesso em: 19 jun. 2024.

AGRADECIMENTOS

Agradecemos à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo financiamento recebido para o desenvolvimento dessa pesquisa (Processo FAPESP [2021/03349-0](#) e Processo FAPESP [2023/13574-7](#))