





Qualidade de dados: análise focada na categoria representacional

Data quality: analysis focused on the representational category



Débora Kaori Yoshitane¹

 <http://lattes.cnpq.br/2099435419438744>
 <https://orcid.org/0009-0005-4214-3602>

Ananda Fernanda de Jesus²

 <http://lattes.cnpq.br/8553935355036462>
 <https://orcid.org/0000-0001-7873-6040>

José Eduardo Santarem Segundo³

 <http://lattes.cnpq.br/5562746387565465>
 <https://orcid.org/0000-0003-3360-7872>

Resumo

A qualidade de dados pode ser dividida em quatro categorias: intrínseca, contextual, representacional e acessibilidade. A qualidade representacional é definida a partir de quão bem estruturados estão os conjuntos de dados, avaliando principalmente aspectos relacionados ao formato e significado desses dados. Esta pesquisa tem como objetivo verificar o estado da arte das discussões a respeito da categoria representacional em pesquisas de qualidade de dados, buscando identificar como o assunto tem sido discutido pela literatura e potenciais temas a serem explorados. Como metodologia, utilizou-se a análise exploratória baseada em um protocolo de pesquisa. Os resultados mostram a preocupação de diversos setores com a qualidade de dados, apontando a qualidade representacional como um fator importante na percepção dos consumidores de dados para tomadas de decisão. Conclui-se que a qualidade representacional de dados ainda é um assunto pouco explorado isoladamente. Conclui-se ainda que a definição do conceito de qualidade de dados representacional, suas dimensões, critérios e métricas ainda podem ser estudados de maneira mais aprofundada.

Palavras-chave: qualidade de dados; qualidade representacional; análise exploratória.

¹ Graduanda em Biblioteconomia e Ciência da Informação, Universidade de São Paulo (USP), Ribeirão Preto, São Paulo, Brasil. Bolsista do Programa Unificado de Bolsas (PUB-USP). deborayoshitane@usp.br.

² Doutoranda em Ciência da Informação, Universidade Estadual Paulista (UNESP), Marília, São Paulo, Brasil. Bolsista da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). af.jesus@unesp.br.

³ Livre Docente em Informação e Tecnologia pela Universidade de São Paulo (USP), Ribeirão Preto, São Paulo, Brasil. Professor Associado, Universidade de São Paulo (USP). santarem@usp.br.

Abstract

Data quality can be divided into four categories: intrinsic, contextual, representational, and accessibility. Representational quality is defined based on how well structured the data sets are, mainly evaluating aspects related to the format and meaning of these data. This research aims to verify the state of the art of discussions regarding the representational dimension in data quality research, seeking to identify how the subject has been discussed in the literature and potential themes to be explored. The methodology used was an exploratory analysis based on a research protocol. The results show the concern of several sectors with data quality, pointing to representational quality as an important factor in the perception of data consumers for decision-making. It is concluded that the representational quality of data is still a subject little explored in isolation, and that the definition of the concept of representational quality, its dimensions, criteria, and metrics can still be further deepened and studied.

Keywords: data quality; representational quality; exploratory analysis.

1 INTRODUÇÃO

Tendo em vista as consequências de problemas de qualidade de dados, como impactos econômicos e sociais, afetando principalmente a tomada de decisão de organizações, a preocupação com a qualidade é uma questão bastante discutida atualmente, especialmente relacionada com a produção e com o consumo de dados em grande quantidade.

A definição adotada com maior frequência na literatura atual de qualidade de dados é associada ao conceito de *fitness for use*, em que a qualidade é entendida a partir de uma abordagem contextual, definindo a qualidade de dados como sinónimo de dados considerados adequados para uso pelos consumidores (Juran, 1988; Zaveri *et al.*, 2012; Langer *et al.*, 2018). Essa definição, centrada no consumidor, também é abordada por Redman (2013), que afirma que uma coleção de dados é de alta qualidade desde que atenda as necessidades do seu consumidor.

Embora os aspectos contextuais tenham um forte impacto, a qualidade de dados pode ser entendida, estudada e avaliada a partir de outras perspectivas, baseadas nas categorias de qualidade estabelecidas por Wang e Strong (1996).

A estrutura da qualidade de dados enquanto uma área de investigação é fortemente influenciada pelas pesquisas de Wang e Strong (1996). Os autores buscaram discutir e definir dimensões, critérios e métricas para o processo de avaliação de qualidade de dados, organizando-os em diferentes categorias que refletem as distintas perspectivas da qualidade dos dados, levando em consideração não apenas a sua acurácia, mas também as necessidades dos usuários. Suas contribuições servem como

base para diversos estudos mais recentes, fato evidenciado pelo grande número de citações de seus trabalhos (Sadiq, 2013, p. 7).

Sendo assim, Wang e Strong (1996) definiram quatro categorias de qualidade de dados segundo aspectos apontados pelos consumidores: intrínseca, contextual, representacional e acessibilidade.

A qualidade representacional, objeto de estudo desta pesquisa, diz respeito ao formato e significado dos dados:

Esses dois aspectos sugerem que, para que os consumidores de dados concluam que os dados estão bem representados, eles devem ser não apenas concisos e consistentemente representados, mas também interpretáveis e fáceis de entender (Wang; Strong, 1996, p. 21).

Embora a qualidade representacional tenha se consolidado, a partir da pesquisa de Wang e Strong (1996), como uma das quatro categorias de qualidade de dados e seja frequentemente mencionada em estudos na área da qualidade de dados, as pesquisas geralmente se concentram nos aspectos intrínsecos e contextuais da qualidade de dados, sendo as dimensões representacional e acessibilidade significativamente menos exploradas pela literatura. Nesse sentido, a elaboração do presente estudo partiu do questionamento: Como é discutida na literatura a dimensão representacional da qualidade de dados?

O objetivo desta pesquisa é verificar, por meio de uma análise exploratória da literatura, com adoção de protocolo de pesquisa, como tem sido discutida a categoria representacional, buscando identificar os temas mais abordados, bem como temas ainda para serem explorados. A próxima seção apresenta os procedimentos metodológicos adotados para a condução da pesquisa.

2 PROCEDIMENTOS METODOLÓGICOS

Considerando seus objetivos e procedimentos metodológicos, esta pesquisa se caracteriza por sua natureza exploratória, tendo como finalidade reunir informações preliminares e aprofundar a compreensão teórica, estabelecendo assim uma base para investigações futuras (Prodanov; Freitas, 2013). Quanto a seus resultados, a pesquisa se caracteriza por seu caráter qualitativo e descritivo.

Para alcançar o objetivo proposto, foi realizada uma análise exploratória da literatura científica, baseada na adoção de um protocolo de pesquisa. Esse protocolo foi elaborado visando orientar a busca, seleção e análise de um *corpus* teórico que permita uma maior compreensão a respeito de como o assunto tem sido elaborado. A análise exploratória é considerada como uma etapa inicial e indispensável para outros estudos mais aprofundados da literatura, podendo seu protocolo ser considerado como protocolo piloto para Revisões Sistemáticas da Literatura e Mapeamentos Sistemáticos da Literatura. O Quadro 1 apresenta o protocolo elaborado para a pesquisa:

Quadro 1 - Protocolo de pesquisa

Protocolo de busca	
Pergunta de pesquisa (principal)	Como é discutida na literatura a dimensão representacional da qualidade de dados?
Objetivos	Identificar o estado da arte das discussões a respeito da dimensão representacional em pesquisas de qualidade de dados
Bases de dados consultada	Web of Science, Taylor & Francis, LISA, LISTA
Estratégia de busca	“Data quality” AND “Representational”
Período abrangido	Sem recorte temporal
Idioma	Inglês
Tipo documental	Artigos em periódicos
Critérios de inclusão	(I) Discute a dimensão representacional da qualidade de dados (I) Foco na dimensão representacional da qualidade de dados (I) Apresenta uma pesquisa aplicada de qualidade de dados com base uma dimensão representacional
Critérios de exclusão	(E) Não está nos idiomas estabelecidos para a pesquisa (E) Não está no formato estabelecido para a pesquisa (E) Apenas menciona a dimensão representacional (E) Não aplica ou não discute o tema da pesquisa (E) Não é possível ter acesso ao documento completo
Formulário de extração	1) Objetivo principal do estudo; 2) Forma como a temática foi abordada; 3) Definições
Data da coleta	Outubro de 2024

Fonte: Elaboração própria.

Com a finalidade de ampliar o *corpus* teórico construído foi adotada ainda a técnica do *snowballing*, que permitiu a identificação e inclusão de referências dos artigos pertinentes ao tema. Apresentados os procedimentos metodológicos adotados, a próxima seção apresenta os resultados obtidos.

3 RESULTADOS E DISCUSSÃO

Com base no protocolo apresentado, 320 artigos foram recuperados. Destes, 294 foram recusados, 16 foram aceitos e 10 eram artigos duplicados. Com o *snowballing*, mais 9 artigos foram incluídos, totalizando 25 artigos para análise (Gráfico 1).



Fonte: Elaboração própria.

Os artigos, então, foram lidos e organizados pelo ano de publicação. O Quadro 2 apresenta o objetivo dos estudos selecionados, seus autores e sua fonte, destacando se foram recuperados pela estratégia de busca ou pelo *snowballing*. Os destaques mostram os trabalhos focados apenas na dimensão representacional.

Quadro 2 - Objetivos dos estudos selecionados

Autores	Objetivo	Fonte
Wang e Strong (1996)	Apresentar um framework com aspectos da qualidade de dados considerados importantes pelos consumidores de dados	Estratégia de busca
Strong, Lee e Wang (1997)	Apresentar uma análise de projetos de qualidade de dados e identificação dos padrões mais comuns de problemas de qualidade	<i>Snowballing</i>
Klein (2001)	Comparar as percepções do usuário em relação à qualidade de dados encontrada na internet e em fontes tradicionais de texto	Estratégia de busca
Calero, Caro e Piattini (2008)	Apresentar uma ferramenta de avaliação de qualidade de dados para portais da web	<i>Snowballing</i>

Caro <i>et al.</i> (2008)	Identificar atributos relevantes para avaliação da qualidade de dados de portais da web	<i>Snowballing</i>
Blake e Mangiameli (2009)	Apresentar métricas para consistência representacional de dados estruturados e não estruturados	<i>Snowballing</i>
Moraga <i>et al.</i> (2009)	Apresentar o desenvolvimento de um modelo de qualidade para portais da web	<i>Snowballing</i>
Salomone, Hyland e Murphy (2011)	Analisar atributos da qualidade de dados em relação a diferentes funções de pessoas que trabalham com dados	<i>Snowballing</i>
Zaveri <i>et al.</i> (2015)	Realizar uma revisão sistemática sobre métodos de avaliação de qualidade de linked data	<i>Snowballing</i>
Heo (2013)	Testar um modelo de avaliação de portais de patrimônio cultural e verificar sua relação com a satisfação do usuário	Estratégia de busca
Leite <i>et al.</i> (2014)	Propor um modelo de avaliação de qualidade de dados em sites de unidades de saúde	Estratégia de busca
Yi, Hua e Sadiq (2016)	Propor um framework para lidar com a inconsistência representacional de dados	Estratégia de busca
Liu <i>et al.</i> (2017)	Investigar se o gênero influencia na avaliação da qualidade de informação em comunidades virtuais com uma perspectiva de expectativa-percepção	Estratégia de busca
Juddoo <i>et al.</i> (2018)	Discutir as dimensões de qualidade de dados no contexto de big data para a indústria da saúde	<i>Snowballing</i>
Langer <i>et al.</i> (2018)	Apresentar uma ferramenta de avaliação de qualidade, o SemQuire	Estratégia de busca
Shang, Weng e Hripcsak (2018)	Apresentar um framework para ajudar na escolha de bases de dados clínicas adequadas para conduzir estudos observacionais	Estratégia de busca
Wahyudi, Farhani e Janssen (2018)	Comparar características de big data com as dimensões da qualidade de dados de organizações de serviços financeiros	<i>Snowballing</i>
Ghasemaghaei e Calic (2019)	Explorar o impacto do uso de big data na qualidade de dados para analisar a qualidade da tomada de decisão de empresas	Estratégia de busca
Kwiatkowska e Pouw (2019)	Propor um framework a partir de uma perspectiva semiótica para qualidade de dados	Estratégia de busca
Mulgund <i>et al.</i> (2020)	Realizar uma revisão sistemática para identificar problemas de qualidade de dados em sites de avaliação de médicos	Estratégia de busca
Jabar <i>et al.</i> (2022)	Investigar e explorar as dimensões de qualidade de dados aplicadas à análise de big data no setor público da Malásia	Estratégia de busca

Zhang e Trace (2022)	Identificar quais dimensões da qualidade de dados os consumidores consideram importantes para avaliar a qualidade de informação dos seus próprios dados de saúde	Estratégia de busca
Matos, Jorge e Sá (2023)	Apresentar uma ferramenta para medir a qualidade das informações de planejamento financeiro estratégico de municípios de Portugal	Estratégia de busca
Priestley, O'Donnell e Simperl (2023)	Apresentar requisitos de qualidade de dados importantes para sistemas de machine learning em cada estágio de desenvolvimento	Estratégia de busca
Timmerman <i>et al.</i> (2023)	Apresentar uma abordagem baseada em custos para analisar e medir a qualidade de dados a partir da consistência representacional	Estratégia de busca

Fonte: Elaboração própria.

A maioria dos artigos recuperados é da área da Ciência da Computação, mas a preocupação com a qualidade de dados não se limita em apenas uma área. Dentre os temas encontrados relacionados à qualidade de dados, destaca-se: a qualidade de dados no contexto do *big data* e do *linked data*; a qualidade de dados médicos e da saúde; a qualidade de dados em sistemas de informação; a qualidade de dados nos setores públicos, financeiros e industriais; e a qualidade de dados da web.

Analisando os objetivos dos estudos, é possível notar que a maioria aborda a qualidade representacional juntamente com as outras categorias, visando discutir diferentes aspectos da qualidade de dados. Mesmo em diversos contextos, grande parte dos trabalhos analisados diretamente baseia suas definições de qualidade no artigo de Wang e Strong (1996), confirmando a influência que esses autores ainda exercem sobre o assunto. De acordo com as dimensões da qualidade representacional (interpretabilidade, facilidade de compreensão, consistência representacional e representação concisa), os autores afirmam que “problemas relacionados ao significado e ao formato surgem na pesquisa de sistemas de banco de dados, na qual o formato é abordado como parte da sintaxe, e o significado, como parte da reconciliação semântica” (Wang; Strong, 1996, p. 21).

Dentre todos os documentos que compõem o *corpus* da pesquisa, apenas três têm como foco principal a qualidade representacional de dados (títulos em destaque no Quadro 2). Ainda assim, estes trabalhos analisam somente a dimensão da consistência representacional.

Yi, Hua e Sadiq (2016) analisaram a inconsistência representacional, apontando o desafio de manter a consistência dos dados devido aos seus diferentes formatos e estruturas, além de serem gerados e representados de forma diferente em diversos países, culturas e empresas.

Quando estes dados distintos são integrados, a inconsistência dos dados torna-se evidente. Por esse motivo, este problema está atraindo cada vez mais a atenção da indústria e da academia, com os dados se tornando cada vez mais massivos e heterogêneos. As pesquisas atuais sobre inconsistência de representação de dados geralmente se concentram em um domínio específico, como nome, endereço etc. (Yi; Hua; Sadiq, 2016, p. 395).

Para solucionar esse problema, as autoras sugerem um framework de três módulos (design, detecção e uniformização de padrões) que gera possíveis padrões a partir de dados apresentados em diversos domínios.

Timmerman *et al.* (2023) aplicaram um experimento com 218 voluntários visando investigar o impacto que a representação consistente e a completude dos dados têm na qualidade. Os resultados mostraram um aumento na usabilidade dos dados ao melhorar sua qualidade por meio da correção de problemas de completude e consistência representacional.

A análise baseada em custos foi implementada na prática por meio de um experimento para a análise de um caso de uso que consistiu em três conjuntos de dados com problemas de integridade e consistência de representação. O experimento confirmou (e quantificou) o aumento esperado da utilidade dos dados devido às melhorias na qualidade dos dados. (Timmerman *et al.*, 2023, p. 10).

De um modo geral, a dimensão representacional é apresentada como um fator essencial para os consumidores de dados: a interpretabilidade, a facilidade de compreensão e a representação concisa e consistente dos dados podem influenciar diretamente nas tomadas de decisão. Uma má qualidade representacional de dados pode resultar em falhas na comunicação, decisões prejudiciais e impactos negativos nos processos de trabalho. Para evitar isso, os estudos propõem definições de padrões adequados ao seu contexto e a documentação das padronizações.

4 CONSIDERAÇÕES FINAIS

Por meio da condução de uma análise exploratória da literatura e da adoção da técnica de *snowballing*, esta pesquisa teve como objetivo identificar de que modo a literatura discute a qualidade representacional de dados. Outro objetivo deste estudo foi verificar a relevância da qualidade representacional, visando compreender seu impacto nos processos de avaliação de qualidade de dados.

Dessa forma, a qualidade representacional de dados mostrou-se como um aspecto importante na percepção dos consumidores de dados, exigindo que estes sejam fáceis de entender e de interpretar, além de serem representados de maneira concisa e consistente. Problemas na qualidade representacional podem trazer prejuízos aos envolvidos, tendo em vista que uma representação falha impede a utilização eficiente dos dados.

A diversidade de contextos dos estudos recuperados pela estratégia de busca demonstra o interesse amplo pela qualidade de dados, principalmente no contexto do big data e dados no contexto da web, no qual o grande volume e a diversidade de fontes dos dados apresentam-se como principais desafios a serem resolvidos na questão da qualidade representacional.

Apesar de sua relevância e impacto na percepção da qualidade em conjuntos de dados, conclui-se que a categoria representacional ainda é pouco explorada pelos pesquisadores, especialmente pesquisas focadas nesta categoria que busquem apresentar definições claras a respeito da qualidade representacional, identificar dimensões e métricas para a sua avaliação, elaborar e testar ferramentas que auxiliem na avaliação deste aspecto de qualidade.

Nesse sentido, a próxima etapa será realizar ajustes no protocolo de pesquisa, visando aplicar uma Revisão Sistemática da Literatura. Ressalta-se a necessidade de um aprofundamento em relação à definição do termo qualidade representacional, visando verificar se esse termo tem sido adotado com significado consistente pelos autores, buscando ainda fornecer maior clareza terminológica a respeito da aplicação do termo. Como estudos futuros, pretende-se explorar mais detalhadamente as dimensões, critérios e métricas relacionados com a qualidade representacional.

Outro aspecto que demanda maior aprofundamento é a potencial relação entre os aspectos representacionais da qualidade de dados e a Ciência da Informação (CI). Ressalta-se que em ambas as áreas o conceito de representação pode assumir diferentes acepções, o que pode ter implicações sobre como essa temática pode ser entendida e discutida no âmbito da CI.

Os resultados desta análise exploratória apontam que a categoria representacional é essencial para uma avaliação completa de qualidade de dados, apesar de ser pouco explorada pela ciência. Ao focar na qualidade representacional, esta pesquisa recupera um assunto ainda a ser desenvolvido em profundidade, destacando suas possíveis definições, dimensões, critérios e soluções para a qualidade representacional dos dados.

REFERÊNCIAS

- BLAKE, Roger H.; MANGIAMELI, Paul. Evaluating the semantic and representational consistency of interconnected structured and unstructured data. **AMCIS 2009 Proceedings**, n. 126, 2009. Disponível em: <https://aisel.aisnet.org/amcis2009/126>. Acesso em: 12 maio 2025.
- CALERO, Coral; CARO, Angélica; PIATTINI, Mario. An applicable data quality model for web portal data consumers. **World Wide Web**, v. 11, p. 465-484, 2008. Disponível em: <https://link.springer.com/article/10.1007/s11280-008-0048-y>. Acesso em: 12 maio 2025.
- CARO, Angélica *et al.* A proposal for a set of attributes relevant for web portal data quality. **Software Quality Journal**, v. 16, p. 513-542, 2008. Disponível em: <https://link.springer.com/article/10.1007/s11219-008-9046-7>. Acesso em: 12 maio 2025.
- GHASEMAGHAEI, Maryam; CALIC, Goran. Can big data improve firm decision quality? The role of data quality and data diagnosticity. **Decision Support Systems**, v. 120, p. 38-49, 2019. Disponível em: <https://dl.acm.org/doi/abs/10.1016/j.dss.2019.03.008>. Acesso em: 12 maio 2025.
- HEO, Misook. User satisfaction with portals: testing for factorial validity and invariance across age groups. **Online Information Review**, v. 37, n. 5, p. 804-820, 2013. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/oir-06-2012-0099/full/html>. Acesso em: 12 maio 2025.

JABAR, Zam Zarina Abdul *et al.* Establishment of big data analytics application model for Malaysian public sector: an expert validation. **EDPACS**, v. 65, n. 4, p. 1-20, 2022.

Disponível em:

<https://www.tandfonline.com/doi/full/10.1080/07366981.2021.1958736>. Acesso em: 12 maio 2025.

JUDDOO, Suraj *et al.* Data governance in the health industry: investigating data quality dimensions within a big data context. **Applied System Innovation**, v. 1, n. 4, p. 43, 2018.

Disponível em: <https://www.mdpi.com/2571-5577/1/4/43>. Acesso em: 12 maio 2025.

JURAN, Joseph Moses. **Quality control handbook**. New York: Mcgraw-Hill, 1988.

KLEIN, Barbara D. User perceptions of data quality: Internet and traditional text sources. **Journal of computer information systems**, v. 41, n. 4, p. 9-15, 2001. Disponível em:

<https://www.tandfonline.com/doi/abs/10.1080/08874417.2001.11647016>. Acesso em: 12 maio 2025.

KWIATKOWSKA, Mila; POUW, Frank. Data quality in secondary data analysis: a case study of ecological data using a semiotic-based approach. *In: INTERNATIONAL CONFERENCE ON DATA SCIENCE, TECHNOLOGY AND APPLICATIONS*, 8., 2019. Prague.

Proceedings [...]. Prague: [S. l.], 2019. v. 1, p. 377-384. Disponível em:

<https://www.scitepress.org/Link.aspx?doi=10.5220/0007978403770384>. Acesso em: 12 maio 2025.

LANGER, André *et al.* SemQuire: assessing the data quality of linked open data sources based on DQV. **Current Trends in Web Engineering**, v. 11153, p. 163-175, 2018.

Disponível em: https://link.springer.com/chapter/10.1007/978-3-030-03056-8_14.

Acesso em: 12 maio 2025.

LEITE, Patrícia *et al.* Towards a model for the measurement of data quality in websites.

New Review of Hypermedia and Multimedia, v. 20, n. 4, p. 301-316, 2014. Disponível

em: <https://www.tandfonline.com/doi/full/10.1080/13614568.2014.968638>. Acesso em: 12 maio 2025.

LIU, Yuewen *et al.* Gender differences in information quality of virtual communities: a study from an expectation-perception perspective. **Personality and individual differences**, v. 104, p. 224-229, 2017. Disponível em:

<https://www.sciencedirect.com/science/article/pii/S0191886916309084>. Acesso em: 12

maio 2025.

MATOS, Sandra; JORGE, Susana; SÁ, Patricia Moura e. Measuring the quality of the strategic financial planning information (Q-FPI) in the local government. **International Journal of Public Administration**, v. 46, n. 10, p. 671-688, 2023. Disponível em:

<https://www.tandfonline.com/doi/full/10.1080/01900692.2021.2011315>. Acesso em:

12 maio 2025.

- MORAGA, Carmen *et al.* SQuaRE-aligned data quality model for web portals. *In: INTERNATIONAL CONFERENCE ON QUALITY SOFTWARE*, 9., 2009. Jeju, Korea (South). **Proceedings [...]**. Jeju, Korea (South): IEEE, 2009. p. 117-122. Disponível em: <https://ieeexplore.ieee.org/document/5381502>. Acesso em: 12 maio 2025.
- MULGUND, Pavankumar *et al.* Data quality issues with physician-rating websites: systematic review. **Journal of Medical Internet Research**, v. 22, n. 9, p. e15916, 2020. Disponível em: <https://www.jmir.org/2020/9/e15916/>. Acesso em: 12 maio 2025.
- PRIESTLEY, Maria; O'DONNELL, Fionntán; SIMPERL, Elena. A survey of data quality requirements that matter in ML development pipelines. **ACM Journal of Data and Information Quality**, v. 15, n. 2, p. 1-39, 2023. Disponível em: <https://dl.acm.org/doi/10.1145/3592616>. Acesso em: 12 maio 2025.
- PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar de. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. 2. ed. Novo Hamburgo: Editora Feevale, 2013.
- REDMAN, Thomas C. Data quality management past, present, and future: towards a management system for data. *In: SADIQ, Shazia (ed.). Handbook of data quality: research and practice*. Berlin: Springer Berlin Heidelberg, 2013. p. 15-40. Disponível em: https://link.springer.com/chapter/10.1007/978-3-642-36257-6_2. Acesso em: 12 maio 2025.
- SADIQ, Shazia. Prologue: research and practice in data quality management. *In: SADIQ, Shazia (ed.). Handbook of data quality: research and practice*. Berlin: Springer Berlin Heidelberg, 2013. p. 1-11. Disponível em: https://link.springer.com/chapter/10.1007/978-3-642-36257-6_1. Acesso em: 12 maio 2025.
- SALOMONE, Sonia; HYLAND, Paul; MURPHY, Glen. Perceptions of data quality dimensions and data roles. *In: AUSTRALIAN AND NEW ZEALAND ACADEMY OF MANAGEMENT CONFERENCE*, 25., 2011, Wellington. **Proceedings [...]**. Wellington: [s.n.], 2011. p. 1-25. Disponível em: <https://eprints.qut.edu.au/48014/>. Acesso em: 12 maio 2025.
- SHANG, Ning; WENG, Chunhua; HRIPCSAK, George. A conceptual framework for evaluating data suitability for observational studies. **Journal of the American Medical Informatics Association**, v. 25, n. 3, p. 248-258, 2018. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/29024976/>. Acesso em: 12 maio 2025.
- STRONG, Diane M.; LEE, Yang W.; WANG, Richard Y. Data quality in context. **Communications of the ACM**, v. 40, n. 5, p. 103-110, 1997. Disponível em: <https://dl.acm.org/doi/10.1145/253769.253804>. Acesso em: 12 maio 2025.
- TIMMERMAN, Yoram *et al.* Cost-based analysis of the impact of data completeness and representational consistency. **Decision Support Systems**, v. 175, p. 114044, 2023.

Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167923623001197>. Acesso em: 12 maio 2025.

WAHYUDI, Agung; FARHANI, Adiska; JANSSEN, Marijn. Relating big data and data quality in financial service organizations. *In: INTERNATIONAL CONFERENCE ON E-BUSINESS, E-SERVICES AND E-SOCIETY*, 17., 2018, Tirana. **Proceedings [...]**. [S. l.]: Springer, 2018. p. 504-519. Disponível em: <https://repository.tudelft.nl/record/uuid:f8bd8560-c503-466e-82d4-069141188061>. Acesso em: 12 maio 2025.

WANG, Richard Y.; STRONG, Diane M. Beyond accuracy: what data quality means to data consumers. **Journal of Management Information Systems**, v. 12, n. 4, p. 5-33, 1996. Disponível em: <http://www.jstor.org/stable/40398176>. Acesso em: 12 maio 2025.

YI, Bingyu; HUA, Wen; SADIQ, Shazia. A pattern-based framework for addressing data representational inconsistency. **Databases Theory and Applications**, v. 9877, p. 395-406, 2016. Disponível em: https://link.springer.com/chapter/10.1007/978-3-319-46922-5_31. Acesso em: 12 maio 2025.

ZAVERI, Amrapali *et al.* Quality assessment for linked open data: a survey. **Semantic Web Journal**, v. 7, n. 1, p. 63-93, 2015. Disponível em: <https://journals.sagepub.com/doi/full/10.3233/SW-150175>. Acesso em: 12 maio 2025.

ZHANG, Yan; TRACE, Ciaran B. The quality of health and wellness self-tracking data: a consumer perspective. **Journal of the Association for Information Science and Technology**, v. 73, n. 6, p. 879-891, 2022. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/asi.24591>. Acesso em: 12 maio 2025.

AGRADECIMENTOS

Agradecemos ao PUB-USP e à FAPESP (Processo 2021/03349-0) pelo financiamento desta pesquisa.



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

Como citar este trabalho:

YOSHITANE, Débora Kaori; JESUS, Ananda Fernanda de; SANTAREM SEGUNDO, José Eduardo. Qualidade de dados: análise focada na categoria representacional. *In: WORKSHOP DE INFORMAÇÃO DADOS E TECNOLOGIA*, 8., 2025, Marília, SP. **Anais [...]**. Marília, SP: Universidade de Marília, 2025. DOI: <http://doi.org/10.22477/viii.widat.221>.