

XXV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - XXV ENANCIB

GT 8 – Dados, Informação e Tecnologia

A QUALIDADE DE DADOS ENQUANTO OBJETO DE ESTUDO DA CIÊNCIA DA INFORMAÇÃO BRASILEIRA: ANÁLISE DE TERMOS CORRELATOS

DATA QUALITY AS AN OBJECT OF STUDY IN BRAZILIAN INFORMATION: ANALYSIS OF RELATED TERMS

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

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

Modalidade: Trabalho Completo

Resumo: a preocupação com a qualidade de conjuntos de dados perpassa diversos aspectos da Ciência da Informação, entretanto, pesquisas focadas em qualidade de dados na área ainda são recentes, especialmente no Brasil. Nesse sentido, partiu-se do questionamento: Como a qualidade de dados tem sido abordada enquanto um objeto de estudo no contexto da Ciência da Informação brasileira? Objetiva-se, a partir do recorte da literatura nacional, apresentar um panorama das abordagens de qualidade de dados na Ciência da Informação. Foi realizada uma análise exploratória da literatura, baseada em protocolo de pesquisa e a análise dos termos correlatos utilizados pelos autores para abordar e discutir a qualidade de dados. Os termos identificados foram agrupados em: 1) tipos de qualidade de dados; 2) sinônimos e termos alternativos para qualidade de dados; 3) termos relacionados a sistematização da qualidade de dados; 4) termos relacionados a procedimentos de qualidade de dados; 5) termos relacionados às áreas e subáreas da Ciência da Informação; 6) instrumentos utilizados no processo de avaliação de qualidade; e 7) termos relacionados aos temas e enfoques dos documentos. Observou-se que a qualidade de dados é discutida especialmente em sua relação com a organização e representação da informação e do conhecimento, com a curadoria digital e com a governança de dados. Destacam-se também os aspectos relacionados com as tecnologias e com o patrimônio cultural. Conclui-se que a qualidade de dados e a Ciência da informação tem um potencial de colaboração mútua que pode ser mais explorado.

Palavras-chave: qualidade de dados; Linked Data; análise terminológica.

Abstract: the search for guaranteeing the quality of data sets covers various aspects of Information Science, however, as approaches to data quality are still recent, especially in Brazil. The research question was: How has the quality of data been addressed as an object of study in the context of Brazilian Information Science? The objective was to identify, based on a clipping of national literature, how the quality of data is addressed as an object of study of Information Science. An exploratory analysis of the literature was carried out, based on the research protocol and the analysis of two correlative terms used by the authors to approach and discuss the quality of data. The terms identified are grouped in: 1) types of data quality 2) synonyms and alternative terms for data quality 3) terms related to systematization of data quality 4) terms related to data quality procedures 5) terms related

to the areas and subareas of Information Science; 6) instruments used in the quality assessment process; e 7) terms related to the topics and approaches of the two documents. Note that data quality is discussed especially in its relationship with the organization and representation of information and knowledge, with digital curation and with data governance. We also highlight aspects related to technologies and cultural heritage. It is concluded that data quality and information science have a potential for mutual collaboration that can be further explored.

Keywords: data quality; Linked Data; terminological analysis.

1 INTRODUÇÃO

As abordagens de qualidade de dados enquanto objeto de estudo da Ciência da Informação tem se ampliado nos últimos anos. Um exemplo da relevância dos dados na Ciência da Informação nacional é a alteração do nome e da ementa de um dos Grupos de Trabalho (GT) da Associação de Pesquisa e Pós-graduação em Ciência da Informação (ANCIB).

O GT8, antes denominado Informação e Tecnologia, passou a ser denominado Dados, Informação e Tecnologia, e teve a sua ementa alterada para a inclusão dos dados: Estudos e pesquisas teórico-práticos sobre e para o desenvolvimento de tecnologias de informação e comunicação que envolvam os processos de coleta, geração, representação, armazenamento, recuperação, disseminação, uso, reúso, gestão, análise, processamento, tratamento, governança, visualização, segurança e preservação **de dados e informação** em ambientes informacionais (ANCIB, 2024, não paginado, grifo nosso).

Jesus, Barité e Santarem Segundo (2024) consideram que o termo qualidade de dados se caracteriza como um conceito poliédrico. Definir um termo como poliédrico implica na compreensão de que, em determinados contextos, a construção dos conceitos especializados se estabelece de forma a que um termo contenha mais de uma acepção (Cabré, 1999). Para os autores a qualidade de dados possui então três acepções principais: como um campo ou um domínio, como um processo e como uma medida ou ponderação. Para essa pesquisa, nos ateremos as definições de qualidade enquanto um domínio e enquanto um processo.

Enquanto campo ou domínios a qualidade de dados se caracteriza como um domínio interdisciplinar desde sua origem, reunindo em seu princípio estatísticos, engenheiros e cientistas da computação, preocupados em discutir formas e elaborar instrumentos que permitam mensurar e melhorar a qualidade dos conjuntos de dados (Batini; Scannapieco, 2016, p. 4, tradução nossa)

Enquanto um processo, a qualidade de dados se reflete no processo de avaliação de qualidade. Esse processo é compreendido por meio da sistematização elaborada por Wang e Strong (1996). A relevância da sistematização do processo de avaliação de qualidade de Wang

e Strong (1996) é reafirmada pela própria literatura da qualidade de dados (Batini; Scannapieco, 2016; Sadiq, 2013). Wang e Strong (1996) apresentam um *framework* de qualidade de dados onde organizam o processo em quatro categorias de qualidade: Intrínseca, contextual, representacional e acessibilidade. Essas categorias agrupam dimensões que são definidas como o aspecto mais abrangente de características dos dados, que organizam problemas semelhantes de qualidade que serão observados visando mensurar a sua qualidade. As dimensões são compostas por critérios que descrevem o atributo específico de qualidade que será avaliado (Wang; Strong, 1996).

Cada critério pode ser mensurado a partir de mais de uma métrica, indicadores que permitem avaliar quantitativamente ou qualitativamente os níveis de qualidade dos dados. As métricas são elaboradas com base nas perspectivas adotadas, nas características das dimensões e nos critérios a serem avaliados, podem possuir um caráter objetivo ou subjetivo (Assaf; Senart; Troncy, 2016; Melo, 2017).

Considerando a importância dos dados no cenário tecnológico atual e a potencial contribuição da Ciência da Informação para a avaliação, melhoria e estudo dos problemas derivados de má qualidade de dados, partiu-se do questionamento: como a qualidade de dados tem sido abordada enquanto um objeto de estudo no contexto da Ciência da Informação brasileira?

Objetiva-se, a partir do recorte da literatura nacional, apresentar um panorama das abordagens de qualidade de dados na Ciência da Informação, para isso propõe-se a realização de uma análise exploratória, baseada em protocolo de pesquisa, e da análise dos termos correlatos utilizados pelos autores para abordar e discutir a qualidade de dados. A próxima seção apresenta os procedimentos metodológicos adotados.

2 PROCEDIMENTOS METODOLÓGICOS

Partiu-se da realização de uma análise exploratória, baseada em protocolo de pesquisa, focada nas publicações de qualidade de dados na Ciência da Informação brasileira, buscando-se identificar como a qualidade de dados é abordada enquanto um objeto de estudo da Ciência da Informação. Para orientar a condução da Análise Exploratória foi estabelecido um protocolo de pesquisa, apresentado no quadro 1

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

Quadro 1 - Protocolo de pesquisa da análise exploratória

Protocolo de pesquisa	
Pergunta de pesquisa principal	Como a qualidade de dados tem sido abordada enquanto um objeto de estudo contexto da Ciência da Informação?
Objetivo	Objetiva-se identificar, a partir do recorte da literatura nacional, como a qualidade de dados é abordada enquanto um objeto de estudo da 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
Critérios 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
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

Fonte: elaborado pelos autores (2025)

Foram observadas, ainda, as referências adotadas pelos autores através da técnica de *Snowballing*, que “consiste em avaliar a lista de referências e a lista de citações de estudos já conhecidos para identificar novos estudos relevantes para a pesquisa” (Silva *et al.*, 2017, p. 29).

Visando analisar como a qualidade de dados se relaciona com outros termos no âmbito da CI, foram coletados os termos correlatos utilizados pelos autores para abordar a qualidade de dados, os termos foram coletados de forma manual e sistematizados em categorias estabelecidas *a posteriori*.

Para identificação de termos relevantes, foram considerados: 1) termos utilizados nas definições de qualidade de dados apresentadas pelos autores; 2) termos utilizados para explicar a relação entre qualidade de dados e ciência da informação; 3) termos que descrevem os objetos de estudo dos processos de avaliação conduzidos ou das ferramentas de avaliação propostas pelos autores.

Os termos passaram por um processo de refinamento e categorização com base na identificação de padrões de aplicação, sendo estabelecidas 7 categorias de termos: 1) tipos de qualidade de dados; 2) sinônimos e termos alternativos para qualidade de dados; 3) termos relacionados a sistematização da qualidade de dados; 4) termos relacionados a

**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

procedimentos de qualidade de dados; 5) termos relacionados às áreas e subáreas da Ciência da Informação; 6) instrumentos utilizados no processo de avaliação de qualidade; e 7) termos relacionados aos temas e enfoques dos documentos. A próxima seção apresenta os resultados obtidos.

3 RESULTADOS

A presente seção apresenta os resultados da análise dos termos correlatos identificados nos estudos aceitos. O quadro 2 apresenta os estudos que compõe o corpus da pesquisa e os termos relacionados a qualidade de dados identificados nesses textos.

Quadro 2 – Termos correlatos identificados nos estudos aceitos

AUTORES	TERMOS CORRELATOS
Bentancourt e Rocha (2012)	qualidade de projeto; metadados; qualidade dos metadados; recuperação da informação; tratamento da informação; base de dados; política de metadados.
Almeida et al. (2016)	anomalias de dados; <i>datacleaning</i> ; limpeza de dados; dados de acidentes aeronáuticos; melhoria na qualidade dos dados; padrão na entrada de dados; validação na entrada de dados.
Fagundes; Macedo; Freund (2017)	governança de dados; dados de qualidade; modelo de ciclo de vida dos dados; "boas práticas de coleta, armazenamento e recuperação"; dados; diretrizes; princípios e boas práticas; preservação digital; dimensões de qualidade; <i>big data</i> .
Melo (2017)	gestão de dados; qualidade aplicada a dados; web semântica; <i>linked data</i> ; <i>linked open data</i> ; dados de qualidade; metodologia de avaliação de qualidade; modelo de qualidade de dados; dimensões de qualidade de dados; problemas de qualidade;
Melo; Botega; Santarem Segundo (2017)	gestão de dados; qualidade aplicada a dados; web semântica; problemas de qualidade; avaliação de qualidade de dados; <i>linked data</i> ; dados de qualidade; padrões; tecnologias; metodologia para avaliação de qualidade; dados conectados; modelo de qualidade; gestão de qualidade de dados; metodologia de avaliação de qualidade de dados.
Melo; Botega; Santarem Segundo (2017)	curadoria de dados; padrões; tecnologias. gestão de qualidade de dados; metodologia de avaliação de qualidade de dados; web semântica; <i>linked data</i> ; <i>linked open data</i> ; qualidade aplicada a dados; modelo de avaliação de qualidade; métricas de qualidade; problemas de qualidade; dados conectados.
Silva et al. (2017)	ontologia; ontologia de domínio; representação de informações; metodologia para avaliação da qualidade de dados; consciência da situação.
Espíndola et al. (2018)	controle da qualidade dos dados; qualidade dos dados; qualidade; governança de dados; dados de qualidade; modelo de ciclo de vida dos dados; "boas práticas de coleta, armazenamento e recuperação"; dados; diretrizes; princípios e boas práticas; preservação digital; processo de qualidade dos dados.
Moura Junior e Aragão (2018)	análise de qualidade de dados; qualidade dos dados; governança corporativa, governança de ti; indicadores de desempenho institucionais; plano de desenvolvimento institucional; qualidade dos dados; meta; indicador.
Piccolo (2018)	garantia de qualidade; garantia da qualidade dos dados; datasus

**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

Juliani et al. (2019)	qualidade dos dados; gestão da informação; governança de dados; boas práticas; catalogação; metadados; organização e representação da informação e do conhecimento; padrão de dados; linguagens documentárias (taxonomias); vocabulários; sistemas de organização do conhecimento; modelo de requisitos; organização e representação da informação.
Moreira et al. (2020)	recuperação de dados; ciclo de vida dos dados; políticas públicas; dados governamentais; dados qualificados; dados abertos; recuperação de dados.
Martins et al. (2021a)	modelo de qualidade de dados; modelo de requisitos; curadoria digital; padrões de qualidade em dados e metadados; representação da informação; catalogações descritivas; vocabulários controlados; regras de catalogação; política de qualidade de dados; objetos digitais; catalogação descritiva; catalogação de assunto, classificação; indexação; análise documental; metadados; padronização e descrição de recursos informacionais, padrões de catalogação; linguagem documentária; tesouro; modelo de requisitos; requisitos de qualidade de dados; padrão de metadados; modelo de metadados; representação e da gestão da informação, política de qualidade de dados; avaliação de qualidade de dados.
Martins et al. (2021b)	modelo de qualidade de dados; curadoria digital; política de qualidade de dados; avaliação de qualidade de dados; requisitos de qualidade de dados.
Piccolo et al. (2021)	curadoria de dados; gestão de dados de pesquisa; garantia de qualidade; dados de pesquisa; dimensões; métricas; curadoria digital; ciclo de vida dos dados.
Coelho Júnior e Lemos (2022a)	patrimônio cultural; indexação; sistemas de recuperação da informação; descrição e publicação de dados; acervos culturais; qualidade dos dados; metadados; catalogação; museologia; coleções digitais; nível da qualidade; dimensões de qualidade.
Coelho Júnior e Lemos (2022b)	avaliação diagnóstica de qualidade de dados; organização da informação; patrimônio cultural; dados com qualidade; padrões de documentação; metadados
Jesus e Santarem Segundo (2022)	critérios de qualidade; dimensões de qualidade; categorias de qualidade, <i>linked data</i> ; web semântica; questões de qualidade.
Turi e Comarela (2022)	lei geral de proteção de dados pessoais; ciência de dados; metrificação da qualidade de dados; modelo de qualidade de dados; dimensões de qualidade de dados; métricas de qualidade dos dados.
Coelho Junior (2023)	patrimônio cultural; indexação; sistemas de recuperação da informação; descrição e publicação de dados; acervos culturais; qualidade dos dados; metadados; catalogação; museologia; coleções digitais; nível da qualidade; dimensões de qualidade.
Dias et al. (2023)	" <i>garbage in, garbage out</i> " (gigo); ciência de dados; painéis de informação; programa dinheiro direto na escola (pdde); tratamento dados.
Lemos e Coelho Junior (2023)	avaliação diagnóstica de qualidade de dados; organização da informação; patrimônio cultural; dados com qualidade; padrões de documentação; metadados; museus.
Lemos e Coelho Junior (2023)	patrimônio cultural; vocabulário controlado; guia de catalogação; práticas de catalogação; organização da informação; representação da informação; metadados; museus.
Jesus e Santarem Segundo (2023a)	vocabulário; melhores práticas; recomendações; descrição formal; descrição da qualidade; <i>linked data</i> ; padrões de metadados; processo de avaliação de qualidade.
Jesus e Santarem Segundo (2023b)	representação da informação. <i>linked data</i> ; dimensões de qualidade; critérios de qualidade; métricas de qualidade; problemas de qualidade.

Fonte: elaborado pelos autores (2025)

A análise dos termos e de suas categorias permitiu observar que, embora forneçam *insights* sobre as abordagens de qualidade de dados na Ciência da Informação, as quatro primeiras categorias reúnem termos mais gerais e menos relacionados com o entendimento de como a qualidade de dados é abordada pela área.

A categoria 1 reuniu os termos que especificam tipos de qualidade de dados, foi composta por apenas dois termos: qualidade de projeto e qualidade de metadados. Já a categoria 2 reúne os termos adotados pelos autores como sinônimos ou termos equivalentes ao conceito de qualidade de dados, tais como: dados de qualidade; qualidade dos dados; qualidade aplicada a dados; dados qualificados; e dados com qualidade. Esses termos refletem certa pluralidade terminológica entorno do termo estudado.

A categoria 3 reúne os termos relacionados a sistematização da qualidade de dados, tais como: dimensões (dimensões de qualidade; dimensões de qualidade de dado); métricas (métricas de qualidade; métricas de qualidade dos dados); problemas de qualidade; garantia de qualidade (garantia da qualidade dos dados); "garbage in, garbage out" (gigo); descrição da qualidade; nível da qualidade; critérios de qualidade; categorias de qualidade; questões de qualidade; anomalias de dados. Esses termos refletem a estruturação da qualidade de dados enquanto processo, seguindo a proposta de Wang e Strong (1996).

Ainda relacionados ao processo de avaliação de qualidade, foram identificados termos que refletem procedimentos de avaliação e melhoria na qualidade de dados, como: limpeza de dados; melhoria na qualidade dos dados; padrão na entrada de dados; validação na entrada de dados; avaliação de qualidade de dados; processo de qualidade dos dados; análise de qualidade de dados; avaliação diagnóstica de qualidade de dados; metrificação da qualidade de dados; tratamento de dados; processo de avaliação de qualidade.

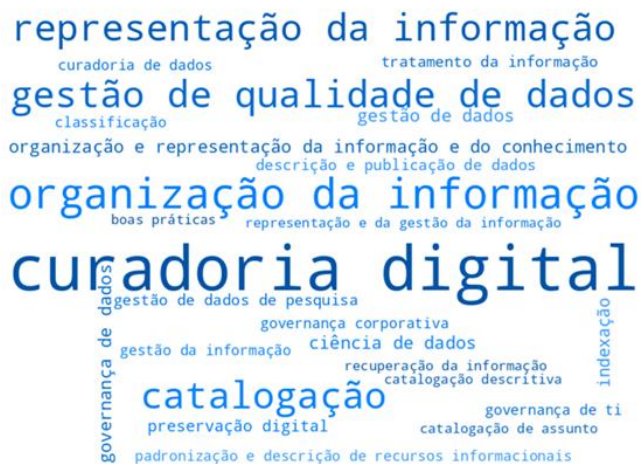
As últimas 3 categorias reúnem os termos relacionados com os enfoques dos estudos, permitindo entender em que contextos é discutida a qualidade de dados na CI; análise dos instrumentos utilizados pela comunidade no processo de avaliação de qualidade e ainda as áreas e subáreas da Ciência da informação no âmbito das quais a qualidade de dados é discutida. As próximas subseções discutem os resultados de cada uma dessas categorias.

3.1 Análise da categoria "5 - termos relacionados às áreas e subáreas da Ciência da Informação"

**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

A categoria 5 reúne os termos relacionados com as áreas e subáreas da CI com as quais a qualidade de dados possui relações interdisciplinares. A figura 2 apresenta a nuvem de palavras dos termos identificados

Figura 2 – Nuvem de palavras da categoria 5



Fonte: elaborado pelos autores (2025)

Como é possível observar na nuvem de palavras, destacam-se nos textos as discussões de qualidade de dados relacionadas à curadoria digital.

A curadoria digital tem ganhado destaque nos últimos anos, construindo-se como uma prática interdisciplinar abrangente que busca estabelecer diretrizes e um conjunto de ações inter-relacionadas para o tratamento e a manutenção do material com valor informacional. Sendo assim, percebe-se que seu objetivo se alinha com a Organização e a Representação da informação demonstrando sua aproximação com a Ciência da Informação (Triques; Arakaki; Castro, 2020, p. 17).

A qualidade de dados, enquanto um processo que busca avaliação e melhoria e conjuntos de dados, se apresenta, portanto, como uma das ações realizadas no âmbito da curadoria digital visando garantir o tratamento e a manutenção dos recursos informacionais digitais. Na análise da figura 2 também é possível observar um destaque de termos relacionados à Organização e Representação da Informação e do Conhecimento. Outros termos com menor destaque reforçam essa relação, tais como catalogação, classificação, tratamento da informação, catalogação descritiva e catalogação de assunto.

Com base nessa análise é possível observar uma relação de colaboração mútua e interdisciplinar entre as áreas, onde a qualidade de dados contribui para a avaliação e melhoria da qualidade dos produtos da Organização e Representação da Informação e do

Conhecimento, como dos dados e metadados presentes nos catálogos, dos vocabulários controlados, da classificação e indexação dos recursos informacionais.

Por outro lado, a qualidade de dados enquanto processo se beneficia de instrumentos da Organização e Representação da Informação e do Conhecimento, seja utilizando-se dos metadados e das descrições para a melhoria do processo de avaliação de qualidade, seja tomando como base para o processo instrumentos de padronização, tais como vocabulários controlados, políticas e melhores práticas.

Destaca-se, ainda que em menor grau, a relação com os aspectos da gestão e da governança de dados.

Segundo Santos e Streit (2018) a governança de dados envolve o processo de instrumentos como ciclo de vida dos dados e o estabelecimento de princípios e melhores práticas voltadas para garantir a estrutura e arquitetura dos dados, a sua segurança e a qualidade.

A governança de dados envolve uma série de decisões visando garantir o controle do acesso, a recuperação, a segurança e a qualidade dos conjuntos de dados ao longo de todo o seu ciclo de vida, envolvendo ainda a explicitação dessas decisões em políticas, ciclos de vida e no estabelecimento da arquitetura dos dados. A figura 3 representa os domínios de decisão da governança de dados:

Figura 3 – Domínios relacionados à governança de dados.



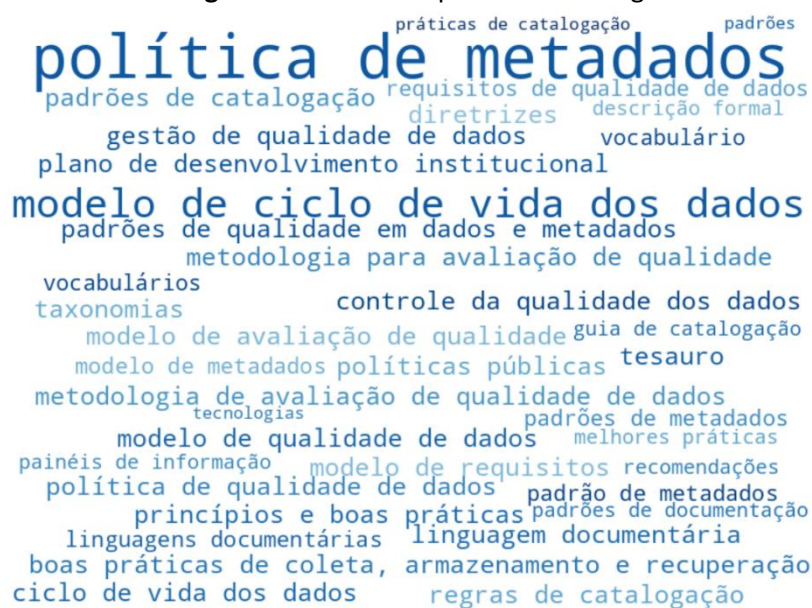
Fonte: Santos e Streit (2018, p. 66)

A relação entre a qualidade de dados a Organização e Representação da Informação e do Conhecimento e a governança de dados também se destaca na análise dos termos agrupados nas categorias 6 e 7.

3.2 Análise da categoria “6 - instrumentos utilizados no processo de avaliação de qualidade”.

A categoria 6 reuniu os termos relacionados aos instrumentos empregados no processo de avaliação de qualidade. Esses instrumentos destacam-se por serem distintos dos encontrados na literatura clássica de qualidade de dados, que se concentra em estabelecer dimensões, apresentar métricas formais e estabelecer instrumentos automáticos/semiautomáticos para auxiliar no processo de avaliação de qualidade. Na CI, a qualidade de dados se destaca pela aplicação de instrumentos de padronização, como boas práticas e diretrizes. A figura 4 apresenta a nuvem de palavras da categoria 6

Figura 4 – Nuvem de palavras da categoria 6



Fonte: elaborado pelos autores (2025)

Analisando os termos da categoria 6 observa-se existe que uma pluralidade de instrumentos, e em especial, de termos utilizados para nomear esses instrumentos. Observa-se a presença de diferentes tipos de instrumentos relacionados à gestão e governança de dados, sendo abordadas políticas, diretrizes, melhores práticas e instrumentos voltados para a governança de dados.

Ressalta-se ainda a questão do ciclo de vida dos dados, que possui um papel importante na governança de dados. A inclusão da qualidade dos dados pode ocorrer em diferentes etapas do ciclo de vida, seja na avaliação inicial de novos dados gerados, na curadoria e preservação dos dados, ou ainda na seleção de novos dados para enriquecimento dos conjuntos de dados. Essa inclusão é importante para garantir o potencial de uso e reúso dos dados e ainda a qualidade dos produtos derivados da utilização desses dados.

Destacam-se ainda entre os instrumentos aqueles relacionados à Organização e Representação da Informação e do Conhecimento, tais como políticas de metadados, padrões de metadados, vocabulários e tesouros.

3.3 Análise da categoria “7 - termos relacionados aos temas e enfoques dos documentos.”

Compreendendo que a qualidade de dados na maioria das vezes é considerada em seu aspecto contextual, no qual a qualidade dos dados depende de seu contexto de aplicação, foram agrupados na categoria 7 os termos relacionados aos temas por meio dos quais se estuda a qualidade de dados na CI. Essa foi a categoria com maior número de termos, considerando termos que se repetem em mais de um artigo. A figura 5 representa a nuvem de palavras da categoria 7.

Figura 5 – Nuvem de palavras da categoria 7



Fonte: elaborado pelos autores (2025)

Como é possível observar existe um grande destaque de metadados, sendo os metadados muitas vezes o principal objeto do processo de avaliação de qualidade na CI, produzidos no âmbito da Organização e Representação da Informação e do Conhecimento, e fundamentais para a recuperação da informação.

Os metadados muitas vezes são abordados de maneira conjunta aos outros termos de destaque, resultando em análise dos metadados relacionados ao patrimônio cultural, metadados publicados como *linked data* e etc.

Destaca-se ainda no quadro a preocupação com a recuperação da informação, objeto de estudo clássico da Ciência da Informação, sendo a qualidade um meio para a melhoria

dessa recuperação, e que muitas vezes também perpassa a questão da qualidade dos metadados. Observa-se na figura o destaque de patrimônio cultural, de museus, acervos e coleções culturais. Também é possível observar a preocupação com qualidade dos dados abertos, dados de pesquisa, dados governamentais, dados no contexto da big data e do *linked data*.

4 CONSIDERAÇÕES FINAIS

A presente pesquisa partiu do questionamento: Como a qualidade de dados tem sido abordada enquanto um objeto de estudo no contexto da Ciência da Informação brasileira? Tendo como objetivo apresentar um panorama das abordagens de qualidade de dados na Ciência da Informação.

Entende-se que esse objetivo foi atingido por meio da análise dos estudos nacionais que abordam a temática e pela sistematização das abordagens adotadas para a discussão do assunto na área, baseada na análise dos termos correlatos identificados.

Em resposta à pergunta de pesquisa, observou-se que enquanto objeto de estudo da CI brasileira, a qualidade de dados é frequentemente abordada em sua relação com a Organização e Representação da Informação e do Conhecimento. A Ciência da Informação possui desde sua estruturação a expertise no desenvolvimento de instrumentos que visam padronizar e estruturar conjuntos de dados, tais como vocabulários, códigos de catalogação e padrões de metadados. Além disso, dados são o produto de processos importantes da área, especialmente no âmbito da Organização e Representação da Informação e do Conhecimento.

Nesse sentido, foi observada a existência de enfoque nos instrumentos e produtos derivados dessa área, tais como metadados, padrões de metadados, vocabulários controlados e padrões de catalogação. Outro destaque é a abordagem da qualidade de dados relacionada a distintos aspectos da curadoria e da preservação digital. Nessas abordagens, processos e instrumentos da qualidade de dados são inseridos como ações realizadas no âmbito da curadoria digital visando garantir o tratamento e a manutenção dos recursos informacionais digitais e ainda a preservação desses recursos a longo prazo.

Da análise conjunta das categorias “5 - termos relacionados às áreas e subáreas da Ciência da Informação”; “6 - instrumentos utilizados no processo de avaliação de qualidade”; e “7 - termos relacionados aos temas e enfoques dos documentos”, observa-se um destaque também para as abordagens de qualidade de dados no âmbito da governança de dados, onde

se discute a sua inclusão como etapa importante do ciclo de vida de dados. Nesse aspecto, destacam-se os instrumentos criados pela CI para auxiliar na qualidade de dados, por meio da criação, adoção e verificação da adequação à melhores práticas, princípios e políticas voltadas para a qualidade dos dados.

Em outro aspecto, a análise dos temas de pesquisa permite observar que se consolidam como objetos da área fenômenos derivados de mudanças tecnológicas, e amplamente relacionados a produção massiva de dados, tais como big data, *linked data*, inteligência artificial, *machine learning*, internet das coisas, entre outros temas. Destacam-se ainda como enfoques os temas relacionados ao patrimônio cultural, como acervos culturais, museus e coleções digitais.

Com base nas análises apresentadas, conclui-se que a qualidade de dados e a Ciência da Informação possuem uma relação interdisciplinar, com potencial de mútua colaboração entre as áreas. A Ciência da informação contribui para a melhoria da qualidade dos dados por meio da elaboração e aplicação de instrumentos, tais como diretrizes, códigos e melhores práticas, enquanto se utiliza de técnicas, processos e instrumentos da qualidade de dados para a melhoria de seus produtos.

Como estudos futuros, pretende-se realizar um aprofundamento dessa relação interdisciplinar entre as áreas, focando em identificar as possibilidades de colaboração mútua entre ambas. Pretende-se ainda ampliar o escopo da presente pesquisa, por meio da análise da literatura internacional.

REFERÊNCIAS

ALMEIDA, Cleibson Aparecido de; SILVA, Leonardo Derckan Rodrigues; SCHILIPACK, Elaine Cristina da Silva; MINERVI, Nivaldo Aparecido. Melhoria na qualidade de dados com a aplicação de. **Atoz: novas práticas em informação e conhecimento**, [s. l.], v. 5, n. 2, p. 72, 2016. Disponível em: DOI: <http://dx.doi.org/10.5380/atoz.v5i2.47303>. Acesso em: 19 dez. 2023.

ANCIB. **Coordenações e Ementas de GT**. [S. l.], 2024. Disponível em: <https://ancib.org/sobre/>. Acesso em: 27 nov. 2024.

ASSAF, Ahmad; SENART, Aline; TRONCY, Raphaël. 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, 1 jul. 2016. Disponível em: DOI: <http://dx.doi.org/10.4018/IJSWIS.2016070104>. Acesso em: 23 out. 2024.

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

BATINI, Carlo; SCANNAPIECO, Monica. **Data quality and Information Quality**: dimensions, principles and techniques. [S. l.]: Springer, 2016. 520 p.

BENTANCOURT, Silva Silvia Maria Puentes; ROCHA, Rafael Port da. Metadados de qualidade e visibilidade na comunicação científica. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 17, n. esp. 2, p. 82-101, 12 dez. 2012. Disponível em: DOI: <https://doi.org/10.5007/1518-2924.2012v17nesp2p82>. Acesso em: 19 dez. 2023.

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

COELHO JÚNIOR, Abeil; LEMOS, Daniela Lucas da Silva. Qualidade de dados em acervos museais: uma avaliação semiautomática para os acervos sob gestão do ibram. *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 22., 2022a, Porto Alegre. **Anais** [...]. Porto Alegre: UFRGS, 2022. p. 1-11. Disponível em: <https://brapci.inf.br/index.php/res/v/202046>. Acesso em: 19 dez. 2023.

COELHO JÚNIOR, Abeil; LEMOS, Daniela Lucas da Silva. Tratamento da informação em acervos culturais: avaliação do uso de vocabulários controlados em coleções museológicas sob gestão do instituto brasileiro de museus. *In*: WORKSHOP DE INFORMAÇÃO, DADOS E TECNOLOGIA, 5., 2022b, Vitória. **Anais** [...]. Vitória: UFES, 2023. p. 235-241. Disponível em: <https://widat2022.ufes.br/wp-content/uploads/2023/04/widat-2022-anais.pdf>. Acesso em: 19 dez. 2023.

DIAS, Guilherme Ataíde; ARAÚJO, Wagner Junqueira de; DINIZ, Adriana Valéria Santos; CÓRDULA, Flavio Ribeiro; COSTA, Paulo Roberto Santos. Garbage In, Garbage Out (GIGO): enfrentando esta máxima nos conjuntos de dados associados ao programa dinheiro direto na escola (pdde). *In*: WORKSHOP DE INFORMAÇÃO, DADOS E TECNOLOGIA, 5., 2023, Brasília. **Anais** [...]. Brasília: Ibict, 2023. p. 1-11. Disponível em: <https://labcotec.ibict.br/widat/index.php/widat2023/article/view/32/60>. Acesso em: 05 dez. 2024.

ESPÍNDOLA, Priscilla Lüdtke; SALM JUNIOR, José Francisco; ROSA, Francisco; JULIANI, Jordan Paulesky. Governança de dados aplicada à ciência da informação. **RDBCI**: Revista Digital de Biblioteconomia e Ciência da Informação, [s. l.], v. 16, n. 3, p. 274-298, 16 ago. 2018. DOI: <http://dx.doi.org/10.20396/rdbci.v16i3.8651080>. Acesso em: 19 jun. 2024.

FAGUNDES, Priscila Basto; MACEDO, Douglas Dyllon Jeronimo de; FREUND, Gislaine Parra. 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, nov. 2017. Disponível em: DOI: <https://doi.org/10.20396/rdbci.v16i1.8650412>. Acesso em: 19 dez. 2023.

JESUS, Ananda Fernanda de; BARITÉ, Mario; SANTAREM SEGUNDO, José Eduardo. Termo qualidade de dados na Ciência da Informação brasileira. *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 24., 2024, Vitória. **Anais**

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

[...]. Vitória: ANCIB, 2024. p. 1-16. Disponível em:
<https://enancib.ancib.org/index.php/enancib/xxivenancib/paper/view/2295>. Acesso em: 29 abr. 2025.

JESUS, Ananda Fernanda de; SANTAREM SEGUNDO, José Eduardo. A descrição formal da qualidade de dados publicados na Web: análise do data quality vocabulary (dqv). **Em Questão**, Porto Alegre, v. 29, n. e-129415, p. 1-34, 2023a. Disponível em: DOI: <http://dx.doi.org/10.1590/1808-5245.29.129415>. Acesso em: 19 nov. 2024.

JESUS, Ananda Fernanda de; SANTAREM SEGUNDO, José Eduardo. Qualidade de dados Linked Data: análise da temática sob a perspectiva da Ciência da Informação. **Informação@Profissões**, Londrina, v. 11, n. 2, p. 153-169, 20 set. 2023b. Disponível em: DOI: [10.5433/2317-4390.2022v11n2p153](https://doi.org/10.5433/2317-4390.2022v11n2p153). Acesso em: 19 dez. 2023.

JESUS, Ananda Fernanda de; SANTAREM SEGUNDO, José Eduardo. 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/xxiienancib/paper/viewFile/811/729>. Acesso em: 19 dez. 2022.

LEMOS, Daniela Lucas da Silva; COELHO JÚNIOR, Abeil. 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, fev. 2023. Disponível em: DOI: <https://doi.org/10.5007/1518-2924.2023.e90510>. Acesso em: 19 dez. 2023.

JULIANI, Jordan Paulesky; MARTIGNAGO, Deisi; MADRID, Grace; SALM JUNIOR, José Francisco; BASTOS, Paulo Ricardo P. de M. Governança de dados aplicada no processo de catalogação. **Revista Brasileira de Biblioteconomia e Documentação**, [s. l.], v. 15, n. 2, p. 81-105, 2019. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/1153>. Acesso em: 19 dez. 2023.

MARTINS, Dalton Lopes; LEMOS, Daniela Lucas da Silva; CARMO, Danielle do; SIQUEIRA, Joyce; OLIVEIRA, Luis Felipe Rosa de. Requisitos de qualidade para dados de agregação em museus: o caso Ibram. *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 21., 2021a, Rio de Janeiro. **Anais** [...]. Rio de Janeiro: ANCIB, 2021a. p. 1-15. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/575>. Acesso em: 19 nov. 2024.

MARTINS, Dalton Lopes; LEMOS, Daniela Lucas da Silva; CARMO, Danielle do; SIQUEIRA, Joyce; OLIVEIRA, Luís Felipe Rosa de. Requisitos de qualidade para dados de agregação em museus: o caso do Ibram. **Tendências da Pesquisa Brasileira em Ciência da Informação**, [s. l.], v. 14, p. 1-25, 2021b. Disponível em: <https://brapci.inf.br/index.php/res/v/197846>. Acesso em: 19 dez. 2023.

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

MELO, Jessica Oliveira de Souza Ferreira; BOTECA, Leonardo Castro; SANTAREM SEGUNDO, José Eduardo. Metodologia de Avaliação de Qualidade para Dados Conectados. **Informação & Tecnologia**, Marília/João Pessoa, v. 4, n. 2, p. 80-101, 20 out. 2017. Disponível em: DOI: <https://doi.org/10.22478/ufpb.2358-3908.2017v4n2.40539>. Acesso em: 19 dez. 2023.

MELO, Jessica Oliveira de Souza Ferreira. **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. Disponível em: <https://repositorio.unesp.br/server/api/core/bitstreams/a4efef73-6c80-4060-b48b-27ddd311409/content>. Acesso em: 19 dez. 2023.

MOREIRA, Fábio Mosso; BISI, Pedro Henrique Santos; BOTECA, Leonardo Castro; SANTAREM SEGUNDO, José Eduardo; SANT'ANA, Ricardo César Gonçalves. A qualidade na recuperação de dados governamentais: um estudo sobre dados de políticas públicas na internet. **Perspectivas em Ciência da Informação**, [s. l.], v. 25, n. 2, p. 103-132, jun. 2020. Disponível em: DOI: <http://dx.doi.org/10.1590/1981-5344/3994>. Acesso em: 19 dez. 2023.

MOURA JUNIOR, Pedro Jácome de; ARAGÃO, Maicon Henrique Ferreira. Metas, ações e indicadores como subsídios para análise da qualidade de dados: uma inversão necessária entre consequentes e antecedentes. *In*: WORKSHOP DE INFORMAÇÃO, DADOS E TECNOLOGIA, 2., 2018, [s. l.]. **Anais [...]** [S. l.: s. n.], 2018. p. 4-10, 2018. Disponível em: <https://labcotec.ibict.br/widat/index.php/widat2023/article/view/132>. Acesso em: 19 nov. 2024.

PICCOLO, Daiane Marcela; TADINI, Antonio Victor Wolf; TEIXEIRA, Heytor Diniz; BOTECA, Leonardo Castro; SANT'ANA, Ricardo César Gonçalves; SANTAREM SEGUNDO, José Eduardo; ALVES, Rachel Cristina Vesu. Qualidade de dados em gestão de dados de pesquisa. **Em Questão**, Porto Alegre, p. 159-184, 7 dez. 2021. Disponível em: DOI: <https://doi.org/10.19132/1808-5245281.159-184>. Acesso em: 19 dez. 2021.

PICCOLO, Daiane Marcela. Qualidade de dados dos sistemas de informação do Datasus: análise crítica da literatura. **Ciência da Informação em Revista**, Maceió, v. 5, n. 3, p. 13-19, 31 dez. 2018. Disponível em: DOI: <https://doi.org/10.28998/cirev.2018v5n3b>. Acesso em: 19 dez. 2023.

SADIQ, Shazia (ed.). **Handbook of data quality: research and practice**. New York: Springer, 2013. 440 p.

SANTOS, Nirian Martins Silveira dos; STREIT, Rosalvo Ermes. O processo decisório de governança de dados. **Brazilian Journal of Information Studies: Research Trends**, [s. l.], v. 2, n. 2, p. 64-73, jan. 2018.

SILVA, Jordana N.; SOUZA, Jessica; OLIVEIRA, Allan César M. de; TAVARES, Maria de Fátima; BOTECA, Leonardo C. 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-WIDAT, 1., 2017, Florianópolis. **Anais [...]**. Florianópolis: UFSC, 2017. p. 1-6. Disponível em:

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

<https://www.widat2017.ufsc.br/wp-content/uploads/2017/09/ST2.2.pdf>. Acesso em: 19 jun. 2024.

TRIQUES, Maria Lígia; ARAKAKI, Ana Carolina Simionato; CASTRO, Fabiano Ferreira de. Aspectos da representação da informação na curadoria digital. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 25, p. 01-21, maio 2020. Disponível em: DOI: <https://doi.org/10.5007/1518-2924.2020.e69898>. Acesso em: 14 abr. 2025.

TURI, Leandro Furlam; COMARELA, Giovanni. Impacto da adequação à lei geral de proteção de dados pessoais na metrificação da qualidade de dados. *In*: WORKSHOP DE INFORMAÇÃO, DADOS E TECNOLOGIA, 5., 2022, Vitória. **Anais [...]**. Vitória: Ufes, 2022. p. 121-126. Disponível em: <https://widat2022.ufes.br/wp-content/uploads/2023/04/widat-2022-anais.pdf>. Acesso em: 19 dez. 2023.

WANG, Richard Y.; STRONG, Diane. 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.

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)