

Acesso aberto**Editor**

Nivaldo Calixto Ribeiro
Universidade Federal de Lavras,
Lavras, Minas Gerais, Brasil.

Pareceristas

Makson de Jesus Reis
Universidade Federal da Bahia, Salvador,
Bahia, Brasil.

Parecerista B com identidade preservada.
Concordou em abrir o parecer sem a sua
identificação.

***Correspondência**

Cristina Marchetti Maia
cristina.maia@unesp.br

Histórico

Recebido em: 29 de maio de 2026.

Aceito em: 18 de agosto de 2026.

Publicado em: 31 de agosto de 2026.

Como referenciar

MAIA, Cristina Marchetti; SILVA, Jéssyca
Maria Santos da; SANTAREM SEGUNDO, José
Eduardo. Competência em dados e ciência de
dados na ciência da informação: panorama da
produção científica e atuação do bibliotecário.
Ciência da Informação Express, Lavras, MG, v.
7, 2026.

DOI: <https://doi.org/10.60144/v7i.2026.201>

Dados da pesquisa

Os conjuntos de dados estão disponíveis no
Repositório Zenodo:

<https://doi.org/10.5281/zenodo.14624517>.

Inteligência artificial

Foi utilizada a ferramenta Claude para auxiliar
na revisão do texto.

Direitos autorais e licença pública

© 2026 Maia, C. M.; Silva, J. M. S. da; Santarem
Segundo, J. E. Este é um artigo de acesso aberto
distribuído sob os termos da Licença [Creative
Commons Atribuição \(CC BY\)](#). Você pode copiar,
distribuir, remixar, transformar e criar a partir deste
material, para qualquer finalidade, inclusive
comercial, desde que seja atribuída a autoria,
indicada a fonte e informadas eventuais alterações,
sem sugerir endosso dos autores.



Competência em dados e ciência de dados na ciência da informação: panorama da produção científica e atuação do bibliotecário

Data and data science competence in information science: overview of scientific output and the librarian's role

Cristina Marchetti Maia*

Doutoranda em Ciência da
Informação. Universidade
Estadual Paulista, Marília, São
Paulo, Brasil.

<https://orcid.org/0000-0001-8442-8386>

Jéssyca Maria Santos da Silva

Mestranda em Ciência da
Informação da Universidade
Estadual Paulista, Marília, São
Paulo, Brasil.

<https://orcid.org/0000-0003-3159-8853>

José Eduardo Santarem Segundo

Doutor em Ciência da
Informação. Universidade
Estadual Paulista, Marília, São
Paulo, Brasil.

<https://orcid.org/0000-0003-3360-7872>

Resumo

Problemática: Considerando a competência em dados como um pilar necessário à execução de projetos de Ciência de Dados, questiona-se: o que tem sido produzido sobre a competência em dados e sua interseção com a Ciência de Dados? **Objetivo:** A pesquisa tem por objetivo investigar a produção científica que relaciona a competência em dados com a Ciência de Dados, no contexto da Ciência da Informação. **Metodologia:** A pesquisa apresenta natureza quali-quantitativa, do tipo descritiva-exploratória, cujos métodos aplicados consistem em revisão bibliográfica, análise bibliométrica e de conteúdo. **Resultados:** A análise bibliométrica demonstrou o crescimento da área ao longo dos anos e apresentou um mapa de colaboração entre os pesquisadores. Já a análise de conteúdo possibilitou traçar um panorama das pesquisas na temática e identificar as fases mais abordadas. **Conclusão:** O trabalho evidenciou onde a atuação do bibliotecário já está consolidada na execução de projetos de Ciência de Dados e expôs as lacunas existentes.

Palavras-chave: Competência em dados; Ciência de Dados; Projeto de Ciência de Dados; Educação em dados.

Abstract

Problem: Considering data literacy as a necessary pillar for the execution of Data Science projects, the following question arises: what has been produced on data literacy and its intersection with Data Science? **Objective:** This research aims to investigate the scientific production relating data literacy to Data Science, within the context of Information Science. **Methodology:** The research has a qualitative-quantitative nature, of a descriptive-exploratory type, whose applied methods consist of a bibliographic review, bibliometric analysis, and content analysis. **Results:** The bibliometric analysis demonstrated the growth of the field over the years and presented a map of researcher collaboration. Content analysis, in turn, made it possible to outline a panorama of research on the topic and identify the most frequently addressed phases. **Conclusion:** The study highlighted where librarian involvement is already consolidated in the execution of Data Science projects and revealed existing gaps.

Keywords: Data literacy; Data Science; Data Science Project; Data education.

1 INTRODUÇÃO

A origem da Ciência de Dados remonta aos estudos de Naur (1974), que propôs o termo Data Science como uma ciência responsável por lidar com dados. Estudos mais recentes como de Donoho (2017) já associa a área a raízes na estatística, computação e na investigação empírica, mas argumenta que a Ciência de Dados constitui um domínio próprio, com métodos e paradigmas específicos. No entanto, outra vertente situa sua consolidação a partir dos anos 2000, impulsionada pelo avanço computacional, pela expansão digital e pelo crescimento do Big Data (Cleveland, 2001; Kitchin, 2014; Cao, 2016).

Dentre os modelos existentes na CD para execução de projetos, destaca-se a estrutura proposta por O'Neil e Schutt (2014) constituída por sete etapas: 1- Coletar dados não processados; 2- Processar os dados, utilizando diferentes ferramentas e linguagens; 3- Limpar os dados, ajustando formatos, removendo duplicatas e tratando valores perdidos ou discrepantes; 4- Realizar análise exploratória para entender e selecionar os dados relevantes para o projeto; 5- Projetar o modelo usando algoritmos, conforme o tipo de problema; 6- Comunicar os resultados por meio de interpretações, visualizações e relatos; 7- Implementar ou prototipar um produto.

Diante disso, os envolvidos na execução desses projetos necessitam de competências específicas para lidar com o armazenamento, gestão, curadoria e preservação dos dados. Além

disso, requerem conhecimentos sobre questões éticas, como o sigilo e a integridade das informações, de forma a assegurar a qualidade do conteúdo para que possam ser trabalhados, avaliados, compartilhados e utilizados de maneira eficiente e adequada em uma argumentação baseada em dados, tanto na tomada de decisões quanto na geração de novos conhecimentos

Nesse contexto, insere-se a competência em dados, entendida como um conjunto de saberes e habilidades voltados à localização, avaliação, uso e compartilhamento éticos de informações (Association of College & Research Libraries - ACRL, 2013), bem como uma base de conhecimento que permite transformar dados em conhecimento acionável, abrangendo desde o desenvolvimento de hipóteses e identificação de problemas até a interpretação, planejamento e monitoramento de informações (Koltay, 2017, p. 9, tradução livre).

Em uma pesquisa desenvolvida por Matos, Condurú e Benchimol (2022), a partir de uma análise bibliométrica, identificou-se que a produção científica de pesquisadores da CI sobre a CD aumentou nos últimos 6 anos, o que indica uma interseção entre os campos como uma tendência de pesquisa em crescimento. Complementarmente, a pesquisa identificou que, dentre os temas que foram recuperados no contexto de dados, a CoInfo representou 4,6% (16 artigos) dos estudos, o que os autores consideraram uma representação significativa e como uma das possibilidades de estudos que correlacionam os dois campos do conhecimento. Outra pesquisa conduzida por Bowler e Shaw (2024) acerca da competência em dados também revela forte conexão com conceitos como o Big Data e a datificação.

A competência em dados constitui um pilar necessário à execução de projetos de Ciência de Dados, pois envolve a compreensão das etapas do processo, a capacidade de lidar com diferentes tipos de dados coletados e gerados, o manuseio consciente e crítico de instrumentos, bem como a construção e interpretação de análises de dados. Diante disso, questiona-se: o que tem sido produzido, no âmbito da Ciência da Informação, sobre a interseção entre a competência em dados e a Ciência de Dados, e quais lacunas e oportunidades essa produção revela para a atuação dos profissionais da informação?

Para responder esse questionamento, a pesquisa teve por objetivo investigar a produção científica que relaciona a competência em dados com a Ciência de Dados, no contexto da

Ciência da Informação, visando traçar um panorama de pesquisas na temática, bem como identificar lacunas e oportunidades para aprimorar a aplicação da competência em dados pelos bibliotecários. Os objetivos específicos são: (1) Analisar a frequência dos artigos através dos anos e realizar um mapeamento da colaboração científica entre os pesquisadores; (2) Verificar quais elementos do processo da Ciência de Dados estão contemplados nos estudos acerca da competência em dados, considerando as etapas de um projeto de CD.

A realização desta pesquisa justifica-se pela necessidade de aprimorar a competência em dados de profissionais da informação e pesquisadores, o que se torna cada vez mais evidente diante do volume crescente de dados gerados e da importância da análise de dados no processo de tomada de decisões.

A pesquisa possui natureza quali-quantitativa, do tipo descritivo-exploratória, cujos métodos consistem em revisão bibliográfica, análise bibliométrica e análise de conteúdo. Para a coleta de dados, foi desenvolvido um protocolo de busca em bases de dados internacionais. A análise quantitativa foi conduzida por meio do software VOSviewer, visando extrair indicadores capazes de mensurar a evolução da produção científica e a colaboração entre os autores da área. Já a análise de conteúdo, realizada após a extração dos dados dos artigos e o preenchimento de planilha no Excel, permitiu correlacionar a competência em dados com a CD.

Acredita-se que a pesquisa pode contribuir para a gestão eficaz de dados em projetos de Ciência de Dados e para o fortalecimento da Ciência da Informação, bem como para o aprimoramento dos serviços de suporte e capacitação voltados ao desenvolvimento da competência em dados dos envolvidos, destacando oportunidades de investigação e de atuação dos bibliotecários.

2 A CIÊNCIA DE DADOS (CD) E A COMPETÊNCIA EM DADOS

A seguir, são apresentados os principais fundamentos da Ciência de Dados, considerando suas etapas de desenvolvimento, bem como sua relação com a competência em dados e as habilidades necessárias para atuação nesse contexto.

2.1 A ciência de dados: origens, consolidação e modelos de processo

A Ciência de Dados é um campo interdisciplinar cuja origem não é consensual na literatura. Há autores que identificam antecedentes mais remotos, ainda no século XX, associando sua formação a discussões sobre tratamento e análise de dados. Nessa perspectiva, Naur (1974) é frequentemente citado como referência inicial do uso da expressão Data Science, enquanto Donoho (2017) sustenta que a área possui raízes históricas ligadas à estatística, à computação e à investigação empírica orientada por dados.

Outra vertente enfatiza sua consolidação mais recente, sobretudo a partir dos anos 2000, em decorrência do avanço da capacidade computacional, da expansão dos ambientes digitais e do crescimento do Big Data, da inteligência artificial e da análise de dados em larga escala (Cleveland, 2001; Kitchin, 2014; Cao, 2016). Nessa leitura, a Ciência de Dados se fortalece como campo voltado à obtenção, tratamento, análise, modelagem, interpretação e uso de dados para produção de conhecimento e apoio à tomada de decisão.

Essas perspectivas são complementares e permitem compreender a Ciência de Dados tanto como resultado de um processo histórico mais amplo quanto como um campo que ganha centralidade no contexto tecnológico contemporâneo. No âmbito da Ciência da Informação, essa aproximação é especialmente relevante, pois o trabalho com dados envolve não apenas modelagem, mas também organização, curadoria, representação, preservação, acesso, qualidade e uso ético da informação.

A literatura também apresenta diferentes modelos para descrever o processo de projetos orientados por dados. O modelo KDD (Knowledge Discovery in Databases) estrutura o processo em etapas como seleção, pré-processamento, transformação, mineração e interpretação (Fayyad; Piatetsky-Shapiro; Smyth, 1996). O CRISP-DM (Cross-Industry Standard Process for Data Mining) organiza o trabalho em compreensão do problema, compreensão dos dados, preparação, modelagem, avaliação e implantação (Shearer, 2000). Além deles, abordagens mais recentes destacam características como colaboração, interatividade, monitoramento e governança.

Neste estudo, adota-se a proposta de O’Neil e Schutt (2014) como referencial analítico principal, não por ser o único modelo possível, mas por oferecer maior detalhamento das etapas do processo e, assim, favorecer a identificação das competências mobilizadas em cada uma delas. Tal escolha é coerente com o objetivo da pesquisa, que não consiste em comparar modelos de Ciência de Dados, mas em examinar como a literatura sobre competência em dados dialoga com etapas concretas desse processo.

Com base nesse referencial, consideram-se as seguintes etapas: identificação do problema, coleta de dados, processamento, limpeza, análise exploratória, modelagem, comunicação e implementação. Além disso, reconhece-se que o processo é atravessado por elementos transversais, como qualidade, documentação, ética, transparência, privacidade e governança.

2.2 Conexão da ciência de dados com a competência em dados

A competência em dados pode ser compreendida como o conjunto de conhecimentos, habilidades e atitudes voltados à localização, compreensão, avaliação, interpretação, gestão, comunicação e uso ético de dados (ACRL, 2013; Koltay, 2015; Koltay, 2017). Sua importância cresce em um contexto no qual os dados assumem papel central na produção de conhecimento, na gestão institucional e na tomada de decisão.

Essa competência não se limita ao domínio técnico de ferramentas. Ela envolve também a capacidade de avaliar a qualidade dos dados, reconhecer limites e vieses, interpretar resultados, comunicar evidências e refletir sobre implicações éticas e sociais de seu uso. Por isso, sua interface com a Ciência de Dados é direta: enquanto a Ciência de Dados organiza métodos e processos de trabalho com dados, a competência em dados fornece a base formativa necessária para participar criticamente desse processo.

Essa relação pode ser observada nas etapas de um projeto de Ciência de Dados. A identificação do problema requer compreender necessidades informacionais; a coleta exige localizar e avaliar fontes; o processamento e a limpeza demandam organização, documentação e curadoria; a análise exploratória pressupõe leitura e interpretação crítica; a modelagem requer compreensão de pressupostos e limites dos modelos; a comunicação envolve tradução,

visualização e uso ético dos resultados; e a implementação exige atenção às consequências práticas e à responsabilidade sobre seu uso.

Assim, a competência em dados pode ser entendida como uma dimensão transversal ao processo de Ciência de Dados. Ela não substitui competências especializadas em estatística ou computação, mas é fundamental para que o trabalho com dados ocorra de forma crítica, reflexiva e socialmente responsável. No caso dos profissionais da informação, essa interface é particularmente relevante. Bibliotecários e outros especialistas podem atuar na formação de usuários, no apoio à gestão de dados de pesquisa, na curadoria, na avaliação de fontes, na organização e documentação dos dados, bem como na mediação de questões éticas e legais. Desse modo, a convergência entre Ciência de Dados e competência em dados constitui um espaço importante de pesquisa, formação e atuação no âmbito da Ciência da Informação.

3 METODOLOGIA

3.1 Delineamento da pesquisa

Trata-se de uma pesquisa de natureza quali-quantitativa, de caráter descritivo-exploratório, desenvolvida por meio de revisão bibliográfica, análise bibliométrica e análise de conteúdo temática. O estudo buscou, de um lado, examinar a evolução da produção científica e a colaboração entre autores e, de outro, analisar qualitativamente como a literatura relaciona competência em dados e Ciência de Dados.

Desde sua concepção, o levantamento foi realizado exclusivamente em bases de dados internacionais. Essa escolha se justifica pelo objetivo de mapear a circulação científica da temática em bases de alcance internacional, considerando que a discussão sobre competência em dados e Ciência de Dados se desenvolve em diálogo com diferentes comunidades acadêmicas.

3.2 Fontes de informação e justificativa das bases

As bases foram selecionadas por sua relevância temática e adequação ao campo da Ciência da Informação. Foram utilizadas Web of Science, Library & Information Science Abstracts (LISA) e

Library, Information Science & Technology Abstracts (LISTA). A LISA e a LISTA foram incluídas por serem reconhecidas na área de Ciência da Informação, enquanto a Web of Science foi adotada por sua abrangência multidisciplinar, pertinente ao caráter transversal do tema investigado. A não inclusão da Scopus constitui uma delimitação metodológica do estudo. Optou-se por combinar uma base multidisciplinar consolidada com duas bases especializadas da área. Ainda assim, reconhece-se que a inclusão da Scopus em um novo estudo futuro, poderia ampliar a cobertura documental.

3.3 Estratégia de busca e critérios de elegibilidade

A busca foi realizada em junho de 2024, com a string: “data science” AND (“data literacy” OR “data competency”), não houve delimitação temporal na recuperação inicial, a fim de permitir a identificação da evolução histórica da produção científica sobre a temática.

Foram aplicados filtros específicos em cada base. Na Web of Science, utilizaram-se filtros de tipo de documento e da categoria Information Science Library Science. Na LISA, foi utilizado o filtro “artigos revisados por pares”. Na LISTA, adotou-se o filtro “revistas acadêmicas analisadas por especialistas”.

Como critérios de inclusão, consideraram-se artigos científicos recuperados nas bases selecionadas, compatíveis com os filtros adotados, relacionados à interseção entre competência em dados e Ciência de Dados no escopo da Ciência da Informação, e disponíveis em português, inglês ou espanhol. Como critérios de exclusão, adotaram-se duplicatas, textos fora do foco temático, publicações incompatíveis com a tipologia documental definida e textos fora dos idiomas selecionados.

A busca inicial resultou em 118 registros: 12 na Web of Science, 95 na LISA e 11 na LISTA. Após remoção de duplicados no gestor de referências Zotero, obteve-se um conjunto de 110 artigos. Para a análise de conteúdo, foi definido um subconjunto inicial de 59 artigos. Desses, 8 foram excluídos por não estarem disponíveis nos idiomas estabelecidos ou por não tratarem efetivamente da temática da competência em dados, resultando em um corpus final de 51 artigos. O percurso de identificação, triagem, elegibilidade e inclusão foi organizado segundo a

lógica do fluxograma PRISMA (*Preferred Reporting Items for Systematic reviews and Meta-Analyses*).

3.4 Procedimento da análise bibliométrica

A análise bibliométrica foi realizada com os 110 artigos obtidos após a exclusão dos duplicados. Os dados foram organizados em planilha do Excel para examinar a evolução da produção ao longo dos anos. Em seguida, os registros foram processados no software VOSviewer para mapear a colaboração entre autores, utilizando-se o filtro de mínimo de dois documentos por autor. Essa etapa permitiu observar tanto a distribuição temporal das publicações quanto a configuração das redes de coautoria.

3.5 Procedimentos da análise de conteúdo

A análise de conteúdo temática foi conduzida sob a perspectiva de Bardin (2010) e incidiu sobre os 51 artigos selecionados ao final da triagem. Para essa etapa, adotou-se o recorte temporal de 2021 a 2024, com o objetivo de concentrar a análise qualitativa na produção mais recente e alinhada às tendências contemporâneas do tema.

A extração dos dados foi realizada em planilha eletrônica, considerando os campos título, autoria, temática e fases da Ciência de Dados identificadas nos estudos. O procedimento envolveu leitura exploratória, codificação, agrupamento em categorias e interpretação dos resultados. Para a categorização das fases, utilizou-se como base o modelo de O'Neil e Schutt (2014). Para fins operacionais, a identificação do problema foi considerada como etapa inicial, e o processamento e limpeza dos dados foram agrupados por sua forte inter-relação. Os códigos analíticos adotados foram:

IP- Identificação do problema: compreensão da necessidade de informação na perspectiva de formação da competência em dados;

Co- Coleta de dados: reconhecimento do tipo de dado e fontes, além da aplicação de técnicas de coleta;

PL- Processamento e Limpeza dos dados: uso de técnicas e ferramentas para armazenar, categorizar e organizar os dados. Inclui atividades de gestão e curadoria de dados.

Ae- Análise exploratória: examinar os dados visando prepará-los para aplicação de técnicas estatísticas. Compreende o uso de tais técnicas e inclui atividades de mineração de dados.

Ma- Modelo algoritmo: compreende empregar a análise preditiva com o uso de algoritmos. Envolve o uso de machine learning;

Cm- Comunicação: aborda interpretação, visualização e uso ético de dados para a tomada de decisão. Considera aspectos de proteção e privacidade dos dados, diretrizes para fazer uma correta citação de dados e orientações sobre licenciamento de dados.

Im- Implementação de um produto: resulta na criação de um produto que gera novos dados.

Além disso, os artigos foram classificados segundo quatro categorias temáticas predominantes: (1) formação e atuação do profissional; (2) educação do usuário/ensino; (3) avaliação da competência em dados; e (4) discussão conceitual.

A análise foi conduzida com base em critérios explícitos e em uma matriz padronizada de extração de dados, buscando garantir maior transparência e coerência ao procedimento. Ainda assim, reconhece-se que a ausência de análise independente por múltiplos pesquisadores pode constituir uma limitação quanto à redução de vieses interpretativos.

3.6 Limitações metodológicas

Entre as limitações do estudo, destacam-se o uso exclusivo de bases internacionais, a não inclusão da Scopus e a possibilidade de que a estratégia de busca não contemple todas as variações terminológicas empregadas pela literatura. Além disso, o recorte de 2021 a 2024 restringe o aprofundamento qualitativo à produção mais recente, embora preserve a análise histórica do conjunto total na etapa bibliométrica. Apesar dessas limitações, os procedimentos adotados foram considerados adequados para mapear a produção científica sobre a temática e identificar tendências, lacunas e possibilidades de atuação no âmbito da Ciência da Informação. Os resultados e discussões são apresentados na seção subsequente.

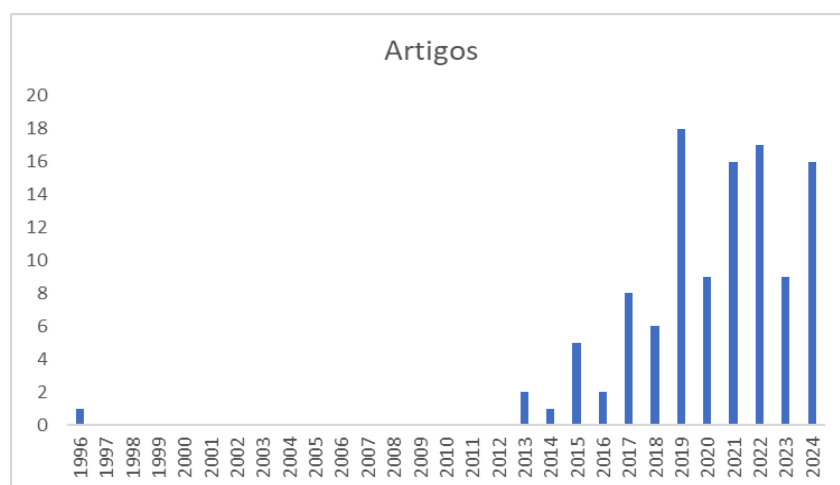
4 RESULTADOS E DISCUSSÕES

4.1. Evolução da produção científica e colaboração entre pesquisadores

As análises a seguir estão em consonância com o objetivo geral desta pesquisa ao investigar a produção científica que relaciona a competência em dados com a Ciência de Dados e atende o primeiro objetivo específico que consiste em analisar a frequência dos artigos ao longo dos anos e realizar um mapeamento da colaboração científica entre os pesquisadores.

A evolução da produção científica sobre competência em dados e CD pode ser acompanhada na figura 1.

Figura 1- Evolução da produção ao longo dos anos, de 1996 a 2024.



Fonte: Dados da pesquisa (2024).

Na Figura 1 é possível visualizar que a temática foi iniciada em 1996, porém mais de uma década não obteve artigos com essas temáticas. De 2013 a 2017 artigos houve um crescimento tímido, com uma média de 3,6 artigos por ano, sendo o ano de 2017 que obteve um número maior com 8 artigos. No ano de 2019 houve um crescimento significativo com 18 artigos, mantendo um equilíbrio entre 2021 e 2022, com 16 artigos e 17 artigos, respectivamente. Observa-se que a média de publicação dos últimos 4 (quatro) anos foi de 14,5 artigos, concluindo-se que a temática está em constante evolução, porém necessita de mais trabalhos na área com essa temática.

Dando continuidade à análise, foi gerado um mapa que demonstra a colaboração dos pesquisadores da área e pode ser visualizado na figura 2:

Figura 2 – Mapa de colaboração de autores.



Fonte: Dados da pesquisa (2024).

Observa-se na figura 2 que a temática possui poucas colaborações, apenas Acker e Bowler possuem uma ligação com mais de dois documentos. Além disso, o conjunto de dados possibilitou identificar os outros pesquisadores que, apesar de possuírem vários artigos na temática, atuam em uma rede de colaboração científica ainda modesta. Destaca-se o autor Tibor Koltay que possui papel de grande relevância nas publicações na área. Com base na figura 3, observa-se que a Ciência da Informação possui poucos trabalhos nessa temática, o que resulta em um pequeno número de colaborações. Sendo assim os pesquisadores desenvolvem trabalhos isoladamente, resultando em agrupamentos menores do que em outras temáticas de pesquisa.

4.2 Temáticas trabalhadas e convergência com a CD

Esta seção apresenta os resultados de uma análise de conteúdo que forneceu dados quali-quantitativos. Primeiramente, é apresentado um levantamento das temáticas presentes nas pesquisas, alinhado ao objetivo geral que consiste em identificar os temas trabalhados na

área, permitindo reconhecer lacunas e oportunidades para aprimorar a aplicação da competência em dados.

Para tal finalidade, foram definidas quatro categorias (Quadro 1) e cada artigo foi encaixado em uma delas, considerando qual área prevalecia no estudo em questão.

Quadro 1 – Categorias de análise dos artigos.

1	Formação e atuação do bibliotecário	Compreensão do contexto profissional, inclui discussões sobre as habilidades e conhecimentos em atuar no desenvolvimento da competência em dados em instituições, bem como no oferecimento de serviços de dados. Também está relacionada aos desafios enfrentados pelas bibliotecas e aborda recursos e ideias relacionados à inovação de práticas nessas unidades.
2	Formação do usuário-ensino	Inclui a análise de práticas voltadas ao ensino da competência em dados, em forma de cursos e <i>workshops</i> , geralmente, promovidas por bibliotecários e educadores. Compreende as reflexões e uso de metodologias de ensino visando o aperfeiçoamento de estratégias pedagógicas para promover essa ação.
3	Avaliação da competência em dados	Inclui mapeamentos dos conhecimentos e habilidades em dados, tanto dos profissionais, como dos professores e estudantes.
4	Discussão conceitual	Trata de pesquisas que buscam identificar o estado atual da pesquisa em competência em dados, por meio de revisões sistemáticas e discussões teóricas da área.

O resultado da categorização é apresentado por ordem decrescente na tabela 1:

Tabela 1 – Temáticas dos estudos.

Categoria	1	2	3	4	Total
Quantidade	30	14	4	3	51

Fonte: Dados da pesquisa (2024).

A predominância de estudos voltados à formação e atuação do bibliotecário, seguida pelos que tratam da educação do usuário, indica não só uma simples distribuição temática, mas uma área ainda em processo de consolidação, na qual a prática precede a teoria. Essa configuração corrobora o argumento de Koltay (2015; 2017) de que a competência em dados não se limita ao domínio técnico de ferramentas, mas se constrói na experiência formativa e na mediação cotidiana do profissional da informação. Além disso, os estudos sobre percepção de bibliotecários acadêmicos sobre a competência em dados em suas funções diárias (Luo; Tang, 2024) e sobre conceitos centrais da biblioteconomia de dados (Ashiq; Warraich, 2022) convergem para o ponto de que competência em dados se consolida cada vez mais como uma postura ética, crítica e reflexiva diante do dado que possui como cerne sua função educacional.

Quanto às práticas pedagógicas, o uso recorrente do storytelling (Adisa et al., 2024) e das abordagens baseadas em arte (Matuk et al., 2024) evidencia, apesar de tímidas, o uso de estratégias diferenciadas de ensino-aprendizagem. Ao contextualizar dados em narrativas, essas abordagens operacionalizam, na prática pedagógica, o próprio argumento teórico da competência em dados como algo que não se limita ao domínio técnico de ferramentas, trazendo à tona justamente as dimensões interpretativa, crítica e comunicacional que a definição de ACRL (2013) e Koltay (2015; 2017) já apontava como centrais.

Muitos estudos mencionam a importância do gerenciamento de dados, sendo que 13 (treze) deles tiveram como foco a GDP e elaboração de plano de gestão de dados. Serviços de apoio à GDP tem grande potencial de atuação do profissional de dados e há uma crescente demanda por profissionais qualificados para atuar nesta área, para capacitar e oferecer serviços especializados aos pesquisadores. O profissional pode auxiliar no planejamento, desenvolver materiais instrucionais, trabalhar em soluções como repositórios institucionais, pois tem conhecimento de instrumentos de representação da informação nesses meios.

Acerca da formação do usuário, a discussão sobre competência algorítmica (Garingan; Pickard, 2021) e sobre educação em dataficação, centrando-se nos recursos educacionais on-line voltados para uma educação sobre a dataficação, com ênfase na proteção de dados (Sander, 2024). Tais estudos evidenciam uma ampliação do escopo da competência em dados,

incorporando dimensões que dialogam com os elementos transversais como ética, transparência, privacidade e governança, que estão presentes nas etapas do processo de Ciência de Dados. Essa convergência reforça o argumento central deste estudo de que a competência em dados não é uma etapa isolada ou uma parte do processo de Ciência de Dados, mas uma dimensão transversal que perpassa cada uma de suas fases, da identificação do problema à implementação corroborando a proposição de que ambos os campos, embora distintos em seus objetos imediatos, são interdependentes na formação de profissionais capazes de atuar de forma crítica, reflexiva e socialmente responsável diante dos dados.

A escassez relativa de estudos sobre avaliação da competência, por sua vez, não deve ser lida apenas como uma lacuna quantitativa, mas como um possível resultado de uma necessidade de discussão conceitual mais profunda sobre a competência em dados que se encontra em processo de amadurecimento.

Dando continuidade, foi realizada uma análise visando atingir o segundo objetivo específico, identificando quais elementos do processo da Ciência de Dados são mencionados nos estudos acerca do ensino da competência em dados, considerando as etapas de um projeto de CD. Utilizou-se como referência a nomenclatura apresentada na seção de Materiais e Métodos e o resultado é demonstrado na tabela 2:

Tabela 2- Fases do projeto em CD nos artigos em competência em dados.

Fases	IP	Co	PL	Ae	Ma	Cm	Im
Quantidade	14	36	24	29	2	41	1

Fonte: Dados da pesquisa (2024).

A concentração dos estudos na fase de comunicação, seguida por coleta e análise exploratória, pode revelar onde a Ciência da Informação já reconhece sua própria expertise dentro do processo de Ciência de Dados, e onde ainda hesita em se posicionar. Essa concentração sugere que a competência em dados, tal como vem sendo tratada empiricamente, ainda opera mais como uma extensão da competência em informação clássica que consiste em localizar, avaliar, organizar e utilizar a informação do que como uma incorporação plena das etapas

computacionalmente mais intensivas do processo, como modelagem e implementação. Vale destacar que alguns artigos não abordaram habilidades específicas em dados para que fossem relacionadas às etapas de CD, pois trataram de uma conceituação mais geral da competência em dados.

A baixa recorrência da etapa de identificação do problema, apesar de sua relevância teórica, expõe uma possível lacuna formativa. Isso porque, se a competência em dados pressupõe a compreensão da necessidade de trabalhar com dados, sua ausência recorrente nos estudos analisados indica que essa dimensão inicial, mais conceitual e menos técnica, ainda não foi suficientemente incorporada às propostas de formação da área, mesmo sendo pré-requisito lógico para todas as etapas seguintes.

Com relação ao processamento e à limpeza dos dados, etapas em que a curadoria e a gestão de dados de pesquisa (GDP) aparecem como território natural de atuação do bibliotecário por ter a expertise de trabalhar com a representação do conteúdo por metadados e o armazenamento voltado à recuperação são, em essência, extensões diretas de competências bibliotecárias tradicionais aplicadas a um novo objeto.

A análise exploratória, por sua vez, marca um ponto de inflexão: é a primeira etapa em que a literatura reconhece explicitamente a necessidade de colaboração interdisciplinar, com estatística, matemática e computação. Isso indica que a competência em dados do bibliotecário não se define pelo domínio técnico dessas ferramentas, mas pela capacidade de atuar criticamente sobre seus resultados, identificando vieses, avaliando inconsistências e questionando implicações. É nesse ponto que a discussão dialoga com Winckelmann (2023) a partir do uso crescente de análise preditiva via inteligência artificial e machine learning amplia o alcance da competência em dados para além da técnica, exigindo vigilância constante sobre impactos discriminatórios e justiça algorítmica, uma preocupação que reposiciona o profissional da informação não como operador de algoritmos, mas como agente de mediação ética desses sistemas.

A predominância da comunicação como competência mais recorrente confirma, em certa medida, que a competência em dados se realiza plenamente na capacidade de traduzir

complexidade em significado compreensível e eticamente responsável para diferentes públicos. Essa constatação retoma a discussão anterior sobre storytelling e abordagens baseadas em arte, sugerindo que a comunicação não é uma etapa isolada, mas o ponto de convergência onde as competências técnicas e interpretativas trabalhadas nas etapas anteriores se tornam socialmente úteis.

Já a baixa incidência da etapa de implementação, reconhecida como não obrigatória pelo próprio modelo de O'Neil e Schutt (2014), não deve ser lida como uma limitação da competência em dados, mas como um indicativo de que essa etapa pode extrapolar o escopo da Ciência da Informação e se ancora em áreas como engenharia de software e Ciência da Computação aplicada.

A análise dos artigos permitiu compreender um pouco como a competência em dados está inserida no contexto da Ciência de Dados, seja por meio de ações de formação como no oferecimento de serviços.

5 CONSIDERAÇÕES FINAIS

Esta pesquisa investigou a produção científica, no âmbito da Ciência da Informação, sobre a interseção entre competência em dados e Ciência de Dados, e as lacunas e oportunidades para a atuação de bibliotecários. Os resultados indicam que os objetivos foram alcançados.

A análise bibliométrica de 110 artigos revelou que, embora os primeiros trabalhos datem de 1996, a temática só ganhou consistência a partir de 2013, com pico em 2019 e média de 14,5 publicações anuais nos últimos quatro anos. Esse crescimento ainda incipiente, reforçado pela baixa colaboração científica entre pesquisadores, evidencia um campo em fase de consolidação. Infere-se, assim, que a interseção entre competência em dados e Ciência de Dados constitui uma linha de pesquisa emergente na Ciência da Informação, o que evidencia um campo em fase de consolidação.

Quanto ao panorama temático, os resultados indicam concentração nas fases de comunicação, coleta e análise exploratória, revelando um padrão de atuação do bibliotecário mais alinhado às competências já consolidadas, como avaliação de fontes, curadoria, mediação e comunicação. O

principal avanço teórico consistiu em articular as etapas do modelo de O'Neil e Schutt (2014) com a literatura sobre competência em dados, expondo lacunas como a escassez de estudos sobre avaliação da competência, a baixa exploração da identificação do problema e a quase ausência de discussões sobre modelagem algorítmica na Ciência da Informação.

Diante disso, recomenda-se que estudos futuros desenvolvam instrumentos para avaliação da competência em dados; investiguem o papel do profissional da informação nas etapas iniciais do processo de Ciência de Dados; e explorem a interface entre competência em dados e questões éticas emergentes, como vieses algorítmicos, transparência e uso responsável de inteligência artificial. Esses caminhos, somados ao fortalecimento de redes de colaboração, indicam que a interseção entre competência em dados e Ciência de Dados constitui um campo fértil e pouco explorado, cuja consolidação depende do engajamento da Ciência da Informação na formação de profissionais capazes de atuar neste campo.

REFERÊNCIAS

ADISA, Ibrahim Oluwajoba; HERRO, Danielle; ABIMBADE, Oluwadara; IRGENS, Golnaz Arastoopour. Engaging elementary students in data science practices. **Information and Learning Sciences**, [S.l.], v. 125, n. 7/8, p. 513-544, 2024. Disponível em: <http://dx.doi.org/10.1108/ILS-06-2023-0062>. Acesso em: 30 ago. 2026.

ASHIQ, Murtaza; WARRAICH, Nosheen Fatima. Librarian's perception on data librarianship core concepts: a survey of motivational factors, challenges, skills and appropriate trainings platforms. **Library Hi Tech**, [S.l.], v. 42, n. 3, p. 849-866, 2022. Disponível em: <http://dx.doi.org/10.1108/LHT-12-2021-0487>. Acesso em: 30 ago. 2026.

ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES. Working Group on Intersections of Scholarly Communication and Information Literacy. **Intersections of Scholarly Communication and Information Literacy: creating strategic Collaborations for a Changing Academic Environment**. Chicago: ACRL, 2013. Disponível em: <https://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Intersections.pdf>. Acesso em: 30 ago. 2026.

BARDIN, Laurence. **Análise de conteúdo**. São Paulo: Edições 70, 2010.

BOWLER, Leanne; SHAW, Charlie. Trends in data literacy, 2018-2023: a review of the literature. **Information Research: an international electronic journal**, Boras, v. 29, n. 2, p. 198-205, 2024.

DOI: <https://doi.org/10.47989/ir292822>. Disponível em:
<https://publicera.kb.se/ir/article/view/55129>. Acesso em: 30 ago. 2026.

CAO, Longbing. Data science and analytics: a new era. **The International Journal of Data Science and Analytics**, [S. l.], v. 1, p. 1–2, 2016. Disponível em:
<https://doi.org/10.1007/s41060-016-0006-1>. Acesso em: 30 ago. 2026.

CLEVELAND, William S. Data science: an action plan for expanding the technical areas of the field of statistics. **International Statistical Review**, [S. l.], v. 69, n. 1, p. 21–26, 2001. DOI: 10.1111/j.1751-5823.2001.tb00477.x. Disponível em:
<https://onlinelibrary.wiley.com/doi/10.1111/j.1751-5823.2001.tb00477.x>. Acesso em: 30 ago. 2026.

DONOHO, David. 50 years of data science. **Journal of Computational and Graphical Statistics**, [S. l.], v. 26, n. 4, p. 745–766, 2017. DOI: 10.1080/10618600.2017.1384734. Disponível em:
<https://www.tandfonline.com/doi/full/10.1080/10618600.2017.1384734>. Acesso em: 30 ago. 2026.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From data mining to knowledge discovery in databases. **AI Magazine**, [S. l.], v. 17, n. 3, p. 37–54, 1996. DOI: 10.1609/aimag.v17i3.1230. Disponível em:
<https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1230>. Acesso em: 30 ago. 2026.

GARINGAN, Dominique; PICKARD, Alison Jane. Artificial intelligence in legal practice: Exploring theoretical frameworks for algorithmic literacy in the legal information profession. **Legal Information Management**, Cambridge, v. 21, n. 2, p. 97–117, 2021. Disponível em:
<https://doi.org/10.1017/S1472669621000190>. Acesso em: 30 ago. 2026.

KITCHIN, Rob. **The data revolution: big data, open data, data infrastructures and their consequences**. London: Sage, 2014.

KOLTAY, Tibor. Data literacy: In search of a name and identity. **Journal of Documentation**, Bingley, v. 71, n.2, p. 401–415, 2015. Disponível em: DOI <https://doi.org/10.1108/JD-02-2014-0026>. Acesso em: 30 ago. 2026.

KOLTAY, Tibor. Data literacy for researchers and data librarians. **Journal of Librarianship and Information Science**, London, v. 49, n. 1, p. 3–14, 2017. Disponível em:
<https://doi.org/10.1177/0961000615616450>. Acesso em: 30 ago. 2026.

LUO, Jiebei; TANG, Rong. Data competency for academic librarians: Evaluating present trends and future prospects. **The Journal of Academic Librarianship**, v. 50, n. 4, p. 102897, 2024. Disponível em: <https://doi.org/10.1016/j.acalib.2024.102897>. Acesso em: 30 ago. 2026.

MATOS, Maurício Torres de; CANDURÚ, Marise Teles; BENCHIMOL, Alegria Célia. A produção científica e o acesso aberto sobre a Ciência de Dados no contexto da Ciência da Informação: estudo. **Páginas a&b: arquivos e bibliotecas**, [S.l.], 3. série, p. 247–248, 2022. Número especial.

Trabalho apresentado na 12º ConfOA, 2021, [Braga]. Disponível em: <https://doi.org/10.21747/21836671/pagnespp15>. Acesso em: 30 ago. 2026.

MATUK, Camillia; VACCA, Ralph; AMATO, Anna; SILANDER, Megan; DESPORTES, Kayla; WOODS, Peter J.; TES, Marian. Promoting students' informal inferential reasoning through arts-integrated data literacy education. **Information and Learning Sciences**, [S. l.], v. 125, n. 3/4, p. 163-189, 2024. Disponível em: <https://doi.org/10.1108/ILS-07-2023-0088>. Acesso em: 30 ago. 2026.

NAUR, Peter. **Concise survey of computer methods**. 3. ed. Lund: Studentlitteratur, 1974.

O'NEIL, Cathy; SCHUTT, Rachel. **Doing Data Science: straight talk from the frontline**. 3. ed. Sebastopol: O'Reilly Media, 2014. Disponível em: https://books.google.com.br/books?hl=pt-BR&lr=&id=ycNKAQAAQBAJ&oi=fnd&pg=PP1&dq=%22data+science%22+Rachel+Schutt&ots=-7D8CQgtwt&sig=dZRUm6bwcW7_xPqnjf7weloZDrw#v=onepage&q&f=false. Acesso em: 30 ago. 2026.

SANDER, Ina. Critical datafication literacy: a framework for educating about datafication. **Information and Learning Sciences**, [S. l.], v. 125, n. 3/4, p. 270-292, 2024. Disponível em: <http://dx.doi.org/10.1108/ILS-06-2023-0064>. Acesso em: 30 ago. 2026.

SHEARER, Colin. The CRISP-DM model: the new blueprint for data mining. **Journal of Data Warehousing**, [S. l.], v. 5, n. 4, p. 13-22, 2000. Disponível em: https://www.academia.edu/42079490/CRISP_DM_The_New_Blueprint_for_Data_Mining_Colin_Shearer_Fall_2000_. Acesso em: 30 ago. 2026.

WINCKELMANN Predictive algorithms and racial bias: a qualitative descriptive study on the perceptions of algorithm accuracy in higher education. **Information and Learning Sciences**, [S. l.], v. 124, n. 9/10, p. 349-371, 2023. Disponível em: <https://doi.org/10.1108/ILS-05-2023-00452023>. Acesso em: 30 ago. 2026.

NOTAS E CRÉDITOS DO ARTIGO

Reconhecimento: não se aplica.

Financiamento: não se aplica.

Conflitos de interesse: não se aplica.

Aprovação ética: Não se aplica.

Manuscrito publicado como preprint: não se aplica.

Contribuições dos autores:

Contribuição	Maia, C. M.	Silva, J. M. S. da	Santarem Segundo, J. E
Concepção do estudo	X		
Conceitualização	X		X
Metodologia	X	X	X
Coleta de dados / investigação	X	X	
Curadoria de dados	X		
Análise dos dados	X	X	
Discussão dos resultados	X		
Visualização (gráficos, tabelas e outros)	X	X	
Rascunho original	X	X	
Revisão e edição final	X		X
Supervisão e administração	X		

Publisher

Universidade Federal de Lavras (UFLA). As ideias expressas neste artigo são de responsabilidade de sua autoria, não representando, necessariamente, a opinião dos editores ou da universidade.