

## Ciência de dados: uma análise de conteúdo da produção científica em Ciência da Informação

Leandro da Conceição Borges  
leandrocb@ufmg.com.br

Mauricio Barcellos Almeida  
mba@eci.ufmg.br

### Resumo

A Ciência de dados apresenta o seu prólogo nas interseções entre a Estatística, Matemática e a Ciência da computação, sendo, na contemporaneidade, uma área que também está presente em instituições do consórcio *iCaucus* de escolas de informação, ou simplesmente *ischools*. Caracterizado como um estudo descritivo acerca dos seus objetivos e exploratório sobre os seus procedimentos técnicos, pautou-se na abordagem qualitativa e quantitativa para levantar características dos textos científicos indexados na Base de Dados em Ciência da Informação, BRAPCI, para verificar como está delineado o estado da arte quanto à produção científica sobre Ciência de dados em uma base nacional na área da informação. Alicerçado na técnica de análise de conteúdo, levantaram-se características de 31 textos científicos indexados na base. Os assuntos levantados nos textos apontaram para o uso de ferramentas tecnológicas, profissionalização, currículo, interseções de áreas, produção científica, metadados, paradigmas e organização da informação. Concluiu-se que, assim como em algumas produções internacionais, a Ciência de dados é bem-vista na área da informação e a sua inserção, como agregadora, pode trazer resultados significativos para ambas. No entanto, diferentemente do cenário internacional, não há nenhum curso em Ciência de dados no Brasil associado às escolas de Ciência da informação.

**Palavras-chave:** Ciência de dados; ciência da informação; produção científica; análise de conteúdo.

## *Data science: a content analysis of scientific production in Information Science*

### Abstract

*Data Science has its prologue in the intersections between Statistics, Mathematics, and Computer Science, and nowadays it is an area that is also present in institutions of the iCaucus consortium of information schools, or simply ischools. Characterized as a descriptive study in terms of its objectives and exploratory in terms of its technical procedures, it was based on a qualitative and quantitative approach to survey the characteristics of the scientific texts indexed in the Base de Dados em Ciência da Informação, BRAPCI, to see how the state of the art in terms of scientific production on Data Science in a national database in the area of information is delineated. The characteristics of 31 scientific texts indexed in the database were surveyed based on the content analysis technique. The subjects raised in the texts pointed to the use of technological tools, professionalization, curriculum, intersections of areas, scientific production, metadata, paradigms, and organization of information. It was concluded that, as in some international productions, Data Science is well regarded in the information field and its inclusion as an aggregator can bring significant results for both. However, unlike the international scenario, there are no data science courses in Brazil associated with information science schools.*

**Keywords:** data science; information science; scientific production; content analysis.



## 1 INTRODUÇÃO

Silva Segundo e Araújo (2019, p. 118) sustentam que os “dados estão por toda parte, e nós produzimos e consumimos sem perceber”. Artefatos como um *smartphone*, por exemplo, produzem uma imensidão de dados, onde somos constantemente “[...] vigiados e percebidos” (Silva Segundo; Araújo, 2019, p. 119). Logo, em uma sociedade cada vez mais orientada por dados, a Ciência de dados nasce como uma disciplina do conhecimento norteadora para a adequação desta questão. Nesse sentido, programas em Ciência de dados tiveram um crescimento significativo para atender à emergência da demanda por profissionais qualificados em postos de trabalho (Zhang; Zeng, 2022), assim como adequações curriculares no contexto acadêmico na contemporaneidade.

Corroborando este olhar, Ortiz-Repiso, Greenberg e Calzada-Prado (2018), em um levantamento realizado em 65 instituições associadas ao consórcio *iCaucus* das *Information Schools*, ou simplesmente *ischools*<sup>1</sup>, constatou que a maioria dessas organizações oferece algum tipo de ensinamento relacionados a dados, principalmente em cursos de mestrado e de graduação.

Partindo desta premissa, o objetivo deste estudo está em investigar o que as produções científicas indexadas na Base de Dados em Ciência da Informação (BRAPCI), uma plataforma brasileira dedicada à coleta, preservação e acesso à informação científica sobre a Ciência da informação, abordam a Ciência de dados em produções indexadas a mesma. Como justificativa ao estudo, produções no cenário internacional têm apresentado pesquisas com fortes ligações da Ciência de dados à Ciência da informação, onde as *ischool* (Zhang *et al.*, 2022; Ortiz-Repiso, Greenberg e Calzada-Prado, 2018) têm ajudado a catapultar cada vez mais o ensino desta área do conhecimento nas modalidades graduação, pós-graduação e especialização. Portanto, a motivação desta pesquisa partiu após verificar a constância dessas ligações de áreas em estudos científicos internacionais, algo que pode ser benéfico no fomento de estudos científicos similares no Brasil.

A pesquisa está dividida em: “Ciência de dados: breve contexto”, onde apresenta um histórico desta área do conhecimento de forma sintética; “Panorama da Ciência de dados no Brasil”, em que, por meio da plataforma e-Mec, do Ministério da Educação, buscou-se entender o delineamento da oferta de cursos de graduação desta área do conhecimento no Brasil; “Metodologia”, com a descrição do passo a passo para se alcançar a “Análise e Discussões dos resultados” propostos; “Considerações finais”, com as apresentações dos ensinamentos proeminentes do estudo e, por fim, as referências utilizadas para o embasamento do estudo.

## 2 CIÊNCIA DE DADOS: BREVE CONTEXTO

Para entender melhor este fenômeno, Morettin e Singer (2020) indicam que muitas instituições internacionais passaram a criar programas sobre Ciência de dados, sendo a sua forma elementar em *Master in Business Administration* (MBA) e conseguinte em cursos de mestrado. Hoje, existem também cursos de graduação e doutorado na área. Inicialmente, estes cursos foram incluídos em escolas de Engenharia e Economia (Morettin; Singer, 2020), mas atualmente é possível verificar a sua ramificação em outras escolas, como as de informação.

O ponto levantado pelos autores é que a Ciência de dados não é uma área nova, pois envolve técnicas ligadas à área da Estatística, com o auxílio da Ciência da computação, sendo muito dos seus modelos e conceitos apresentados como disponíveis nos dias de hoje, ou mesmo foram esquecidos na literatura há décadas (Morettin; Singer, 2020). O uso desses modelos

---

<sup>1</sup> São instituições pertencentes a um consórcio internacional que visam apresentar a interseção entre a informação, a tecnologia e a sociedade, com associações com outras áreas do conhecimento e a cultura. As *ischools* são tidas como um fenômeno recente, portanto, fomentam discussões sobre a sua legitimidade na contemporaneidade (Wu *et al.*, 2011). Borges e Oliveira (2011) informam que a institucionalização das *ischools* ocorreu em 2005, estando as mesmas presentes em cinco continentes pelo mundo (Américas, Ásia, África, Europa e Oceania), compondo atualmente mais de 100 instituições associadas ao *iCaucus*.

estatísticos não era aplicado, até então, por conta das limitações impostas pela computação. Alguns exemplos que poderiam ilustrar a afirmação são o modelo em árvore, utilizado na década de 1980, e os algoritmos de suporte vetorial, que são teorias presentes na Ciência de dados, mas são oriundas da Estatística (Morettin; Singer, 2022). Urs e Minhaj (2022) corroboram a essa visão inicial e incluem nesse rol de áreas a Tecnologia da Informação e o *Business* (Negócios), além de instituições como as *ischools* para a popularização da área atualmente. Já Carvalho, Menezes e Bonidia (2024) acrescentam a Matemática como uma área que é elementar para a Ciência de dados, junto da Estatística e da Ciência da computação.

Alguns pontos, creditado por Urs e Minhaj (2022), sobre a Ciência de dados precisam ser destacados: 1 – existem discussões entre os pares sobre a origem do campo, onde há uma ala que acredita ser uma área germinada na Estatística, mas há discordâncias, com afirmações de que a mesma é um termo que foi apropriado pela Ciência da computação. 2 - Jeff Wu é tido como o primeiro pesquisador a utilizar o termo “Ciência de dados” sendo um nome alternativo à Estatística, isto em 1985. 3 - Pesquisadores como Leo Breiman, John Chambers e Willian Swain Cleveland eram entusiastas que a Estatística tivesse um horizonte amplo, indo além do aspecto clássico e teórico (Urs; Minhaj, 2022).

Portanto, surgiu uma ênfase mais aprofundada na preparação e apresentação dos dados, indo além da modelagem estatística, assim como a valorização da previsão em vez da inferência. Logo, Cleveland sugeriu uma nova área do conhecimento, oriunda da Estatística, com o *modus operandi* da Ciência da computação, inserindo aspectos da mineração de dados, chamada Ciência de dados (Urs; Minhaj, 2022).

Em linhas gerais, pode-se dizer que a jornada da Ciência de dados começou com a ideia de processamento de dados, evoluindo para a descoberta de conhecimento em bancos de dados e o uso de ontologias para melhorar a extração de informações da web semântica. A análise de dados é vista como a pedra fundamental do campo, envolvendo a mineração de dados, a descoberta de conhecimento, o aprendizado de máquina e a análise descritiva. O crescimento exponencial dos dados é um desafio para todas as áreas do conhecimento e, também, para as organizações, levando-as a explorarem o potencial dos dados para obter *insights* valiosos (Urs; Minhaj, 2022).

Dessa forma, buscando o olhar para a Ciência da informação, Matos, Condurú e Benchimol (2022) situam a Ciência de dados como uma disciplina que surge no início da década dos anos 2000, cuja inclinação está na análise de dados digitais, em um contexto crescente de robustez dos dados, que afetam diferentes ambientes organizacionais e acadêmicos. Em linhas gerais, segundo os autores, a Ciência de dados possui como premissa “[...] a criação de técnicas, tecnologias e teorias para a utilização de dados provenientes de fontes heterogêneas e complexas” (Matos; Condurú; Benchimol, 2022, p. 2). Tal fato é importante para criação de mecanismos para tomada de decisão, além de agregar valor à informação.

Nesse sentido, pode-se considerar a Ciência de dados como uma disciplina interdisciplinar, cuja justificativa de sua existência está na preocupação com os dados, “[...] desde a captura e representação, passando pelo armazenamento, segurança, análise e disseminação, com forte componente tecnológico” (Matos; Condurú; Benchimol, 2022, p. 19). Nesse sentido, existem semelhanças entre os objetivos da Ciência da informação e da Ciência de dados, sendo a Ciência da computação um balizador, pois esta área do conhecimento seria a responsável por disponibilizar os aportes tecnológicos. Quando juntas essas disciplinas ajudariam no “[...] crescimento da produção científica [...], bem como possibilidades de atuação dos cientistas da informação em atividades relacionadas à [Ciência de dados]” (Matos; Condurú; Benchimol, 2022, p. 19). Com este mote, o cenário dos cursos de graduação em Ciência de dados no Brasil será apresentado na seção a seguir.

### 3 PANORAMA CIÊNCIA DE DADOS NO BRASIL

O mapeamento aqui demonstrado foi realizado durante o mês de janeiro de 2024 na base e-Mec<sup>2</sup>. O objetivo deste levantamento consiste em verificar como está delineado o cenário dos cursos de graduação em Ciência de dados no Brasil. O olhar pelos cursos de graduação ocorreu pelos mesmos já estarem em funcionamento no país, quando comparados aos cursos de pós-graduação e, também, por apresentarem um número maior de cursos.

Em buscas na Plataforma Sucupira<sup>3</sup>, não se obteve nenhum resultado de cursos *stricto sensu*<sup>4</sup>, ainda que, em buscas no *Google* por programas de pós-graduação na área, foram obtidos resultados de instituições como a Pontifícia Universidade Católica de Campinas, Instituto Federal de São Paulo, Pontifícia Universidade Católica do Rio de Janeiro, Estácio, XP, Cruzeiro do Sul, na modalidade *lato sensu*<sup>5</sup>, corroborando Morettin e Singer (2020).

Cabe também a menção de que existem iniciativas *stricto sensu* no Brasil, como os cursos oferecidos pela Fundação Getúlio Vargas<sup>6</sup>, Universidade Federal da Bahia<sup>7</sup>, Universidade de São Paulo<sup>8</sup> e Universidade de Fortaleza<sup>9</sup>. Ainda que estes programas sejam projetados para capacitar profissionais e pesquisadores para atuarem em áreas que demandam análise intensiva de dados e desenvolvimento de soluções inovadoras, o ponto aqui levantado é que nenhum curso apresenta na sua nomenclatura o nome “Ciência de dados” unicamente. Neste caso, há sempre uma “disputa” com outras áreas do conhecimento em seus títulos ou ênfase, mas não somente em Ciência de dados.

A busca na base e-MEC consistiu primeiramente na aba “Consulta textual” da plataforma, na opção cursos de graduação, e pesquisou-se pelo nome “Ciência de dados”, conforme indicação da Figura 1.

Figura 1 – Busca por cursos de Ciência de dados no Brasil

The image shows the e-MEC website interface. At the top, there are navigation tabs: e-MEC, Consultar Cadastro, Documentos de Apoio ao Sistema, Inscrição para BASIS, Regulação / Avaliação, Declaração de Regularidade de Curso, and Perguntas Frequentes. Below this is the header 'Cadastro Nacional de Cursos e Instituições de Educação Superior' and 'Cadastro e-MEC'. A banner for 'Ação Premiada 14º Concurso Inovação na Gestão Pública Federal' is visible. The main content area contains a disclaimer about the data's accuracy and a search interface. The search interface has three tabs: 'Consulta Avançada', 'Consulta Textual' (selected), and 'IES Extintas'. A search box labeled 'Nome do Curso' contains the text 'ciência de dados'. A 'Pesquisar' button is located to the right of the search box.

Fonte: e-Mec (2024).

Foram recuperados 73 cursos, sendo que alguns ainda não estão iniciados, e outros estão extintos, conforme apresentação do Quadro 1.

<sup>2</sup> Base mantida pelo Ministério da Educação do Brasil. Disponível em: <https://emec.mec.gov.br/emec/nova>. Acesso em: 18 jan. 2024.

<sup>3</sup> Disponível em: <https://sucupira-v2.capes.gov.br/sucupira4/>. Acesso em: 12 jan. 2024.

<sup>4</sup> Mestrado e Doutorado.

<sup>5</sup> Especialização.

<sup>6</sup> Disponível em: <https://emap.fgv.br/curso/doutorado>; Acesso em: 12 set. 2024.

<sup>7</sup> Disponível em: <https://www.conre3.org.br/portal/ufba-tera-primeiros-programas-de-mestrado-e-doutorado-em-estatistica-e-ciencia-de-dados-do-brasil/>. Acesso em: 12 set. 2024.

<sup>8</sup> Disponível em: <https://saocarlos.usp.br/pos-graduacao-gratuita-na-usp-mestrado-profissional-com-enfase-em-ciencia-de-dados/>. Acesso em: 12 set. 2024.

<sup>9</sup> Disponível em: <https://unifor.br/web/pos-graduacao/-/ciencia-de-dados-e-inteligencia-artificial-conheca-a-nova-linha-de-pesquisa-oferecida-pela-unifor>. Acesso em: 12 set. 2024.

**Quadro 1** – Panorama dos cursos de graduação em Ciência de dados no Brasil

Estado	Nome	Modalidade	Grau	Situação	Instituição
São Paulo	Universidade de São Paulo	Presencial	Bacharelado	01/01/2009	Pública
Brasília	Centro Universitário do Instituto de Educação Superior de Brasília	Presencial	Bacharelado	26/02/2018	Privada
São Paulo	Universidade Cruzeiro do Sul	À distância	Tecnológico	01/02/2019	Privada
São Paulo	Universidade Católica de Santos	À distância	Tecnológico	01/02/2019	Privada
São Paulo	Universidade Cidade de São Paulo	À distância	Tecnológico	01/02/2019	Privada
São Paulo	Universidade de Franca	À distância	Tecnológico	01/02/2019	Privada
Rio Grande do Sul	Centro Universitário da Serra Gaúcha	À distância	Tecnológico	01/02/2019	Privada
Paraná	Centro Universitário Dom Bosco	À distância	Bacharelado	23/04/2019	Privada
Santa Catarina	Centro Universitário Estácio de Sá	À distância	Tecnológico	26/08/2019	Privada
São Paulo	Centro Universitário Estácio Ribeirão Preto	À distância	Tecnológico	01/10/2019	Privada
Espírito Santo	Universidade Vila Velha	À distância	Tecnológico	03/02/2020	Privada
São Paulo	Centro Universitário das Américas	Presencial	Bacharelado	07/02/2020	Privada
Rio Grande do Sul	Centro Universitário Ritter dos Reis	À distância	Bacharelado	10/02/2020	Privada
São Paulo	Universidade Anhembi Morumbi	Presencial	Bacharelado	10/02/2020	Privada
São Paulo	Centro Universitário Braz Cubas	À distância	Tecnológico	10/02/2020	Privada
Paraná	Centro Universitário Internacional	À distância	Tecnológico	10/02/2020	Privada
Rio de Janeiro	Fundação Getúlio Vargas	Presencial	Bacharelado	10/02/2020	Privada
São Paulo	Universidade Nove de Julho	À distância	Tecnológico	11/02/2020	Privada
São Paulo	Pontifícia Universidade Católica de São Paulo	Presencial	Bacharelado	17/02/2020	Privada
Santa Catarina	Universidade do Vale do Itajaí	À distância	Tecnológico	02/03/2020	Privada

Rio de Janeiro	Universidade Estácio de Sá	À distância	Tecnológico	16/03/2020	Privada
Paraíba	Universidade Federal da Paraíba	Presencial	Bacharelado	08/06/2020	Pública
Paraná	Universidade Positivo	À distância	Tecnológico	08/08/2020	Privada
São Paulo	Faculdade de Tecnologia de Santana de Parnaíba	Presencial	Tecnológico	31/08/2020	Privada
Pernambuco	Centro Universitário Joaquim Nabuco de Recife	À distância	Tecnológico	16/09/2020	Privada
Mato Grosso do Sul	Universidade Anhanguera	À distância	Tecnológico	08/02/2021	Privada
Paraíba	Centro Universitário de João Pessoa	Presencial	Tecnológico	22/02/2021	Privada
São Paulo	Pontifícia Universidade Católica de Campinas	Presencial	Bacharelado	22/02/2021	Privada
São Paulo	Universidade de Sorocaba	Presencial	Bacharelado	25/02/2021	Privada
Rio Grande do Sul	Pontifícia Universidade Católica do Rio Grande do Sul	Presencial	Bacharelado	03/03/2021	Privada
Ceará	Universidade Federal do Ceará	À distância	Tecnológico	16/07/2021	Pública
Mato Grosso do Sul	Centro Universitário Anhanguera Pitágoras	À distância	Tecnológico	02/08/2021	Privada
Rio de Janeiro	Centro Universitário IBMEC	Presencial	Bacharelado	02/08/2021	Privada
Rio de Janeiro	Centro Universitário Anhanguera Pitágoras	À distância	Tecnológico	10/08/2021	Privada
São Paulo	Centro Universitário Anhanguera Pitágoras	À distância	Tecnológico	27/08/2021	Privada
Paraíba	Centro Universitário de João Pessoa	À distância	Tecnológico	01/09/2021	Privada
São Paulo	Centro Universitário Fundação Santo André	Presencial	Bacharelado	14/02/2022	Privada
Minas Gerais	Pontifícia Universidade Católica de Minas Gerais	Presencial	Bacharelado	21/02/2022	Privada
Rio Grande do Sul	Faculdade Senac de Porto Alegre	À distância	Presencial	07/03/2022	Privada

Brasília	Centro Universitário de Brasília	Presencial	Bacharelado	07/03/2022	Privada
Paraná	FAE Centro Universitário	Presencial	Bacharelado	07/03/2022	Privada
Brasília	Faculdade Capital Federal	Presencial	Tecnológico	10/03/2022	Privada
São Paulo	Universidade Anhembi Morumbi	Presencial	Bacharelado	18/03/2022	Privada
<b>CURSOS EXTINTOS OU NÃO INICIADOS</b>					
Paraná	Pontifícia Universidade Católica do Paraná	À distância	Tecnológico	Extinto	Privada
Rio Grande do Sul	Faculdade de Tecnologia Senac Pelotas	Presencial	Tecnológico	Extinto	Privada
Santa Catarina	Universidade do Oeste de Santa Catarina	Presencial	Tecnológico	Extinto	Privada
São Paulo	Universidade de Araraquara	Presencial	Tecnológico	Extinto	Privada
São Paulo	Universidade Presbiteriana Mackenzie	À distância	Tecnológico	Não iniciado	Privada
São Paulo	Faculdade de Tecnologia Rubens Lara	Presencial	Tecnológico	Não iniciado	Privada
São Paulo	Universidade de São Paulo	Presencial	Bacharelado	Não iniciado	Pública
Santa Catarina	Universidade da Região de Joinville	À distância	Tecnológico	Não iniciado	Privada
Minas Gerais	Universidade de Uberaba	À distância	Tecnológico	Não iniciado	Privada
São Paulo	Faculdade de Tecnologia de Ourinhos	Presencial	Tecnológico	Não iniciado	Pública
Brasília	Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa de Brasília	Presencial	Bacharelado	Não iniciado	Privada
Paraná	Centro de Ensino, Ciência, Tecnologia do Paraná	À distância	Tecnológico	Não iniciado	Privada
Rio de Janeiro	Instituto INFET	À distância	Tecnológico	Não iniciado	Privada

São Paulo	Fundação Universidade Virtual do Estado de São Paulo	À distância	Bacharelado	Não iniciado	Pública
Paraná	Faculdade Integrada CESUMAR de Curitiba	Presencial	Bacharelado	Não iniciado	Privada
Minas Gerais	Faculdade XP Educação	À distância	Bacharelado	Não iniciado	Privada
Mato Grosso do Sul	Faculdade CESUMAR	À distância	Bacharelado	Não iniciado	Privada
São Paulo	Faculdade de Tecnologia de Adamantina	Presencial	Tecnológico	Não iniciado	Pública
Rio de Janeiro	Instituto INFET	Presencial	Tecnológico	Não iniciado	Privada
São Paulo	Universidade de Sorocaba	À distância	Bacharelado	Não iniciado	Privada
Rio Grande do Sul	Universidade de Santa Cruz do Sul	Presencial	Tecnológico	Não iniciado	Privada
São Paulo	Universidade Santa Cecília	Presencial	Tecnológico	Não iniciado	Privada
Rio de Janeiro	Centro Universitário IBMEC	Presencial	Bacharelado	Não iniciado	Privada
Bahia	Centro Universitário SENAI CIMATEC	À distância	Bacharelado	Não iniciado	Privada
Paraíba	Universidade Federal da Paraíba	Presencial	Bacharelado	Não iniciado	Pública
Rio Grande do Sul	Universidade La Salle	À distância	Tecnológico	Não iniciado	Privada

Fonte: e-Mec (2024).

Explicando o Quadro 1, de forma geral, 43 instituições possuem cursos ativos, quatro estão extintos, 22 não iniciados e quatro com repetições pertencentes à Universidade do Oeste de Santa Catarina. Sessenta e uma instituições são classificadas como privadas, sendo quatro com cursos extintos, e 17 com cursos não iniciados. Portanto, a contagem final dos cursos ativos ligados a instituições privadas são 40. Em relação às instituições públicas, são contabilizados oito cursos, sendo cinco não iniciados e três em atividade. Vinte e três cursos são à distância, e 20 presenciais.

Os cursos são divididos em bacharelados e tecnológicos. Segundo Bondan e Bardagi (2008, p. 583), os cursos tecnológicos “[...] foram criados para responder às novas tendências do mercado que cada vez mais exige profissionais com perfil diferenciado, dotado de competências e habilidades específicas”. Os formandos desta modalidade de ensino estão aptos a ingressarem em cursos de especialização, mestrado e doutorado (Bondan; Bardagi, 2008). Dos cursos listados, 28 são classificados como bacharelado, mas dez ainda não foram iniciados. Nesse sentido, 18 estão ativos. Quarenta e um são tecnológicos, sendo 12 não iniciados, e quatro extintos. Vinte e cinco estão ativos.

O primeiro curso credenciado é o da Universidade de São Paulo, no ano de 2009, indo ao encontro da afirmação de Matos, Condurú e Benchimol (2022), e com isso, pode-se inferir que, no Brasil, o ensino em Ciência de dados ainda é incipiente. O curso da Universidade de São



Paulo é associado ao curso de Estatística da instituição. Ainda consta na planilha oferecida pelo e-Mec que a instituição tem um curso de Ciência de dados que ainda não foi iniciado.

O IBMEC apresenta dois cursos: um iniciado e outro não, assim como o INFNET, que tem dois cursos não iniciados, sendo um presencial e um à distância. A Universidade Federal da Paraíba apresenta dois cursos, sendo um iniciado, e o outro não iniciado. Por fim, o Quadro 2 apresenta o percentual de cursos ativos por estados do Brasil.

**Quadro 2** – Percentual de cursos ativos de Ciência de dados ativos no Brasil

Estado	Quantitativo	Percentual
Brasília	3	7%
Ceará	1	2%
Espírito Santo	1	2%
Mato Grosso do Sul	2	5%
Minas Gerais	1	2%
Paraíba	3	7%
Paraná	4	9%
Pernambuco	1	2%
Rio de Janeiro	4	9%
Rio Grande do Sul	4	9%
Santa Catarina	2	5%
São Paulo	17	41%
<b>Total</b>	<b>43</b>	<b>100%</b>

Fonte: dados de pesquisa (2024).

O Quadro 2 indica que o estado de São Paulo é o que possui o maior número de cursos ativos de Ciência de dados no Brasil. Os demais estados compõem percentuais menores quando comparados com São Paulo. Os cursos levantados apresentam ênfase geral em Ciência de dados, inteligência analítica, inteligência artificial, *machine learning*, dados para negócios e Estatística. Cabe lembrar que a Universidade Federal de Minas Gerais irá oferecer um curso de bacharelado em Ciência de dados em 2024. No entanto, a informação ainda não consta na base de dados do Ministério da Educação. A Universidade Federal do ABC também irá oferecer a modalidade bacharelado em Ciência de dados em 2025. Por fim, é importante a menção de que nenhum desses cursos possuem associações com escolas de informação, diferente do cenário internacional.

#### 4 METODOLOGIA

Trata-se de uma pesquisa caracterizada acerca dos seus objetivos como descritiva, e exploratória sobre os seus procedimentos técnicos (Gil, 2008). Em relação à abordagem da pesquisa, pode ser classificada como qualitativa e quantitativa, pois serão levantadas características dos textos científicos e o número de produções.

Para a resolução dos aspectos qualitativos presentes nas produções levantadas, aplicou-se a análise de conteúdo de Bardin (2011), que consiste em um conjunto de técnicas para realizar uma arguição, levando em consideração os procedimentos, meios sistemáticos e objetivos para a descrição das mensagens, utilizando indicadores, quantitativos ou não, para o levantamento de impressões sobre a produção e recepção dessas mensagens (Bardin, 2011).

Portanto, primeiramente prepararam-se os materiais a serem analisados, e foram realizadas a codificação por meio da análise temática, a categorização e a interpretação desses dados.

O *corpus* da pesquisa serão os documentos recuperados da base BRAPCI, aplicando, nesse sentido, a técnica de pesquisa documental. Os resultados foram tabulados no *software* Microsoft Excel.

Foram realizadas quatro buscas, pois a realização de estratégias mais elaboradas não trouxe resultados significativos. A primeira busca foi com o termo em português: “Ciência de dados”, que recuperou 48 resultados significativos. A segunda estratégia foi com o termo em inglês “*Data Science*”, que recuperou 45 resultados. A terceira estratégia foi a associação dos dois termos, utilizando o operador booleano AND, ou seja, “*Data Science*” AND “Ciência de dados”, que recuperou 28 resultados. E, por fim, a quarta estratégia utilizou o operador booleano OR com os dois termos, “*Data Science*” OR “Ciência de dados”, que recuperou cinco resultados.

Consideraram-se os termos em português e em inglês por serem mais usuais quando se pesquisa sobre Ciência de dados no cenário nacional, com a delimitação temporal (2014-2024), ou seja, dez anos, e a tipologia artigo. Cabe lembrar que antes buscou-se auxílio no Tesouro Brasileiro de Ciência da Informação<sup>10</sup> e no dicionário de Biblioteconomia e Arquivologia (Cunha; Cavalcanti, 2008), e não foram encontrados termos similares para Ciência de dados.

Todos os dados foram dispostos em uma planilha em Excel, cujos campos adicionais para entender a informação foram: ano, tipologia, periódico, país de publicação, assunto e nome do trabalho publicado. Dos 48 resultados da primeira estratégia, nove eram duplicatas, portanto, 39 foram considerados aptos para a análise.

Da segunda, foram excluídos 29 resultados por se tratarem de duplicatas da primeira estratégia, sendo, portanto, considerados 16 resultados, que somados computam 55 resultados. Da terceira (28 resultados) e quarta estratégia (cinco resultados), todos eram duplicatas, e nenhum resultado foi aproveitado. Após outra análise, 24 resultados foram descartados por não se adequarem ao escopo da pesquisa (55-24=31). Nesse sentido, 31 documentos foram considerados aptos e os resultados serão dispostos a seguir.

## 5 ANÁLISE E DISCUSSÃO DOS RESULTADOS

No Quadro 3, será apresentada a categorização baseada na análise de conteúdo. Realizou-se a leitura dos textos, fazendo-se sobressair os principais aspectos do seu conteúdo. Todo o levantamento considerou aspectos que definem a essência dos textos. Salienta-se que, em um texto, poderia ter mais de um assunto. Por este motivo, algumas referências serão apresentadas mais de uma vez, mas não levando em consideração a categoria, que não é repetitiva.

Quadro 3 - Categorias

Categorias	Subcategoria1	Subcategoria2
Tecnologia	Desenvolvimento para tratar dados.	Quantidade massiva de dados
	Disseminação dos dados	
	Gestão e curadoria de dados	Habilidades e competência de uso de dados
	Representação da informação	Tesouro
	Ciclo de vida dos dados	Cibernética e <i>e-Science</i>
	Complexidade de implantação	Mudanças (culturais e tecnológicas)
	Práticas abertas	Impactos positivos

<sup>10</sup>Disponível em: [http://sitehistorico.ibict.br/publicacoes-e-institucionais/tesouro-brasileiro-de-ciencia-da-informacao-1/copy\\_of\\_TESAUROCOMPLETOFINALCOMCAPA24102014.pdf](http://sitehistorico.ibict.br/publicacoes-e-institucionais/tesouro-brasileiro-de-ciencia-da-informacao-1/copy_of_TESAUROCOMPLETOFINALCOMCAPA24102014.pdf). Acesso em: 13 jan. 2024.

	Uso e reuso de dados	Capacitação
<b>Profissionalização</b>	Alunos	Alfabetização de dados
	Profissionais	Alfabetização de dados
		Empresarial
		Jornalismo
		Ciência da informação
		Humanidades digitais
		Bibliotecário
		Cientista de dados
<b>Currículo</b>	Cursos de graduação	Biblioteconomia
		Arquivologia
	Cursos de mestrado	Biblioteconomia
<b>Interseção</b>	Áreas	Ciência de dados
		Ciência da informação
		<i>Big data</i>
		<i>e-Science</i>
		Ciência da computação
		Visualização de dados
		Biblioteconomia
		Mineração
		Jornalismo
		Alfabetização de dados
		Saúde
		Estatística
		Matemática
<b>Ferramentas de dados</b>	<i>Google Dataset Search</i>	
	<i>VosViewer</i>	
	Ciência de redes	
	Python	
	R	
<b>Produção científica</b>	Periódicos	Títulos
		Ciência da informação
		Temáticas em comum
		Técnicas e ferramentas
		Ferramentas e recuperação
		<i>Big data</i>
		Ciência de dados
	Interseção de áreas	
	Estudos europeus	
<b>Metadados</b>	Interoperabilidade	
	Competência em dados	
	Tratamento	
	Governança	
	Prática de dados	
	SKOS	
	<i>Bibframe</i>	
<b>Paradigma</b>	<i>Big data</i>	
	Tecnologia da informação	
	Metadado	
	Desenvolvimento científico	
<b>Organização da informação</b>	Taxonomia	

Fonte: dados de pesquisa (2024).

A categoria **Tecnologia** presente no estudo de Delbianco *et al.* (2023) enfatizam que o desenvolvimento e o uso tecnológico têm propiciado o surgimento de uma quantidade massiva de dados diariamente. Na visão de Pessoa, Sousa e Cardoso (2020), a tecnologia da informação contribuiu para que um novo paradigma surgisse, pautado na questão da disseminação dos dados que são apresentados nos ambientes informacionais.

Para Costal, Sales e Zattar (2020), as Tecnologias da Informação e Comunicação são propícias para a gestão, curadoria dados e a necessidade de desenvolvimento de habilidades e competências específicas para o seu uso. Partindo deste princípio, o estudo de Ramalho (2017), de forma preliminar, apresenta a proposta de representação do Tesouro Brasileiro de Ciência da Informação no formato *Simple Knowledge Organization System* (SKOS) por meio de trechos da categoria “Tecnologias da Informação e Comunicação”. Em linhas gerais, pode-se dizer que estas tecnologias possibilitaram que produtores e consumidores de informação pudessem interagir de forma mútua, ajudando na formulação de produtos ou serviços de informação (Mello Filho; Araújo Júnior, 2021). Com isso, a cibernética e a *e-Science*, por exemplo, têm representado um elo importante na dinamicidade e fluidez na composição do ciclo de vida dos dados (Reis; Sena, 2021).

A complexidade de implantação dos recursos propiciados pela gestão de dados está alinhada as mudanças culturais e tecnológicas nas práticas informacionais (Henning *et al.*, 2018). Nesse sentido, quando essas implicações estão de acordo com as práticas abertas, há a possibilidade de maior impacto de descobertas e o incentivo da inovação científica (Monteiro; Sant’ana, 2019), como é o caso do repositório institucional da Fundação Oswaldo Cruz, o ARCA (Pedroso; Lima; Assef Neto, 2017). Esta discussão traz luz para o uso e reuso de dados em diversos setores da sociedade e a capacitação dos seus profissionais (Costal; Sales; Zattar, 2020) para a atuação no mercado de trabalho.

Em relação à **profissionalização**, o estudo de Martín González e Iglesias Rodríguez (2021) apresenta alguns olhares sobre profissões no gerenciamento de dados. Nele, é proposto um treinamento de alunos de nível superior em alfabetização de dados, começando do zero, apontando a identificação, localização, visualização e recuperação de dados, levando em consideração a Ciência de dados e os dados abertos.

Delbianco *et al.* (2023) também trazem luz para alfabetização de dados e a capacitação de estudantes e profissionais da informação no contexto de uso e reuso de dados na prática científica e empresarial, visando a melhor compreensão e aplicação dos dados.

Já Antón Bravo e Serrano Tellería (2020) apontam o jornalismo de dados para tratar dos dados abertos e outras fontes de informações em dados, com o aporte de usabilidade, reuso e métodos científicos com o intuito de extração e descoberta de dados estruturados complexos. Tal questão também está entrelaçada com a alfabetização de dados, assim como a visualização de dados.

Para Mello Filho e Araújo Júnior (2021), os objetos de fronteira, que, segundo os autores, são artefatos que estão limítrofes ao contexto social e permitem a facilitação da comunicação com outras conjunturas, podem ser um campo de atuação dos formados em Ciência da informação.

Lemos, Martins e Souza (2023) apresentam o profissional da informação como elemento importante para a criação de terminologias para a estruturação da informação de forma crível, a partir de modelos semânticos para a representação da informação, em diferentes sistemas de informação, desembocado na área das Humanidades digitais.

Reis e Sena (2021) discutem a competência profissional do bibliotecário de universidades com o uso dos dados, sendo o mesmo um elo importante para a pesquisa científica.

Costal, Sales e Zattar (2020) visam apresentar as habilidades do prisma técnico e tecnológico que o bibliotecário precisa adquirir para desenvolver a competência em dados, ou seja, é necessário ter noções relacionadas aos padrões de metadados, o conhecimento de

recursos de dados, aprendizado e uso da linguagem de programação, dentre outras competências interdisciplinares da Ciência de dados.

Poderíamos definir que as profissões aqui apresentadas compõem o rol dos chamados “cientistas de dados”, independentemente do nicho mercadológico. Com a sua divulgação em 2008 e definição em 2012, a terminologia cientista de dados foi tida como uma das melhores profissões posicionadas no mercado de trabalho no século XXI, fazendo com que surgisse muitas demandas de estudo para a sua formação (Curty; Serafim, 2016).

Quanto ao **currículo**, na pesquisa de Delbianco *et al.* (2023), há a pretensão de se criar uma ementa mais bem fundamentada para a disciplina de Ciência de dados no Brasil, para os cursos de graduação em Biblioteconomia e Arquivologia. Foram levantadas 24 instituições e identificaram-se 37 cursos, cujas ementas de disciplinas poderiam incorporar a Ciência de dados. Os autores concluíram que há escassez de disciplinas voltadas para esse viés nas áreas analisadas e que, na realidade brasileira, seria mais assertivo incluir a alfabetização de dados nas disciplinas de Ciência de dados.

Costal, Sales e Zattar (2020), para levantar a competência em dados na formação acadêmica do bibliotecário, foram verificadas ementas dos cursos de mestrado em Biblioteconomia de instituições credenciadas à *American Library Association*. As autoras concluíram que existe interseção da competência em dados com a competência informacional, algo que pode contribuir para prática profissional do bibliotecário, de caráter técnico ou tecnológico.

Sobre a **interseção de áreas**, Matos, Condurú e Benchimol (2022) apontam como positiva a complacência da área com a Ciência da informação. Para os autores, isso ajuda no crescimento da área e, conseqüentemente, no aumento da produção científica de ambas.

Logo, Pessoa, Sousa e Cardoso (2020) acreditam que a Ciência de dados traz para a Ciência da informação uma problemática que envolve a questão dos dados e a sua disponibilidade em diversos suportes informacionais, pois se faz necessário o esforço em tratar, usar e reutilizar os dados abertos na perspectiva da necessidade dos seus usuários. Já na visão de Fernandes (2020), o levantamento da questão do desenvolvimento de métodos, ferramentas e práticas em Ciência de dados impactaria a Ciência da informação. O autor prossegue demonstrando que existem implicações de caráter epistemológico entre as áreas, que sinalizam para uma oportunidade de trocas entre as mesmas.

Mello Filho e Araújo Júnior (2021) demonstram que os assistentes virtuais, ou *chatbots*, são peças-chaves na integração comunicacional entre os profissionais da Ciência da informação com a Ciência de dados, uma vez que essas ferramentas de inteligência artificial para a criação dos assistentes virtuais são utilizadas para a organização da informação. Reis e Carvalho (2022) discutem a relação da Ciência de dados com a Ciência da informação, de modo a entender quais paradigmas da ciência estão envoltos nesta interseção, chegando até a *e-Science*, para estudar o comportamento do volume de dados disponibilizados pela web, que esbarra no quarto paradigma da ciência proposto por Thomas Kuhn.

Vale e Cristovão (2023) indicam que a Ciência de dados e a Ciência da informação possuem aspectos interdisciplinares que, quando se relaciona com a análise de rede de dados, podem ajudar a sociedade na resolução de problemas. Para Martínez Musiño (2021), as duas áreas são recentes, quando comparadas com outras áreas do conhecimento, e usufruem das Tecnologias da Informação e Comunicação para a análise do grande volume de dados e informações disponíveis. Portanto, são áreas que dispõem de um caráter intra, multi e transdisciplinar para a atuação no tratamento dos dados e informações em larga escala.

Buscando reunir e fundamentar questões interdisciplinares da Ciência da informação, Ciência de dados e da Ciência da computação, Lemos, Martins e Souza (2023) apresentam, através de um modelo teórico-conceitual, uma proposta para contribuir na representação da informação com qualidade para a produção de conhecimento cultural em rede. Analisando a formação em Ciência de dados nos Estados Unidos, Curty e Serafim (2016) apontam interseções

com a Estatística, Matemática e a Computação, incluindo a programação e a modelagem avançada. Para este ponto, ao candidato que pretende entrar nesses cursos, é primordial a noção de funcionamento das mesmas, caracterizando-se como um pré-requisito ao aluno.

Reis e Sena (2021) discutem a Biblioteconomia de dados e a Ciência de dados na perspectiva da *e-Science*, com a finalidade de esclarecer o trabalho do bibliotecário de dados em bibliotecas universitárias, que, em linhas gerais, precisa de educação continuada para acompanhar os avanços propostos pela organização do conhecimento. Rautenberg e Carmo (2019), ao pesquisar o *Big data* e as ferramentas para a extração de informações para a tomada de decisão, indicam que, no processo de evolução da tríade dados, informação e conhecimento, surge uma nova disciplina para a atuação profissional, a Ciência de dados. Nesse sentido, os autores elucidam as diferenças entre *Big data* e Ciência de dados, onde a primeira está voltada para infraestrutura do *hardware* e de serviços de computação em nuvem, e a segunda apresenta maior alinhamento com os *softwares* que transformam os dados em informação, e isso tende a ajudar os tomadores de decisão nas suas práticas laborais.

Pinto *et al.* (2019) acreditam que a Ciência de dados, junto do *Big data*, é impulsionador de algumas práticas empresariais que envolvem o melhoramento de capacidades operacionais, estratégicas e no desempenho financeiro de corporação. Ainda assim, o *Big data* apresenta alguns desafios quanto ao seu uso. No entanto, a interseção das áreas é interessante para ajudar no desempenho da área da mineração, neste caso, a disciplina associada à Geologia.

Falcão *et al.* (2024), ao pesquisarem a área da mineração no Brasil, situam o *Big data* e como o seu desenvolvimento ajudou no surgimento da Ciência de dados. Melo e Rockembach (2019) analisam os conceitos e delimitações do *Big data* como uma área emergente que envolve as Tecnologias de Informação e Comunicação e ligações interdisciplinares, sendo este processo uma conexão com os profissionais da informação.

Para Silva (2019), a visualização de dados é uma subárea da Ciência de dados. A subárea foi desenvolvida com o tempo utilizando pesquisas com análise cognitivas básicas por meio de apresentações de evidências em informação. Antón Bravo e Serrano Tellería (2020), ao apresentarem o jornalismo de dados, indicam também a visualização de dados e a alfabetização de dados como áreas de interseção.

Por fim, Sobral, Lima e Sobral (2021) apontam a área da saúde interligada com a tecnologia, principalmente com aplicações de Ciência de dados, e sinalizam para a participação da Ciência da informação nesse processo.

Sobre as **ferramentas de dados**, Pinto e Amaral (2020) verificaram o *Google Dataset Search*, que se constitui em uma ferramenta lançada pelo *Google* em 2018, que busca identificar, indexar e disponibilizar na rede mundial de computadores, os *datasets*, que são os conjuntos de dados massivos. Além de contribuir para a produção científica na área de Ciência de dados, o estudo concluiu que a ferramenta é promissora quanto a sua usabilidade, alinhamento internacional de interoperabilidade de dados e metadados, além de que já existem repositórios brasileiros indexados pelo *Google Dataset Search*.

Moraes e Kafure (2020) apontam como as técnicas e as ferramentas de bibliometria e ciência de redes podem ser úteis no mapeamento de uma área do conhecimento ou assunto de pesquisa. Os resultados indicaram que a sumarização, visualização e análise de redes possibilitam combinar elementos para a compreensão da informação e do conhecimento na área estudada. As ferramentas utilizadas na pesquisa se mostraram eficazes na identificação das autorias e sobre o que estão discutindo a produção científica em questão. O estudo utilizou a ferramenta de visualização de dados *VosViewer*.

Por fim, Rodríguez Sánchez, Rivero González e Soca Lozano (2023), ao verificar a análise de sentimentos por meio de *tweets* das vacinas cubanas contra a covid-19, a Soberana 02 e a Abdala, utilizaram as linguagens de programação *Python* e *R*, com bibliotecas específicas para Ciência de dados. O estudo concluiu a predominância dos sentimentos de medo e confiança nas vacinas foram predominantes, sendo o primeiro um pouco abaixo da segunda.

Em relação à **produção científica**, Ribeiro (2020) verificou as abordagens do *Big data* e da Ciência aberta em periódicos de Ciência da informação e que a temática vem ganhando espaço na área. Ao realizar um estudo bibliométrico e altimétrico na base de dados *Dimensions*, Gontijo, Hamanaka e Araújo (2021) levantaram alguns títulos para a realização do acoplamento bibliográfico. Concluíram que, durante o recorte definido pela pesquisa, os títulos com mais artigos publicados sobre a temática gestão de dados científicos foram: *International Journal of Digital Curation*, *Journal of eScience Librarianship* e *Data Science Journal*.

Matos, Condurú e Benchimol (2022), ao analisarem a interseção da produção científica em Ciência de dados na Ciência da informação, sinalizam para temáticas em comum entre as áreas e o acesso livre das publicações ajudam na maior penetrabilidade desses títulos e, também, das instituições. Esta interação caracteriza-se como uma tendência de crescimento, permitindo uma conversa profícua entre as áreas. Sobral, Lima e Sobral (2021) verificaram a produção científica sobre hospitais relacionados à Ciência de dados na *Web of Science*. A pesquisa concluiu, dentre outras coisas, que os títulos de periódicos com maior relevância quanto à temática de estudo integram o grupo das editoras *Springer* e *Elsevier*.

Melo e Rockembach (2019) verificaram a Arquivologia e a Ciência da informação, com o viés do *Big data* e as suas implicações na atuação profissional em arquivos digitais. Levantaram a produção científica com a temática *Big data* e cientista de dados, visando verificar quais foram os artigos mais publicados. Por fim, refletem sobre o *Big data* e a sua importância para o trabalho dos profissionais da informação. Moraes e Kafure (2020) constataram que a utilização das técnicas bibliométricas e as ferramentas da ciência de redes foram eficazes na descoberta dos pesquisadores em Ciência de dados, assim como os assuntos que são debatidos na sua produção científica.

Bufrem *et al.* (2016), ao verificarem a produção internacional sobre a ciência orientada a dados, por meio da *e-Science* e da Ciência de dados, constataram que, de forma geral, os termos de maior destaque foram *distributed computer systems*, *grid computing* e *Big data*. Quando analisada na área da Biblioteconomia, as temáticas de maior ênfase foram: *digital library* e *open access*, indicando que há uma polarização da disponibilização do acesso à informação científica em meio digital.

Martínez Musiño (2021), por meio da informetria, verificou a interseção da Ciência de dados e da Ciência da informação em produções científicas nos idiomas inglês, italiano, francês, português e japonês. Dentre alguns pontos, constatou-se a natureza prática das duas áreas, sendo, segundo o autor, áreas emergentes, que se debruçam nas tecnologias de informação e comunicação para a análise de grandes volumes de dados.

O estudo de Pinto e Amaral (2020) busca somar com a produção científica na área de Ciência de dados, mais especificamente na parte de ferramentas de armazenamento e recuperação de dados *on-line*, uma vez que há escassez de produção desta natureza. O estudo situa o *Google Dataset Search*. Henning *et al.* (2018), ao apresentarem os princípios FAIR, citam estudos europeus que atestam a importância destes princípios em áreas do conhecimento, principalmente quando relacionados às necessidades e aos exemplos de aplicação.

Quanto aos **metadados**, Pinto e Amaral (2020), ao analisarem o *Google Dataset Search*, apontam para a não disponibilidade dos padrões de metadados, mesmo quando são adotados de forma igualitária pela plataforma. Em outras palavras, o ponto levantado pelos autores é a interoperabilidade dos padrões de metadados. Costal, Sales e Zattar (2020) levantam, na pesquisa realizada, os resultados apontados para as habilidades técnicas e tecnológicas que estão associadas aos padrões de metadados, como uma das competências em dados para a formação de bibliotecários.

Bufrem *et al.* (2016) trazem os metadados de 5.427 documentos recuperados na base de dados *Web of Science*, com as práticas de dados na Ciência da informação. Rautenberg e Carmo (2019) informam que uma das contribuições da Ciência de dados pode estar no

tratamento dos dados e metadados, assim como a sua exploração para servir de subsídios para a produção de informação precisa no processo decisório.

Pedroso, Lima e Assef Neto (2017) acreditam que a governança dos metadados do repositório institucional da Fundação Oswaldo Cruz, juntamente com a classificação e recuperação dos objetos digitais nele depositados, está associada a uma complexidade institucional e metodológica, uma vez que a sua produção científica é diversa. Ramalho (2016) analisa o *Bibframe* que é um modelo de metadados que possibilita a descrição relacional entre os recursos, além da sua associação com o *Linked Data*, Ciência de dados, publicação ampliada e web semântica.

Ramalho (2017) apresenta o SKOS, um metadado de representação formal da estrutura básica e conteúdo de Sistemas de Organização do Conhecimento em ambientes digitais. A pesquisa analisa a aplicação deste padrão de metadados no Tesouro Brasileiro de Ciência da Informação.

No que se refere a contextos contemporâneos, Pinto *et al.* (2019) relacionam alguns desafios ao *Big data*, dentre eles novos **paradigmas**, pois, ao lidarem com dados complexos, novas arquiteturas podem surgir. Pessoa, Sousa e Cardoso (2020) elucidam que a tecnologia da informação ajudou a trazer um paradigma que está preocupado com a disseminação dos dados presentes em diversos ambientes informacionais. Com isso, a Ciência de dados traz para a Ciência da informação problemáticas de análise, tratamento, uso e reuso de dados a partir das necessidades informacionais dos usuários. Portanto, o paradigma dos dados em colaboração com a Ciência da informação pode ajudar na disponibilidade dos dados governamentais abertos.

Ramalho (2016), ao concluir o seu estudo, atesta que o *Bibframe* contribui para a mudança de paradigma sobre como são encarados os processos de controle, troca e compartilhamento de recursos informacionais, ajudando a reposicionar as bibliotecas na atualidade, fortalecendo as atividades de identificação e formalização de recursos disponibilizados *on-line*. Ramalho (2017) acredita que o desenvolvimento da ciência está relacionado com a evolução da sociedade, seus paradigmas científicos e pesquisas. Nesse sentido, os conhecimentos científicos, pautados pelo empirismo, alicerça os seus paradigmas. Por meio do desenvolvimento tecnológico, há os quatro paradigmas da ciência, onde o primeiro se adequa aos fenômenos naturais; o segundo, ao uso de modelos; o terceiro, ao ramo computacional, e o quarto, ao volume de dados. Por fim, entende-se que esses paradigmas, até chegarem a *e-Science*, são precisos para estudar a contemporaneidade a partir dos dados.

Mello Filho e Araújo Júnior (2021) apresentam uma proposta de **organização da informação** em taxonomias para o desenvolvimento de assistentes virtuais, que com os objetos de fronteira podem ajudar os profissionais da Ciência da informação em um processo sinérgico com a Ciência de dados.

## 6 CONSIDERAÇÕES FINAIS

A técnica de análise de conteúdo possibilita “destrinchar” um documento e verificar todos os assuntos possíveis em uma pesquisa científica ou acadêmica. Buscou-se neste estudo entender como a Ciência de dados está sendo abordada em estudos indexadas na BRAPCI, em diferentes associações.

Cabe aqui mencionar que a BRAPCI é uma iniciativa louvável da área da informação no Brasil, neste caso, falando da sua mantenedora e, também, dos esforços em manter em acesso aberto às informações lá mantidas. Portanto, o desenvolvimento de pesquisas, tendo a mesma como agregadora, contribui para o seu fortalecimento e reconhecimento.

Os resultados apresentados têm a Ciência de dados como uma disciplina parceira da Ciência da informação, contribuindo para se fazer entender as suas técnicas, mecanismos, estudos, ferramentas e profissionalizações. Ainda assim, existe um *gap*, que não vai ao encontro das pesquisas com os cursos oferecidos no Brasil.



Apesar de demonstrações positivas na interseção das duas áreas, constatou-se que não existe nenhum curso de Ciência de dados associado às escolas de informação nacionais. Logo, estudos como este e os que indicam disciplinas com o viés de dados aos cursos de Ciência da informação, Biblioteconomia e Arquivologia são importantes para atestarem que esta carência precisa ser suprida de alguma forma, pois dados e informação estão interligados para gerar conhecimento.

Conclui-se que, no cenário internacional de produção científica e ensino, a lógica entre a Ciência de dados e a Ciência da informação faz sentido, pois as áreas estão buscando complementos para entender a existência de ambas, sendo estes documentos um resultado que validam as suas experiências. No Brasil, ainda que incipiente, estas associações são necessárias para que o fomento, mesmo que tardio, aconteça, fazendo com que a Ciência da informação e a Ciência de dados sejam de fato agregadas, e não desassociadas.

## REFERÊNCIAS

ANTÓN BRAVO, A.; SERRANO TELLERÍA, A. Periodismo de datos: de las técnicas de las ciencias Sociales a las habilidades de la ciencia de datos. **HTXT.net**, Barcelona, n. 20, p. 41-54, 2020.

Disponível em: <https://raco.cat/index.php/Hipertext/article/view/361650/463121>. Acesso em: 20 jan. 2024.

BARDIN, L. **Análise de conteúdo**. Lisboa: Edições 70, 2011.

BONDAN, A. P.; BARDAGI, M. P. Comprometimento profissional e estressores percebidos por graduandos regulares e tecnológicos. **Paidéia**, Ribeirão Preto, v. 18, n. 41, p. 581-590.

Disponível em:

<https://www.scielo.br/j/paideia/a/RjmrDZPRTqfZnzMgrSkYf9r/?format=pdf&lang=pt>. Acesso em: 10 dez. 2023.

BORGES, L. C.; OLIVEIRA, M. A produção científica sobre ischools em programas de pós-graduação em ciência da informação no Brasil. **Múltiplos Olhares em Ciência da Informação**, Belo Horizonte, v. especial, p. 1-15, 2021. Disponível em:

<https://periodicos.ufmg.br/index.php/moci/article/view/37080>. Acesso em: 12 set. 2022.

BRAPCI. **Base de Dados em Ciência da Informação**. 2024. Disponível em:

<https://brapci.inf.br/#/>. Acesso em: 2 jan. 2024.

BUFREM, L. S. *et al.* Produção internacional sobre ciência orientada a dados: análise dos termos data Science e e-science na Scopus e na Web of Science. **Informação e Informação**, Londrina, v. 21, n. 2, p. 40-67, maio/ ago. 2016. Disponível em:

<https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/26543/20114>. Acesso em: 12 jan. 2024.

CARVALHO, A. C. P. L. F.; MENEZES, A. G.; BONÍDIA, R. P. **Ciência de dados: fundamentos e aplicações**. São Paulo: LTC, 2024.

COSTAL, M.; SALES, L. ZATTAR, M. Competência em dados: habilidades na atuação e formação do bibliotecário. **Biblos**, Rio Grande, v. 34, n. 2, p. 52-71, jul./dez. 2020. Disponível em:

<https://periodicos.furg.br/biblos/article/view/11809>. Acesso em: 13 jan. 2024.

CUNHA, M. B. da; CAVALCANTI, C. R. de O. **Dicionário biblioteconomia e arquivologia**. Brasília, DF: Brinquet Lemos, 2008.



CURTY, R. G.; SERAFIM, J. da S. A formação em ciência de dados: uma análise preliminar do panorama estadunidense. **Informação e Informação**, Londrina, v. 21, n. 2, p. 307-328, maio/ago. 2016. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/27945/20195>. Acesso em: 10 jan. 2024.

DELBIANCO, N. R. *et al.* Alfabetização de dados no contexto da ciência da informação: análise dos cursos de graduação no Brasil. **Encontros Bibli**, Florianópolis, v. 28, p. 1-29, 2023. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/94094>. Acesso em: 10 jan. 2024.

e-MEC. **Cadastro Nacional de Cursos de Instituições de Educação Superior**. 2024. Disponível em: <https://emec.mec.gov.br/emec/nova>. Acesso em: 18 jan. 2024.

FALCÃO, L. C. de J. *et al.* Uso de deep learning para a construção de um modelo de recuperação da informação aplicado para o setor de mineração no Brasil. **Em questão**, Porto Alegre, v. 30, p. 1-30, 2024. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/135550/90651>. Acesso em: 29 jan. 2024.

FERNANDES, J. H. C. Interloquções bibliográficas e epistemológicas entre a ciência de dados e a ciência da informação. **Ciência da Informação**, Brasília, DF, v. 49, n. 3, p. 232-242, set./ dez. 2020. Disponível em: <https://revista.ibict.br/ciinf/article/view/5655/5295>. Acesso em: 20 jan. 2024.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GONTIJO, M. C. A.; HAMANAKA, R. Y.; ARAÚJO, R. F. de. Research data management: a bibliometric and altimetric study based on Dimensions. **Iberoamerican Journal of Science Measurement and Communication**, Tallinn, v. 1, n. 3, p. 1-19, 2021. Disponível em: <https://ijsmc.pro-metrics.org/index.php/i/article/view/16/17>. Acesso em: 10 jan. 2024.

HENNING, P. C. *et al.* Desmistificando os princípios FAIR: conceito, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 19., 2018, Marília. **Anais eletrônicos** [...] Marília: UNESP, 2018. p. 5047-5067. Disponível em: <https://cip.brapci.inf.br//download/103243>. Acesso em: 12 jan. 2024.

LEMOS, D. L. S.; MARTINS, D. L.; SOUZA, R. R. Organização e representação da informação e do conhecimento em contextos informacionais: uma proposta de um modelo teórico-conceitual para a qualidade de objetos culturais digitais. **Fronteiras da Representação do Conhecimento**, Belo Horizonte, v. 3, n. 2, p. 86-124, 2023. Disponível em: <https://cip.brapci.inf.br//download/228265>. Acesso em: 12 jan. 2024.

MARTÍN GONZÁLEZ, Y.; IGLESIAS RODRÍGUEZ, A. Alfabetização de dados: projetando um novo cenário de treinamento para o contexto universitário. **Revista Ibero-Americana em Ciência da Informação**, Brasília, DF, v. 14, n. 1, p. 318-330, jan./ abr. 2021. Disponível em: <https://cip.brapci.inf.br//download/153349>. Acesso em: 13 jan. 2024.

MARTÍNEZ MUSIÑO, C. La informetría y el análisis del discurso aplicados a la producción científica em la ciencia de datos y ciencia de la información. **e-Ciencias de la Informacion**, San

José, v. 11, n. 2, p. 151-175, 2021. Disponível em:

<https://www.redalyc.org/journal/4768/476867752017/html/>. Acesso em: 10 jan. 2024.

MATOS, M. T.; CONDURÚ, M. T.; BENCHIMOL, A. C. Interseções na produção científica da ciência da informação e ciência de dados. **Acervo**, Rio de Janeiro, v. 35, n. 2, p. 1-18, maio/ago. 2022. Disponível em:

<https://revista.an.gov.br/index.php/revistaacervo/article/view/1804/1739>. Acesso em: 10 dez. 2023.

MELLO FILHO, L. L. de; ARAÚJO JÚNIOR, R. H. de. Objetos de fronteira: um diálogo entre a ciência da informação e a ciência de dados. **Encontros Bibli**, Florianópolis, v. 26, p. 1-22, 2021. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/77247/45518>. Acesso em: 13 jan. 2024.

MELO, J. F.; ROCKEMBACH, M. Arquivologia e ciência da informação na era do big data: perspectivas de pesquisa e atuação profissional em arquivos digitais. **Prisma.com**, Porto, n. 39, p. 14-28, 2019. Disponível em: <https://lume.ufrgs.br/handle/10183/204101>. Acesso em: 12 jan. 2024.

MONTEIRO, E. C. de S. de A.; SANT'ANA, R. C. G. Repositório de dados na infraestrutura de pesquisa: adoção dos princípios FAIR. **Ciência da Informação**, Brasília, DF, v. 48, n. 3, p. 347-353, set./dez. 2019. Disponível em: <https://revista.ibict.br/ciinf/article/view/4878/4448>. Acesso em: 12 jan. 2024.

MORAES, L. L. de.; KAFURE, I. Bibliometria e ciência de dados: um exemplo de busca e análise de dados da Web of Science (WoS). **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 18, p. 1-20, 2020. Disponível em: <https://www.scielo.br/j/rdbci/a/WkSBdJB9zNjc7zhx7CHqBcJ/?format=pdf&lang=pt>. Acesso em: 12 jan. 2024.

MORETTIN, P. A.; SINGER, J. M. **Introdução à ciência de dados: fundamentos e aplicações**. São Paulo: Universidade de São Paulo, 2020. Disponível em: <https://www.ime.usp.br/~jmsinger/MAE0217/cdados2020mar.pdf>. Acesso em: 20 dez. 2023.

ORTIZ-REPISO, V.; GREENBERG, J.; CALZADA-PRADO, J. A cross-institutional analysis of data-related curricula in information science programmes: a focused look at the ischools. **Journal of Information Science**, v. 44, n. 6, 8 Jan. 2018. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/0165551517748149?journalCode=jisb>. Acesso em: 2 jan. 2024.

PEDROSO, M. de M.; LIMA, J. da C.; ASSEF NETO, V. B. Ciência de dados aplicados ao Arca: desenvolvimento e disponibilização de ferramentas para recuperação da informação no repositório institucional da Fundação Oswaldo Cruz. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, Rio de Janeiro, n. 11, p. 1-5, nov. 2017. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/1417/pdf1417>. Acesso em: 15 jan. 2024.

PESSOA, L. G. de S. B.; SOUSA, M. R. F. de; CARDOSO, T. L. de C. Aplicação de dados governamentais abertos à luz da ciência da informação. **Ciência da Informação**, Brasília, DF, v. 49, n. 3, p. 69-81, set./dez. 2020. Disponível em: <https://revista.ibict.br/ciinf/article/view/5520/5331>. Acesso em: 10 jan. 2024.

PINHEIRO, L. V. R.; FERREZ, H. D. **Tesouro brasileiro de ciência da informação**. Brasília, DF: IBICT, 2014.

PINTO, A. L.; AMARAL, E. D. Google Dataset Search: visão geral e perspectivas para indexação e disponibilização de conjunto de dados científicos abertos. **Ciência da Informação**, Brasília, DF, v. 49, n. 3, p. 173-187, set./ dez. 2020. Disponível em: <https://revista.ibict.br/ciinf/article/view/5505/5321>. Acesso em: 15 jan. 2024.

PINTO, V. A. *et al.* Interdisciplinarity in data Science over big data: findings for mining industry. **Informação e Sociedade**, João Pessoa, v. 29, n. 4, p. 61-74, out./ dez. 2019. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/47536/29295>. Acesso em: 12 jan. 2024.

PLATAFORMA Sucupira. 2024. Disponível em: <https://sucupira-v2.capes.gov.br/sucupira4/>. Acesso em: 12 jan. 2024.

RAMALHO, R. A. S. Bibframe: modelo de dados interligados para bibliotecas. **Informação e Informação**, Londrina, v. 21, n. 2, p. 292-306, maio/ ago. 2016. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/26425/20130>. Acesso em: 12 jan. 2024.

RAMALHO, R. A. S. Representação SKOS da categoria tecnologia da informação e comunicação do tesouro brasileiro de ciência da informação: um estudo preliminar. **Tendências Brasileiras em Ciência da Informação**, [s. l.], v. 10, n. 2, p. 1-13, ago./ dez. 2017. Disponível em: <https://revistas.ancib.org/index.php/tpbci/article/view/417/416>. Acesso em: 15 jan. 2024.

RAUTENBERG, S.; CARMO, P. R. V. do. Big data e ciência de dados: complementariedade conceitual no processo de tomada de decisão. **Brazilian Journal of Information Science: Research Trends**, Marília, v. 13, n. 1, p. 56-67, 2019. Disponível em: <https://revistas.marilia.unesp.br/index.php/bjis/article/view/8315/5641>. Acesso em: 12 jan. 2024.

REIS, M. de J.; CARVALHO, T. de. Ciência de dados e ciência da informação: evolução e paradigmas da ciência. **Informação em pauta**, Fortaleza, v. 7, p. 1-21, 2022. Disponível em: <http://www.periodicos.ufc.br/informacaoempauta/article/view/78490/227477>. Acesso em: 14 jan. 2024.

REIS, M. J.; SENA, N. C. S. Biblioteconomia de dados e ciência de dados no contexto da e-science. **Revista Fontes Documentais**, Aracajú, v. 4, p. 51-64, 2021. Disponível em: <https://cip.brapci.inf.br//download/193856>. Acesso em: 12 jan. 2024.

RIBEIRO, N. C. Big data em periódicos da área de ciência da informação: uma abordagem voltada para a ciência aberta e a ciência de dados. **Revista Informação na Sociedade Contemporânea**, Natal, v. 4, p. 1-15, 2020. Disponível em: <https://periodicos.ufrn.br/informacao/article/view/22452/13544>. Acesso em: 12 jan. 2024.

RODRÍGUEZ SÁNCHEZ, A.; RIVERO GONZÁLEZ, L. M.; SOCA LOZANO, S. Análisis de sentimientos y tópicos sobre las vacunas cubanas Soberana 02 y Abdala em Twitter. **Revista Cubana de Información en Ciencias de la Salud**, Havana, v. 34, 2023. Disponível em: <https://acimed.sld.cu/index.php/acimed/article/view/2261/pdf>. Acesso em: 12 jan. 2024.

SILVA, F. C. C. Visualização de dados: passado, presente e futuro. **LIINC em revista**, Rio de Janeiro, v. 15, n. 2, p. 205-223, nov. 2019. Disponível em: <https://revista.ibict.br/liinc/article/view/4812/4325>. Acesso em: 14 jan. 2024.

SILVA SEGUNDO, S. J.; ARAÚJO, W. J. Curadoria e ciclo de vida dos dados. In: DIAS, G. A.; OLIVEIRA, B. M. J. F. (org.). **Dados científicos: perspectivas e desafios**. João Pessoa: Editora UFPB, 2019. p. 113-152. Disponível em: <https://www.editora.ufpb.br/sistema/press5/index.php/UFPB/catalog/book/359>. Acesso em: 12 set. 2024.

SOBRAL, N. V.; LIMA, G. L. de Q.; SOBRAL, A. S. P. de M. Produção científica sobre hospitais no contexto da ciência de dados: um estudo a partir da Web of Science. **Encontros Bibli**, Florianópolis, v. 26, n. esp., p. 1-16, 2021. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/78824/47431>. Acesso em: 12 jan. 2024.

URS, S. R.; MINHAJ, M. Evolution of data science and its education in ischools: an impressionistic study using curriculum analysis. **Journal of the Association for Information Science and Technology**, 28 Mar. 2022. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.24649>. Acesso em: 10 jan. 2024.

VALE, L.; CRISTOVÃO, H. M. Visualização de informação sobre preços de medicamentos da base de dados abertos da ANVISA com auxílio de análise de redes de informação. **Revista Ibero-Americana de Ciência da Informação**, Brasília, DF, v. 16, n. 1, p. 206-225, jan./ abr. 2023. Disponível em: <https://cip.brapci.inf.br//download/219744>. Acesso em: 12 jan. 2024.

WU, D. *et al.* The state of ischools: an analysis of academic research and graduate education. **Journal of Information Science**, v. 38, n. 1, 19 Dec. 2011. Disponível em: <https://journals.sagepub.com/doi/10.1177/0165551511426247>. Acesso em: 12 jan. 2024.

ZHANG, Y.; ZENG, M. L. Creating an undergraduate data science program in a disciplinary and institutional context in an ischool. **Proceedings of the Association for Information Science and Technology**, 14 Oct. 2022. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/epdf/10.1002/pr2.747>. Acesso em: 10 jan. 2024.