

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

**GT 7 – Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação**

**TESES E DISSERTAÇÕES DA CIÊNCIA DA INFORMAÇÃO: ANÁLISE DE DADOS DA CAPES**

***INFORMATION SCIENCE THESES AND DISSERTATIONS: A CAPES DATA ANALYSIS***

**Francis Bento Marques** – Universidade Federal de Minas Gerais (UFMG)

**Benildes Coura Moreira dos Santos Maculan** – Universidade Federal de Minas Gerais (UFMG)

**Renato Rocha Souza** – Fundação Getúlio Vargas (FGV)

**Modalidade: Trabalho Completo**

**Resumo:** Este estudo apresenta uma análise abrangente das teses e dissertações defendidas nos programas de pós-graduação em Ciência da Informação no Brasil, com base nos dados abertos da CAPES, entre 2004 e 2022. Foram tratados dados sobre programas, docentes, discentes e trabalhos defendidos, com uso de técnicas de ciência de dados e análise bibliométrica apoiadas em scripts em Python. Os resultados revelam crescimento da área em número de programas e produção científica, mas também evidenciam desigualdades regionais, baixa internacionalização e falhas na padronização dos dados. A pesquisa contribui com uma metodologia de curadoria de dados educacionais e oferece subsídios para avaliação institucional e formulação de políticas públicas voltadas ao fortalecimento da área.

**Palavras-chave:** ciência da informação; dados abertos governamentais; análise bibliométrica.

**Abstract:** This study presents a comprehensive analysis of theses and dissertations defended in graduate programs in Information Science in Brazil, based on open data from CAPES between 2004 and 2022. Data on programs, faculty, students, and defended works were processed using data curation techniques and bibliometric analysis supported by Python scripts. The results reveal growth in the number of programs and scientific output in the field, but also highlight regional disparities, low internationalization, and standardization issues in the data. The research contributes a methodology for educational data curation and offers insights for institutional evaluation and public policy development aimed at strengthening the field.

**Keywords:** information science; open government data; bibliometric analysis.

### 1 INTRODUÇÃO

O crescimento na geração e disseminação de dados digitais tem alterado os processos decisórios e as formas de interação social, especialmente nos contextos acadêmico, institucional e governamental (Belluzzo, 2018). Esse fenômeno, frequentemente associado ao termo *big data*, representa não apenas um desafio, mas também uma oportunidade para a

pesquisa científica, a inovação e a formulação de políticas públicas (Deepa *et al.*, 2022). A capacidade de extrair valor desses conjuntos de dados depende do uso de métodos analíticos e do fortalecimento de práticas informacionais que assegurem a sua qualidade, utilidade e impacto (Graves; Hendler, 2014; Corsar; Edwards, 2017).

No campo da educação superior, os dados abertos governamentais têm se consolidado como recursos estratégicos. A disponibilização pública de informações por meio de plataformas como o Portal de Dados Abertos do Governo Federal permite à sociedade acompanhar a execução de políticas públicas e oferece subsídios relevantes para pesquisadores e gestores (Attard *et al.*, 2015). No contexto da pós-graduação brasileira, a Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) tem promovido o acesso a seus dados por meio de seu Plano de Dados Abertos (CAPES, 2017), viabilizando análises sobre programas, docentes, discentes e produções acadêmicas.

Entre os dados disponibilizados, destacam-se as teses e dissertações como expressões diretas da atividade científica nos programas de pós-graduação. Esses trabalhos constituem fontes privilegiadas para a compreensão da produção do conhecimento e da dinâmica de formação acadêmica, especialmente em áreas como a Ciência da Informação (CI), cuja consolidação no Brasil está fortemente vinculada ao desenvolvimento da pós-graduação *stricto sensu* (Viana; Oliveira, 2016).

Entretanto, a literatura aponta a existência de lacunas nos estudos que analisam a produção científica da área. São recorrentes as abordagens focadas apenas em artigos publicados em periódicos e eventos, o que compromete a visibilidade das pesquisas desenvolvidas nos programas (Moreira; Ferneda, 2020; Silva, 2021). Além disso, estudos anteriores apontam a necessidade de ampliar o escopo de análise para além de instituições específicas e períodos delimitados (Queiroz; Noronha, 2004), bem como de compreender as correlações entre temas de pesquisa, linhas dos programas, formação dos docentes e redes de colaboração (Lança; Amaral; Gracioso, 2018; Souza; Maricato, 2022; Telmo, 2019).

Neste contexto, o presente artigo apresenta uma análise bibliométrica da produção científica da pós-graduação em CI no Brasil, com base nos dados abertos da Capes referentes às teses e dissertações defendidas entre 2004 e 2022. Para isso, foram criadas estratégias de curadoria, organização e análise de dados, com o objetivo de mapear indicadores da área, identificar lacunas nos registros oficiais e oferecer subsídios à avaliação institucional e à formulação de políticas públicas.

## **2 FUNDAMENTAÇÃO TEÓRICO-METODOLÓGICA**

A pesquisa fundamenta-se em quatro bases teórico-metodológicas principais: a comunicação científica e o papel das teses e dissertações como fontes de informação; o histórico e as políticas da pós-graduação no Brasil; os estudos bibliométricos; e as questões associadas à curadoria e qualidade dos dados abertos.

A comunicação científica é compreendida como processo essencial para o avanço do conhecimento e a consolidação das comunidades científicas. Conforme apontam Galdino, Garcia e Amaral (2022), trata-se de um fenômeno dinâmico, em constante transformação, que abrange tanto a produção quanto a disseminação e o uso da informação científica. Nesse processo, as teses e dissertações ocupam lugar de destaque, por registrarem resultados de pesquisas originais conduzidas sob orientação acadêmica e integrarem o acervo informacional da pós-graduação. Melo, Trinca e Maricato (2021) destacam que esses documentos refletem diretamente as linhas de pesquisa dos programas e, por isso, constituem fontes legítimas para estudos de avaliação e mapeamento da área.

O desenvolvimento da pós-graduação no Brasil tem sido influenciado por políticas públicas formuladas principalmente pela Capes, cuja atuação compreende o fomento, a avaliação e a regulação dos cursos de mestrado e doutorado. De acordo com Almeida (2017), a pós-graduação brasileira passou por diferentes fases, com expansão significativa a partir da década de 1990 e intensificação da avaliação da qualidade na década seguinte. A produção científica é um dos pilares do sistema de avaliação da Capes, que considera indicadores como número de titulados, publicação de artigos, internacionalização e impacto da formação acadêmica.

No caso específico da CI, sua consolidação como área de conhecimento e atuação profissional se concretiza, em grande parte, pelos resultados obtidos nos programas de pós-graduação. Segundo Viana (2016), a análise das temáticas abordadas nas pesquisas da área permite compreender não apenas o desenvolvimento do campo, mas também suas articulações com outras disciplinas e suas transformações epistemológicas. Queiroz e Noronha (2004) reforçam que a caracterização do perfil dos programas, docentes e linhas de pesquisa é fundamental para a avaliação e o fortalecimento da área.

Os estudos bibliométricos são tomados como principal abordagem metodológica da presente pesquisa. Conforme Grácio (2020, p. 20), a bibliometria se apoia em “conceitos, teorias e procedimentos da CI, da Sociologia da Ciência, da Matemática, da Estatística e da

Computação”, com o objetivo de mensurar a produção científica. Esses estudos permitem o mapeamento de indicadores como volume de publicações, autoria, coautoria, colaboração entre instituições, e perfil temático da produção científica (Araújo; Alvarenga, 2011; Schweitzer; Rodrigues, 2013). Ainda segundo Melo, Trinca e Maricato (2021), os estudos bibliométricos são fundamentais para a compreensão das relações estabelecidas no processo de produção e comunicação da ciência.

A análise bibliométrica, entretanto, depende da qualidade e da estrutura dos dados utilizados. A política de dados abertos no Brasil foi formalizada pelo Decreto nº 8.777/2016, que estabelece diretrizes para a disponibilização pública de informações em formato aberto e sob licença livre. A Capes, em conformidade com esse decreto, desenvolveu sucessivos Planos de Dados Abertos (PDA). Esses planos orientam a seleção e disponibilização de conjuntos de dados, respeitando a Lei Geral de Proteção de Dados (LGPD), que impõe cuidados específicos no tratamento de dados pessoais (Capes, 2022; Brasil, 2016; Frazão; Oliva; Tepedino, 2019). Embora a LGPD não impeça a publicação de dados abertos, ela exige que esses sejam tratados com medidas de segurança, como criptografia, controle de acesso e políticas de proteção (Gregori, 2020; Piurcosky *et al.*, 2019). A CAPES disponibiliza atualmente 57 conjuntos de dados, do período de 1987 a 2024. Esses dados permitem acesso a informações sobre periódicos, bolsas, docentes, programas, orçamento e, especialmente, as teses e dissertações defendidas.

No âmbito da disponibilização de Dados Abertos Governamentais (DAG), a Capes permite análises bibliométricas e institucionais com base em dados oficiais (Ubaldi, 2013; Capes, 2024). No entanto, estudos como o de Torino, Trevisan e Vidotti (2019) apontam que esses dados ainda não atendem aos níveis mais avançados de abertura preconizados por Berners-Lee (2012), permanecendo no terceiro nível das cinco estrelas. Corsar e Edwards (2017) destacam que a confiabilidade, utilidade e impacto dos dados abertos estão diretamente relacionados à sua qualidade, o que inclui aspectos como acurácia, completude, padronização e atualidade. Por sua vez, Di Vaio, Hassan e Alavoine (2022) apontam que o uso de ferramentas como programação, inteligência artificial e bibliometria pode ampliar o potencial analítico dos dados, desde que sejam observados os princípios de curadoria e as práticas de organização da informação.

### 3 METODOLOGIA

A pesquisa adotou abordagem exploratória e descritiva, de natureza predominantemente quantitativa, fundamentada na análise de dados abertos da CAPES. Foram aplicados procedimentos bibliométricos combinados a estratégias de curadoria, tratamento e visualização de dados, organizados em sete etapas

#### 3.1 Coleta de Dados

Foram utilizados dados da CAPES organizados no eixo temático Avaliação da Pós-Graduação Stricto Sensu, composto por oito divisões e 35 conjuntos de dados (2022). Selecionaram-se quatro conjuntos: 1) Catálogo de Teses e Dissertações, 2) Docentes, 3) Discentes e 4) Programas de Pós-Graduação Stricto Sensu no Brasil. Do Catálogo de Teses e Dissertações, foram utilizados 19 arquivos CSV (2004–2022), excluindo-se dados anteriores a 2004 por não conterem informações sobre docentes e discentes. O conjunto de Docentes também reuniu 19 arquivos CSV no mesmo intervalo, contendo dados sobre titulação, vínculo e dedicação, mas sem considerar a produção intelectual individual. O conjunto de Discentes forneceu igualmente 19 arquivos CSV, com informações sobre nacionalidade, ingresso, situação acadêmica e instituição vinculada. Por fim, o conjunto dos Programas incluiu 10 arquivos CSV, disponíveis apenas a partir de 2013, com dados institucionais e classificatórios.

O quadro 1 ilustra a cobertura de alguns campos-chave, evidenciando os desafios de qualidade apontados.

**Quadro 1** – Cobertura de Campos Relevantes no DataFrame Consolidado

Campo	Descrição	Observação
<b>NM_DISCENTE</b>	Nome do discente	Inconsistências de grafia exigiram normalização.
<b>NM_ORIENTADOR</b>	Nome do orientador	Valores ausentes significativos em períodos mais antigos.
<b>DS_RESUMO</b>	Resumo do trabalho	Registros com resumos incompletos ou ausentes.
<b>DC_PALAVRA_CHAVE</b>	Palavras-chave	Muitas vezes ausentes ou não padronizadas.
<b>ID_PESSOA</b>	Identificador único	Ausente na maioria dos registros, um desafio para a rastreabilidade.

**Fonte:** Elaborado pelos autores (2025).

Ao todo, foram coletados 67 arquivos CSV: 19 de teses, 19 de docentes, 19 de discentes e 10 de programas, além de quinze arquivos de metadados em PDF, que auxiliaram na correta interpretação e manipulação dos dados. A organização seguiu dois blocos temporais: DF1 (2004–2012) e DF2 (2013–2022), refletindo o crescimento progressivo no volume e na complexidade das bases de dados.

### **3.2 Tratamento dos Dados**

O tratamento incluiu padronização, limpeza e harmonização dos campos, realizado no Google Colab PRO com linguagem Python e bibliotecas específicas (pandas, os, unidecode, time). Foram identificadas inconsistências como valores ausentes, duplicidades e variações de grafia. Aplicaram-se normalização de caracteres, padronização de caixa, substituição de valores nulos, categorização e filtragens, resultando em DF1 (2004–2012) com 41 campos e DF2 (2013–2022) com 56/58 campos, posteriormente harmonizados para viabilizar análises comparativas.

### **3.3 Consolidação de Dados**

A consolidação ocorreu em cinco fases:

1. Integração dos DataFrames DF1 e DF2 no conjunto unificado DF3\_trabalhos\_BR;
2. Padronização e compatibilização estrutural;
3. Tratamento de dados ausentes e conversão de codificação para UTF-8;
4. Aplicação de procedimentos idênticos aos dois blocos temporais;
5. Harmonização de campos e seleção de variáveis relevantes, assegurando integridade e coerência.

### **3.4 Criação do DataFrame**

Os registros de DF1 e DF2 foram integrados no DF3\_trabalhos\_BR, com ajuste de variáveis, padronização de nomenclaturas, adaptação de identificadores únicos, eliminação de colunas redundantes e inclusão de variáveis compatíveis com o modelo final. A integração foi realizada por concatenação sequencial, com verificação de integridade e tratamento de eventuais lacunas, resultando em um conjunto final com 1.232.870 registros e 24 colunas.

### **3.5 Análise Exploratória**

A análise exploratória foi conduzida em cinco fases: exame estrutural das variáveis, verificação e tratamento de ausências, diagnóstico de duplicatas e inconsistências textuais, padronização terminológica e organização de três conjuntos complementares (DF4 – programas, DF5 – docentes, DF6 – discentes). Incluiu estatísticas descritivas, gráficos e tabelas para identificação de tendências temporais, padrões regionais e correlações entre variáveis.

### 3.6 Recorte de Dados

O recorte para a área de CI foi realizado a partir de DF3\_trabalhos\_BR e conjuntos complementares, filtrando-se pelo código 60700009 e descrição “Ciência da Informação”. O processo resultou em quatro DataFrames específicos, cujos totais de registros são: DF3\_trabalhos\_CI resultou em 4.934 registros; DF4\_programas\_CI, em 220 registros; DF5\_docentes\_CI, em 6.057 registros; e DF6\_discentes\_CI, em 4.926 registros, após ajustes e validações, conforme mostra o quadro 2.

**Quadro 2** – Número de registros finais para a área de Ciência da Informação (2004-2022)

DataFrame	Conteúdo	Total de Registros
DF3_trabalhos_CI	Teses e Dissertações	4.934
DF4_programas_CI	Programas de Pós-Graduação	220
DF5_docentes_CI	Docentes vinculados	6.057
DF6_discentes_CI	Discentes vinculados	4.926

Fonte: Elaborado pelos autores (2025).

### 3.7 Visualização dos Dados

A etapa de visualização estruturou os resultados em dois blocos: o panorama nacional da pós-graduação stricto sensu e o recorte específico da área de CI, ambos abrangendo o período de 2004 a 2022.

Foram utilizados quatro conjuntos principais em cada bloco: DF3 (trabalhos), DF4 (programas), DF5 (docentes) e DF6 (discentes). A seleção das visualizações para este artigo priorizou gráficos e tabelas capazes de evidenciar o posicionamento da CI no contexto nacional, reduzindo o número originalmente produzido no estudo completo. As representações foram geradas a partir de *scripts* em Python, com bibliotecas como *pandas* e *matplotlib*, garantindo padronização visual e reprodutibilidade.

O quadro 3 mostra as etapas sequenciais de modo sintetizado.

**Quadro 3** – Fluxograma das etapas metodológicas

Etapa	Descrição	Entrada/Saída (DataFrames)
<b>1. Coleta de Dados</b>	Extração de arquivos CSV do Portal de Dados Abertos da Capes.	Entrada: Portal da Capes. Saída: 67 arquivos CSV brutos (Teses, Docentes, Discentes, Programas).
<b>2. Tratamento dos Dados</b>	Padronização, limpeza, harmonização de campos e remoção de inconsistências.	Entrada: CSVs brutos. Saída: Dados tratados e normalizados em dois blocos temporais.
<b>3. Consolidação</b>	Integração das bases de dados em um conjunto estruturado.	Entrada: Blocos tratados. Saída: DataFrames consolidados DF1 (2004-2012) e DF2 (2013-2022).
<b>4. Criação dos DataFrames</b>	Validação, organização final e junção dos blocos em um DataFrame único.	Entrada: DF1 e DF2. Saída: DF3_trabalhos_BR (1.232.870 registros).

<b>5. Análise Exploratória</b>	Identificação de padrões, tendências, tratamento de ausências e validação.	Entrada: DF3_trabalhos_BR, DF_programas, DF_docentes, DF_discentes. Saída: DataFrames validados.
<b>6. Recorte dos Dados</b>	Filtragem dos registros para isolar os dados da área de CI.	Entrada: DataFrames gerais. Saída: DF3_trabalhos_CI, DF4_programas_CI, DF5_docentes_CI, DF6_discentes_CI.
<b>7. Visualização dos Dados</b>	Apresentação gráfica dos resultados e indicadores para análise.	Entrada: DataFrames da CI. Saída: Gráficos e tabelas com os resultados da pesquisa.

**Fonte:** Elaborado pelos autores (2025).

O quadro 3 sintetiza as sete etapas sequenciais, desde a coleta inicial até a visualização dos resultados. Esse encadeamento metodológico foi estruturado de modo a assegurar a rastreabilidade e a integridade dos dados em todas as fases, permitindo a reprodução do processo por outros pesquisadores e a adaptação da metodologia a diferentes contextos. A descrição detalhada de cada etapa, associada ao uso de ferramentas e procedimentos padronizados, visou garantir a consistência dos indicadores gerados e a aderência aos princípios da ciência aberta.

#### **4 RESULTADOS E ANÁLISES**

Esta seção apresenta os principais achados, combinando a descrição dos indicadores com interpretações alinhadas à literatura estudada. São expostos os resultados referentes ao panorama da pós-graduação nacional e ao recorte específico da área de CI, acompanhados de comparações e perfis institucionais. Os resultados são apresentados em quatro subseções, a saber: 1) Cobertura e completude dos campos, 2) Quantidade final de registros após cruzamentos, 3) Panorama nacional e da CI e 4) Perfil de programas, docentes e discentes.

##### **4.1 Cobertura e completude dos campos**

Os campos recuperados nas bases de dados apresentaram diferentes taxas de preenchimento, com ausência ou inconsistência em variáveis como identificadores únicos, resumos e nomes de orientadores. A tabela 1 apresenta uma síntese desses dados.

**Tabela 1** – Campos recuperados e características de preenchimento (dados da área de CI)

<b>Campo</b>	<b>Observações</b>
Grau acadêmico	Apresenta valores ausentes em alguns registros, corrigidos na curadoria
Área de conhecimento	Ausências residuais preenchidas
Nome do discente	Variações ortográficas persistentes
Resumo	Ausências significativas; preenchimento manual em casos específicos
Palavras-chave	Muitas incompletas ou genéricas ou inexistentes
Nome do orientador	Falhas pontuais de preenchimento
Identificador único de docente/discente	Campo ausente em boa parte dos registros

**Fonte:** elaboração própria com dados CAPES (2004–2022).

#### 4.2 Quantidade final de registros após cruzamentos

Após os cruzamentos e filtrações, foram obtidos os DataFrames consolidados DF3\_CI, DF4\_CI, DF5\_CI e DF6\_CI. As quantidades finais (tabela 2) evidenciam pequenas diferenças entre registros de discentes e trabalhos.

**Tabela 2** – Quantidade final de registros após recorte e cruzamentos (área de CI)

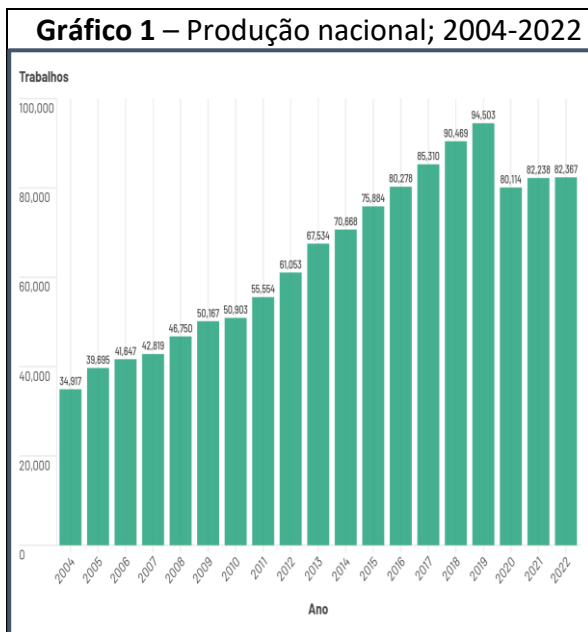
DataFrame	Registros finais	Observações
DF3_trabalhos_CI	4.934	Teses e dissertações
DF4_programas_CI	220	Programas validados sem inconsistências ou campos vazios
DF5_docentes_CI	6.057	Registros completos e consistentes
DF6_discentes_CI	4.926	Diferença em relação a DF3 por variação de grafia, discentes não titulados e programas multidisciplinares

Fonte: elaboração própria a partir dos dados do estudo (CAPES, 2004–2022).

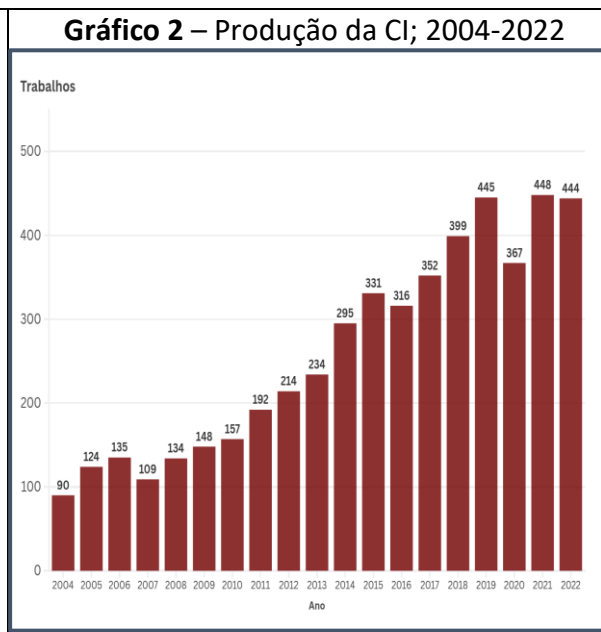
#### 4.3 Panorama nacional e da CI

No período de 2004 a 2022, o Brasil registrou 1.232.870 teses e dissertações defendidas na pós-graduação *stricto sensu* (gráfico 1). No mesmo período, a área de CI totalizou 4.934 registros (gráfico 1). Em ambos os contextos, a trajetória foi de crescimento consistente até 2019, com queda em 2020 e recuperação parcial nos anos seguintes.

O gráfico 1 mostra que houve um crescimento contínuo na produção entre 2004 e 2019, passando de 34.917 para 94.503 trabalhos, um aumento de aproximadamente 170,7%, destacando os anos de 2005, 2012 e 2013, que registraram os maiores acréscimos percentuais (13,68%, 9,89% e 10,61%, respectivamente). Em 2020, a produção caiu 15,23%, totalizando 80.114 registros, provavelmente devido à pandemia de COVID-19. Nos anos seguintes, observou-se uma recuperação parcial: aumento de 2,65% em 2021 e estabilidade em 2022, com variação de apenas 0,16%. A área com o maior volume de produções acadêmicas é Ciências Humanas, seguida pelas Ciências da Saúde e Ciências Sociais Aplicadas, que, juntas, concentram 46,19% de toda a produção no período 2004-2022.



Fonte: Marques (2025, p. 261).



Fonte: Marques (2025, p. 309).

Analisando o gráfico 2, nota-se que em 2004, foram defendidos 90 trabalhos em CI (1,82% do total), com crescimento nos anos seguintes: 124 (2,51%) em 2005 e 135 (2,74%) em 2006. Após uma leve queda em 2007 (109; 2,21%), os números voltaram a subir: 134 (2,72%) em 2008, 148 (3,00%) em 2009, e 157 (3,18%) em 2010. O crescimento continuou com 192 (3,89%) em 2011 e 214 (4,34%) em 2012. A expansão se intensificou a partir de 2013: 234 (4,74%), 295 (5,98%) em 2014 e 331 (6,71%) em 2015. De 2016 a 2019, os números seguiram em alta: 316 (6,40%), 352 (7,13%), 399 (8,09%) e 445 (9,02%). Em 2020, houve queda para 367 (7,44%), possivelmente também devido à pandemia da COVID-19, com posterior recuperação: 448 (9,08%) em 2021 e estabilidade em 2022 com 444 trabalhos (9,00%). Observa-se que a Universidade Federal de Minas Gerais (UFMG) lidera em número absoluto (13,84%), seguida pela Universidade de Brasília (UnB) (10,40%) e pela Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP) (9,93%), a Universidade Federal da Paraíba (UFPB) (6,75%), a Universidade Federal de Santa Catarina (UFSC) (6,08%) e a Universidade Federal do Rio de Janeiro (UFRJ) (6,00%).

Comparativamente, pode-se dizer que ambos os contextos demonstram uma tendência geral de trajetória de expansão interrompida pela pandemia, com posterior recuperação. A diferença está na escala da produção e na intensidade do crescimento: a CI, embora minoritária, apresentou um crescimento percentual mais expressivo e consistente ao longo dos anos, alinhada, mas, mais vigorosa, em termos relativos, que o panorama nacional. Assim, o comportamento da produção acadêmica da CI, ao mesmo tempo que acompanha

tendências nacionais, demonstra resiliência frente a eventos disruptivos e reforça a importância de políticas de fomento e mecanismos de avaliação que combinem indicadores quantitativos e qualitativos para sustentar o desenvolvimento da área.

#### 4.4 Perfil de programas, docentes e discentes

Entre 2004 e 2022, nacionalmente, existem 503 programas ativos, distribuídos de forma desigual entre as regiões. Por exemplo: Norte (36; 7%), Nordeste (84; 17%), Centro-Oeste (43; 9%), Sudeste (248; 49%) e Sul (92; 18%). Em contrapartida, na área CI são 27 programas ativos, que refletem as mesmas desigualdades regionais, concentrados principalmente no Sudeste e Nordeste: Norte (1; 3,7%), Nordeste (8; 29,6%), Centro-Oeste (1; 3,7%), Sudeste (13; 48,2%) e Sul (4; 14,8%). A tabela 3 mostra uma síntese.

**Tabela 3** – Distribuição regional dos programas de pós-graduação no Brasil e na área de CI (2004–2022)

Região	Programas – Brasil (n)	% Brasil	Programas – CI (n)	% CI
Sudeste	248	49,0	13	48,2
Sul	92	18,0	4	14,8
Nordeste	84	17,0	8	29,6
Centro-Oeste	43	9,0	1	3,7
Norte	36	7,0	1	3,7
<b>Total</b>	<b>503</b>	<b>100</b>	<b>27</b>	<b>100</b>

Fonte: elaboração própria com dados CAPES (2004–2022).

Ao longo do período analisado, a pós-graduação em CI no Brasil apresentou crescimento contínuo no número de discentes, docentes e programas acadêmicos, acompanhada por aumento progressivo de titulados, refletindo a consolidação da área como campo de pesquisa e formação. A distribuição dos programas, porém, permanece desigual, com maior concentração em determinadas regiões e instituições historicamente consolidadas, como USP, UFMG, UnB, UNESP e UFPB, que combinam alta produtividade acadêmica, diversidade temática e capacidade de captação de recursos.

No conjunto de dados dos docentes da CI registraram-se 766 docentes, a maioria nascida entre 1960 e 1979, configurando um quadro docente maduro, com baixa renovação geracional (aqueles nascidos após 1990). O título de Doutorado predomina, reforçando o alto nível de qualificação exigido pela área. A concentração etária de titulação situa-se entre 30 e 45 anos, com pico aos 36 anos, e declínio progressivo após os 48. A análise institucional mostra polos formadores relevantes, liderados por USP, UFRJ-IBICT e UFMG, além de participação de universidades estrangeiras. No campo formativo, quase metade dos docentes titulou-se em

CI, seguida por áreas afins como Comunicação, Educação, História e Engenharia de Produção, revelando caráter multidisciplinar. O cenário aponta para uma base consolidada e altamente qualificada, mas com desafios de rejuvenescimento e distribuição regional equilibrada.

A titulação discente concentra-se na faixa etária de 27 a 45 anos, com média de 24 anos no mestrado e 36 no doutorado, refletindo trajetórias acadêmicas e profissionais diversas. A presença de programas acadêmicos e profissionais com perfis distintos amplia as possibilidades de inserção no mercado e na pesquisa, indicando fortalecimento institucional e maior reconhecimento da área no contexto científico e tecnológico.

De forma integrada, os resultados apresentados evidenciam que a pós-graduação em CI no Brasil, entre 2004 e 2022, consolidou-se como um campo em expansão contínua, sustentado por um corpo docente altamente qualificado, pela diversificação de programas acadêmicos e profissionais e pelo crescimento do número de titulados. A análise revela que, embora haja concentração geográfica e institucional na formação e atuação docente, a área mantém caráter multidisciplinar, dialogando com diferentes campos do conhecimento. Ao mesmo tempo, identificam-se desafios estruturais, como a renovação geracional do quadro docente e a distribuição regional equilibrada dos programas, tendo em vista ampliar a presença e a infraestrutura em regiões menos representadas. Esses aspectos, quando considerados em conjunto, permitem compreender não apenas o estágio atual de maturidade da área, mas também apontar caminhos para seu fortalecimento, expandindo a capacidade formativa, a inserção social e o impacto científico da CI no cenário nacional.

## **5 CONSIDERAÇÕES FINAIS**

Este estudo apresentou uma análise da produção científica da pós-graduação em CI no Brasil, com base nos dados abertos disponibilizados pela CAPES para o período de 2004 a 2022. Ele buscou apresentar um panorama da área no contexto nacional, ressaltando as potencialidades analíticas proporcionadas pelo uso de dados abertos e a importância da curadoria nesse processo. A partir de procedimentos de curadoria e análise bibliométrica, foi possível evidenciar o crescimento expressivo da área, especialmente no número de programas e de trabalhos defendidos.

Os resultados também destacaram a persistência de desigualdades regionais e falhas na padronização dos dados, como a ausência de identificadores únicos e inconsistências nos registros. Tais questões reforçam a necessidade de aprimoramentos na qualidade das bases

de dados públicas, de modo a assegurar análises mais precisas e confiáveis sobre a produção científica brasileira. Como resposta a esse ponto, foi instituído formalmente pela Capes, em 2022 (Capes, 2023), a iniciativa da Governança de Dados da Pós-Graduação (GoPG), que busca qualificar e integrar as informações do sistema nacional, promovendo maior transparência e confiabilidade dos dados. Embora seja um avanço institucional relevante, sua implementação ainda está em curso, sendo necessário acompanhar seus efeitos na melhoria das bases utilizadas em outros estudos bibliométricos.

Como continuidade a este trabalho, sugere-se a realização de investigações que aprofundem a análise qualitativa das temáticas e linhas de pesquisa predominantes nos programas de pós-graduação em CI, bem como estudos comparativos com outras áreas do conhecimento, a fim de ampliar a compreensão sobre o posicionamento e a evolução da área no sistema nacional de pós-graduação.

#### **AGRADECIMENTOS**

Agradecimento ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio à pesquisa (Processo 307765/2023-7, autora BCMSM; e autor RRS).

#### **REFERÊNCIAS**

ALMEIDA, Karla Nazareth Corrêa de. **A pós-graduação no Brasil: história de uma tradição inventada**. 2017. 213f. Tese (Doutorado em Educação) – Faculdade de Educação, Universidade Estadual de Campinas, Campinas, 2017.

ARAÚJO, Ronaldo Ferreira; ALVARENGA, Lídia. A bibliometria na pesquisa científica da pós-graduação brasileira de 1987 a 2007. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 16, n. 31, p.51-70, jan./jun. 2011.

ATTARD, Judie; *et al.* A systematic review of open government data initiatives. **Government information quarterly**, [S. l.], v. 32, n. 4, p. 399-418, 2015.

BELLUZZO, Regina Célia Baptista. **A competência em informação no Brasil: cenários e espectros**. São Paulo: ABECIN Editora, 2018.

BERNERS-LEE, Tim. **5 [stars] open data**. [S. l.], 2012. Disponível em: <http://5stardata.info/en/>. Acesso em: 10 jan. 2023.

BRASIL. Decreto nº 8.777, de 11 de maio de 2016. **Política de Dados Abertos do Poder Executivo Federal**. Disponível em: <https://encurtador.com.br/t8tSq>. Acesso em: 4 mar. 2025.

BRITO, Kellyton dos Santos et al. Is Brazilian open government data actually open data?: An analysis of the current scenario. **International Journal of E-Planning Research (IJEPR)**, v. 4, n. 2, p. 57-73, 2015.

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

CAPES. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Plano de dados Abertos**. Brasília: CAPES, 2017.

CAPES. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Perguntas e respostas sob aspectos da LGPD**. Brasília: CAPES, 2022.

CAPES. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. Relatório Imersivo Supucpira 2.0: dados e sistemas da pós-graduação brasileira. Brasília: CAPES, 2023.

CORSAR, David; EDWARDS, Peter. Challenges of open data quality: More than just license, format, and customer support. **Journal of Data and Information Quality (JDIQ)**, [S. l.], v. 9, n. 1, p. 1-4, 2017.

DEEPA, Natarajan *et al.* A survey on blockchain for big data: approaches, opportunities, and future directions. **Future Generation Computer Systems**, 2022.

DI VAIO, Assunta; HASSAN, Rohail; ALAVOINE, Claude. Data intelligence and analytics: A bibliometric analysis of human–Artificial intelligence in public sector decision-making effectiveness. **Technological Forecasting and Social Change**, [S. l.], v. 174, p. 121201, 2022.

FRAZÃO, Ana; OLIVA, Milena Donato; TEPEDINO, Gustavo. **Lei geral de proteção de dados pessoais e suas repercussões no direito brasileiro**. Thomson Reuters Brasil, 2019.

GALDINO, Rosangela; GARCIA, Leonardo Guimarães; AMARAL, Roniberto Morato do. Contribuições da bibliometria ao tratamento de dados institucionais não sistematizados de produção científica: o caso do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP). **Brazilian Journal of Information Science: Research trends**, Marília, v. 16, n. 2, p. 1-21, jul./dez. 2022.

GRÁCIO, Maria Cláudia Cabrini. **Análises relacionais de citação para a identificação de domínios científicos**: uma aplicação no campo dos Estudos Métricos da Informação no Brasil. São Paulo: Editora UNESP, 2020.

GRAVES, Alvaro; HENDLER, James. A study on the use of visualizations for open government data. **Information Polity**, [S. l.], v. 19, n. 1-2, p. 73-91, 2014.

GREGORI, Maria Stella. Os impactos da Lei Geral de Proteção de Dados Pessoais na saúde suplementar. **Revista de Direito do Consumidor**, p. 171-196, 2020.

LANÇA, Tamie Aline; AMARAL, Roniberto Morato; GRACIOSO, Luciana Souza. Multi e interdisciplinaridade nos programas de pós-graduação em Ciência da Informação brasileiros. **Perspectivas em Ciência da Informação**, [S. l.], v. 23, p. 150-183, 2018.

MARQUES, Francis Bento. Explorando a pós-graduação em Ciência da Informação no Brasil: um estudo baseado nos dados abertos da CAPES. 2025. 455f. Tese (Doutorado em Gestão & Organização do Conhecimento, área de concentração Ciência da Informação) - Programa de Pós-Graduação em Gestão & Organização do Conhecimento da Escola de Ciência da Informação da Universidade Federal de Minas Gerais. Belo Horizonte, 2025.

MELO, João Henrick Neri de; TRINCA, Tatiane Pacanaro; MARICATO, João de Melo. Limites dos indicadores bibliométricos de bases de dados internacionais para avaliação da Pós-

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

Graduação brasileira: a cobertura da Web of Science nas diferentes áreas do conhecimento. **TransInformação**, Campinas, v. 33, n. 1, p. 1-16, jan./jun. 2021.

MOREIRA, Jonathan Rosa; FERNEDA, Edberto. Produção científica nos programas de pós-graduação nas áreas de informação no Brasil. **Informação & Informação**, [S. l.], v. 25, n. 4, p. 142-168, 2020.

PIURCOSKY, Fabrício Peloso et al. A lei geral de proteção de dados pessoais em empresas brasileiras: uma análise de múltiplos casos. **Suma de Negócios**, [S. l.], v. 10, n. 23, p. 89-99, 2019.

QUEIROZ, Fernanda Mendes; NORONHA, Daisy Pires. Temática das dissertações e teses em ciência da informação no Programa de Pós-Graduação em Ciências da Comunicação da USP. **Ciência da Informação**, [S. l.], v. 33, p. 132-142, 2004.

SCHWEITZER, Fernanda; RODRIGUES, Rosângela Schwarz. Produção científica em áreas multidisciplinares: educação a distância no Brasil. **Em Questão**, Porto Alegre, v. 19, n. 1, p. 156-172, jan./jun. 2013.

SILVA, Sônica Mônica da. **A produção científica da Ciência da Informação no Brasil: análise do período de 2010 a 2020 nas bases de dados Web of Science e Scopus**. 2021. 106f. Dissertação (Mestrado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2021.

SOUZA, Kelma Patrícia de; MARICATO, João de Melo. Motivações para a formação de redes de colaboração em bancas examinadoras de defesas de mestrado e doutorado. **InCID: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, v. 12, n. 2, p. 173-193, fev. 2022.

TELMO, Flávia de Araújo. **Análise de redes sociais de colaboração em bancas de defesa de doutorado na pós-graduação em ciência da informação**. 2019. 160f. Dissertação (Mestrado em Ciência da Informação) – Centro de Ciências Sociais e Aplicadas, Universidade Federal da Paraíba, João Pessoa, 2019.

TORINO, Emanuelle; TREVISAN, Gustavo Lunardelli; VIDOTTI, Silvana Aparecida Borsetti Gregório. Dados abertos CAPES: um olhar à luz dos desafios para publicação de dados na web. **Ciência da Informação**, Brasília, v. 48, n. 3, p. 38-46, 2019.

UBALDI, Barbara. Open Government Data: towards empirical analysis of Open Government Data Initiatives. **OECD Working Papers on Public Governance**, [S. l.], n. 22, p. 1-61, 2013.

VIANA, Andreia Soares. **Temáticas das teses dos Programas de Pós-Graduação em Ciência da Informação nível seis na Capes**. 2016. 154f. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2016.

VIANA, Andreia Soares; OLIVEIRA, Marlene. Teses dos programas de pós-graduação em ciência da informação nível seis na CAPES em relação às linhas de pesquisa. **Biblionline**, [S. l.], v. 12, n. 2, p. 117-131. 2016.