

Interseções na produção científica da ciência da informação e ciência de dados

Intersections in the scientific production of library and information science and data science /
Intersecciones en la producción científica de la ciencia de la información y la ciencia de datos

Maurício Torres de Matos

Mestre em Ciência da Informação pela Universidade Federal do Pará (UFPA), Brasil.

mauriciotmatos@gmail.com

Marise Teles Condurú

Doutora em Ciências do Desenvolvimento Socioambiental pela UFPA. Professora da Faculdade de Biblioteconomia, do Mestrado em Ciência da Informação e do Mestrado Profissional em Gestão de Recursos Naturais e Desenvolvimento Local da UFPA, Brasil.

marise@ufpa.br

Alegria Celia Benchimol

Doutora em Ciência da Informação pelo convênio entre Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) e Universidade Federal do Rio de Janeiro (UFRJ). Professora do Programa de Pós-Graduação em Ciência da Informação da UFPA e do Programa de Pós-Graduação em Diversidade Sociocultural do Museu Paraense Emílio Goeldi, Brasil.

alegria.benchimol@gmail.com

RESUMO

Este artigo objetiva analisar a relação entre a produção científica da ciência da informação e da ciência de dados utilizando a bibliometria, com a análise de 232 artigos, entre 1996 e 2021. Destaca-se a correlação entre elas pelo crescimento da produção, temáticas em comum, significativa produção em acesso aberto e ampla difusão de periódicos e instituições. Conclui-se que a interseção entre as disciplinas parece ser uma tendência em crescimento, que pode ser percebida pelo aumento da produção interligada nos últimos anos, permitindo maior diálogo.

Palavras-chave: ciência da informação; ciência de dados; estudo bibliométrico.

ABSTRACT

This article aims to analyze the relationship of scientific production between library and information science and data science using bibliometrics, with the analysis of 232 articles dated between 1996 and 2021. The correlation between them is highlighted by the growth of production, themes in common, significant production in open access, and wide dissemination of journals and institutions. It was concluded that the intersection between the disciplines seems to be a growing trend, which can be seen by the increase in interconnected production in recent years, allowing for greater dialogue.

Keywords: information science; data science; bibliometric study.

RESUMEN

Este artículo tiene como objetivo analizar la relación de producción científica entre la ciencia de la información y la ciencia de datos utilizando la bibliometría, con el análisis de 232 artículos entre 1996 y 2021. Se destaca la correlación entre ellos por el crecimiento de la producción, temas en común, producción significativa en acceso abierto y amplia difusión de revistas e instituciones. Se concluyó que la intersección entre las disciplinas parece ser una tendencia creciente, lo que se puede ver por el aumento de la producción interconectada en los últimos años, lo que permite un mayor diálogo.

Palabras clave: ciencia de la información; ciencia de datos; estudio bibliométrico.

Introdução

A ciência de dados (CD) é uma disciplina surgida no início da década de 2000 e voltada à análise de dados digitais em um contexto tecnológico onde há vasto volume e variedade de dados produzidos em um ritmo acelerado. Para Coneglian et al. (2017), esse contexto gerou grandes mudanças em métodos de análises de dados, tanto em ambientes organizacionais quanto acadêmicos.

Como objetivo, a ciência de dados dispõe-se à criação de técnicas, tecnologias e teorias para utilização de dados provenientes de fontes heterogêneas e complexas para o auxílio à tomada de decisões e à criação de valor para a informação (Cao, 2016). Trata-se de um campo interdisciplinar que, na busca de seus objetivos, preocupa-se com dados desde a captura e representação, passando pelo armazenamento, segurança, análise e disseminação, com forte componente tecnológico.

Considerando a definição de Saracevic (1999, p. 1.051) para ciência da informação (CI) como “a ciência e prática que lida com a efetiva coleta, armazenamento, recuperação e uso da informação”, verificam-se semelhanças entre o objeto de estudo desta e a ciência de dados (CD). A interdisciplinaridade que caracteriza a CI, em especial com a ciência da computação, além da necessidade que essa ciência apresenta de manter-se atualizada com os avanços tecnológicos, permite cogitar a possibilidade de crescimento da produção científica estudando interseções entre as disciplinas, bem como possibilidades de atuação dos cientistas da informação em atividades relacionadas à CD.

Assim, levantou-se como questão de pesquisa saber: como se dá a interseção da produção científica da ciência da informação com a ciência de dados, de modo a identificar possíveis campos de estudo em comum, bem como verificar a produção de acesso aberto relacionada à temática?

Para responder a essa problemática temos como objetivo geral analisar a relação da produção científica entre a ciência da informação e a ciência de dados, elencando como objetivos específicos: a) identificar instituições, países, pesquisadores e revistas científicas de acesso aberto que mais abordaram o tema; b) verificar tendências de estudo com interseções entre a CI e CD.

A pesquisa foi realizada a partir de padrões bibliométricos básicos de publicações científicas que apresentam correlação entre a ciência da informação e a ciência de dados, disponibilizadas nas bases de dados da Web of Science e Scopus, visando identificar e medir avanços e tendências, a partir dos temas mais frequentes, volume de publicações, autores mais prolíficos, revistas de acesso aberto, instituições e países com maior produção científica. Não foram

utilizadas limitações temporais na pesquisa, a fim de incluir o maior número possível de resultados para a amostra pesquisada, de modo que os dados contemplem o período de 1996 a 2021. A pesquisa também buscou analisar o cenário internacional da temática estudada, visando a uma melhor identificação de tendências e posterior comparação com a produção brasileira, conforme será discutido nos resultados.

A fim de fundamentar a discussão dos resultados obtidos, discute-se a seguir aspectos teóricos da CI e CD.

Sobre a ciência de dados e a ciência da informação

As últimas décadas apresentaram avanços tecnológicos em um ritmo sem precedentes na história humana, sobretudo nas tecnologias de informação e comunicação (TIC). Esses avanços geraram novos hábitos e modos de interação das pessoas com a tecnologia, resultando na geração de um vasto volume e variedade de informação a um compasso acelerado que, embora tenha tornado o acesso à informação mais democrático e mais simples de várias maneiras, criou desafios para o uso, mediação, organização e recuperação da informação, sobretudo na distinção entre informações relevantes para as necessidades de um usuário e aquelas que são dispensáveis.

Por conta desse contexto, por volta do começo dos anos 2000 surgem os primeiros trabalhos em uma nova disciplina, denominada ciência de dados. Esse campo da ciência, caracterizado pela interdisciplinaridade, objetiva a obtenção de valor a partir de dados através de processos de busca de conhecimento sobre eles, de modo a obter informações factíveis à tomada de decisões (Cao, 2016).

Há diferentes pontos de vista e definições sobre ciência de dados. Dhar (2013), Gelman et al. (2014 apud Lin Wang, 2018), Provost e Fawcett (2013), por exemplo, associam essa disciplina à análise de dados e à extração de informação e conhecimento a partir destes. Cao (2018) ressalta que vários termos, como *analytics*, *big data* e ciência de dados, estão extremamente interconectados e são muitas vezes confundidos entre si. Tal definição, como corroboram Frické (2015) e Lowrie (2017), apontam para a CD mais como uma área profissional do que como ciência, visto que carece de uma definição curricular precisa, sistemas teóricos próprios e estruturas disciplinares, dependendo em grande parte de métodos e ferramentas definidos por outras ciências.

Em linha similar, Cleveland (2001) destaca que a CD tende a focar apenas em aspectos computacionais de dados, o que limita e pode comprometer seus resultados. Para o autor, existe potencial para o envolvimento de mais áreas,

pois a análise de dados envolve mais do que apenas tecnologia, de modo que ele sugere que os trabalhos da área devem dedicar-se à investigação multidisciplinar, no desenvolvimento de modelos e métodos para dados e na construção de teorias.

A ciência da informação, ao contrário, é uma disciplina com aspectos práticos e teóricos consolidados, cujo objeto de estudo (a informação) relaciona-se diretamente com o conceito de dados e que, como pode ser evidenciado nos resultados desta pesquisa, estuda uma quantidade significativa dos temas comumente relacionados à CD. Acredita-se, assim, que há a possibilidade de relações interdisciplinares mais completas desta com a CD, considerando o objetivo principal da ciência de dados, a produção de informação e conhecimento a partir da análise de dados (Provost; Fawcett, 2013). Essa possibilidade é também abordada por Virkus e Garoufallou (2020).

As transformações na CI em função das mudanças tecnológicas e das discussões acerca da ciência de dados já podem ser sentidas de maneira prática. Nos Estados Unidos, por exemplo, grande parte dos programas de pós-graduação em ciência da informação sofreu impactos em função de avanços na tecnologia de informação, alterando os cursos oferecidos e buscando maior foco em prover conhecimentos tecnológicos aos estudantes, incluindo cursos relacionados à ciência de dados, de modo a atender às habilidades e aos conhecimentos que os profissionais da informação precisam para atuar de acordo com as necessidades atuais de bibliotecas e agências de informação (Hu, 2013).

Essa tendência também é percebida por Lin Wang (2018), com a ressalva de que para que a ciência da informação aproveite as oportunidades apresentadas pela ciência de dados é necessário que se esclareça a relação entre as duas disciplinas, de modo que a CD possa produzir mais benefícios à CI. Como o autor destaca, a recíproca é verdadeira, uma vez que teorias, modelos, técnicas e processos da CI, uma ciência mais consolidada que a CD, podem gerar grandes contribuições ao desenvolvimento desta última.

Como destaca Pinheiro (2006), a ciência da informação, por sua natureza interdisciplinar, encontra-se em constante transformação de suas relações interdisciplinares, resultando em novas relações epistemológicas que modificam os limites da área. O foco da CI em processos de representação da informação e do conhecimento e de seus registros, face às inovações tecnológicas nesses processos e registros, parece levar a área a uma nova mudança de paradigma. Tecnologias como redes móveis, *smartphones*, *tablets*, armazenamento em nuvem etc. – e as imensas massas de dados delas resultantes – mudaram os ciclos da informação, os suportes informacionais e as relações com registros de informação (Souza et al., 2013), sendo necessário para a CI adaptar-se a essa nova realidade.

Somado a isso, o cenário de avanços tecnológicos em conjunto com mudanças na percepção da ciência e nos regimes de produção do conhecimento – fatos originários do movimento de ciência aberta (Smart et al., 2019) – levaram à discussão quanto à abertura e ao compartilhamento de dados científicos a fim de proporcionar a criação de novos conhecimentos, maior participação científica e possibilidades de inovação. A ciência de dados, como área tecnológica e dependente de dados digitais (especialmente dados abertos) em seus objetos de pesquisa, relaciona-se aos objetivos da ciência aberta e dos dados abertos devido à importância da reprodutibilidade e da transparência de dados e da necessidade de inovação para a área; além disso, os avanços constantes e a demanda de aprimoramento contínuo tornam o compartilhamento de conhecimento essencial para o desenvolvimento da ciência de dados.

Frente ao exposto, torna-se relevante pesquisar o estado da produção científica no campo da ciência da informação correlacionada à ciência de dados, de modo a entender a influência e as relações existentes entre as duas disciplinas, bem como verificar a produção de acesso aberto relacionada à temática. A avaliação e a análise da produção e comunicação científicas têm como um de seus principais instrumentos a bibliometria, que permite o estudo da produção científica de modo a identificar e medir avanços e tendências da ciência.

Dentre os instrumentos formais de comunicação científica, os periódicos científicos consistem em um dos principais meios para divulgação do conhecimento, devido ao menor tempo para publicização em comparação a outros meios e à credibilidade decorrente da avaliação por pares (Ferreira, 2010). Assim, a avaliação de publicações desses periódicos permite estudar a produção científica por áreas ou assuntos, podendo-se assim identificar e estimar tendências e impactos na ciência.

A necessidade de avaliar e analisar a produção e comunicação científicas levaram, no ano de 1922, ao surgimento da “bibliografia estatística”, desenvolvida por Wyndham Hulme, com a nomenclatura moderna de “bibliometria”, cunhada em 1934 por Paul Otlet (Medeiros; Vitoriano, 2015; Momesso; Noronha, 2017), consistindo em uma área de estudos voltada à quantificação da produção e uso da informação registrada. A técnica consiste na principal ferramenta para medição do conhecimento científico a partir da análise de publicações, tendo assim sido adotada como metodologia para o desenvolvimento desta pesquisa, como detalhado a seguir.

Metodologia

O estudo consiste em uma análise bibliométrica de publicações correlatas entre ciência da informação e ciência de dados. Buscou-se analisar a produção internacional sobre o tema, a fim de possibilitar uma análise mais ampla das tendências de estudo em comum entre as áreas. De forma similar, não foram adotados filtros temporais para a coleta, resultando em uma amostra que contém publicações entre o período de 1996 e 2021, ressaltando-se que a busca foi realizada no começo do ano de 2022.

A bibliometria consiste em uma técnica amplamente utilizada pela CI com o propósito de investigar processos de comunicação, sobretudo da comunicação científica, quantificando-os. Os estudos bibliométricos desenvolveram-se a partir de leis empíricas já existentes, como as leis de Lotka, Zipf, Bradford etc. (Braga, 1973; Guedes; Borschiver, 2005 apud Medeiros; Vitoriano, 2015).

Uma vez que um dos objetivos da pesquisa é a análise dos autores mais produtivos na produção científica correlata entre as duas áreas, a lei de Lotka, que afirma que o número de autores que fazem n contribuições em um determinado campo científico é aproximadamente $1/n^2$ daqueles que fazem uma só contribuição (Alvarado, 2002; Cândido; Garcia; Campos; Tambosi Filho, 2018; Urbizagastegui, 2008), possui especial relevância para o estudo realizado.

A fim de alcançar os objetivos propostos, dividiu-se a metodologia em cinco etapas. A primeira consistiu na identificação e seleção de bases de dados, resultando na escolha de *Web of Science* e *Scopus*. Essas plataformas foram escolhidas por serem referências da comunicação científica nas mais diversas áreas de estudo, além de consistirem em bases internacionais que permitem uma observação da produção científica internacional de forma mais aproximada.

Na segunda etapa, definiram-se os parâmetros e estratégias de busca nas bases de dados. Foram empregados os descritores “*data science*”, “*e-science*”, “*big data*” e “*data mining*”, sempre em conjunção com “*information science*”. Uma vez que as bases pesquisadas concentram publicações de vários países, optou-se pela utilização de descritores em inglês devido à grande utilização dessa língua para publicações internacionais. O período de cobertura da pesquisa não foi limitado, uma vez que se visou ao estudo das primeiras publicações convergentes entre CI e CD, bem como sua tendência no decorrer do tempo. A pesquisa limitou-se a publicações de artigos científicos cuja área de publicação fosse a CI e foram encontradas publicações no período entre 1996 e 2021.

Na pesquisa, encontrou-se o total de 232 artigos distribuídos entre as bases de dados selecionadas, após se desconsiderar as duplicações entre as bases.¹

Dos 232 artigos mencionados, a base Scopus apresentou maior número de resultados com 156 títulos, enquanto 124 foram encontrados na Web of Science, totalizando 48 títulos em ambas as bases, conforme representado na Figura 1.

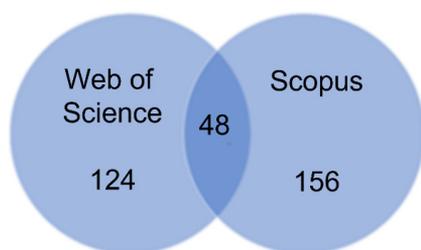


Figura 1 – Distribuição da produção científica encontrada por base de dados pesquisada. Fonte: dados da pesquisa, 2022.

Na terceira etapa, executou-se o tratamento dos dados, mediante consolidação dos resultados das diferentes bases utilizadas, exportação e tabulação das informações. Os resultados exportados continham as informações: título da publicação, resumo, ano, autores, periódico e instituição dos autores, tabuladas para a análise. Os resultados da amostra foram organizados utilizando planilhas eletrônicas, editadas a partir do software Excel e da ferramenta on-line Planilhas Google, ambos utilizados para tratamento, cálculos e posterior geração de gráficos relacionados à amostra.

Na quarta etapa, realizou-se a análise e depuração de resultados, identificando a partir dos títulos, resumos e palavras-chave aqueles artigos que se adequavam à temática, parâmetros e escopo do estudo, para isso eles deveriam conter, em um ou mais desses campos, os descritores utilizados na etapa anterior, de modo a identificar que se tratavam de trabalhos da CI abordando CD. Nessa etapa identificaram-se ainda documentos duplicados (presentes em mais de uma das bases), de modo a desconsiderar a duplicidade.

Por fim, na quinta etapa, foi realizada a interpretação dos resultados, identificando o número de publicações por cada autor e instituições encontrados, o quantitativo de publicações por país (considerando o país de origem aquele relacionado à instituição) e o quantitativo por periódico. Foi feita ainda a classificação

¹ O conjunto de dados, com a lista de artigos componentes da amostra pesquisada, está disponível em: <https://zenodo.org/record/6670108>.

posterior quanto ao tipo de acesso dos periódicos da amostra, ou seja, se o periódico era de acesso aberto, acesso por assinatura ou acesso misto (híbrido).

Em casos de coautoria, foram consideradas as informações individuais de cada autor e de sua instituição para cálculo de quantitativos. Nos casos de artigos produzidos por autores da mesma instituição, esta e seu país foram considerados uma única vez.

A partir das palavras-chave dos artigos, foi feita a análise quanto aos temas mais abordados, de modo a se estimar as temáticas e assuntos mais comuns das relações encontradas entre CD e CI, conforme apresentado nos resultados.

Conexões da produção científica entre ciência da informação e ciência de dados

Das bases de dados pesquisadas, percebeu-se maior concentração das publicações na Scopus, de modo que 67,24% dos artigos do universo total estão presentes nessa base. No entanto, as publicações presentes na Web of Science não são irrelevantes, visto que 53,44% dos artigos encontram-se na base.

A fim de delinear um panorama da produção científica em que a temática representasse conexões entre ciência da informação e ciência de dados, observou-se que o primeiro trabalho encontrado, de acordo com o escopo da pesquisa, data de 1996 e está indexado na Web of Science. Trata-se do artigo “Historical perspectives on information science” de Cornish Tao, que, embora publicado antes da origem formal da área de ciência de dados, lida com o uso de mineração de dados (*data mining*) para descobrimento de conhecimento em bases de dados; dessa forma, considerando que a busca automatizada de informação e geração de conhecimento consiste em um dos campos de estudo da CD, bem como da CI, o trabalho citado representa uma possibilidade de interseção entre as duas ciências, motivo pelo qual foi incluído nos resultados.

O próximo artigo encontrado data do ano 2000, o “Informetrics: a new area of quantitative studies”, que relaciona a mineração de dados e a recuperação da informação. Verifica-se, no período de 2000 a 2013, um número um tanto irregular de artigos e uma média baixa de publicações (3,17 por ano), à exceção do ano 2006 (com seis artigos encontrados) e 2008 (com 11 artigos identificados). Nota-se, a partir do ano de 2014, um maior número de publicações (média de aproximadamente 24,12 artigos/ano) e um crescimento acentuado de publicações correlacionando temas da CI com aqueles da CD (Figura 2).

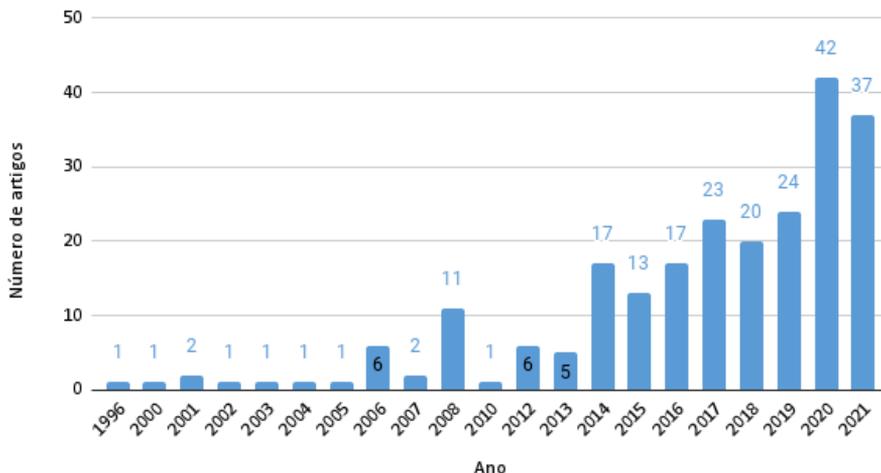


Figura 2 – Distribuição da produção científica encontrada por ano de publicação. Fonte: dados da pesquisa, 2022.

Quanto à autoria, identificou-se um total de 569 autores distintos. Os autores com maior número de publicações são apresentados no Quadro 1.

Quadro 1 – Autores com maior número de publicações, de 1996 a 2021

Número de artigos	Autor	País
4	MUSIÑO, C. M (Dokumenta Consultoría e Integraciones)	México
3	JONES, Kyle M. L. (Indiana University)	Estados Unidos
	IMRAM, M. (The University of Queensland)	Austrália
	PALMER, C. L. (University of Illinois)	Estados Unidos
	SHAH, C. (University of Washington)	Estados Unidos
	VANSCOY, A. (University at Buffalo)	Estados Unidos
	ZHANG, Y. (Wuhan University)	China

Fonte: dados da pesquisa, 2022.

O autor com maior número de publicações identificadas foi Celso Martínez Musiño (Dokumenta Consultoría e Integraciones) com quatro artigos, seguido

de seis autores com três e de 38 autores com dois. Assim, 523 publicaram um único artigo. Verifica-se, desse modo, ampla difusão de autores dentro do universo total de publicações. Essa dispersão pode ser explicada por se tratar de uma temática interdisciplinar, que permite uma variedade significativa de conteúdos e áreas do conhecimento, além de a ciência de dados consistir em uma área ainda recente. A Lei de Lotka, que utiliza a metodologia de contagem considerando autor e coautor(es), parece não ser aplicável para a amostra, o que pode indicar que a correlação entre as duas áreas ainda se encontra em um estágio pouco avançado.

Quanto às origens dos autores, verificou-se um total de 42 países, com maior número de publicações nos Estados Unidos (76), seguido por China (32), Brasil (25), Itália (18) e Reino Unido (13). Entende-se que a busca por descritores de língua inglesa, ainda que permita ter uma ideia do panorama geral internacional da pesquisa científica relacionada ao tema pesquisado, favorece países cuja língua nativa é o inglês. Ainda assim, é interessante destacar que o Brasil aparece com número significativo de pesquisas, indicando uma tendência de temas da CD na CI nacional. Parece, ainda, haver uma disposição à divulgação internacional dessas pesquisas, considerando a quantidade encontrada, sendo o país com o terceiro maior número de contribuições, significativamente à frente de alguns países de língua inglesa nativa.

Em relação às instituições às quais os autores estão vinculados, identificaram-se ao todo 268, entre universidades, entidades privadas, órgãos governamentais e instituições de pesquisa. No Quadro 2 são mostradas as instituições com maior número de publicações.

Quadro 2 – Instituições acadêmicas com maior número de publicações

Instituição	n. de artigos publicados
Indiana University	8
University of Illinois	7
Wuhan University	7
Universidade de Brasília	6
Drexel University	5
Nanjing University	5
University of Texas	5

Fonte: dados da pesquisa, 2022.

Dentre as instituições com publicações identificadas na amostra observada, destaca-se a Universidade de Indiana, com oito artigos. Também se destacam as Universidades de Illinois e de Wuhan (com sete artigos cada) e a Universidade de Brasília, com seis publicações. Percebe-se que entre as sete instituições com maior número de publicações, a maioria (quatro) é localizada nos Estados Unidos, e duas são instituições chinesas. Das entidades brasileiras, destaca-se a Universidade de Brasília, com seis artigos, aparecendo como a quarta instituição com maior número de publicações. Assim como mencionado nos resultados referentes a autores, o aparecimento de uma instituição brasileira entre aquelas com maior número de publicações mostra uma associação significativa do cenário de pesquisa em CI com a CD e temas relacionados.

Os artigos da amostra foram publicados em um total de 111 periódicos distintos. No Quadro 3 são apresentados os 12 periódicos com maior número de publicações.

Quadro 3 – Periódicos científicos com maior número de publicações

Periódico	n. de artigos publicados
<i>Proceedings of the Association for Information Science and Technology</i>	19
<i>Journal of Information Science</i>	17
<i>Digital Libraries: The Era of Big Data and Data Science, IRCDL 2000</i>	10
<i>Journal of the Association for Information Science and Technology</i>	9
<i>Journal of Documentation</i>	6
<i>Advances in Classification Research Online</i>	6
<i>Perspectivas em Ciência da Informação</i>	5
<i>Online Information Review</i>	5
<i>Ciência da Informação</i>	5
<i>Library Hi Tech</i>	5
<i>Scientometrics</i>	5
<i>Journal of Library Science in China</i>	5

Fonte: dados da pesquisa, 2022.

O periódico com maior número de publicações, totalizando 19 (8,19% do total), é o *Proceedings of the Association for Information Science and Technology*. Também compõem a lista: *Journal of Information Science* (com 17 publicações, representando 7,33% do total), *Digital Libraries: The Era of Big Data and Data Science, IRCDL 2000* (com dez publicações), *Journal of the Association for*

Information Science and Technology (nove publicações), *Journal of Documentation* e *Advances in Classification Research Online* (seis publicações cada) e os periódicos *Perspectivas em Ciência da Informação*, *Online Information Review*, *Ciência da Informação*, *Library Hi Tech*, *Scientometrics* e *Journal of Library Science in China*, com cinco publicações cada.

Os 12 periódicos com maior número de títulos corresponderam a 37,5% do total de publicações. Um grupo de quarenta periódicos tem mais de uma publicação (162 artigos, ou 69,83%) e um total de 71 periódicos tem apenas um artigo (30,60%), mostrando, assim, uma concentração considerável de artigos no primeiro grupo e uma ampla difusão de periódicos com publicações sobre a temática de estudo.

Do universo de 111 periódicos distintos nas publicações analisadas, apenas 45 (40,54%) são periódicos de acesso aberto e 27 possuem acesso misto (com algumas publicações de acesso aberto e outras com limitações de acesso). Dos artigos componentes da amostragem da pesquisa, conforme se observa na Figura 3, um número significativo de 77 artigos (33,19%) foi publicado em periódicos de acesso aberto e 59 (25,43%) em periódicos de acesso misto. Percebe-se, assim, que os periódicos de acesso aberto são prolíficos em relação à temática pesquisada, ainda que representem um percentual de publicações significativamente menor do que os periódicos de acesso restrito ou misto. Em um tema como o da pesquisa, em que a acessibilidade à informação é essencial para sua evolução e para o desenvolvimento de inovações, tal limitação é chamativa.

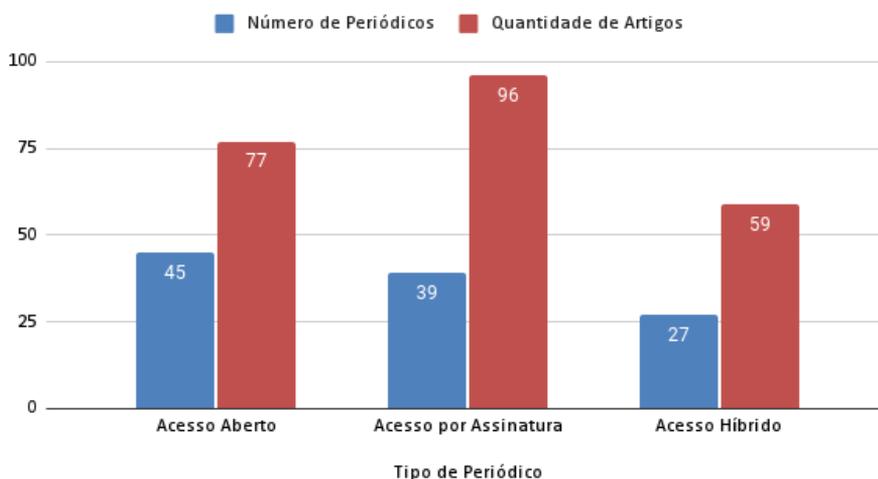


Figura 3 – Distribuição do número de artigos encontrados e periódicos por tipo de acesso. Fonte: dados da pesquisa, 2022.

No universo de periódicos de acesso aberto encontrados, uma única revista concentrou 26,25% do total de publicações e os 11 periódicos com maior número de publicações correspondem a 66,25% do total (Quadro 4). Isso evidencia uma grande concentração de artigos em um número pequeno de revistas.

Quadro 4 – Periódicos científicos de acesso aberto com maior número de publicações

Periódico	n. de artigos publicados
<i>Advances in Classification Research Online</i>	6
<i>Perspectivas em Ciência da Informação</i>	5
<i>Journal of Library Science in China</i>	5
<i>Ciência da Informação</i>	5
<i>Library Philosophy and Practice</i>	4

Fonte: dados da pesquisa, 2022.

Os periódicos de acesso aberto encontrados pela pesquisa corresponderam a 40,54% das revistas; 24,32% equivaleram a revistas de acesso híbrido e 35,13%, com acesso por assinatura. Verifica-se, assim, um percentual considerável de periódicos de acesso aberto compondo o universo da pesquisa.

É significativo que, dentre os periódicos de acesso aberto com maior número de publicações, dois títulos sejam brasileiros, *Perspectivas em Ciência da Informação* e *Ciência da Informação*, ambos com cinco artigos. Além disso, a maior parte das publicações de autores brasileiros foi realizada em periódicos de acesso aberto.

Os resultados apresentados parecem indicar a existência de fortes possibilidades para pesquisas relacionando a CI e a CD, em áreas de estudo diversas, como discutido a seguir.

Tendências de estudos correlatos entre a ciência da informação e a ciência de dados

Os assuntos e áreas abordadas pelos artigos apresentam grande variedade, com um número significativo de interseções da CI com administração, educação, medicina etc. Para tanto, foram abordadas ainda aplicações ou teorias da ciência de dados.

A partir da amostra obtida, podem-se identificar tendências de estudos correlacionando a CI e a CD, especialmente na análise de papéis que profissionais da informação podem exercer no contexto das atividades relacionadas a dados,

ou mudanças curriculares na CI decorrentes do advento da CD, sendo esses os temas mais encontrados na amostra, com 34 artigos tratando dos assuntos. Além disso, foram encontrados 22 artigos dedicados à análise de domínio de temas da CI relacionados à CD.

É possível perceber também possibilidades de aplicações das pesquisas na comunicação científica, especialmente para métricas de produção científica (foram encontrados 28 artigos relacionados a métricas).

Dois temas de interesse comum entre as duas áreas apresentaram número significativo de estudos na amostra pesquisada, sendo eles a mineração de dados (26 artigos tratam diretamente desse tema) e o *big data* (24 artigos). Percebe-se no resultado uma tendência à utilização da mineração de dados e do *big data* relacionados a assuntos voltados a métricas (como auxiliares no processo de exploração da informação), organização e representação do conhecimento, em que as técnicas da CI aparecem como norteadoras, a fim de permitir a melhor utilização da informação nos campos citados, como também foram encontrados diversos artigos buscando compreender os domínios desses dois assuntos na área da CI.

Estudos voltados à competência em informação no contexto de dados e à gestão do conhecimento são significativos na amostra, com 16 e 15 artigos, respectivamente. Também se percebeu uma tendência a estudos voltados às subáreas de classificação (12 artigos identificados) e 11 artigos referentes a bibliotecas digitais, iSchools e aplicações voltadas a bibliotecas.

Outros temas também foram abordados nos artigos, como representação do conhecimento, estudos de usuário, *e-science*, sistemas de informação e de recomendação, dados abertos, representação da informação e aplicações da ciência de dados nos campos do *design* e da recuperação da informação.

No Gráfico 4 são apresentados os temas mais abordados em relação ao total da amostra de pesquisa.

Entende-se, considerando a quantidade limitada de artigos relacionando os dois campos de estudo e os diferentes enfoques encontrados, mesmo em temas similares, que esses achados representam mais possíveis tendências de estudos correlacionados do que efetivas conexões entre as áreas. Ao mesmo tempo, a diversidade da amostra estudada pode significar a existência de amplas possibilidades de colaborações entre a ciência da informação e a ciência de dados.

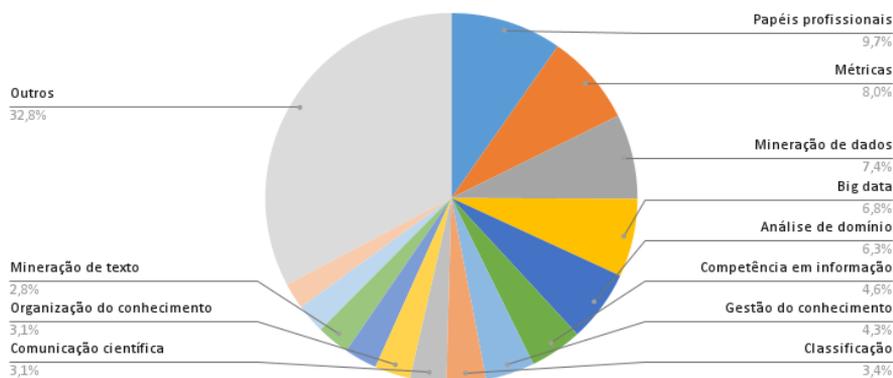


Figura 4 – Percentual dos principais temas encontrados em relação ao total da amostra. Fonte: dados da pesquisa, 2022.

Ao mesmo tempo em que a diversidade de temas parece apontar para uma abordagem inconstante e não sistemática entre CI e CD, também indica que as possibilidades de interseções entre elas são amplas.

Considerações finais

A análise desenvolvida, a partir dos estudos bibliométricos apresentados, permite a observação de aspectos relevantes acerca da temática da ciência de dados produzida por pesquisadores da ciência da informação, cuja produção científica aumentou nos últimos seis anos, podendo indicar um maior interesse da CI nos temas relacionados à prática da CD.

A interseção entre os campos da CI e da CD parece ser uma tendência de pesquisa em crescimento, o que pode ser percebido pelo aumento da produção interligada nos últimos anos e pela quantidade de países de origem dos artigos encontrados. Embora grande parte dos artigos tenha sido produzida por instituições dos Estados Unidos, entende-se que esse resultado se relaciona diretamente à metodologia utilizada, de busca por descritores em inglês e em bases de dados relacionadas a essa língua, visto que quantidade significativa da produção científica não é realizada em inglês, no entanto, por outro lado, essa metodologia nos permitiu analisar um panorama global.

A quantidade significativa de artigos abordando as mudanças relacionadas à CD nos currículos dos cursos de CI e biblioteconomia ou o surgimento de novas áreas de atuação dos profissionais da informação em decorrência da CD

denota que as transformações introduzidas pela evolução tecnológica atual e o surgimento da CD vêm causando interesse à CI.

Essas mudanças e novas tendências de pesquisa vêm se consolidando e devem permitir o surgimento de novas possibilidades de pesquisa, outras correlações com a ciência de dados e possivelmente mudanças nas relações interdisciplinares da ciência da informação com áreas tecnológicas, sobretudo com a ciência da computação, baseando-se na abordagem crescente de objetos de estudo da CD pela CI, tais como *big data*, mineração de dados, armazenamento em nuvem, redes sociais etc.

Na pesquisa, identificou-se ampla variedade de temas de interface entre ciência da informação e ciência de dados, podendo, assim, indicar possibilidades de estudos que correlacionem os campos de conhecimento, tais como papéis dos profissionais, métricas de comunicação científica, mineração de dados, *big data*, análise de conteúdo, competência em informação e gestão do conhecimento.

Entende-se que a quantidade amostral ainda é concisa, apesar da tendência crescente identificada. Há, assim, potencial para o desenvolvimento de pesquisas que explorem de forma mais completa as perspectivas de relação entre as áreas, sobretudo considerando as similaridades entre os objetos de estudo e a natureza interdisciplinar de ambas. A pesquisa permitiu identificar algumas dessas possibilidades, visando contribuir para a discussão na ciência da informação a respeito de seus novos papéis na denominada era da informação e dos dados digitais.

O papel cada vez mais central da informação e do conhecimento na sociedade contemporânea faz com que a ciência da informação adote novos papéis, permitindo maior diálogo com outras disciplinas, pela própria essência interdisciplinar do campo. Essa conjuntura reforça ainda a importância da ciência aberta nesse ambiente de inovação como facilitadora ao compartilhamento de pesquisas, dados e mesmo códigos de computador, promovendo maior cooperação entre pesquisadores e público para o desenvolvimento informacional e científico.

Referências

- ALVARADO, R. U. A Lei de Lotka na bibliometria brasileira. *Ciência da Informação*, Brasília, v. 31, n. 2, p. 14-20, ago. 2002. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652002000200002&lng=en&nrm=iso. Acesso em: 27 maio 2022.
- BRAGA, G. M. Relações bibliométricas entre frente de pesquisa (*research front*) e revisões da literatura: estudo aplicado à ciência da informação. *Ciência da Informação*, Brasília, v. 2, n. 1, p. 9-26, 1973. Disponível em: <http://revista.ibict.br/ciinf/article/view/20>. Acesso em: 27 maio 2022.

- CÂNDIDO, R. B.; GARCIA, F. G.; CAMPOS, A. L. S.; TAMBOSI FILHO, E. Lei de Lotka: um olhar sobre a produtividade dos autores na literatura brasileira de finanças. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, v. 23, n. 53, p. 1-15, set./dez. 2018. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v-23n53p1/37239>. Acesso em: 27 maio 2022.
- CAO, L. Data science: a comprehensive overview. *ACM Computing Surveys*, v. 50, n. 3, jun. 2018. Disponível em: <https://dl.acm.org/doi/10.1145/3076253>. Acesso em: 27 maio 2022.
- _____. Data science and analytics: a new era. *International Journal of Data Science and Analytics*, v. 1, n. 1, p. 1-2, abr. 2016. Disponível em: <https://link.springer.com/article/10.1007/s41060-016-0006-1>. Acesso em: 27 maio 2022.
- CLEVELAND, W. S. Data science: an action plan for expanding the technical areas of the field of statistics. *International Statistical Review*, v. 69, n. 1, p. 21-26, abr. 2001. Disponível em: <https://doi.org/10.1111/j.1751-5823.2001.tb00477.x>. Acesso em: 27 maio 2022.
- CONEGLIAN, C. S.; DIEGER, R.; SANTARÊM SEGUNDO, J. E.; CAPTREZ, M. O papel estratégico da web semântica no contexto do big data. In: WORKSHOP DE INFORMAÇÃO, DADOS E TECNOLOGIA, 1., Santa Catarina, 2017. Anais ... Santa Catarina: Universidade Federal de Santa Catarina, 2017. Disponível em: <https://repositorio.ufsc.br/handle/123456789/180289>. Acesso em: 27 maio 2022.
- DHAR, V. Data science and prediction. *Communications of the ACM*, v. 56, n. 12, p. 64-73, 2013. Disponível em: <https://dl.acm.org/doi/10.1145/2500499>. Acesso em: 27 maio 2022.
- FERREIRA, A. G. C. Bibliometria na avaliação de periódicos científicos. *DataGramaZero: Revista de Ciência da Informação*, v. 11, n. 3, 2010. Disponível em: <https://brapci.inf.br/index.php/res/v/7/110>. Acesso em: 27 maio 2022.
- FRICKÉ, M. Big data and its epistemology. *Journal of the Association for Information Science and Technology*, v. 66, n. 4, p. 651-661, 2015. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/asi.23212>. Acesso em: 27 maio 2022.
- HU, S. Technology impacts on curriculum of Library and Information Science (LIS): a United States (US) perspective. *Libres: Library and Information Science Research Electronic Journal*, v. 23, n. 2, p. 1-9, set. 2013. Disponível em: <https://www.libres-ejournal.info/1033/>. Acesso em: 27 maio 2022.
- LIN WANG. Twinning data science with information science in schools of library and information science. *Journal of Documentation*, v. 74, n. 6, p. 1.243-1.257, 2018. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/JD-02-2018-0036/full/html>. Acesso em: 27 maio 2022.
- LOWRIE, I. Algorithmic rationality: epistemology and efficiency in the data sciences. *Big Data & Society*, v. 4, p. 1-13, jan./jun. 2017. Disponível em: <https://journals.sagepub.com/doi/full/10.1177/2053951717700925>. Acesso em: 27 maio 2022.
- MEDEIROS, J. M. G.; VITORIANO, M. A. V. A evolução da bibliometria e sua interdisciplinaridade na produção científica brasileira. *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, v. 13, n. 3, p. 491-503, set./dez. 2015. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8635791>. Acesso em: 27 maio 2022.
- MOMESSO, A. C.; NORONHA, D. P. Bibliométrie ou bibliometrics: o que há por trás de um termo? *Perspectivas em ciência da informação*, v. 22, n. 2, abr./jun. 2017. Disponível em: <https://www.scielo.br/j/pci/a/X4xTbMZNdVXV3mNsYFRnSbQ>. Acesso em: 27 maio 2022.
- PINHEIRO, L. V. R. Ciência da informação: desdobramentos disciplinares, interdisciplinaridade e transdisciplinaridade. In: GONZÁLEZ DE GÓMEZ, M. N.; ORRICO, E. G. D. (org.). *Políticas de memória e informação: reflexos na organização do conhecimento*. Natal: Editora UFRN, 2006. p. 111-141. Disponível em: <http://ridi.ibict.br/handle/123456789/18>. Acesso em: 27 maio 2022.
- PROVOST, F.; FAWCETT, T. Data science and its relationship to big data and data-driven decision making. *Big Data*, v. 1, n. 1, p. 51-59, 2013. Disponível em: <https://www.liebertpub.com/doi/10.1089/big.2013.1508>. Acesso em: 27 maio 2022.
- SARACEVIC, T. Information science. *Journal of the American Society for Information Science*, v. 50, n. 12, p. 1.051-1.063, 1999. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/epdf/10.1002/%28SICI%291097-4571%281999%2950%3A12%3C1051%3A%3AAID-ASI2%3E3.o.CO%3B2-Z>. Acesso em: 27 maio 2022.

- SMART, P. et al. Open science and open innovation in a socio-political context: knowledge production for societal impact in an age of post-truth populism. *R&D Management*, v. 49, n. 3, p. 279-297, 2019. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/radm.12377>. Acesso em: 27 maio 2022.
- SOUZA, R. R.; ALMEIDA, M. B.; BARACHO, R. M. A. Ciência da informação em transformação: big data, nuvens, redes sociais e web semântica. *Ciência da Informação*, v. 40, n. 2, p. 159-173, 2013. Disponível em: <https://revista.ibict.br/ciinf/article/view/1379>. Acesso em: 27 maio 2022.
- URBIZAGASTEGUI, R. A produtividade dos autores sobre a Lei de Lotka. *Ciência da Informação*, v. 37, n. 2, p. 87-102, maio/ago. 2008. Disponível em: <https://revista.ibict.br/ciinf/article/view/1214>. Acesso em: 27 maio 2022.
- VIRKUS, S.; GAROUFALLOU, E. Data science and its relationship to library and information science: a content analysis. *Data Technologies and Applications*, v. 54, n. 5, p. 643-663, 2020. Disponível em: <https://doi.org/10.1108/DTA-07-2020-0167>. Acesso em: 27 maio 2022.

Recebido em 18/11/2021

Aprovado em 15/6/2022