Bibliotecário de dados: enfoques e perspectivas nacionais de estudos na Ciência da Informação

Raimunda Fernanda dos Santos[a] *e João Victor Moraes Sena*[b]

1. Introdução

A partir do século XXI, com o desenvolvimento da Web, o barateamento e acesso aos computadores pessoais e dispositivos móveis, o mundo conectou-se de forma jamais imaginada. Foi observada uma explosão de dados gerados por usuários em todos os lugares, a cada minuto por câmeras de monitoramento, GPS, fotos, vídeos, *posts*, e-mails, mensagens entre outros. Segundo um relatório produzido pela multinacional americana IDC, especializada em armazenamento de dados, (tradução livre) de 2005 até 2020, o universo digital crescerá por um fator de 300 (trezentos), de 130 (cento e trinta) exabytes para 40.000 (quarenta mil) exabytes, ou 40 (quarenta) trilhões de gigabytes (mais de 5.200 gigabytes para cada homem, mulher e criança em 2020). De agora até 2020, o universo digital dobrará a cada dois anos.

Tais informações evidenciam que a produção de dados é exorbitante pelo acesso massivo da Web. De acordo com Manyika et al. (2011, p.1), esse fenômeno é conceituado como Big Data, que se configura como "conjuntos de dados que estão

a Doutora em Ciência da Informação pela Universidade Federal da Paraíba – UFPB. Professora na Universidade Federal do Rio de Janeiro – UFRJ. E-mail: fernanda@facc.ufrj.br. ORCID: https://orcid.org/0000-0002-7750-3269.

b Graduando em Biblioteconomia e Gestão de Unidades de Informação pela Universidade Federal do Rio de Janeiro – UFRJ. E-mail: joaosena7@gmail.com.

além da capacidade de captura, armazenamento e análise das ferramentas de software típicas".

Com o aumento exponencial de sua produção, os dados tornaram-se um ativo de suma importância para as organizações, que passaram a dar atenção ao tratamento dos dados, fundamentando a base para as tomadas de decisões. Portanto, surge a necessidade de profissionais qualificados para lidar com a consequência dos dados e das informações geradas, tornando cada vez mais necessário os profissionais capacitados(as) que tenham habilidades para realizar as atividades que esse novo paradigma exige. Esses profissionais são os(as) cientistas de dados, os(as) analistas de dados, *business intelligence* (inteligência de negócios), engenheiro(a) de dados, bibliotecário(a) de dados, cada um com sua especialidade, mas com sua devida e importante contribuição para as organizações.

No senso comum, a sociedade tem uma visão do(a) bibliotecário(a) de que somente atua com livros ou documentos físicos, isolado dentro de uma biblioteca organizando estantes de livros. Porém a profissão passou por grandes evoluções durante as décadas, rompendo com barreiras conservadoras e atualizando sua atuação em novos ambientes e com as novas tecnologias. A partir dessas atualizações, o(a) bibliotecário(a) adquire novas habilidades e competências para atuar além da forma tradicional e uma das novas formas de atuação é a biblioteconomia de dados.

De acordo com Semeler (2017), a Biblioteconomia de Dados é uma disciplina que aplica a Ciência da Informação, e-Science[1] e Data Science[2] em Unidades de Informação. Seguindo essa linha de raciocínio, o(a) bibliotecário(a) de dados, contribui diretamente na área de dados, tanto na Ciência de Dados, quanto na Gestão de Dados. Conforme Corrêa (2016, pág. 391) "a coleta, organização e preservação da informação tem sido confiada a bibliotecários(as) de diferentes setores que trabalham com os fluxos de informação; a articulação dos planos de gestão de dados é simplesmente uma manifestação moderna destas funções". Diante disso, faz-se o seguinte questionamento: Como a Biblioteconomia e a Ciência da Informação estão discutindo aspectos relativos a (ao) bibliotecário(a) de dados?

A partir do questionamento feito anteriormente esta pesquisa tem como objetivo analisar a produção científica nacional sobre os aspectos concernentes à atuação do(a) bibliotecário(a) no contexto dos dados.

A relevância desta pesquisa decorre, em linhas gerais, da necessidade de compreender como os pesquisadores da área de Biblioteconomia e Ciência da Informação têm discutido a temática, considerando a sua relevância para a formação de futuros profissionais bibliotecários(as). Além disso, esta pesquisa também evidencia o status das pesquisas sobre o tema,

1 Compreendida como "a base para o pensamento das ciências orientadas ao uso intensivo dos dados" (SEMELER, 2017, p. 7).

2 "Conjunto de habilidades, métodos, técnicas e tecnologias da Estatística e da Ciência da Computação, utilizadas para extrair conhecimento de dados e para a criação de produtos e serviços a partir de dados" (SEMELER, 2017, p. 7).

elucidando os principais enfoques dos estudos sobre Biblioteconomia de dados.

2. Ciência da Informação e Ciência de Dados

Durante a Segunda Guerra Mundial um grande número de pessoas passou a trabalhar em processos de coleta, seleção, processamento e disseminação de informações que fossem relevantes para o esforço de ganhar a guerra (FREIRE, 2006, p.10). A partir desse fato, o conceito de Ciência da Informação (CI) surgiu após os problemas gerados pelo volume e pelo valor das informações produzidas durante e após o conflito. A Ciência da Informação está preocupada com o corpo de conhecimentos relacionados à origem, à coleção, à organização, ao armazenamento, à recuperação, à interpretação, à transmissão, à transformação e à utilização da informação (BORKO, 1968, p. 3).

De acordo com Saracevic (1996, p.42) a Ciência da Informação consiste em uma área interdisciplinar, estando inexoravelmente ligada à tecnologia da informação. Ainda de acordo com o autor, a Ciência da Informação é um campo dedicado às questões científicas e à prática profissional, voltadas para os problemas da efetiva comunicação do conhecimento e de seus registros entre os seres humanos, no contexto social, institucional ou individual do uso e das necessidades de informação.

A Ciência de Dados, por sua vez, consiste em um campo interdisciplinar que envolve três domínios de conhecimento: Programação de Computadores; Estatística e Matemática; e

Domínio do Conhecimento (RAUTENBERG; CARMO, 2019). Com esses domínios, os cientistas de dados devem saber extrair informação e conhecimento úteis a partir de grandes volumes de dados estruturados ou não, resolvendo problemas de negócio e melhorando a tomada de decisão. Esse profissional deve desenvolver diversas habilidades relacionadas aos domínios do conhecimento supracitados como, por exemplo, análises estatísticas, mineração de dados, uso de aprendizado de máquina, ferramentas de visualização de dados (criação de gráficos e painéis com *Tableau*, *Power BI*), habilidades em programação (*Python*, R, SQL) e uso de *frameworks* de *Big Data*, como *Hadoop* e *Apache Spark*. Suas principais responsabilidades são desenvolver, testar e manter algoritmos de aprendizado de máquina; realizar análise exploratória de dados detectando e tratando anomalias nos dados; fazer inferência e interpretação das análises e apresentar os resultados. Objetivando dar prosseguimento a essas discussões, a seguir serão apresentados aspectos concernentes à Biblioteconomia de Dados.

2.1 Biblioteconomia de dados

A biblioteconomia de dados é um termo que vem sendo discutido desde a década de 1960 na América do Norte e Europa, ou seja, não é um termo novo (SEMELER; PINTO, 2020). Segundo os autores, a biblioteconomia de dados tem ênfase na criação de serviços relacionados ao uso de dados digitais, repositórios de dados, curadoria de dados e gestão de dados nas bibliotecas, além de atividades ligadas ao gerenciamento e à

curadoria de todos os tipos de dados, sendo seu foco o tratamento, a gestão e a curadoria de dados de pesquisas em qualquer disciplina científica (SEMELER, 2017, p.65).

Uma segunda definição encontra-se no site da *Australian National Data Service* (ANDS), porém mais restritiva quanto à atuação do(a) bibliotecário(a) de dados, sendo "[…] profissionais bibliotecários engajados na gestão de dados de pesquisa, usando dados de pesquisa como recurso e dando suporte a pesquisadores nas suas atividades" (ANDS, 2018, tradução nossa). Também se ressalta que a biblioteconomia de dados é um campo interdisciplinar, sendo uma interseção entre Ciência da Informação, *e-Science* e Ciência de dados (SEMELER; PINTO, 2020).

Cabe ressaltar que a biblioteconomia de dados não deve ser somente restrita aos dados de pesquisas (provenientes de pesquisas científicas), diante do fato de que as organizações utilizam dados em seus modelos de negócio e atuação para geração de informação e tomadas de decisão. Portanto, o papel do(a) Bibliotecário(a) de Dados é essencial para que haja um uso efetivo da informação (neste caso, codificada em conjuntos de dados) pelas organizações (como as governamentais), no processo de busca por soluções de problemas (CALLONI, 2020).

Considerando a incipiência e atualidade do tema, bem como o fato de que a Biblioteconomia de Dados não deve ser restrita apenas aos dados de pesquisa, faz-se necessário compreender de que maneira os pesquisadores da área de Biblioteconomia e

Ciência da Informação têm discutido o tema na literatura nacional. Para tanto, serão utilizados os procedimentos metodológicos descritos a seguir.

3. Procedimentos metodológicos

Esta pesquisa é caracterizada como bibliográfica, exploratória, descritiva com abordagem qualitativa. Para atender o objetivo geral deste estudo, foi realizada a recuperação, análise e mapeamento da produção científica nacional sobre a temática a partir das pesquisas bibliográfica, exploratória, descritiva com abordagem qualitativa. Com isso, foram realizadas buscas na Base de Dados em Ciência da Informação (BRAPCI) e na Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), no mês de maio de 2021, visando recuperar as produções científicas nacionais através do emprego dos seguintes descritores em ambas as fontes citadas anteriormente: "Biblioteconomia de Dados" e "Bibliotecário de Dados".

A seleção das bases de dados BRAPCI e BDTD se deu pelo fato de que a primeira agrega as produções científicas publicadas em artigos de periódicos da área de Biblioteconomia e Ciência da Informação, atualmente disponibilizando 19.255 (dezenove mil duzentos e cinquenta e cinco) textos publicados em 57 (cinquenta e sete) periódicos nacionais impressos e eletrônicos. A segunda fonte, por sua vez, integra e dissemina, em um só portal de busca, os textos completos de Teses e Dissertações defendidas nas instituições brasileiras de ensino e pesquisa em diversas áreas do

conhecimento, dentre elas a Biblioteconomia e a Ciência da Informação.

Seguindo essa linha de raciocínio, o total de produções científicas recuperadas na BRAPCI mediante o emprego dos termos citados anteriormente foi no total de 105 (cento e cinco) artigos de periódicos. A maioria dessas produções científicas recuperadas (oitenta), apesar de apresentar pelo menos um dos termos pesquisados, não se articulavam com este estudo, tendo seus conteúdos dedicados à avaliação de bases de dados; estudos bibliométricos realizados em bases de dados; bem como mapeamento do uso de metadados. Desse total, apenas 25 (vinte e cinco) artigos de periódicos possuíam articulação direta com a Biblioteconomia de Dados e o(a) Bibliotecário(a) de Dados, essas produções foram consideradas nas análises desta pesquisa.

Em nítido contraste, na BDTD foram recuperadas apenas duas produções científicas, sendo uma Dissertação e uma Tese sobre a temática citada anteriormente. Nessa perspectiva, o corpus final para a realização da análise da produção científica nacional foi constituído de 27 (vinte e sete) trabalhos, sendo 25 (vinte e cinco) artigos de periódicos, uma Dissertação e uma Tese.

Como procedimento para a realização da análise dessas pesquisas utilizou-se o método da análise de conteúdo, possibilitando tornar compreensíveis os principais aspectos enfatizados nos estudos sobre a temática. Essa análise foi realizada manualmente através de uma leitura técnica baseada na

norma ABNT NBR 12676, que trata especificamente sobre métodos para análise de documentos e determinação dos seus assuntos. Para tanto, foram considerados, especialmente: título e subtítulo, resumo, palavras-chave, introdução, ilustrações, diagramas, tabelas e explicações, resultados, considerações finais e referências bibliográficas das produções científicas analisadas.

Com isso, foi possível identificar o status da pesquisa sobre o tema, bem como evidenciar os principais enfoques, tendências e perspectivas de estudos sobre o assunto, aspectos que são elucidados na seção seguinte.

4. Análise da produção científica nacional sobre Biblioteconomia de Dados

Diante das análises realizadas, verificou-se que as produções científicas nacionais sobre Bibliotecário de Dados estão distribuídas em um intervalo de tempo entre 2004 a 2020, havendo uma incidência de publicações do ano de 2016 em diante. Os anos de 2019 e 2020 ganham destaque como os períodos que apresentaram maior quantidade de publicações científicas, conforme é possível observar no Quadro 1.

Seguindo a linha de raciocínio discutida no referencial teórico desta pesquisa, acredita-se que a incidência de publicações nos últimos anos acerca da temática pode ser decorrente da necessidade emergente da atuação do(a) Bibliotecário(a) nesse cenário, haja vista que o seu perfil contempla diversas competências e habilidades que podem contribuir para o ciclo de

vida dos dados, mediante aplicação de estratégias em diferentes contextos.

Quadro 1 – Quantidade de publicações por ano

Ano e publicação	Quantidade
2004	1
2012	1
2014	1
2015	1
2016	3
2017	2
2018	3
2019	8
2020	7

Fonte: Autores.

Os periódicos científicos eletrônicos e anais de eventos técnico-científicos se configuram como fontes importantes para a comunicação científica, tornando disponível para a comunidade os resultados das pesquisas em diferentes áreas do conhecimento. Com base nos resultados da pesquisa realizada na BRAPCI, também foi possível perceber os periódicos da área de Ciência da Informação cujos trabalhos analisados encontram-se publicados, são eles: Múltiplos olhares em Ciência da Informação; Informação e Informação; Ciência da Informação; Em Questão; Biblos; Revista ACB: Biblioteconomia em Santa Catarina; Reciis: Revista Eletrônica de Comunicação, Informação e Inovação em Saúde; Revista Digital de Biblioteconomia e Ciência da

Informação (RDBCI); Perspectivas em Ciência da Informação; Encontros Bibli; Informação em Pauta; AtoZ: novas práticas em Informação e Conhecimento; Anais do Encontro Regional de Estudantes de Biblioteconomia, Documentação, Gestão e Ciência da Informação (2018). O Quadro 2 apresenta essas informações de maneira sintetizada e organizada com base na incidência das produções.

Quadro 2 – Quantidade de produções por canais de publicação

Fontes das publicações	**Quantidade de produções**
Informação e Informação	6
Ciência da Informação	4
Em Questão	3
Reciis- Rev. Eletr. Comun. Inf. Inov. Saúde	2
RDBCI: Rev. Dig. Bibliotecon. Cienc. Inf.	2
Múltiplos olhares em Ciência da Informação	1
Biblos	1
Revista ACB: Biblioteconomia em Santa Catarina	1
EREBD 2018	1
Perspectivas em Ciência da Informação	1
Encontros Bibli	1
Inf. Pauta	1
AtoZ	1

Fonte: Autores.

Observa-se que o periódico que possui maior número de publicações sobre Bibliotecário de Dados é Informação e

Informação, periódico científico eletrônico do Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual de Londrina (PPGCI/UEL) com um total de seis publicações acerca da temática. Em seguida, a revista Ciência da Informação, periódico científico do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) aparece no quadro com um destaque de três publicações sobre a temática. O periódico "Em Questão" (PPGCI/UFRGS), por sua vez, apresenta o total de três publicações sobre Bibliotecário de Dados e, em seguida, aparecem as demais revistas citadas anteriormente.

Verificou-se ainda que a maioria das produções científicas recuperadas e analisadas concentram-se, em linhas gerais, em investigar questões concernentes à gestão de dados de pesquisa. Embora tenha sido mencionado anteriormente que a Biblioteconomia de dados não deve ser restrita apenas aos dados de pesquisas (provenientes de pesquisas científicas), constatou-se que os pesquisadores da área de Ciência da Informação têm dedicado os seus estudos a esse enfoque temático.

Além das produções científicas disponíveis em periódicos científicos eletrônicos e anais de eventos técnico-científicos, durante a busca e recuperação das produções científicas na BDTD foi possível identificar dois trabalhos provenientes de Programas de Pós-Graduação em Ciência da Informação de duas Universidades do Brasil, são elas: Universidade Federal de Santa Catarina (UFSC) e Universidade Federal de Sergipe (UFS).

O Quadro 3 apresenta os enfoques temáticos das publicações e os respectivos autores cujos trabalhos foram analisados mediante a realização da análise de conteúdo. Em linhas gerais, é possível visualizar a diversidade de enfoques temáticos nas produções científicas analisadas, uma vez que foram identificadas 12 (doze) temas nas pesquisas analisadas. Percebe-se que a maioria das publicações nacionais (22%) concentram-se em investigar aspectos concernentes à gestão de dados de pesquisa, sobretudo no que diz respeito às atividades do(a) Bibliotecário(a) e as competências necessárias para a sua atuação nesse campo. Essas pesquisas ressaltam a importância de esse(a) profissional atuar desde as fases que antecedem a pesquisa até a pesquisa finalizada, preocupando-se com a organização, tratamento, gestão e preservação dos dados e com os direitos do acesso, uso e local de armazenamento.

Quadro 3 – Enfoques temáticos das produções científicas analisadas

Enfoques temáticos	**Autores**
Gestão de Dados de Pesquisa	Santa Anna; Dias; Maculan (2019) Carvalho e Queiroz (2019) Sales et al. (2019) Lima; Pinto e Farias (2019) Sales e Sayão (2012) Corrêa (2016) Resende e Bax (2020)
Repositórios Digitais de Dados de Pesquisa	Sayão; Sales (2016) Tartarotti; Dal'Evedove; Fujita (2019) Paganine; Amaro (2020)

Enfoques temáticos	Autores
Práticas de gerenciamento e publicação de dados no contexto da Web Semântica com dados abertos interligados	Silveira (2020) Guandaline e Santos (2018)
Proteção e privacidade de dados em Unidades de Informação	Vignoli et al. (2020) Serra (2015)
Governança de Dados	Espíndola et al. (2018)
Dados de pesquisa no contexto da Biblioteconomia de Dados	Semeler e Pinto (2019)
Tratamento de dados terminológicos	Xavier et al. (2004)
Mineração de dados como estratégia para a composição dos catálogos de periódicos *open access*	Silva; Queiroz e Simões (2019)
Mineração de dados científicos	Ferreira e Corrêa (2020)
Modelos de busca, acesso, recuperação e apropriação da informação na Web de Dados	Palleta e Gonzales (2019)
A atuação do bibliotecário no tratamento dos dados provenientes da *e-Science*	Costa e Cunha (2014)
e-Science e *Data Science* no contexto da Ciência orientada a dados	Bufrem et al. (2016)

Fonte: Autores.

Em consonância de evidências, 13% das produções científicas recuperadas tratam sobre repositórios de dados de pesquisa, os quais são considerados como sistemas de recuperação da informação contemporâneos voltados para apoiar

o armazenamento, tratamento, recuperação, acesso e uso dos dados de pesquisa.

De outro modo disposto, 9% das produções científicas recuperadas tratam sobre as práticas de gerenciamento e publicação de dados na Web semântica com dados abertos interligados, perspectivas que devem ser consideradas nos sistemas contemporâneos. Por outro lado, 9% dos trabalhos recuperados preocupam-se em investigar questões relacionadas à proteção e privacidade de dados em unidades de informação.

Os demais trabalhos recuperados e analisados possuem os seguintes enfoques: dados de pesquisa no contexto da Biblioteconomia de dados; tratamento de dados terminológicos; mineração de dados como estratégia para a composição de catálogos de periódicos *Open Access*; mineração de dados científicos; modelos de busca, recuperação; acesso e apropriação da informação na web de dados; a atuação do Bibliotecário(a) no tratamento dos dados provenientes da *e-Science*; *e-Science* e *Data Science* no contexto da Ciência orientada a dados.

Diante do exposto, é possível observar que as produções científicas evidenciam a importância da atuação do(a) Bibliotecário(a), além do contexto dos dados de pesquisa, nas práticas de governança, gerenciamento, mineração, tratamento, organização, busca, acesso, uso, reúso e ligação dos dados, sobretudo no contexto da Web Semântica.

5. Considerações finais

Ao analisarmos as produções recuperadas, foi possível perceber o status das pesquisas nacionais sobre Bibliotecário de Dados. Constatou-se um grande foco nos últimos anos com relação às pesquisas voltadas para a gestão de dados de pesquisa e repositórios digitais de dados de pesquisa; bem como para as práticas de gerenciamento e publicação de dados no contexto da Web Semântica com dados abertos interligados.

Nesse sentido, constata-se a necessidade de investigação e apresentação das competências e habilidades do(a) Bibliotecário para o desenvolvimento de práticas concernentes à alfabetização de dados; bem como a organização, tratamento e recuperação de dados, sobretudo no contexto da Web semântica e dos Sistemas híbridos de Organização do Conhecimento. Esses enfoques podem ser trabalhados de maneira verticalizada por profissionais e pesquisadores da área em pesquisas futuras.

Referências

BORKO, H. Information Science: What is it? **American Documentation**, v.19, n.1, p.3-5, Jan. 1968.

CALLONI, R. Bibliotecário de Dados. In: Silva, F.C.C. **O perfil das novas competências na atuação bibliotecária**. Florianópolis, SC: Rocha Gráfica e Editora, 2020. p. 339-362.

CORRÊA, F. C. O papel do bibliotecário na gestão de dados científicos. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v.14, n.3, p. 387-406, 2016.

FREIRE, G. Ciência da informação: temática, história e fundamentos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.11, n.11, p. 6-19, 2006.

MANYICA, J.; CHUI, M.; BROWN, B.; BUGHIN, J.; DOBBS, R.; ROXBURGH, C.; BYERS, A. H. **Big data**: The next frontier for innovation, competition, and productivity, 2011.

RAUTENBERG, S.; CARMO, P. R. Big Data e Ciência de Dados: complementariedade conceitual no processo de tomada de decisão. **BJIS**: Research Trends. v.1, n.13, p.56-67, 2019.

SARACEVIC, T. Ciência da Informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jul. 1996.

SEMELER, A. R. Ciência da Informação em contextos de e-Science: Bibliotecários de Dados em tempos de Data Science. 2017. **Tese** (Doutorado em Ciência da Informação) - Universidade Federal de Santa Catarina, 2017.

SEMELER, A. R.; PINTO, A. L. Data librarianship as a field study. **Transiformação**, Campinas, v. 32, out. 2020.