

UNIVERSIDADE DE SÃO PAULO
ESCOLA DE COMUNICAÇÃO E ARTES
DEPARTAMENTO DE INFORMAÇÃO E CULTURA

NATHALIA CASTILHO ESCARLATE

Biblioteconomia de Dados e a atuação do profissional da informação

São Paulo
2020

NATHALIA CASTILHO ESCARLATE

Biblioteconomia de Dados e a atuação do profissional da informação

Trabalho de conclusão de curso de graduação em Biblioteconomia, apresentado ao Departamento de Informação e Cultura da Escola de Comunicação e Artes da Universidade de São Paulo.

Orientador: Prof. Dr. Francisco Carlos Paletta

São Paulo

2020

Catálogo na Publicação
Serviço de Biblioteca e Documentação
Escola de Comunicações e Artes da Universidade de São Paulo
Dados inseridos pelo(a) autor(a)

Escarlate, Nathalia Castilho
Biblioteconomia de Dados e a atuação do profissional
da informação / Nathalia Castilho Escarlate; orientador,
Francisco Carlos Paletta. - São Paulo, 2020.
70 p.: il.

Trabalho de Conclusão de Curso (Graduação) -
Departamento de Informação e Cultura / Escola de
Comunicações e Artes / Universidade de São Paulo.
Bibliografia

1. Biblioteconomia de Dados. 2. Bibliotecário de
dados. 3. Ciência de Dados. 4. Big Data. I. Paletta,
Francisco Carlos. II. Título.

CDD 21.ed. - 020

Nome: ESCARLATE, Nathalia Castilho

Título: Biblioteconomia de Dados e a atuação do profissional da informação

Trabalho de conclusão de curso em Biblioteconomia apresentado à Escola de Comunicação e Artes da Universidade de São Paulo.

Aprovado em: 23 de Outubro de 2020.

Banca Examinadora

Prof. Dr. Francisco Carlos Paletta (Orientador)

Instituição: Universidade de São Paulo - Escola de Comunicação e Artes –
Departamento de Informação e Cultura

Julgamento: _____

Profa. Dra. Cibele Araújo Camargo Marques dos Santos

Instituição: Universidade de São Paulo - Escola de Comunicação e Artes –
Departamento de Informação e Cultura.

Julgamento: _____

Prof. Dr. Ivan Claudio Pereira Siqueira

Instituição: Universidade de São Paulo - Escola de Comunicação e Artes –
Departamento de Informação e Cultura.

Julgamento: _____

AGRADECIMENTOS

A origem deste trabalho remete ao ano de 2017. Na palestra “Curadoria digital de dados de pesquisa”, organizada na Universidade de São Paulo pelo Prof. Dr. Luis Fernando Sayão, encontrei o objetivo profissional de atuar como bibliotecária de dados. Depois disso, orientei tanto meus estudos quanto o trabalho profissional para relacionar a biblioteconomia à tecnologia e aos dados, sempre que possível.

Atualmente, encerro este ciclo da graduação e inicio meu exercício laboral. Agradeço às diversas pessoas que contribuíram direta e indiretamente para a minha formação acadêmica e profissional.

Agradeço à minha família, que me incentivou, encorajou e deu todo o suporte possível durante essa jornada com tantos desafios. Seu apoio foi fundamental para o meu fortalecimento e crescimento.

Agradeço às lições dos gestores que marcaram a minha trajetória profissional, cujo aprendizado foi essencial para o meu desenvolvimento.

Agradeço aos meus amigos, em especial a André Bayeux, Iago Fernandes de Sousa, Monique Rocha, Priscila Mansoldo, Simone Andriani, Tadeu Inocêncio Freitas, Tibério Cruz e Vinicius Motta, com quem sempre conversei sobre as possibilidades de futuro profissional, das incertezas, dos riscos a serem tomados e dos desafios da vida, conversas que me fizeram acreditar até no impossível.

Agradeço aos meus professores que fizeram com que eu me apaixonasse pela profissão, tanto como bibliotecária quanto como analista de sistemas da informação.

Agradeço ao meu orientador, Prof. Dr. Francisco Carlos Paletta, pelo incentivo, pela orientação, por ter acreditado e valorizado o tema proposto.

Agradeço à Universidade de São Paulo por todo este processo de formação acadêmica que impactou a minha transformação como estudante, profissional e cidadã.

RESUMO

ESCARLATE, Nathalia C. **Biblioteconomia de Dados e a atuação do profissional da informação**. 2020. 70 p. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia) – Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 2020.

A evolução da tecnologia proporcionou a produção de dados em grande volume, de modo que sua variedade se tornasse relevante na sociedade devido ao seu potencial de aperfeiçoar o processo de tomada de decisões, aumentar a produtividade e viabilizar o desenvolvimento de soluções, serviços e produtos aplicados a diversas esferas da sociedade. Este fenômeno vem a ser conhecido como Big Data. Este aglomerado de dados demanda que profissionais se apropriem de saberes e práticas para o tratamento e extração do valor das informações, visando ao benefício de indivíduos e organizações no processo de tomada de decisão. Nesse sentido, tem-se a Biblioteconomia de Dados, uma próspera área de atuação que combina as técnicas tradicionais da Biblioteconomia e da Ciência de Dados para tratar os dados oriundos do Big Data. Com tal premissa, o presente trabalho tem o objetivo de apresentar a definição de Biblioteconomia de Dados, as habilidades e competências do profissional da área, e as suas esferas de atuação e formação. Tendo em vista o objetivo exposto, faz-se uma revisão de literatura a fim de apresentar conceitos fundamentais para a compreensão da Biblioteconomia de Dados, sendo eles dado, informação e conhecimento, a apresentação de conceitos e impactos do Big Data, a apresentação da Ciência de Dados como disciplina complementar à Biblioteconomia de Dados. Essas reflexões podem contribuir para que se compreenda que o bibliotecário de dados deve se apropriar cada vez mais das demandas da sua era, podendo desempenhar um papel de protagonista nesse processo.

Palavras-chave: Biblioteconomia de Dados. Bibliotecário de dados. Ciência de Dados. Big Data.

ABSTRACT

ESCARLATE, Nathalia C. **Data librarianship and the role of the information professional**. 2020. 70 p. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia) – Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 2020.

The evolution of technology has promoted the production of data in large volumes so that become relevant in society due to its potential to improve the decision-making process, increase productivity, and enable the development of solutions, services, and products in diverse spheres of society, the phenomenon that comes to be known as Big Data. The agglomerate of data produced by the Big Data requires that professionals appropriate themselves of knowledge and practises to treat this data in order to extract their value, vising to benefit individuals and organisations in the decision-making process. In this sense, there is Data Librarianship, a flourishing area of expertise that combines traditional techniques of Librarianship and Data Science to treat data from Big Data, and with this premise, the objective is to present the Data Librarianship, its definitions and the skills and competencies of that professional, and its areas of education and training. Giving the exposed objective, a bibliographic survey is made in order to present fundamental concepts for understanding Data Librarianship, as the concept of data, information and knowledge, a presentation of Big Data concepts and impacts, and the presentation of Data Science as a complementary discipline to Data Librarianship. These concepts contribute to the understanding that the Data Librarian should be appropriating themselves of the demands of the Age of Data, and they can perform the main role in these processes.

Keywords: Data Librarianship. Data Librarian. Data-Science. Big Data.

LISTA DE SIGLAS

BRAPCI	Base de Dados em Ciência da Informação
ISTA	Information Science & Technology Abstracts
LISA	Library & Information Science Abstracts
LISTA	Library, Information Science & Technology Abstracts
IASSIST	International Association for Social Science Information Service and Technology
RBBD	Revista Brasileira de Biblioteconomia e Documentação

LISTA DE ILUSTRAÇÕES

Figura 1 – Visualização integrada do Big Data.....	22
Figura 2 – Diagrama de Venn da Ciência de Dados	33
Figura 3 – Diagrama de Venn da Biblioteconomia de Dados no contexto de e- Science.....	56
Figura 4 – O curso de Ciência de Dados e de Informação e suas características técnicas	58

LISTA DE TABELAS

Tabela 1 – Tipos de cientistas de dados	32
Tabela 2 – Habilidades do bibliotecário de dados	41
Tabela 3 – Planejamento e controle de gerenciamento de dados.....	47
Tabela 4 – As etapas da análise de dados e as contribuições do profissional da informação.....	50
Tabela 5 – Contribuições do bibliotecário de dados nas etapas da gestão de dados	54

SUMÁRIO

1	INTRODUÇÃO	11
1.1	Objetivos gerais.....	11
1.2	Objetivos específicos.....	11
1.3	Metodologia	12
2	DADO, INFORMAÇÃO E CONHECIMENTO: CONCEITOS	14
2.1	Dados	15
2.2	Informação.....	16
2.3	Conhecimento	18
3	BIG DATA: DEFINIÇÕES E IMPACTOS	19
3.1	Definindo o Big Data.....	19
3.1.1	Volume.....	20
3.1.2	Velocidade	20
3.1.3	Variedade	21
3.2	O desenvolvimento das tecnologias da informação propulsoras do Big Data 22	
3.3	Da economia de dados à 4ª revolução	23
4	BIBLIOTECONOMIA DE DADOS.....	27
4.1	Ciência de Dados	28
4.1.1	Definindo a Ciência de Dados	29
4.1.2	Subáreas da Ciência de Dados	31
4.1.3	Desafios e perspectivas da Ciência de Dados.....	34
4.2	Biblioteconomia de dados.....	35
5	BIBLIOTECÁRIO DE DADOS: O PROFISSIONAL DA INFORMAÇÃO	39
5.1	Habilidades e competências.....	39
5.2	Áreas de atuação	42
5.2.1	Literacia de dados.....	43
5.2.2	Governança de dados.....	46
5.2.3	Análise de dados	48
5.2.4	Gestão de dados e repositórios de pesquisa	51
6	A FORMAÇÃO EM BIBLIOTECONOMIA DE DADOS.....	57
7	CONCLUSÃO	60
	REFERÊNCIAS.....	61

1 INTRODUÇÃO

A evolução da tecnologia proporcionou a produção de dados em grande volume, de modo que sua variedade se tornasse relevante na sociedade devido ao seu potencial para aperfeiçoar o processo de tomada de decisões, aumentar a produtividade e viabilizar o desenvolvimento de soluções, serviços e produtos aplicados a diversas áreas e contextos, tanto na esfera corporativa, quanto na acadêmica ou governamental. Consequentemente, a produção de dados tem se prestado a diversos propósitos na contemporaneidade (MARQUESONE, 2016).

Considerando-se a premissa do protagonismo e da demanda da usabilidade de dados em diversos contextos da sociedade, destaca-se a Biblioteconomia de Dados como uma proeminente área de atuação do profissional da informação. Surgem novas demandas por informações de dados e diversas oportunidades desencadeadas pelo Big Data¹.

A Biblioteconomia de Dados é um ramo da Biblioteconomia que se preocupa com a demanda de indivíduos e organizações por informações de dados, que surge em meio às revoluções tecnológicas dos últimos tempos. O profissional deste ramo se vale de ferramentas e dispositivos desenvolvidos no âmbito das recentes transformações tecnológicas para atuar. A fim de compreender suas possibilidades de inserção no mercado, faz-se necessário mapear a área de atuação, o perfil, as habilidades e competências exigidas do bibliotecário de dados.

1.1 Objetivos gerais

Tem-se como objetivo geral identificar na literatura existente a definição de Biblioteconomia de Dados, assim como as possíveis áreas de atuação do profissional desta área, o bibliotecário de dados.

1.2 Objetivos específicos

Delimitam-se os seguintes objetivos específicos que norteiam o mapeamento proposto para este trabalho:

¹ A definição de Big Data será abordada no capítulo 4, intitulado “Big Data: definições e impactos”.

- a) apresentar a definição de Biblioteconomia de Dados;
- b) mapear as áreas de atuação do bibliotecário correlatas ao Big Data e à Ciência de Dados (*Data Science*);
- c) realizar o levantamento das habilidades e competências exigidas para a atuação profissional nessas áreas de atuação;
- d) explorar as perspectivas da Biblioteconomia de Dados;
- e) refletir sobre as novas exigências da formação em Biblioteconomia para o aprimoramento do profissional bibliotecário de dados e sua atuação no mercado de trabalho atual.

1.3 Metodologia

Visando à identificação das áreas de atuação do bibliotecário de dados, realizou-se uma pesquisa exploratória, uma vez que esse tipo de investigação contribui para “[...] o aprimoramento de ideias ou [a] descoberta de instituições” (GIL, 2004, p. 41).

A pesquisa foi delimitada em duas etapas que guiaram o seu desenvolvimento. A primeira, voltada ao referencial teórico, apresentado nos capítulos dois e três; a segunda, dedicada à proposta da pesquisa, inserida nos capítulos quatro a seis. Uma vez estabelecida essa estrutura do trabalho, realizou-se o levantamento de fontes de informação bibliográficas em três etapas.

O método de investigação de fontes de informação se deu com o emprego de bases de dados, pesquisa da bibliografia nos materiais pertinentes encontrados e busca nos portais Google Scholar e Portal de Busca Integrada da Universidade de São Paulo, a fim de identificar possíveis recursos informacionais não indexados nas bases de dados selecionadas.

A primeira etapa do levantamento bibliográfico destinou-se à compreensão de conceitos, fundamentos de dado, informação e conhecimento (capítulos dois). Já a segunda etapa foi dedicada aos materiais que tratam especificamente da Biblioteconomia de Dados (capítulos quatro a seis). Em ambas etapas, o levantamento foi realizado nas seguintes bases de dados: Base de Dados em Ciência da Informação (BRAPCI), Revista Brasileira de Biblioteconomia e Documentação (RBBD), Information Science & Technology Abstracts (ISTA), Library & Information Science Abstracts (LISA) e Library, Information Science & Technology Abstracts (LISTA). Foram identificados artigos, teses, anais e livros publicados no interior do recorte temporal compreendido

entre os anos 2015 e 2020. A busca se valeu das palavras-chave “Análise de dados”, “Bibliotecário de dados”, “Biblioteconomia de dados”, “Ciência de Dados”, “Big Data”, “Data Librarian”, “Data librarianship”, “Data science”, “Data Science” AND “Librarian”, “Data Science” AND “Librarianship”, dentre outras.

A terceira etapa se baseou na pesquisa de conceitos, do histórico, de impactos e das definições de Big Data e Ciência de Dados (capítulo três e seção 4.2, respectivamente). Para tanto, valeu-se do levantamento bibliográfico realizado nas bases de dados JSTOR, Emerald Insights, Data Science Journal, Journal of Big Data, Elsevier e IEEE Explorer, que identificou artigos, teses, anais e livros publicados entre os anos de 2015 e 2020, que continham as palavras-chave “Big Data” e “Data Science”, relacionando-as, por meio de operadores booleanos, aos termos “history”, “definition”, “value”, “concept”, “perspective”, dentre outros.

Por fim, para atingir o objetivo geral ora proposto, o trabalho foi dividindo em seis capítulos, para além desta introdução e da conclusão. O capítulo dois apresenta as definições de “dado”, “informação” e “conhecimento” como termos-chave tanto para a Biblioteconomia quanto para o Big Data e a Ciência de Dados.

O capítulo três expõe a definição de Big Data como um fenômeno informacional e aborda as tecnologias da informação que proporcionaram o seu desenvolvimento, bem como o seu impacto na economia e na era dos dados. O capítulo quatro reflete sobre as competências do bibliotecário para a atuação nas demandas de informações de dados, apresentando seu histórico, definição e papel. Além disso, aponta a complementaridade deste setor com o da Ciência de Dados, motivo pelo qual se apresentam também a definição de Ciência de Dados, suas subáreas, desafios e perspectivas.

No capítulo cinco, trata-se do tema central do trabalho, com uma discussão acerca das áreas de desenvolvimento da Biblioteconomia de Dados. O capítulo seis discorre sobre as perspectivas do mercado de trabalho da Biblioteconomia de Dados, formação acadêmica e iniciativas existentes que contribuem para o crescimento da área. Por fim, no capítulo sete, apresentam-se as considerações finais e a conclusão.

2 DADO, INFORMAÇÃO E CONHECIMENTO: CONCEITOS

A sociedade no fim do século XX passou por transformações que foram alavancadas com o crescente desenvolvimento das Tecnologias da Informação e Comunicação (TICs), fazendo com que o papel da informação e da comunicação se tornassem cada vez mais relevantes para os processos de transformação da sociedade (KERBAUY, 2009).

O impacto dessas mudanças introduziu, *a posteriori*, a expressão “sociedade informacional”, definida por Castells (1999, p. 65) como uma “forma específica de organização social em que a geração, o processamento e a transição da informação tornam-se as fontes fundamentais de produtividade e poder devido às novas condições tecnológicas surgidas nesse período histórico”. Este novo modo de organização social influenciou tanto a economia quanto a sociedade, desencadeando um novo paradigma, o da tecnologia da informação. As relações promovidas neste contexto são fortemente mediadas pela interação do indivíduo com a informação, tendo como principal suporte a tecnologia, especialmente pela lógica das redes (WERTHEIN, 2000). A tecnologia proporcionou a convergência de diferentes áreas do conhecimento, estimulando processos cada vez mais multidisciplinares e colaborativos, em diversas camadas da sociedade, ao nível global (CASTELLS, 1999).

Mais recentemente, o desenvolvimento da tecnologia da informação, combinado à eletrônica e à interconectividade de dispositivos e indivíduos na internet, contribuiu para “o crescimento, a disponibilidade e o uso exponencial de informações digitais estruturadas e não estruturadas” (BARRETO, 2014, p. 5). Eis o que promoveu o surgimento do fenômeno conhecido como Big Data, caracterizado pelo elevado volume de informação, complexidade e velocidade de coleta, processamento e compartilhamento de dados (BARRETO, 2014). O Big Data vem transformando o modo pelo qual indivíduos e organizações interagem com a informação e o conhecimento, articulando um:

[...] novo comportamento para lidar com a informação, considerando a nova visão de produção, distribuição e consumo de conteúdos digitais de grande volume e necessidade de processamento online. As configurações do chamado “Big Data” representam uma mudança no modelo de formações dos estoques e fluxos da informação. Sem ser uma nova tecnologia, ou mesmo um novo conceito, o termo ‘Big Data’ é usado para descrever o crescimento, a disponibilidade e o uso exponencial de informações digitais estruturadas e não estruturadas. É caracterizado: pelo volume de informação, sua complexidade, variedade e pela velocidade de ordenação destes dados em uma superestrutura de informação (BARRETO, 2014, p. 5).

O Big Data desencadeou transformações por meio dos dados, elemento constituinte do fluxo da informação. Este pode ser representado pela tríade dado-informação-conhecimento, que exemplifica a transformação de dados em informação, e da informação em conhecimento (ROWLEY, 2007). O impacto deste fluxo pode ser percebido em diversas camadas da sociedade, motivo pelo qual é possível caracterizar o período contemporâneo como “a era dos dados”.

Em razão das transformações no modo como os indivíduos se informam e consolidam conhecimento, em decorrência da explosão informacional ocasionada por avanços tecnológicos, faz-se necessário definir as noções de “dados”, “informação”, e “conhecimento”, que mais uma vez estão no palco das mudanças da sociedade.

2.1 Dados

A definição de “dados” não é consensual. De toda forma, busca-se aqui apresentar brevemente algumas das vertentes que a expressão assume, com o objetivo de traçar uma interpretação que tangencie a Ciência da Informação.

Etimologicamente, a palavra “dado” se origina da palavra latina *datum* (LINDBERG, 2012 apud HJØRLAND, 2018), que significa “dar”. O sentido desta palavra remete à ideia de algo que foi dado pela natureza ao homem. Jensen (1950 apud HJØRLAND, 2018) apresenta a perspectiva de algo que foi “tomado” ou selecionado da natureza por cientistas com base naquilo que estava sendo observado por eles. Tal interpretação se assemelha à proposta por Machlup, que define “dados” como coisas que foram dadas a “analistas, investigadores e solucionadores de problemas; eles podem ser números, palavras, sentenças, registros, suposições; qualquer coisa data, não importa em qual forma e origem” (1984 apud HJØRLAND, 2018, tradução nossa).

“Dados” também podem se apresentar como “[...] substitutos simbólicos, linguísticos, matemáticos ou outros, que são aceitos como atributos para representação de pessoas, objetos, eventos e conceitos.” (BURCH *et al.*, 1983 apud SANTOS; SANTANA, 2015, p. 201). Do ponto de vista da informática, podem ser compreendidos como “um byte de informação codificando números, texto, imagem, som, vídeo [...]” (MOREIRA; CARVALHO; HORVÁTH, 2019, p. 17, tradução nossa). Contudo, há críticas a essa definição, que apontam não haver vantagens em reduzir todo o universo e as obras culturais, sociais e científicas a 0 e 1 (FURNER, 2016).

A classificação de dados é realizada com base na identificação de diversas propriedades que eles possuem, propriedades que os diferem uns dos outros. Essencialmente, pode-se dividir os dados em duas grandes categorias: escalar e composto. Dados escalares se caracterizam por apresentarem valores individuais, enquanto dados compostos, como o nome já indica, são aqueles que combinam diversos valores em um único objeto (DALE; WALKER, 1992).

Os dados escalares possuem caráter qualitativo e quantitativo, de tipo numérico, booleano e textual. Podem ser ordinais e contínuos. Por outro lado, os dados compostos podem ser não-estruturados, semiestruturados e estruturados (DALE; WALKER, 1992; JOHNSON, 2019). Os dados de tipo não-estruturado são aqueles que, em um contexto de armazenamento e recuperação, não possuem restrições à forma do objeto de dado, mas podem possuir padrões em comum. Esses tipos podem ser, por exemplo, datasets, tabelas de símbolos ou registros, imagens etc.

Em meio às diversas definições existentes e às tentativas de definir o conceito de dados, especialmente diante de transformações tecnológicas e informacionais que contribuem para fomentar discussões e suposições equívocas sobre essa temática, Hjørland (2018) propõe-se a definir “dados” a partir do conceito explicitado por Max Kaase. O autor estabelece que “dados são informações sobre as propriedades de unidades de análises”², fazendo-se compreender que se trata de elementos primários cujo agrupamento pode ser materializado como informação, a qual, ao ser aplicada a um contexto, gera conhecimento (SEMIDÃO, 2014).

2.2 Informação

O conceito de informação mostra-se cada vez mais central na contemporaneidade. O seu destaque faz luz não somente por conta dos avanços tecnológicos desde a Segunda Guerra Mundial, mas também por ter se tornado nato-digital (CAPURRO; HJØRLAND, 2007).

O entendimento do conceito de informação está atrelado ao seu significado e uso em cada domínio do conhecimento, fazendo com que apresente diferentes conotações (CAPURRO; HJØRLAND, 2007). Disso resulta a dificuldade de obter um conceito universalmente inequívoco, especialmente pela atual abundância informacional que

² “Data are information on properties of units of analysis.” (KAASE apud HJØRLAND, 2018).

marca a “era da informação” ou a “sociedade da informação” (BAWDEN; ROBINSON, 2012) .

A informação apresenta-se como “[...] um fenômeno humano. Envolve indivíduos transmitindo e recebendo mensagens no contexto de suas ações” (CAPURRO; HJØRLAND, 2007, p. 161). Ademais, ela “[...] comporta um elemento de sentido. É um significado transmitido a um ser consciente por meio de uma mensagem inscrita em um suporte espacial-temporal: impresso, sinal elétrico, onda sonora etc.” (LE COADIC, 1994, p. 5).

Embora não haja uma concepção universal do conceito, um modelo que aproxima diversas áreas é o da teoria matemática, com caráter quantitativo, apresentada por Claude Shannon em 1948. A informação toma a forma de uma mensagem a ser transmitida a partir de uma fonte que precisa de um meio para alcançar o seu destinatário, sofrendo ou não ruídos nesse processo (CAPURRO; HJØRLAND, 2007). Esta interpretação interessante é criticada por tratar de um modelo de informação que pode ser encarado como um processo de transmissão de mensagens.

De outro modo, a perspectiva da Ciência da Informação apresenta a noção discutida com duas facetas. A primeira, é a da informação como um objeto ou coisa; a segunda, a da informação dotada de caráter subjetivo, que carrega um significado que depende da interpretação cognitiva (CAPURRO; HJØRLAND, 2007).

Como objeto ou coisa, pode até mesmo se confundir com dados, pois essa faceta pode apresentar caráter quantitativo, tal como a descrição de um documento do qual indica-se a quantidade de páginas, número de palavras, dentre outros aspectos que contribuem para o entendimento do documento em si (CAPURRO; HJØRLAND, 2007). Contudo, ao analisar a carga e o significado informacional do documento, a informação sobre a sua relevância é mais difícil de se extrair, pois “[...] o significado é, entretanto, determinado nos contextos social e cultural” (CAPURRO; HJØRLAND, 2007, p. 193–194).

Como portadora de significado, tal como na explanação de Le Coadic (1994, p. 5), constitui-se como “[...] um conhecimento inscrito (gravado sob a forma escrita (impressa ou numérica), oral ou audiovisual [...]) [que] comporta um elemento de sentido. É um significado transmitido a um ser consciente por meio de uma mensagem inscrita em um suporte espacial-temporal”, cujo objetivo é a transmissão de conhecimento.

2.3 Conhecimento

O conceito de conhecimento é dotado de complexidade, uma vez que suas definições permeiam diversas áreas de estudo. Em verdade, mesmo dentro de cada domínio de análise a noção é amplamente discutida ao se tratar de sua natureza, formas de representação e manifestação (ROWLEY, 2007). Em termos genéricos, o conhecimento pode ser compreendido como a informação processada, tratada e dotada de significado. Trata-se de

[...] informação avaliada e organizada na mente humana, de modo a poder ser utilizada propositadamente. Neste sentido, o termo pode ser equiparado ao termo entendimento. No entanto, é geralmente usado de forma imprecisa para descrever fenômenos aos quais o termo 'informação' é muitas vezes associado (FEATHER; STURGES, 2003, p. 342 apud SEMIDÃO, 2014, p. 125).

Tal perspectiva corrobora com o fato de que o conhecimento sempre referencia informações (ROWLEY, 2007) e contempla entendimento, capacitação, experiências, habilidade e valores. Atrelado à informação, ele ainda pode ser interpretado como “uma informação ‘refinada’, apresentada na forma de uma estrutura maior, e ambos, informação e conhecimento, podem estar internalizados na mente de alguém ou encapsulados em algum tipo de documento” (BAWDEN; ROBINSON, 2012, p. 73, tradução nossa).

Alguns autores acreditam que o conhecimento adquire dois sentidos: o primeiro, do que é sabido; o outro, do que é conhecido. O conhecimento “sabido” faz referência àquele que é adquirido, mas não se torna familiar ao indivíduo, tal como o conhecimento tácito. Alguém sabe sobre algo, mas não obteve experiência o suficiente acerca dele. Já o conhecimento “conhecido” se dá quando o indivíduo obtém esta experiência (FRICKÉ, 2019).

Uma vez que o indivíduo adquire conhecimento e experiência sobre algo, desenvolve a inteligência, que pode ser utilizada para produzir novos dados, informações e conhecimentos acerca de um tema.

3 BIG DATA: DEFINIÇÕES E IMPACTOS

O desenvolvimento das tecnologias da informação proporciona novas formas de lidar com a informação no cotidiano, no qual os dados – a partícula elementar da informação – estabelecem-se como protagonistas de novos processos para a tomada de decisão e a geração de conhecimento. Estima-se que a produção de dados global ultrapassou a marca de 16 *zettabytes*³ em 2020 (CAVANILLAS; CURRY; WAHLSTER, 2016).

Essa grande produção implicou o avanço tecnológico para aprimorar a capacidade de processamento, armazenamento e compartilhamento de dados de diversas formas a uma velocidade acelerada, alimentando, assim, o fenômeno conhecido como Big Data (CAESARIUS; HOHENTHAL, 2018). O Big Data vem proporcionado substanciais transformações no modo com que se analisa a informação a partir dos dados, o que promove “uma forma completamente nova de produção de conhecimento” que produz “insights que anteriormente eram considerados impossíveis” (CAESARIUS; HOHENTHAL, 2018, p. 2) em diversas áreas de atuação e contextos da sociedade.

Para compreender como isso se dá, cumpre explorar a definição do Big Data, suas aplicações, valor, implicações e perspectivas na contemporaneidade.

3.1 Definindo o Big Data

O Big Data pode ser compreendido como um fenômeno desencadeado pelas recentes transformações tecnológicas ocorridas na sociedade. Ele possui grande potencial informacional para aperfeiçoar o processo de tomada de decisões, aumentar a produtividade e viabilizar o desenvolvimento de soluções, serviços e produtos em diversos contextos da sociedade (MARQUESONE, 2016), seja na esfera corporativa, seja na acadêmica ou governamental.

A definição do fenômeno é amplamente debatida na literatura de modo que se encontram diversas formas de apresentá-lo (BIHL; YOUNG II; WECKMAN, 2016; CONEGLIAN; GONÇALEZ; SANTARÉM SEGUNDO, 2017). Contudo, há um consenso, especialmente entre as áreas da tecnologia da informação, no sentido de apresentar o

³ Zettabyte é uma “[u]nidade de medida de informação equivalente a 10²¹ bytes” (DICIONÁRIO PRIBERAM DA LÍNGUA PORTUGUESA, 2019).

Big Data sucintamente a partir de suas principais características, quais sejam, volume, velocidade e variedade (os “três Vs”).

3.1.1 Volume

O volume se destaca como a característica mais significativa do Big Data, pois, em decorrência do avanço das tecnologias da informação (OUSSOUS *et al.*, 2018), dados podem ser extraídos de diversas fontes de informação (TAURION, 2014 apud CONEGLIAN; GONÇALEZ; SANTARÉM SEGUNDO, 2017). Isso provoca um estado em que “a geração e acumulação de dados está além da capacidade de processamento, ao ponto em que os usuários são sobrecarregados pelo mesmo” (BIHL; YOUNG II; WECKMAN, 2016, p. 3).

A IDC estima que o volume de dados em 2025 será de até 175 zettabytes (REINSEL; GANTZ; RYDNING, 2018). Se este montante fosse armazenado em DVDs, por exemplo, ter-se-ia uma fila deste suporte apta a dar duas voltas em torno na Terra. Ou então, se fosse realizado o download desta quantidade de dados, o processo demoraria 1.8 bilhões de anos para ser completado.

3.1.2 Velocidade

A velocidade caracteriza-se pela rapidez com que os dados são gerados e processados visando à extração de informação (OUSSOUS *et al.*, 2018). Este aspecto do Big Data está intrinsecamente relacionado ao desenvolvimento das tecnologias da informação que se retroalimentam. Isso porque quanto maior a capacidade de captura e processamento de um imenso volume de dados, maior será a necessidade de desenvolver novas metodologias de processamento e comunicação para tratar dessa avalanche (AL-SALIM *et al.*, 2018), bem como desenvolver infraestruturas robustas para comportar o aumento exponencial deste processo.

A velocidade no fluxo de captura e processamento de dados de um servidor até um usuário final é crucial para a maximização do valor do dado coletado, considerando que o atraso desse processo pode ser extremamente prejudicial à análise da informação e da tomada de decisão (AL-SALIM *et al.*, 2018). Para exemplificar este ponto, basta imaginar que em uma região montanhosa existam sensores IoT que indiquem tremores e movimentações na terra. Caso os dados coletados dos sensores demorem para serem

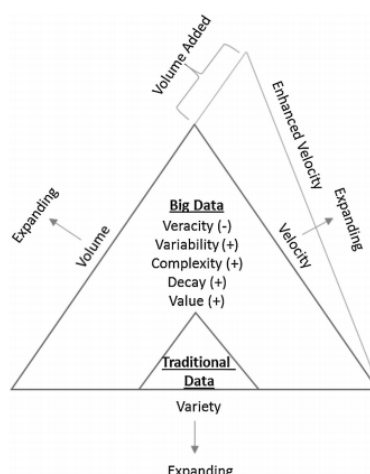
enviados a um servidor, e o servidor também demore para tratar e compartilhar esses dados para técnicos de monitoramento, haverá um atraso na sinalização de um possível terremoto na região, com o comprometimento das vidas de milhares de pessoas. De outro modo, no cenário em que essa comunicação e tratamento de dados seja extremamente rápida e eficaz, os desastres podem ser previstos e as medidas podem ser tomadas antecipadamente.

3.1.3 Variedade

A variedade contempla as diferentes formas de obtenção de dados, bem como a diversidade de formatos dos mesmos (AL-SALIM *et al.*, 2018). Esta característica expressa a riqueza do Big Data em potencializar o uso de “qualquer coisa” como dado com características peculiares e distintas. Obtêm-se dados em formato numérico, textual, audiovisual obtidos a partir de sensores, satélites, aplicativos, dentre outras fontes (AL-SALIM *et al.*, 2018; BIHL; YOUNG II; WECKMAN, 2016) que, quando traduzidas para permitir a interpretação e compreensão humana, podem gerar grande valor informacional.

Esses “três Vs” do Big Data contribuem para a melhor compreensão dos desafios tecnológicos e humanos que se enfrenta para proporcionar novas formas de lidar com dados a fim de processos diversos tipos de processos (CAVANILLAS; CURRY; WAHLSTER, 2016). O trabalho de Lee (2017) apresenta uma forma de visualização integrada do Big Data, que contempla as duas dimensões do fenômeno, uma de causa e outra de efeito, e estabelece a relação entre elas conforme a ilustração abaixo.

Figura 1 - Visualização integrada do Big Data



Fonte: Lee (2017)

A primeira dimensão é representada pelos aspectos “volume”, “velocidade” e “variedade”, cujo crescimento afeta a dimensão secundária, caracterizada pelos aspectos “veracidade”, “variabilidade”, “complexidade”, “decaimento” e “valor”. Dados tradicionais são referenciados na visualização com um subconjunto do Big Data por apresentarem também o volume, a velocidade e a variedade atuando em menor escala no seu desenvolvimento. O desenvolvimento de cada uma dessas três dimensões afeta as demais, o que exprime a relação do avanço da tecnologia para com o próprio desenvolvimento do Big Data.

3.2 O desenvolvimento das tecnologias da informação propulsoras do Big Data

O Big Data se desenvolve na medida em que as tecnologias da informação e comunicação se transformam e aprimoram a capacidade do homem e da máquina de capturarem cada vez mais rápido um grande volume de dados variados. O fenômeno não é algo novo. Uma análise linha do tempo, que pode ser dividida em três períodos, aponta que tipo de desenvolvimento permitiu que o Big Data assumisse as proporções atuais. Lee (2017) caracteriza essas fases da seguinte forma:

- a) **Big Data 1.0** (1994-2004): marcado pelo advento do e-commerce e pela capacidade de mineração de dados não estruturados da Web. Destaque para a capacidade de descobrir e analisar o perfil comportamental de usuários na internet. Necessidade de aumentar a capacidade de armazenamento de dados gerados pelas ciências.

- b) **Big Data 2.0** (2005 – 2014): marcado pela Web 2.0 e pelas mídias sociais. Há a capacidade dos usuários de interagirem com a internet e contribuírem com os seus próprios conteúdos.
- c) **Big Data 3.0** (2015 -): marcado pelas tecnologias da internet das coisas (internet of things – IoT) e a sua capacidade de gerar dados em diferentes suportes, como imagens, áudios ou vídeos a partir de dispositivos e sensores que interagem com o mundo externo. Combina dados de todas as fases anteriores.

Analizadas as fases do Big Data e os aspectos tecnológicos que marcaram o seu desenvolvimento, apresentam-se agora os principais domínios que originam os dados que compõem o Big Data:

- a) **Dados sociais:** dados e informações gerados a partir da experiência humana; envolvem de dados analógicos a eletrônicos e digitais no formato de fotografias, vídeos, tweets, likes, uploads em redes sociais etc.;
- b) **Dados transacionais:** gerados por transações, como pagamento de contas, compras online, registro de clientes, anotação de pedido, dentre outros;
- c) **Dados de máquinas:** gerados a partir de máquinas, sensores, internet das coisas; são coletados do mundo físico e digital em diversas áreas, como saúde, jogos, trânsito, agricultura, astronomia (CLOUDMOYO, 2016; DEVLIN; ROGERS; MYERS, 2012).

Esses dados podem ser gerados, obtidos e transformados em diversos formatos e suportes, de vídeos a logs de computador, histórico de compras etc. Tendo em vista que podem ser originados por meio de diversas atividades humanas e não humanas, compreende-se sua capacidade para produzir uma avalanche de informações que contribuem para o aumento ininterrupto do Big Data. Há imenso potencial do Big Data para a geração de informação e valor para as instituições em diversos setores e contextos da sociedade, tema que será abordado na próxima sessão, que se refere à chamada economia *data-driven* (“orientada pelos dados”).

3.3 Da economia de dados à 4ª revolução

Conforme visto anteriormente, o Big Data proporciona grandes possibilidades de obtenção de dados de diversas naturezas em variados formatos, que contribuem para o

desenvolvimento de novos insights, produtos e soluções em diferentes contextos e organizações (CAVANILLAS; CURRY; WAHLSTER, 2016). Tal aspecto é atribuído ao valor do Big Data que, em algumas definições, constitui uma quarta característica do fenômeno.

As instituições agora são dotadas de uma nova perspectiva de análise que lhes permite proporcionar mudanças em diferentes esferas da sociedade em níveis socioeconômicos e ambientais (AL-SALIM *et al.*, 2018; CAVANILLAS; CURRY; WAHLSTER, 2016; MARQUESONE, 2016). O impacto da tecnologia na sociedade transformou o modo como as instituições lidam e aprendem com dados. A revista *The Economist* defendeu que os dados são o novo petróleo do mundo; que, devido à tecnologia, tornou-se uma fonte extremamente abundante, lucrativa e de rápido crescimento que tem transformado a competitividade do mercado (THE ECONOMIST, 2017a).

Em outra matéria, a revista destacou que:

Dados são para este século o que o petróleo foi para o último: um motor de crescimento e mudança. Os fluxos de dados criaram uma infraestrutura, novos negócios, novos monopólios, novas políticas e – crucialmente – uma nova economia. A informação digital é distinta de qualquer outro recurso anterior; ela é extraída, refinada, valorizada, comprada e vendida em diferentes formas. Ela mudou as regras para o mercado e demanda novas práticas aos reguladores [...] (THE ECONOMIST, 2017b, tradução nossa).

A popularidade dos dados no mercado fez com que se desenvolvesse uma economia de dados (*data-driven*), que gera produtos, sistemas, softwares, serviços e inteligência a partir da criação, produção, distribuição, venda e consumo de produtos de dados (CAO, 2018). De acordo com Cao (2018), esta economia de dados pode ser classificada em sete tipos de negócios:

- a) *data-enabling*: negócios que são guiados e desenvolvidos absolutamente por dados; estão no cerne da nova geração de economia e inteligência de dados; são voltados aos processos de tratamento de dados, da aquisição ao gerenciamento e manipulação;
- b) *data-intensive*: guiam negócios tradicionais a se orientarem pelos dados, explorando o valor que eles podem oferecer para a inovação, a capacitação do aumento das oportunidades de negócio e de produtividade;

- c) negócios de Tecnologia da Informação (TI) clássicos: transformação de empresas tradicionais por meio de renovação produzida pela orientação de dados;
- d) negócios tradicionais que carecem de dados: “datatificação” e transformação digital de negócios tradicionais;
- e) negócios não digitais: mercado em que se realiza a digitalização de negócios visando ao aumento da eficiência de processos, como serviços jurídicos, de cartório, entre outros;
- f) serviços de dados: negócios recentes que transformam dados em serviços, focando na construção, fornecimento e manutenção de serviços de dados, a exemplo do chatbot, de coleta de dados de comportamento de consumidor para e-commerce etc.;
- g) economia do futuro: possibilidades de inovação dos negócios de dados que se combinam com as áreas de negócio pré-existentes.

De modo geral, a economia de dados visa à transformação digital e computacional das instituições por meio da indústria dos dados, melhorando questões de aparatos tecnológicos e potencializando análises por meio o uso de dados em diferentes contextos organizacionais, tornando serviços e produtos mais inteligentes e conectados. O Big data contribui para que as empresas se capacitem a criar diferentes negócios, desenvolvam novos produtos, serviços e aperfeiçoem as operações e fluxos existentes (LEE, 2017).

A implementação e uso do Big Data nas instituições pode trazer benefícios às mesmas em relação à economia de custos, aumento da qualidade, serviços e produtos (com maior capacidade de personalização dos mesmos, de acordo com o público alvo), bem como a melhoria no processo de tomada de decisão (LEE, 2017).

No âmbito científico, o Big Data contribui para o desenvolvimento da e-Science, modo de se “realizar” a ciência em colaboração global por meio de instituições conectadas virtualmente, que realizam trabalhos no ambiente digital, compartilhando pesquisas e informações (SEMELER, 2017). Diversas disciplinas, especialmente as correlatas às ciências exatas e naturais, contribuem para a e-Science e, conseqüentemente, para o próprio crescimento do Big Data, pois sua produção e compartilhamento são orientados por dados, o que acarreta grandes volumes de dados que requerem uma infraestrutura robusta para o seu próprio desenvolvimento e cooperação global (SEMELER, 2017).

Os impactos do Big Data se refletem em diversos âmbitos da sociedade, capacitando mudança de práticas, métodos, produtos e serviços, cujo caráter exige que os indivíduos se tornem cada vez mais letrados neste universo para enfrentar situações cotidianas de tomadas de decisões por meio de tais recursos.

Tendo em mente que o profissional da informação é um dos principais agentes atuantes nas questões de mediação e organização de informação, verifica-se a seguir o papel deste profissional nesta era dos dados.

4 BIBLIOTECONOMIA DE DADOS

Apesar do desenvolvimento do Big Data ter sido alavancado pela tecnologia e afetar as relações dos indivíduos para com dados e informação, o Big Data não é um fenômeno tão recente quanto a sua popularização faz parecer. Bihl, Young II e Weckman enfatizam que “os problemas de Big Data e de sobrecarga informacional existem há séculos, em diferentes formas. Várias pessoas, no decorrer da história, examinaram formas de coletar, ordenar, agrupar, categorizar e indexar os dados disponíveis aos mesmos” (2016, p. 3, tradução nossa). Durante a história moderna, por exemplo, a humanidade enfrentou problemas correlatos à explosão informacional com a acelerada produção, circulação e disseminação de textos científicos e acadêmicos, especialmente com a criação de universidades e com o nascimento da imprensa, graças à invenção de Gutemberg (BIHL; YOUNG II; WECKMAN, 2016; ROSENBERG, 2003). Nesse período ocorreu uma volumosa e vertiginosa produção, circulação e disseminação de textos científicos e acadêmicos, fazendo com que os profissionais da informação fossem protagonistas das soluções postas pelos desafios informacionais.

As explosões informacionais ocorreram em diversos momentos da história, e práticas, conhecimentos e tecnologias analógicas, eletrônicas e digitais se desenvolveram para contribuir para a organização da informação, especialmente no campo da Biblioteconomia e da Documentação. Na atualidade, ao nos depararmos com o Big Data, “da mesma forma que as bibliotecas tradicionalmente facilitaram acessos a documentos, agora profissionais da informação podem facilitar acesso a dados, mesmo que dados não necessariamente se enquadr[em] no mesmo formato de documentos que bibliotecas ofereciam.” (STUART, 2011 apud KOLTAY, 2015, p. 402).

É conhecido que “profissionais da informação podem estar vinculados a qualquer etapa do ciclo de informação, pois possuem capacidade para trabalhar com eficiência e eficácia no tratamento da informação em qualquer tipo de organização” (CONEGLIAN; GONÇALEZ; SANTARÉM SEGUNDO, 2017, p. 131). Pode-se defender, ainda, que “bibliotecas e bibliotecários são exclusivamente adequados para trabalhar com Big Data. Bibliotecas têm uma longa tradição de serem as primeiras a adotar tecnologias, e com o Big Data não deveria ser exceção [...]” (HOY, 2014, p. 7, tradução nossa).

Ora, para trabalhar com o Big Data faz-se necessário possuir uma gama de conhecimentos multidisciplinares tanto da Biblioteconomia, especializada no tratamento de qualquer tipo de informação, independente do suporte, quanto da Ciência de Dados,

que reúne as práticas tecnológicas que permitem navegar pelos Big Data. Essas áreas são complementares para atuarem frente ao Big Data. Contudo, verifica-se que apesar de o conhecimento de domínio estar presente na Ciência de Dados, os cientistas de dados apresentam maior dificuldade em desenvolver uma “perspectiva focada nas necessidades informacionais de gestores e administradores” (CONEGLIAN; GONÇALEZ; SANTARÉM SEGUNDO, 2017, p. 129) e usuários, abrindo espaço para que o profissional da informação se aproprie dessa demanda. Ele pode se apresentar como fundamental ao “buscar a sinergia entre as necessidades dos gestores e a expertise dos profissionais da computação” (CONEGLIAN; GONÇALEZ; SANTARÉM SEGUNDO, 2017, p. 129).

Por outro lado, entende-se que o conhecimento tecnológico e estatístico provindo da Ciência de Dados é um gap na formação do bibliotecário, fazendo com que o mesmo, em muitos momentos, seja desconsiderado por companhias e instituições para tratar suas demandas de informações de dados. Verifica-se, na literatura, a existência de discussões sobre a Ciência de Dados ser um uma subárea da Ciência da Informação (MARCHIONINI, 2016). O termo “cientista de dados” não é utilizado para designar um tipo de profissional, “pois se compreende que tanto o profissional da informação, quanto o profissional da computação estão exercendo as funções do cientista de dados, porém, dividindo tais funções conforme as suas habilidades” (CONEGLIAN; GONÇALEZ; SANTARÉM SEGUNDO, 2017, p. 139).

[A]lguns programas de treinamento em ciência de dados podem ser alojados em escolas de informação, contudo, é mais estratégico ver a ciência da informação como um componente essencial da ciência de dados para que o campo emergente possa se beneficiar da diversidade de perspectivas que a colaboração interdisciplinar oferece (MARCHIONINI, 2016, p. 5).

Porque a Ciência de Dados é complementar às habilidades e competências do bibliotecário de dados, a próxima seção dedica-se a apresentá-la.

4.1 Ciência de Dados

As mudanças proporcionadas pelo Big Data, tanto em relação às suas próprias demandas tecnológicas quanto às oportunidades de análise e descobertas, conduziram à valorização da Ciência de Dados para tratar de desafios que surgiram com o aumento

exponencial de dados, especialmente em meio aos círculos acadêmicos e industriais (ZHU; XIONG, 2015a, 2015b). Apesar de ser uma área de atuação que vem ganhando grande destaque na atualidade (ZHU; XIONG, 2015b), o termo “Ciência de Dados” foi utilizado pela primeira vez na década de 1990 para cunhar a preocupação e necessidade de tratar dados de diversas áreas da ciência a partir de tecnologias e habilidades específicas que fomentam a extração de conhecimento a partir de dados (ZHU; XIONG, 2015a). Este campo já possuía o condão de “revelar as características e/ou estruturas dos fenômenos naturais, humanos e sociais a partir da utilização de dados.” (SEMELER, 2017, p. 69).

4.1.1 Definindo a Ciência de Dados

A Ciência de Dados pode ser compreendida, de modo geral, coma a ciência que estuda dados, que se consolida a partir de três perspectivas que a caracterizam: o foco no seu objeto de estudo, os dados; o enfoque nos seus processos; e seu caráter transdisciplinar.

Quanto ao seu perfil transdisciplinar, pode-se constatar que as principais disciplinas que contribuem para a estruturação da Ciência de Dados são a Computação, a Estatística e a Ciência da Informação, acompanhadas de conhecimentos provindos das Ciências Sociais, da Gestão e da Comunicação (CAO, 2018).

No que tange aos processos, a Ciência de Dados respalda-se na pirâmide “Dado-Informação-Conhecimento-Inteligência-Sabedoria”. Cada elemento apresenta-se como o guia para essa disciplina, que estipula que o cientista de dados deve “‘pensar com sabedoria’, ‘compreender o domínio’, ‘gerir dados’, ‘computar com dados’, ‘descobrir conhecimento’, ‘comunicar com *stakeholders*’, ‘agir a partir de insights’, ‘entregar produtos’” (CAO, 2018, p. 31). Tal conjunto de habilidades enfatiza o caráter interdisciplinar da Ciência de Dados, pois, para desenvolver os seus processos, faz-se necessário conhecer diversas áreas. Cada um desses aspectos da formação deste profissional será visto brevemente a seguir, com base em Cao (2018):

- a) “Pensar com sabedoria” faz referência aos resultados e desafios da Ciência de Dados em transformar dados em conhecimento. Esse aspecto diferencia tal disciplina das demais pelo fato de esperar que ela produza insights baseados em evidência, dotados de valor, que anteriormente não eram visíveis, o que só vem a acontecer com grande experiência e conhecimento dos cientistas de dados;

- b) “Compreender o domínio” trata do conhecimento necessário para compreender o domínio e o contexto dos dados se são trabalhados;
- c) “Gerir dados” aponta para uma das principais características da Ciência de Dados, a capacidade de aquisição, tratamento, armazenamento, busca, indexação e recuperação de dados. Esse aspecto está atrelado à qualidade, segurança, ética e privacidade de dados;
- d) “Computar com dados” faz referência à manipulação de dados com base em hipóteses, aplicando técnicas que vão desde a coleta de dados à análise estatística e visualização. Esse aspecto destaca as ramificações da Ciência de Dados, pois as atividades desempenhadas nesse processo estão atreladas a competências de engenharia de dados, estatística, design e (CAO, 2018);
- e) “Descobrir conhecimento” trata da manipulação de dados de alto nível, que faz uso de predições, clusterização, identificação de padrões e classificações para descobrir o conhecimento escondido a respeito de um problema, seja a sua solução, causa ou efeito;
- f) “Comunicar com stakeholders” refere-se aos processos de comunicação necessários para o desenvolvimento e conclusão de um projeto de Ciência de Dados, especialmente porque um projeto dessa natureza não envolve só cientistas de dados, mas também pessoas com diferentes papéis e incumbências, tais como analistas de dados, gerentes e executivos. Esse aspecto da Ciência de Dados reforça a necessidade de o profissional ter uma visão holística de suas tarefas e projetos, desenvolvendo a comunicação entre as partes (CAO, 2018);
- g) “Entregar produtos de dados” indica a capacidade de prover conhecimento e achados em descobertas a partir da aplicação de técnicas (mineração de dados, machine learning, etc.) e desenvolvimento de projetos analíticos, softwares, aplicativos, sites, sistemas etc.;
- h) “Agir a partir de insights” refere-se à capacidade de gerar insights a partir de dados e exige um alto nível de expertise para que seja possível “[...] capturar com eficácia os mecanismos, dinâmicas, princípios e fatores determinantes intrínsecos e fundamentais refletidos no mundo dos dados.” (CAO, 2018, p. 43, tradução nossa).

Vistos os principais aspectos que contribuem para compreender a Ciência de Dados, entende-se que ela exige conhecimentos, técnicas e habilidades que estão além

dos conhecimentos computacionais e estatísticos. Requer-se uma gama de conhecimentos oriundos de diferentes disciplinas para que seja possível a efetiva realização do trabalho de Ciência de Dados.

A Ciência de Dados é uma disciplina em formação. Ela trilha um percurso para firmar a sua consolidação como uma nova ciência. Tem como foco métodos de observação e racionalização de dados, desenvolvimento de teorias e tecnologias que contribuem tanto para a identificação de dados com credibilidade quanto para o oferecimento de suporte a outras áreas de pesquisa, de modo a viabilizar a aquisição de conhecimentos valiosos a partir de dados (ZHU; XIONG, 2015a).

Apesar de recente, a área tem causado grande impacto nos meios laborais e acadêmicos, porque a Ciência de Dados é vista como essencial para o aperfeiçoamento da qualidade dos produtos e serviços da indústria e organizações. No Brasil, destaca-se como a 5ª profissão emergente, inclusive à frente de Engenharia de Dados (LINKEDIN, 2020). Recorrentemente, aparece nos rankings das profissões mais procuradas (CAO, 2018) e como a profissão do século 21 (DAVENPORT; PATIL, 2012). Contudo, o mercado ainda se depara com dificuldades na hora de encontrar cientistas de dados qualificados, o que se reflete também na necessidade de a Ciência de Dados ganhar maior espaço na academia, com disciplinas e cursos próprios (GOYAL *et al.*, 2020).

4.1.2 Subáreas da Ciência de Dados

As possibilidades de trabalho do cientista de dados são vastas em razão das variadas combinações de habilidades e técnicas deste profissional (SEMELER, 2017), além da constante demanda do mercado de trabalho e da academia para o tratamento de grandes volumes dados. As atividades dos cientistas de dados em uma organização podem ser categorizadas em segmentos de atuação que cobrem a realização de tarefas gerenciais até técnicas. Cao (2018) apresenta seis áreas de atuação, sendo elas:

- a) arquitetura: arquitetos de bancos de dados, de redes, de nuvens, de sistemas, solução e segurança;
- b) engenharia: engenharia de dados, de software, aplicativos, sistemas, rede, programador, desenvolvedor;

- c) cientista de dados: modelador de dados, analista de descoberta de conhecimento, analistas de machine learning, de estatística, de otimização, de visualização, de insights, minerador de dados, pesquisador quantitativo etc.;
- d) estrategistas: analistas de negócios (business analytics), analistas de inteligência de negócio (business intelligence), etc.;
- e) administradores: gerente de requisitos, gerente de projetos, gerente de qualidade, bibliotecário, gerente de redes, gerente de dados etc.;
- f) executivos: *Chief Data Officer* (CDO), *Chief Analytics Officer* (CAO), *Chief Data Scientist* (CDS), *Chief of Information Officer* (CIO), *Chief Technology Officer* (CTO) e o *Data Scientist Lead*.

A exploração dos campos de atuação do cientista de dados é apresentada também por Zacharias Voulgaris (2014 apud SEMELER, 2017), que classifica essas funções com base nas habilidades ou campos do conhecimento requeridos para seu desempenho, conforme sintetizado por Alexandre Semeler no quadro a seguir.

Tabela 1 - Tipos de cientistas de dados

Tipo	Descrição
<i>Data developers</i>	O objetivo de um <i>data developer</i> é adquirir informações úteis a partir de dados. Geralmente, concentra-se em questões técnicas de análise e gestão de dados. Seu trabalho envolve o tratamento de dados para organizá-los em grandes bases de dados, preparando-os para consulta, análise e visualização dos resultados.
<i>Data researchers</i>	São geralmente do mundo acadêmico e demonstram uma sólida experiência em estatísticas ou em qualquer ciência que empregue estatísticas (cientistas sociais).
<i>Data creative</i>	Geralmente, possuem experiência acadêmica e são excepcionalmente bons com tecnologias (<i>designers</i> de <i>software</i> para governança e análise de dados), aprendizado de máquina e proficientes em alguma linguagem de programação.
<i>Data business people</i>	Trabalham com negócios e, geralmente, lideram equipes especializadas em <i>Data Science</i> . Seu foco é o aumento das receitas de uma empresa.
<i>Mixed/generic Data scientists</i>	São como os <i>data business people</i> , mas sem a ampla experiência ou o foco em negócio.

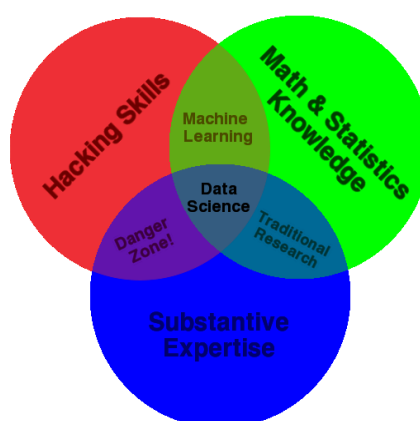
Fonte: Voulgaris apud Semeler (2017)

Verifica-se que o desempenho de atividades *data-driven* contempla diferentes áreas de atuação, fazendo com que diversos caminhos possam ser trilhados em uma carreira no mundo dos dados. A magnitude do impacto do Big Data explora o potencial de aprendizados intrínsecos e extrínsecos a dados, reforçando o caráter multidisciplinar da Ciência de Dados. Sendo uma nova área de atuação que tem se destacado no mercado, o cientista de dados precisa satisfazer requisitos para desempenhar a sua profissão, com formação acadêmica e desenvolvimento de habilidades que contribuem para a consolidação da sua qualificação, o que será abordado na sequência.

A graduação em Ciência de Dados é recente, com poucos programas dessa natureza sendo oferecidos pelas instituições universitárias, enquanto a formação especializada é encontrada majoritariamente nos cursos de pós-graduação (SONG; ZHU, 2016). Geralmente, os cientistas de dados são oriundos de áreas correlatas às ciências exatas e naturais, tais como a ciência da computação, estatística, matemática, informática, engenharia e física (CAO, 2018). Contudo, independente da formação formal do indivíduo, o que se espera dele é que possua conhecimento e experiência para trabalhar e solucionar problemas com dados (SONG; ZHU, 2016).

A relação interdisciplinar da Ciência de Dados pode ser expressa pelo diagrama de Venn da Ciência de Dados, proposto por Conway (2010), reproduzido abaixo.

Figura 2 - Diagrama de Venn da Ciência de Dados



Fonte: Conway (2010)

Para explorar a relação, posta na imagem acima, da tríade que compõe a Ciência de Dados, eis os principais aspectos do diagrama:

- a) *hacking skills*: referem-se aos conhecimentos da ciência da computação relativos à lógica da programação, ao processamento e limpeza de dados, dentre outros.

Conway (2010) destaca que um bom cientista de dados não precisa necessariamente ter cursado ciência da computação ou cursos correlatos, mas precisa apresentar um sólido domínio de manipulação de algoritmos;

- b) *math & statistics knowledge*: conhecimentos necessários para a extração de insights por meio da formulação de hipóteses que podem ser comprovadas cientificamente com testes e metodologias específicas a partir de ferramentas matemáticas e estatísticas (SEMELER, 2017);
- c) *substantive expertise*: refere-se à capacidade de aplicar os conhecimentos hackers e matemáticos a uma área do conhecimento específica.

As principais preocupações da Ciência de Dados estão relacionadas às suas teorias fundamentais, ou seja, à exploração da natureza dos dados e suas tecnologias, conforme explanado por Zhu e Xiong (2015a) e brevemente referidas a seguir:

- a) teorias fundamentais: trata-se de questões relativas à relação de dados com a análise de dados, estruturação de dados com base em sua tipologia e desenvolvimento de métodos de pesquisa de dados;
- b) exploração da natureza dos dados: relativa à classificação dos dados, tamanho, credibilidade, crescimento e impacto na sociedade;
- c) tecnologia e aplicações: desenvolvimento de metodologia de pesquisa científica baseada em dados de modo que contribua para a eficiência e os resultados em experimentos de diferentes áreas do conhecimento; aperfeiçoamento de aplicações e técnicas de análise visando à eficiência para com o Big Data.

Em suma, a Ciência de Dados apresenta-se como uma disciplina que combina conhecimentos e habilidades multidisciplinares que, com o desenvolvimento de teorias e tecnologias inovadoras, é capaz de contribuir para a solução de problemas e a descoberta de novas informações, a partir de fatos e evidências de diferentes áreas do conhecimento (ZHU; XIONG, 2015a).

4.1.3 Desafios e perspectivas da Ciência de Dados

A Ciência de Dados trilha um percurso para firmar a sua consolidação como uma nova ciência. Tem como foco métodos de observação e racionalização de dados, desenvolvimento de teorias e tecnologias que contribuem tanto para a identificação de

dados com credibilidade quanto para o oferecimento de suporte a outras áreas de pesquisa, de modo a viabilizar a aquisição de conhecimentos valiosos a partir de dados (ZHU; XIONG, 2015a). Os impactos do desenvolvimento dessa disciplina são expostos por Cao, quando este infere que:

A Ciência de dados está estimulando a geração de uma nova profissão, conhecida como profissão de dados, que se refere a qualquer carreira, trabalho ou responsabilidade relacionada a dados. Essa profissionalização orientada a dados é abrangente e sem precedentes [...] [I]mpulsiona a criação de uma variedade de novos cargos de dados, maturidade capacitacional em relação a dados, e competência organizacional [...]. O surgimento da profissão de dados desafia fundamentalmente os sistemas educacionais existentes. A ciência de dados se apresenta com uma nova disciplina e, como resultado, estão surgindo novos cursos de ciência de dados, novos assuntos, metodologia de entregas, ensino e aprendizado de tecnologia, relacionamentos e produtos educacionais [...] (2018, p.70-71, tradução nossa).

Cao (2018, p. 70) também postula que a inovação orientada a dados "está criando formas e aspectos de produtos de dados. Esses produtos de dados estão [...] mudando da economia focada na internet para a economia de dados em termos de novos caminhos socioeconômicos de estudo, trabalho, entretenimento e viagem", demonstrando a sua capacidade de transformação nas esferas socioeconômicas.

4.2 Biblioteconomia de dados

A Biblioteconomia de Dados pode ser compreendida como uma disciplina que estuda e aplica métodos voltados à prática de coleta, tratamento, gestão, recuperação, compartilhamento e preservação de dados (KELLAM; PETER, 2011; WANG, 2018). Ela utiliza técnicas, recursos e saberes da Biblioteconomia e da Ciência de Dados (SEMELER, 2017) para criar e oferecer serviços e produtos de dados a usuários e instituições de acordo com as suas necessidades de informações de dados (GOBEN; ZILINSKI; BRINEY, 2016; RICE; SOUTHALL, 2016; SEMELER, 2017).

Entende-se que para continuar firmando a sua relevância frente às necessidades por dados informacionais da sociedade, em especial as desencadeadas pelo Big Data, "bibliotecários precisam cada vez mais tornarem-se *data-savvy*⁴ para adquirirem um

⁴ A expressão "savvy" indica conhecimento prático, experiência, habilidade (CAMBRIDGE DICTIONARY, 2020). Um profissional data-savvy, assim, é um profissional que possui experiência, habilidade e saber prático no trato dos dados.

conhecimento mais profundo a respeito do ciclo de vida de dados de pesquisa para aprimorar os serviços que oferecem” (BARBARO, 2016, p. 25). De fato, com as mudanças proporcionadas e exigidas pela era dos dados, o desenvolvimento de habilidades específicas para a gestão de informação e extração de conhecimento vem sendo um fator cada vez mais valorizado no mercado de trabalho, especialmente pelo valor que este saber tem agregado à informação (CAVANILLAS; CURRY; WAHLSTER, 2016; CONEGLIAN; GONÇALEZ; SANTARÉM SEGUNDO, 2017).

Neste contexto de transformações informacionais, o profissional da informação se destaca, pois “[...] recebeu formação específica para trabalhar com dados, informação e conhecimento, bem como sua mediação” (VALENTIM, 2014, p. 140 apud CONEGLIAN; GONÇALEZ; SANTARÉM SEGUNDO, 2017, p.131). Este elemento contribuiu para o surgimento de uma proeminente área de atuação do profissional da informação que se propõe a engajar-se e atuar dentro desse universo informacional proporcionado pelo Big Data, a Biblioteconomia de Dados. Na literatura em inglês, encontram-se variações do termo. Mencionam-se “Data Librarianship” (JOHNSON, 2019; HYAMS, 2008; SEMELER; PINTO; ROZADOS, 2017), “Databrarianship” (KELLAM; THOMPSON, 2016; STEPHANIE L. MAATTA, 2013), “Data library” (WANG *et al.*, 2016) e “Data Science librarianship” (EKSTRØM, 2019). Contudo, o termo também varia conforme o foco de atuação deste profissional, fazendo com que a área também venha a ser conhecida como “E-science Librarianship” (SEMELER; PINTO, 2019; XIA; WANG, 2014) e até mesmo “Big Data Library” (WANG *et al.*, 2016).

A origem desta área remonta à década de 1960 com o surgimento de arquivos de dados na Europa e com o oferecimento de serviços de dados em bibliotecas norte-americanas em decorrência do desenvolvimento de pesquisa computadorizada e de mudanças da metodologia de pesquisa das ciências sociais, cuja faceta adotou caráter mais científico, com métodos qualitativos para tratamento estatístico e análise de populações e descrição, explicação e predição do comportamento das mesmas. Tal aspecto estimulou o compartilhamento e reutilização de dados por motivos de economia e eficiência na recuperação de informação (RICE; SOUTHALL, 2016; KOLTAY, 2017; RICE; SOUTHALL, 2016; SEMELER, 2017), demandando processos inerentes à criação de grandes repositórios de dados de pesquisa, muitas vezes mantidos por universidades e instituições governamentais que poderiam ser utilizados por diversos profissionais.

O papel do profissional da informação inserido neste contexto modificou-se rapidamente frente às demandas de seus usuários. Inicialmente, o seu trabalho estava

centralizado no serviço de referência, recuperando e compartilhando dados conforme a demanda de pesquisadores. Contudo, devido à crescente complexidade das fontes e tipos de dados que foram surgindo, este profissional precisou se adaptar e tornar-se familiarizado com diferentes técnicas de manipulação de dados para atender às diferentes demandas de seus usuários, tais como recomendar coleções, divulgar e compartilhar novos dados para fins de pesquisa, auxiliar os pesquisadores em problemas com os quais viriam a se deparar ao lidar com os dados e utilizar a tecnologia como um de seus principais aliados para tratar essas questões (RICE; SOUTHALL, 2016).

Em 1974 se realizou um dos principais eventos da recente história da Biblioteconomia de Dados, a “Conferência sobre Serviços e Programas de Dados em Bibliotecas” (*Conference on Data and Program Library Services*), ocorrida em Toronto, no Canadá (AFFELT, 2015), na qual a profissão *data librarian* foi oficialmente apresentada pela International Association for Social Science Information Service and Technology (IASSIST) (SEMELER, 2017). Embora a Biblioteconomia de Dados tenha ganhado maior destaque recentemente, ela não é uma disciplina nova, pois desde o princípio aplica tradicionais conhecimentos e técnicas biblioteconômicas aos conjuntos de dados. Isso envolve saberes tais como a “catalogação e a organização de materiais; a preservação e a curadoria; serviços de usuários e de referência [...], aquisição e desenvolvimento de coleções, [...] implementação de serviços de referência aos usuários [...]” (SEMELER, 2017, p. 66–67).

A Biblioteconomia de Dados estimula a descoberta e a criação de novas habilidades. Contribui para que as atividades da área geral da Biblioteconomia sejam repensadas e adaptadas para o universo dos dados, fazendo com que surjam novas práticas para o atendimento a usuários com a criação de serviços, consultoria e treinamentos voltados ao gerenciamento, à curadoria e ao compartilhamento de dados (HYAMS, 2008; SEMELER, 2017). As práticas da Biblioteconomia de Dados contribuem para facilitar o processo de descoberta e recuperação de dados, provendo e garantindo a qualidade de dados, bem como a “reutilização [dos dados] ao longo do tempo por meio de atividades que incluem a autenticação, arquivamento, gerenciamento e representação” (GORDON-MURNANE, 2012, p. 33).

Agentes da área se preocupam com a constante descoberta de novos dados:

Uma das principais tarefas da Biblioteconomia de Dados é ser *upstream* (GOLD, 2007a), o que significa que ela não deve apenas se preocupar com informação publicada, mas também com fontes potenciais de dados, sendo o seu objetivo

compreender como os diferentes tipos de dados podem ser capazes de gerar informação útil [...] (XIA; WANG, 2014; KOLTAY, 2017 apud SEMELER, 2017, p. 67).

O modo como os indivíduos passaram a se relacionar com dados, recentemente, contribuiu para o desenvolvimento da Biblioteconomia de Dados, que “ajuda pessoas a encontrar, utilizar e entender dados” (JOHNSON, 2019, p. 17) e que, diante de um grande volume de diferentes tipos de dados, ordena esse ecossistema, beneficiando ao máximo o seu uso por indivíduos e instituições. Para tanto, o bibliotecário de dados deve se atualizar continuamente tanto a respeito das técnicas biblioteconômicas quanto daquelas da Ciência de Dados, áreas que lhe proveem um rico e poderoso instrumental para extrair o máximo de valor e aproveitamento dos dados, a partir das demandas de seus usuários. Com isso, ele tem condições para participar ativamente dos processos e transformações da era dos dados (GORDON-MURNANE, 2012; SEMELER, 2017).

Nesse sentido, cumpre explorar as habilidades e competências que contribuem para que o agente desta área obtenha êxito no seu exercício profissional.

5 BIBLIOTECÁRIO DE DADOS: O PROFISSIONAL DA INFORMAÇÃO

A era dos dados vem proporcionado mudanças em diversas camadas da sociedade, transformando o modo como pessoas e organizações se relacionam com dados que cada vez mais fazem parte do cotidiano.

Em meio à inundação desses elementos, faz-se necessário o desenvolvimento de tecnologias, técnicas e estudos que permitam a criação, armazenamento, gerenciamento, curadoria, tratamento, compartilhamento e preservação de dados, bem como a criação de serviços e produtos de dados. Essas tarefas vêm sendo atribuídas a diferentes profissionais, como cientistas de dados, profissionais da tecnologia da informação, administradores e profissionais da informação, a exemplo do bibliotecário de dados (ECLEVIA *et al.*, 2019; SEMELER; PINTO; ROZADOS, 2017).

5.1 Habilidades e competências

A Biblioteconomia de Dados apresenta-se como uma área com profissionais de formação e experiências diversas, que podem atuar com diferentes níveis de especialização (de forma especializada ou generalista) e provendo distintos serviços e produtos de dados (FEDERER, 2018). Ela mescla os conhecimentos, habilidades e instrumentos da Ciência de Dados (ROY; SHOLLER, 2018) com “os princípios, práticas e recursos da Biblioteconomia tradicional aplicada aos dados. Isso envolve praticamente todos os estágios do trabalho tradicional do bibliotecário: aquisição (desenvolvimento de coleção), organização (catalogação e metadados) e a implementação de serviços de usuários” (SEMELER; PINTO; ROZADOS, 2019, p. 6, tradução nossa).

As habilidades dos bibliotecários de dados contemplam diversas frentes de trabalho que surgem em decorrência das demandas de usuário e instituições, o que reforça o caráter multidisciplinar e frutífero da disciplina.

Os bibliotecários de dados, entre outras habilidades, devem se preocupar com avaliação e retenção de dados, promoção de dados, marketing de dados, conscientização de dados, coordenação de práticas com dados em toda unidade e instituição, habilidades de negociação e habilidades de gerenciamento de expectativas de dados. (SEMELER, 2017, p. 77).

A variedade dos dados disponíveis na atualidade é um fator que contribui para o valor que dados ganharam recentemente. Com tal premissa, os bibliotecários de dados

“deve[m] possuir as habilidades necessárias para trabalhar com quaisquer tipos de dados, sejam eles observacionais, computacionais ou experimentais” (SEMELER, 2017, p. 77).

Bibliotecários tradicionalmente possuem habilidades e competências para trabalhar com coleta, organização, gerenciamento e preservação de dados (KIM, 2016). Para atuar como bibliotecários de dados, devem ajustar e aplicar suas habilidades para o novo contexto de dados (THOMAS; URBAN, 2018), pois as demandas de usuários e instituições requerem que eles se apropriem de habilidades que venham a ser pouco conhecidas e desenvolvidas nas escolas de Biblioteconomia. Para contribuírem genuinamente, devem rapidamente se tornar *data experts*, dominando os fluxos de trabalho do ciclo de vida dos dados e fornecendo serviços de dados (BARBARO, 2016).

Ele deve compreender como as investigações sobre dados modificam as práticas e as teorias que fundamentam a Biblioteconomia. Além de deter o papel de ajudar pesquisadores a coletar, localizar e recomendar recursos de dados apropriados, como o uso de provedores de serviços de dados de pesquisa, os bibliotecários de dados podem oferecer serviços de consultoria sobre o preparo e a aquisição de dados para análise de questões de pesquisa em disciplinas específicas (SEMELER, 2017, p. 77).

Verificam-se, na literatura, diversas compilações de habilidades do bibliotecário de dados (BARBARO, 2016; CALZADA PRADO; MARZAL, 2013; SEMELER, 2017). Algumas possuem foco em uma área de atuação específica, tal como curadoria de dados ou serviço de referência de dados. É o caso do guia “Oportunidades de Engajamento de Dados” (*Data Engagement Opportunities*), desenvolvido por Megan Sapp Nelson e Abigail Gobin, da Association of College & Research Libraries (ACRL), que lista 25 habilidades que podem ser transferidas da atividade bibliotecária para cada etapa do ciclo de vida dos dados de pesquisa (GOBIN; NELSON, 2018). Federer *et al.* (2020) compilaram uma quantidade abrangente dessas habilidades e as dispuseram nos termos da tabela 2.

Tabela 2 - Habilidades de bibliotecários de dados

Habilidade	Habilidade específica
<i>Data skills</i> (Habilidades com dados)	Plano de gestão de dados e fluxo de dados; Padrões e curadoria de dados e metadados; Compartilhamento e reuso de dados; Citação de dados
Habilidades computacionais	Literacia computacional Design de banco de dados Familiaridade com linguagens de programação relevantes como R e Python Machine learning e mineração de dados ou de texto Visualização de dados
Pesquisa e conhecimento de assuntos	Compreensão geral sobre a relevância científica ou do assunto de interesse; Design de pesquisa e de fluxos de trabalho; Estatística; Responsabilidade pela condução de pesquisa e ética de pesquisa Métodos de reprodutibilidade
Habilidades tradicionais de bibliotecas	Consulta e referência Pesquisa de literatura Comunicação acadêmica Bibliometria Treinamento e instruções Avaliação
Habilidades para desenvolvimento de programas e serviços	Entrevista e avaliação de habilidades para compreender as necessidades institucionais Escopo e planejamento para sustentabilidade Disposição para assumir falhas Habilidades de comunicação e marketing
Habilidades interpessoais	Ensino de habilidades de ciência em equipe Empreendedorismo Habilidade de defesa
Habilidades de aprendizado ao longo da vida	Flexibilidade e adaptabilidade Mindset “antropológico” Lógica e solução de problema Design thinking Computational thinking

Fonte: Federer *et al.* (2020)

Um bibliotecário de dados não precisa possuir todas as habilidades expostas na tabela 2 para ser bem sucedido (FEDERER *et al.*, 2020). Na verdade, “nem todos os bibliotecários estão interessados em se tornarem cientistas de dados” (BARBARO, 2016, p. 26) e nem todas as instituições vão solicitar todas as habilidades listadas, pois devem analisar criticamente quais são as necessidades e habilidades requeridas de acordo com os objetivos da instituição (FEDERER *et al.*, 2020).

De qualquer modo, tais habilidades contribuem para o oferecimento de serviços de dados em instituições, o que cria a oportunidade de estabelecer novos laços com novos grupos de usuários (MARTIN, 2012). Além disso, permitem dar suporte aos

usuários como criadores, aprendizes e colaboradores com acesso a software, fontes de dados, treinamentos, dentre outros recursos que venham a contribuir para sanar suas necessidades, cobrindo todas as etapas do ciclo de vida dos dados (GUSS, 2016). Outrossim, permitem que se levem em consideração “as questões socioculturais associadas com a coleta e uso de dados” (MARCHIONINI, 2016, p. 5).

A possibilidade de oferecer serviços para todos contribui para o incentivo e suporte ao ensino da literacia quantitativa para usuários que não tenham uma formação dessa natureza, tal como estudantes de cursos de STEM (Statistic, Technology, Engineering and Mathematics) (GUSS, 2016). Ademais, constata-se diversas vantagens e benefícios de oferecer serviços de dados para diferentes grupos e estruturas departamentais de uma instituição, como, por exemplo, o fortalecimento da relação de confiança entre os membros (GUSS, 2016).

A variedade de habilidades do bibliotecário de dados reflete o potencial de atuação desse profissional frente às demandas de indivíduos e instituições. Entende-se que cada habilidade pode ser desenvolvida em diferentes níveis, conforme as exigências do trabalho e da área de atuação do bibliotecário de dados.

5.2 Áreas de atuação

Bibliotecários de dados podem atuar em diferentes tipos de instituições, realizando diversas formas de tarefas, conforme a natureza e o objetivo da organização na qual se encontra. Enquanto houver necessidades de trabalhar com dados, independentemente da finalidade, haverá oportunidade para este profissional.

Na atualidade, as instituições acadêmicas apresentam uma grande demanda de trabalho com dados, o que as torna um dos principais palcos de atuação do bibliotecário especialista na área (FEDERER, 2018; SEMELER, 2017). Não obstante, é inegável que o mercado também tem lidado de forma intensa com dados, para a extração de valor e de conhecimento sobre eles (CONEGLIAN; GONÇALEZ; SANTARÉM SEGUNDO, 2017). Verifica-se, ainda, que a ebulição dessas demandas em diferentes camadas da sociedade faz com que esteja em curso um processo de alfabetização das crianças e de adolescentes na relação com os dados, mediação de que o bibliotecário também pode fazer parte (FONTICHIARO; OEHRLI, 2016).

Diante a existência de diversos cenários de atuação, analisam-se, com maior detalhe, algumas dessas potenciais áreas de atuação para o bibliotecário de dados: a

literacia de dados, a governança de dados, a análise de dados e a gestão de dados e repositórios de dados de pesquisa.

5.2.1 Literacia de dados

As mudanças proporcionadas pelo avanço tecnológico e do Big Data refletem-se diretamente sobre o modo como os indivíduos lidam com dados no dia a dia, seja no âmbito acadêmico, organizacional, governamental e até mesmo pessoal. Isso torna necessária uma base de conhecimento comum para leitura, interpretação e argumentação a partir de dados (WOLFF *et al.*, 2016).

A preocupação em aprender, interpretar e tomar decisões com dados no cotidiano diz respeito à literacia de dados, um campo de estudo recente, que pode ser compreendido como “[...] um conjunto de habilidades específicas e base de conhecimento que empodera indivíduos a transformar dados em informação e em conhecimento para tomada de decisão ao lhes permitir acessar, interpretar, avaliar criticamente, gerenciar e utilizar dados eticamente.” (KOLTAY, 2016, p. 10, tradução nossa)⁵.

Entende-se que a “alfabetização de dados” pode ser enraizada na cultura educacional dos indivíduos desde cedo. Dados estão presentes em nosso cotidiano, seja em gráficos exibidos em jornais e sites de notícias, estudos, ou até mesmo em aplicativos e sites de encontros pessoais, dentre outros. Nestes ambientes os indivíduos tornam-se capazes de compreender melhor, por meio da literacia de dados, como esses produtos são gerados, e também um senso crítico para avaliar a credibilidade e seriedade dos mesmos (BOWEN; BARTLEY, 2014; FONTICHIARO; OEHRLLI, 2016).

A criação de uma cultura e educação baseada em literacia de dados requer um grande esforço por parte de profissionais e educadores. Para que seja possível alfabetizar indivíduos no universo dos dados, é preciso que esses sujeitos possuam uma base de conhecimentos digitais ou estejam alfabetizados digitalmente (FONTICHIARO, OEHRLLI, 2016). A literacia digital refere-se à habilidade de usufruir das ferramentas digitais com o máximo de competência, eficácia e ética para fins de saciar suas

⁵ “[...] a specific skill set and knowledge base, which empowers individuals to transform data into information and into actionable knowledge by enabling them to access, interpret, critically assess, manage and ethically use data.” (KOLTAY, 2016, p. 10).

necessidades informacionais nas esferas pessoal, civil e profissional (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS, 2019).

A difusão da literacia digital possui obstáculos motivados pelas desigualdades de gênero, econômicas, sociais e educacionais. Elas tendem a se acirrar uma vez que diversos aspectos da vida estão migrando para o ambiente digital, tal como educação à distância, o e-banking, o e-commerce, as ofertas de trabalho, dentre outros, aumentando a exclusão digital (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS, 2019).

Este processo demonstra a importância do letramento digital. As literacias digitais se fazem essenciais inclusive para que o indivíduo seja capaz de analisar criticamente o conteúdo digital ao qual é exposto, uma vez que a internet propicia um ambiente que dificulta a validação da veracidade e confiabilidade de informações (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS, 2019). Nesse quesito, a literacia de dados contribui para que o sujeito desenvolva suas habilidades no universo das redes e mantenha sua autonomia intelectual.

Como aponta a Federação Internacional de Associações de Bibliotecas, as bibliotecas, em conjunto com diversos atores,

sempre desempenharam um papel crucial para com a literacia de dados ao prover acesso público à internet e outras ferramentas tecnológicas [...] [e] deram um passo lógico buscando desenvolver a literacia digital daqueles que não dominam serviços digitais e tecnologia, que se sentem desconfortáveis online, ou que não entendem o impacto – positivo ou negativo – do que eles fazem em suas vidas digitais. Bibliotecários estão cientes de que a habilidade de usar a tecnologia facilita significativamente os esforços profissionais, pessoais e sociais das pessoas (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS, 2019, tradução nossa).

Esse princípio de profunda colaboração dos bibliotecários para com a literacia digital da sociedade complementa o fazer bibliotecário em relação à literacia de dados, especialmente no meio escolar e acadêmico (FONTICHIARO; OEHLI, 2016; WOLFF *et al.*, 2016). A literacia de dados apresenta-se como uma tendência em bibliotecas acadêmicas e destaca-se como um serviço. De fato, espera-se que bibliotecários de dados exerçam o papel de educadores de literacia de dados (KOLTAY, 2017), pois as habilidades da literacia de dados se fazem essenciais para os indivíduos, independente do grau educacional dos mesmos. E o bibliotecário de dados possui os conhecimentos e competências requeridos, pelo seu entendimento sobre a importância e o impacto dos dados na sociedade, mas também pela sua capacidade de identificar fontes de dados e

coletá-los, bem como interpretá-los, analisá-los, manipulá-los e utilizá-los (CALZADA PRADO; MARZAL, 2013).

Existem diversas metodologias para implementação de um currículo de literacia de dados, (BOWEN; BARTLEY, 2014; CARLSON; JOHNSTON, 2015; FONTICHIARO; OEHRLE, 2016; KOLTAY, 2015; RIDSDALE *et al.*, 2015). Há, inclusive, propostas sob a ótica do letramento freiriano, conforme desenvolvido por Tygel e Kirsch (2015). Estes autores destacam que para que seja possível a literacia de dados, os indivíduos necessitam de habilidades técnicas provindas das áreas da computação, matemática e estatística. Com isso, tornam-se capazes de analisar, utilizar e produzir dados criticamente, que podem ser assimilados com a aplicação das etapas de investigação, tematização, problematização e sistematização, que contribuem para que os indivíduos sejam aptos a realizar a leitura, processamento, comunicação e produção de dados.

Considerando a aplicação de um currículo para a literacia de dados sob uma ótica mais detalhada em relação às disciplinas de dados, destaca-se a estrutura proposta por Ridsdale *et al.* (2015), que elaboraram uma matriz das competências com os seguintes temas:

- a) *Conceptual framework*: aquisição de conhecimento e compreensão das aplicações de dados;
- b) *Data collection*: conhecimento acerca da exploração e avaliação de fontes de dados de credibilidade;
- c) *Data management*: conhecimento de organização, manipulação, conversão, criação, curadoria e preservação de dados;
- d) *Data evaluation*: conhecimento em ferramentas para análise de dados, criação de planos de análise de dados, compreensão e interpretação de dispositivos para visualização de dados (tabelas, gráficos etc.), usabilidade de dados para identificação de problemas, capacidade de criação e organização de dispositivos para visualização de dados etc.;
- e) *Data application*: desenvolvimento de pensamento, cultura de dados, ética, referenciamento, compartilhamento para tomada de decisão baseada em dados.

Este tipo de letramento vem ganhando cada vez mais relevância frente ao impacto das exposições aos dados às quais os indivíduos vêm sendo expostos na atual era (CALZADA PRADO; MARZAL, 2013). A alfabetização de dados vem recentemente sendo tratada na literatura e aos poucos sendo implementada em diferentes instituições,

tal como escolas (FONTICHIARO; OEHLI, 2016), universidades, instituições de pesquisa (KOLTAY, 2017) e organizações do mercado e indústria (BRODSKY, 2019).

Percebe-se, assim, que o bibliotecário de dados pode desempenhar um papel essencial na literacia de diferentes grupos de indivíduos e instituições, contribuindo para a orientação dos mesmos em meio à inundação de dados proporcionada pelo Big Data e e-Science, capacitando-os para a análise, interpretação e apropriação de dados em tomadas de decisão.

5.2.2 Governança de dados

A governança de dados se apresenta como uma potencial área de atuação para o profissional da informação (KOLTAY, 2016). Ela pode ser compreendida como “uma ação multidisciplinar que tem como finalidade tratar os dados como insumos ativos e tangíveis nas organizações” (ESPÍNDOLA *et al.*, 2018, p. 276), por meio da determinação de “políticas, padronizações, processos, papéis, responsabilidades e tecnologias para melhor acompanhar e monitorar os dados gerados, armazenados, utilizados e eliminados na organização” (ESPÍNDOLA *et al.*, 2018, p. 276).

A governança de dados respalda-se nos princípios da integridade, transparência e auditabilidade, acompanhando os processos de gestão de dados na organização “como forma de garantir que os dados gerados nos processos estejam alinhados aos objetivos organizacionais. Além disso, monitora e acompanha os dados para que possam ser utilizados de forma eficaz e eficiente no atendimento dos objetivos organizacionais traçados pela alta administração” (ESPÍNDOLA *et al.*, 2018, p. 281). Ela contribui para que as organizações assegurem a qualidade de dados para a tomada de decisão, desenvolvimento de produtos e serviços (ESPÍNDOLA *et al.*, 2018; KOLTAY, 2016).

Os princípios da governança de dados podem ser aplicados para qualquer tipo de dado, tanto para organizações de negócios quanto para organizações acadêmicas e científicas (ESPÍNDOLA *et al.*, 2018), cujas práticas envolvem “processos e controles para garantir que as informações no nível de dados sejam confiáveis e únicas (não redundantes)” (ASSIS, 2018, p. 54). Apesar de se apresentar como uma área pouco explorada na literatura, bem como nas próprias instituições, é parte essencial para assegurar a qualidade de dados, contribuindo também para os processos de análise e gestão de dados de instituições (AUSTIN BOOTH; HENDRIX, 2015; KOLTAY, 2016).

A governança de dados é aplicada conforme o foco da instituição, percorrendo questões relativas à privacidade, políticas, normas e estratégias, qualidade de dados, arquitetura/integração, *data warehouse*, *business intelligence*, privacidade, *compliance* e segurança, alinhamento com as estratégias de TI e negócio (BARATA, 2015). Tem como responsabilidade o “gerenciamento de ativos de dados [...] planejamento, supervisão, controle e uso dos dados” (BARATA, 2015, p. 33).

As práticas de planejamento e controle do gerenciamento de dados contemplam diversas atividades, conforme indicado na tabela abaixo.

Tabela 3 – Planejamento e controle de gerenciamento de dados

Grupo	Atividade
Planejamento	Identificação de necessidades estratégicas de dados
	Desenvolvimento e manutenção da estratégia de dados
	Estruturação de papéis e funções no gerenciamento dos dados
	Identificação e nomeação dos profissionais que exercerão os papéis no contexto dos dados (por exemplo, administradores de dados, administradores de metadados, entre outros)
	Estabelecimento da função de gestor de dados e orientações do trabalho desta função
	Desenvolvimento e aprovação das políticas, padrões e procedimento dos dados
	Revisão e aprovação da arquitetura de dados
	Planejamento dos projetos de gestão e serviços de dados
	Estimativa do valor dos ativos de dados e os cursos associados ao seu gerenciamento
Grupo	Atividade
Controle	Supervisão das áreas e profissionais relacionados ao gerenciamento de dados
	Coordenação das atividades de GD
	Gerenciamento e resolução dos problemas de dados
	Monitoramento e garantia da conformidade regulatória, contemplando políticas padrões e arquitetura de dados
	Supervisão e gestão de projetos e serviços de dados
	Comunicação e promoção do valor dos ativos de dados

Fonte: Barata (2015, adaptação nossa)

A aplicação da governança de dados em instituições conta com o uso de *frameworks* que guiam e auxiliam a implementação de tarefas e determinação de seus *stakeholders*. A exemplo desses dispositivos, tem-se a Matriz RACI, Situation Analysis Canvas, Process Model Canvas, dentre outros.

Em relação à atuação do bibliotecário de dados nos processos de governança de dados, tem-se em mente que ele realiza atividades e dispõe de conhecimentos correlatos a essa prática, das atividades administrativas às técnicas, tais como tratar de políticas de licença de uso e de direitos autorais, manter a instituição alinhada à LDPD, desenvolver política de coleção, identificar e determinar fontes de informação e de dados de credibilidade, desenvolver normas e padronização para com os dados da instituição, criar ferramentas de controle e recuperação de coleções (como catálogos, tesouros, vocabulário controlado, dentre outras) que auxiliam no processo de gestão de dados da instituição (BARATA, 2015; DOSSANTOS, 2015; ESPÍNDOLA *et al.*, 2018; KOLTAY, 2016; PIZARRO; DAVOK, 2008; SANTOS, 2010).

Para a realização de atividades correlatas à governança de dados, verifica-se que as habilidades de programação não se mostram essenciais em um primeiro momento, mas se ressalta a importância de obter conhecimento acerca de fontes de informação, representação do conhecimento, desenvolvimento de coleções, administração e gerenciamento, literacia de dados, dentre outras, aplicadas ao contexto da instituição (BARATA, 2015; DOSSANTOS, 2015; ESPÍNDOLA *et al.*, 2018; KOLTAY, 2016; PIZARRO; DAVOK, 2008; SANTOS, 2010).

A contribuição do bibliotecário de dados para a governança de dados é tal que é possível fazer a seguinte analogia:

no mundo do Big Data, profissionais de governança de dados precisarão pensar em si mesmos como se fossem bibliotecários de dados, auxiliando usuários a encontrar o melhor dado disponível para trabalhadores do conhecimento. Eles irão, então, estabelecer processos para permitir que esses trabalhadores do conhecimento identifiquem o dado mais importante para a conformidade e a qualidade, de modo a movê-los para o segundo nível (DOSSANTOS, 2015, tradução nossa).

Por isso, entende-se que o bibliotecário de dados tem condições de desempenhar um papel essencial para a governança de dados. O seu conhecimento acerca de dados e tecnologia pode guiar a instituição para que se beneficie da melhor forma possível de dados de qualidade e credibilidade com o respaldo deste especialista.

5.2.3 Análise de dados

A análise de dados, presente no Big Data, é um dos fatores-chave para a compreensão do impacto deste fenômeno. Ela apresenta-se como uma área que

viabiliza a obtenção de novos insights e informações para a tomada de decisão, constituindo-se como parte fundamental de processos de inovação e desenvolvimento (CAO, 2018). A análise é um processo intermediário entre o dado e a informação, traduzindo o significado dos dados e contribuindo para a agregação de valor a eles.

A análise de dados caracteriza-se pela capacidade de investigar um conjunto de dados e responder a perguntas específicas acerca deles, de modo a deles extrair valor. No contexto do Big Data, verificam-se diversas formas de realizar esta tarefa (CAO, 2018; CONEGLIAN; GONÇALEZ; SANTARÉM SEGUNDO, 2017; MOREIRA; CARVALHO; HORVÁTH, 2019). Essas formas são representadas por modelos nos quais os dados passam por um processo que se inicia na coleta e termina na exposição e compartilhamento. Tal processo vem a ser conhecido como o ciclo de vida dos dados.

O processo de análise de dados respalda-se em diversas metodologias que variam entre si em relação à quantidade de etapas necessárias para a obtenção dos dados em sua forma bruta até a extração de valor dos mesmos (MOREIRA; CARVALHO; HORVÁTH, 2019). Dentre as diversas metodologias existentes, destacam-se as KDD e CRISP-DM, desenvolvidas com o foco acadêmico e industrial, respectivamente. O processo KDD constitui-se de nove etapas, que vão do aprendizado sobre o domínio até a utilização do conhecimento descoberto, enquanto a metodologia CRISP-DM é composta por seis etapas, que começam com o processo de entendimento do negócio até o lançamento da solução encontrada (MOREIRA; CARVALHO; HORVÁTH, 2019). Independentemente da metodologia de análise de dados adotada, compreende-se que, no geral, exige-se que se saiba sobre o domínio ou contexto no qual os dados estão inseridos, as técnicas da ciência de dados, a ciência da informação e o design para o tratamento e comunicação sobre os dados (CAO, 2018).

A análise de dados contribui para a obtenção de respostas acerca do passado, presente e futuro que eles venham a apresentar, contribuindo para que se compreenda o que aconteceu, o que está acontecendo, e o que pode acontecer, por meio de análise descritiva, preditiva e prescritiva, respectivamente (CAO, 2018). Por isso, a análise de dados tem se destacado, nos últimos anos, por contribuir para responder a perguntas estratégicas e a tomada de decisões estratégicas de maneira a produzir resultados igualmente estratégicos, independentes do contexto (AUSTIN BOOTH; HENDRIX, 2015). O exercício da análise de dados se apresenta como “uma oportunidade para bibliotecas e bibliotecários abordarem de forma proativa a cultura, o talento, o custo e

questões de dados em programas institucionais de análise de dados” (AUSTIN BOOTH; HENDRIX, 2015, p. 697).

Coneglian, Gonzalez e Santarém Segundo (2017) apresentam as contribuições do profissional da informação no ciclo de vida dos dados nas seis etapas de análise propostas por Bugembe (2016), indicadas na tabela a seguir.

Tabela 4 - As etapas da análise de dados e as contribuições do profissional da informação

Etapas	Objetivo	Contribuição do profissional da informação
(1) Fonte	Determinar a fonte dos dados a serem utilizadas	Identificação das melhores e principais fontes de informação de qualquer natureza.
(2) Captura e armazenamento	Determinar os dados necessários para a realização da análise	Propriedade e precisão para a indicação da temporalidade dos dados a serem obtidos.
(3) Processamento e fusão	Estruturação dos dados para fusão de informação	Propriedade para a identificação das estruturas e padrões para conversão de massas de dados em informações estruturadas; Aplicação de conceitos de representação da informação nos dados visando maior visibilidade informacional sobre eles.
(4) Acesso	Validação com gestores para com a definição dos dados a serem tratados	Disponibilização e compartilhamento dos dados para com outros departamentos e stakeholders.
(5) Análise	Construção de algoritmos e definição das estatísticas a serem utilizadas pela análise	Direcionamento do rumo e determinação das variáveis a serem consideradas pela análise; Colaborar com o profissional da computação.
(6) Exposição	Apontar e apresentar os resultados obtidos na análise realizada	Intermediar e expor com maior propriedade os resultados atingidos com a análise de dados.

Fonte: Coneglian, Gonzalez, Santarém Segundo (2017, adaptação nossa).

A participação do bibliotecário de dados no processo de análise de dados enfatiza tanto o seu caráter multifacetado quanto a sua capacidade de interagir com diferentes tipos de profissional. Essa característica é essencial para a sua atuação com o Big Data, pois este é um fenômeno interdisciplinar. Assim, envolver profissionais de diferentes áreas trabalhando em conjunto com um objetivo em comum, para extrair valor e conhecimento de dados, torna-se mais efetivo (CONEGLIAN; GONÇALEZ; SANTARÉM SEGUNDO, 2017).

Entende-se que o bibliotecário de dados não necessariamente precisa assumir o papel de analista de dados, mas pode oferecer suporte por meio do ensino das ferramentas que auxiliam na realização na análise de dados (tais como estatística, linguagens de programação, softwares de documentação, dentre outros). O profissional da informação precisa ter conhecimento sobre esses dispositivos a fim de conseguir reconhecer as origens dos dados processados (GOBEN; ZILINSKI; BRINEY, 2016). Uma vez que o bibliotecário de dados apresenta domínio acerca de análise de dados, o mesmo pode oferecer serviços e produtos, tanto para bibliotecas quanto para instituições de diversas naturezas, agregando valor aos dados e também ao seu próprio exercício profissional (GOBEN; ZILINSKI; BRINEY, 2016).

Entende-se que o bibliotecário de dados pode se apropriar de qualquer etapa do processo de análise de dados, uma vez que o mesmo adquira e desenvolva as habilidades e competências necessárias. Estas, somadas aos seus conhecimentos oriundos da Biblioteconomia e da Ciência da Informação, só vêm a agregar ao seu papel como analista de dados e, portanto, à sua melhor inserção no mercado de trabalho.

5.2.4 Gestão de dados e repositórios de pesquisa

A gestão de repositórios de dados de pesquisa se destaca como umas das áreas de atuação que acompanha o bibliotecário de dados desde o surgimento da disciplina em decorrência dos grandes volumes de dados oriundos das ciências sociais mantidos em repositórios de bibliotecas públicas, acadêmicas, e em arquivos de órgãos governamentais (CORRALL; KENNAN; AFZAL, 2013; KELLAM; THOMPSON, 2016; RICE; SOUTHALL, 2016; SEMELER, 2017). O desenvolvimento das práticas de armazenamento de dados ordenados se configurou no conceito de repositórios de dados, que podem ser compreendidos como uma robusta infraestrutura digital que viabiliza a gestão de dados de pesquisa. Eles têm como objetivo o armazenamento, organização, preservação e compartilhamento

[...] da produção intelectual, científica e/ou artística de uma instituição (repositórios institucionais), área do conhecimento (repositórios temáticos) ou mesmo de uma comunidade não necessariamente vinculada a uma instituição de ensino e pesquisa [...] e dos dados provenientes de pesquisas científicas (repositórios de dados de pesquisa), que podem ser compartilhados entre os pesquisadores de uma comunidade científica, visando a seu uso e reuso no contexto da Ciência Aberta (VECHIATO *et al.*, 2017, p. 9).

Repositórios de dados em instituições acadêmicas e científicas se destacam, nesse contexto, devido à produção de dados de pesquisa e dos benefícios da utilização, reutilização e compartilhamento de dados gerados por essas instituições. Eles contribuem para “capacitar os pesquisadores a formular novos tipos de indagações, hipóteses e a usar métodos analíticos inovadores no estudo de questões críticas para a ciência e para a sociedade” (SAYÃO; SALES, 2015, p. 5).

Os dados de pesquisa apresentam-se como “[...] produtos de pesquisa heterogêneos e contextualizados conforme as disciplinas do conhecimento. Nesse sentido, os dados de pesquisa devem ser válidos, compartilháveis, heterogêneos e contextualizáveis dentro de uma comunidade científica” (COSTA; DA CUNHA, 2014; PINTO; SEMELER, 2019, p. 122). Eles possuem a missão de garantir o acesso contínuo e aberto “aos resultados de pesquisa que se manifestam na forma de dados, e que são considerados parte importante do patrimônio digital da humanidade” (SAYÃO; SALES, 2016, p. 96). Visam, assim, à garantia de um ambiente dinâmico e flexível com o objetivo de dar apoio “à execução dos processos de gestão de dados de pesquisa.” (SAYÃO; SALES, 2016, p. 99).

Os repositórios de dados podem ser institucionais, orientados por objetos, disciplinares ou multidisciplinares (SAYÃO; SALES, 2016), e se formam conforme as necessidades das instituições e organizações na qual estão inseridos.

Repositórios de dados institucionais são aqueles gerenciados “no âmbito de uma instituição acadêmica, como universidades ou institutos de pesquisa, e são voltados para arquivar dados que são, geralmente, provenientes unicamente das atividades acadêmicas dessas instituições” (SAYÃO; SALES, 2016, p. 101), todavia, tendo em vista as diversas atividades de pesquisa desenvolvidas por instituições acadêmicas, compreende-se que seus repositórios podem ser disciplinares ou multidisciplinares.

Repositórios de dados orientados por objetos caracterizam-se por constituir coleções de dados que “são resultados de projetos de pesquisa ou resolução de problemas específicos” (SAYÃO; SALES, 2016, p. 104).

Os benefícios de se constituir um repositório de dados estão atrelados à aspectos relacionados à instituição, comunidade e produção científica, promovendo um novo status à produção acadêmica. Os benefícios podem se dão pela visibilidade dos dados, facilidade para compartilhamento, disponibilidade, curadoria e reutilização, segurança e preservação de dados, credibilidade de autoria conferida e dimensionada, insumo para a memória científica, desenvolvimento e verificação de indicadores de qualidade e

produtividade da instituição, criação de novos serviços de informação e formação de redes de repositórios de dados (SAYÃO; SALES, 2016).

Nesse sentido, os repositórios de dados de pesquisa configuram uma nova perspectiva às práticas de gestão de dados de pesquisa, perspectiva que se faz necessária para a contínua qualidade das práticas e repositórios de pesquisa. Ela se faz essencial para o registro e a preservação do fazer acadêmico, apresentando-se como uma tarefa “cada vez mais elementar para as bibliotecas acadêmicas” (PINTO; SEMELER, 2019, p. 122), e deve acontecer durante todo do ciclo de vida dos dados (PARTLO; SYMONS; CARLSON, 2015). É como Sayão e Sales (2016, p. 92) destacam:

As ações que coletivamente permeiam o ciclo de vida dos dados de pesquisa vêm sendo chamadas de gestão de dados de pesquisa, e se consolidam, de maneira ideal, em consonância com padrões de ampla aceitação, condicionantes de domínios disciplinares, requisitos estabelecidos pelos pesquisadores e, quando existentes, políticas institucionais e diretrizes de alcance nacional e internacional.

Em relação ao ciclo de vida dos dados, compreende-se que cumpre que haja uma infraestrutura tecnológica que os compreenda em sua totalidade, tal como nos repositórios digitais de dados de pesquisa, que vieram a se tornar cada vez mais populares em decorrência das exigências das “agências financiadoras de pesquisa e [d]a incorporação pelas instituições de pesquisa e pela sociedade em geral dos valores e princípios preconizados pela Ciência Aberta” (SAYÃO; SALES, 2016, p. 92).

A gestão de dados de pesquisa segue diferentes modelos desenvolvidos por diferentes instituições, de acordo com a natureza de suas práticas e necessidades (ARAÚJO *et al.*, 2019). Sayão e Sales (2016), por exemplo, apresentam um modelo de gestão de dados constituído por cinco etapas: (1) captura de dados, (2) catalogação das coleções de dados, (3) arquivamento e preservação, (4) interoperabilidade e (5) recuperação, acesso e reuso. O modelo DCC, do Data Curation Centre, a sua vez, que tem como objetivo promover abordagens de ciclo de vida para a gestão de materiais digitais (HIGGINS, 2008), constitui-se a partir das principais práticas de gestão de dados, tais como descrição e representação da informação, planejamento de preservação, participação da comunidade, curadoria e preservação (HIGGINS, 2008). De diferente modo, o modelo de gestão de dados de pesquisa do arquivo de dados do Reino Unido compõe-se por seis etapas, quais sejam, (1) planejamento de pesquisa, (2) coleta de dados, (3) processamento e análise de dados, (4) publicação e compartilhamento de

dados, (5) preservação de dados e (6) reutilização de dados (LI; DRESSEL; HERSEY, 2018).

As atividades realizadas por bibliotecários na gestão de dados de pesquisa, conforme o modelo do arquivo do Reino Unido, são apresentados na tabela a seguir, que indica as contribuições do bibliotecário em cada uma das etapas do ciclo.

Tabela 5 - Contribuições do bibliotecário de dados nas etapas da gestão de dados

Etapas da gestão de dados	Contribuições do bibliotecário
Planejamento de pesquisa	Conscientizar sobre benefícios do planejamento de pesquisa; Auxiliar usuários e pesquisadores em relação aos requisitos das agências e programas de pesquisa Criar o planejamento de gestão de dados.
Coleta de dados	Auxiliar com fontes e aquisição de dados Auxiliar com documentação do processo de coleta e organização de arquivos.
Processamento e análise de dados	Auxiliar com a aquisição e licença de uso de ferramentas de análise de dados; Oferecer treinamentos de ferramentas de análise de dados, de análise estatística e linguagens de programação que auxiliam na análise de dados; Orientar sobre a documentação de dados e metadados de modo a garantir que dados sejam encontrados por outros pesquisadores.
Publicação e compartilhamento de dados	Orientação em relação a requisitos para publicação e compartilhamento de dados em fontes relevantes; Fornecer ferramentas para compartilhamento e sincronização de dados com colaboradores Auxílio e orientação em relação a direitos, políticas e leis vigentes sobre dados.
Preservação de dados	Garantir que formatos de arquivos sejam utilizáveis ou migrados para novos formatos conforme a tecnologia se desenvolve; Gerenciar metadados para os dados do repositório; Garantir e realizar práticas de segurança, tal como rotinas de backup, versionamento, dentre outras.
Reutilização de dados	Orientar usuários e pesquisadores em relação às boas práticas de citação de dados; Compilar e monitorar as métricas de citação e impacto da instituição.

Fonte: Li, Dressel, Hersey (2018, p. 69–70, tradução e adaptação nossa)

As práticas relativas à gestão de dados exigem conhecimentos e práticas oriundas dos saberes administrativos e biblioteconômicos (SAYÃO; SALES, 2016), bem como de conhecimentos da informática. Por isso, o bibliotecário de dados deve se preocupar com a gestão dos repositórios de dados de pesquisa, e ainda, visando ao oferecimento de produtos e serviços diversos, podem atuar também com a descoberta, manipulação,

análise e visualização de dados disponíveis em repositórios de dados de pesquisa no contexto da e-Science, como assinala Semeler.

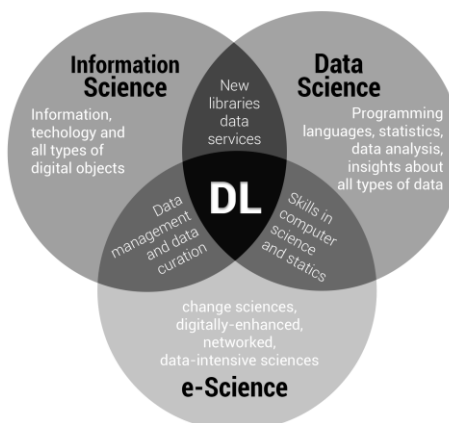
A gestão de dados de pesquisa é uma tarefa que deve ser assumida por bibliotecários de dados. Os dados de pesquisa são parte essencial do registro acadêmico científico, e a gestão dos dados de pesquisa é cada vez mais vista como uma tarefa elementar para bibliotecas acadêmicas. A gestão de dados de pesquisa é essencial para assegurar a transparência da pesquisa científica, preservar dados e permitir a reutilização e a reanálise de dados, garantindo os avanços do conhecimento científico (SEMELER, 2017, p. 80)

A e-Science apresenta-se como a forma de se fazer ciência a nível global, com pesquisadores e instituições conectados em redes colaborativas que, com o uso de dispositivos tecnológicos de diversas escalas (SEMELER, 2017), “assume a transformação das ciências em direção ao uso intensivo de grandes volumes de dados.” (SEMELER, 2017, p. 48).

A Biblioteconomia de Dados, no âmbito da e-Science, é exaustivamente explorada por Semeler e Pinto (SEMELER, 2017, 2017; SEMELER; PINTO, 2019; SEMELER; PINTO; ROZADOS, 2017), que afirmam que o foco da disciplina “está direcionado à criação de serviços e/ou produtos relacionados ao uso e consumo de dados de pesquisa” (PINTO; SEMELER, 2019, p. 115), os quais “são parte essencial do registro acadêmico científico e a gestão dos dados de pesquisa é cada vez mais vista como uma tarefa elementar para bibliotecas acadêmicas” (SEMELER; PINTO, 2019, p. 122).

Tudo isso configura um novo fazer para o bibliotecário de dados. Neste contexto, exige-se que ele tenha conhecimentos e habilidades que competem às áreas da Ciência de Informação, e-Science e *Data Science*. Esta relação com as demais áreas pode ser representada por meio do diagrama de Venn reproduzido abaixo.

Figura 3 – Diagrama de Venn da Biblioteconomia de Dados no contexto de e-Science



Fonte: Semeler, Pinto (2019, p. 26)

Acredita-se que há diversas oportunidades para bibliotecários de dados atuarem neste campo, uma vez que se verifica uma tendência mundial na produção, armazenamento e compartilhamento de dados de pesquisa que precisam crescer de modo ordenado, a fim de que instituições e pesquisadores usufruam ao máximo dessa produção.

6 A FORMAÇÃO EM BIBLIOTECONOMIA DE DADOS

O campo de atuação da Biblioteconomia de Dados, embora recente, tem diversas oportunidades de atuação conforme o Big Data se expande e continua a oferecer diferentes demandas, serviços e produtos de dados a indivíduos e instituições. Conforme visto anteriormente, a área é dotada de saberes, técnicas e habilidades tanto da Biblioteconomia quanto da Ciência de Dados, os quais capacitam o profissional da informação a atuar nesse universo de dados. Para tanto, exige-se que ele continue a se apropriar e a se atualizar constantemente quanto aos conhecimentos e práticas relacionados a dados e informação, especialmente pelo fato de a tecnologia se transformar de modo ininterrupto.

Em relação aos conhecimentos biblioteconômicos, “embora haja valor nas habilidades dos bibliotecários, há necessidade de um novo conjunto de habilidades para aprimorar o engajamento e suporte para o gerenciamento e exploração de dados” (KIM, 2016, p. 162). Assim, sendo necessário que haja investimentos no currículo de programas de Biblioteconomia para que se possam contemplar serviços de dados (THOMAS; URBAN, 2018). Compreende-se, também, que para a aquisição das habilidades e competências oriundas da Ciência de Dados, importa que bibliotecários busquem por cursos e formações complementares que lhes provejam do conhecimento acerca desse domínio.

Apesar da carência dos cursos de graduação, há grande oferta de cursos e programas de pós-graduação em Ciência de Dados (CAO, 2018). O crescente interesse em Ciência de Dados e tecnologia faz com que haja uma grande disponibilidade de cursos de programação, gerenciamento de banco de dados, mineração de dados, modelagem de dados, análise estatística, visualização de dados, dentre outros, que podem capacitar os bibliotecários para tratar as demandas técnicas de dados (CAO, 2018). No bojo desses cursos, destacam-se o “Data Scientist Training for Librarians” (DST4L, 2020) e o projeto “Library Carpentry”, que têm como objetivo ensinar habilidades de software e dados para a comunidade de bibliotecas e instituições de informação (LIBRARY CARPENTRY, 2020).

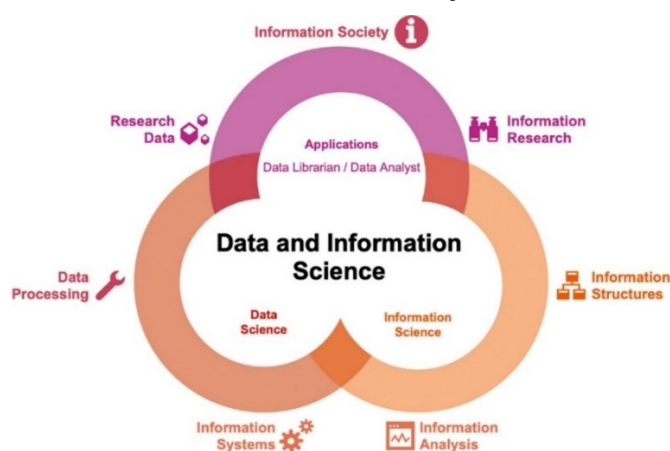
Nesse sentido, entende-se que a reformulação e implementação de currículos de graduação passam por processos burocráticos e administrativos de diversas ordens, dificultando as mudanças nos cursos que venham a atender de imediato as necessidades e mudanças do próprio mercado de trabalho (FÜHLES-UBACH *et al.*,

2019). Apesar disso, a existência de disciplinas e currículos voltados à Biblioteconomia de Dados em escolas de Biblioteconomia e de Ciência da Informação ao nível da graduação vêm mudando aos poucos. Verifica-se uma maior tendência para o surgimento de cursos extracurriculares e de pós-graduação, especialmente nos Estados Unidos (KIM, 2016; MCCAFFREY; GIESBRECHT, 2016).

No âmbito exclusivo da educação formal, destaca-se a iniciativa da Universidade Técnica de Colônia (TH Köhl), na Alemanha, que entre 2017 e 2018 começou um processo de reformulação dos cursos de graduação com a análise crítica dos professores do Instituto de Ciência da Informação, em colaboração com estudantes e com 20 representantes de bibliotecas públicas, científicas e especiais do país (FÜHLES-UBACH *et al.*, 2019). Nesse processo, verificaram o perfil, as competências, as habilidades de bibliotecários, dentre outros aspectos, que resultaram em dois cursos de graduação: “Biblioteca e Comunicação Digital” (*Library and Digital Communication*) e “Ciência de Dados e de Informação” (*Data and Information Science*)⁶ (FÜHLES-UBACH *et al.*, 2019).

O curso de “Ciência de Dados e de Informação” combina as práticas da Ciência da Informação e da Ciência de Dados, permitindo que os estudantes escolham entre dois caminhos de especialização, o do mercado, como analistas de dados, ou o da academia, como bibliotecários de dados. Estas novidades vieram a “equipar os estudantes com os requisitos do trabalho do moderno mundo orientado por dados informacionais” (FÜHLES-UBACH *et al.*, 2019, p. 258), conforme ilustrado na figura abaixo.

Figura 4 - O curso de Ciência de Dados e de Informação e suas características técnicas



Fonte: Fühles-Ubach *et al.* (2019)

⁶ Os cursos podem ser consultados no endereço eletrônico: https://www.th-koeln.de/en/academics/data-and-information-science-bachelors-program_53882.php.

No grupo de “Ciência de Dados” são ensinadas habilidades técnicas relacionadas ao processamento dados, tais como identificação, análise, modelagem, mineração e extração de dados e textos. Para isso, ensina-se sobre programação em Python, desenvolvimento e administração de software, dentre outros conhecimentos que tangem aos sistemas de informação (FÜHLES-UBACH *et al.*, 2019). Já no grupo de “Ciência da Informação” são ensinados conhecimentos e técnicas de organização da informação, de estrutura de informação em processos de pesquisa, indexação automática, arquitetura da informação, design e avaliação de sistema de recuperação da informação (FÜHLES-UBACH *et al.*, 2019).

A sua vez, o conjunto das “Aplicações” tem como propósito o ensino da aplicabilidade dos dados às principais áreas de atuação do bibliotecário de dados, no mercado e na academia. Os estudantes aprendem sobre a análise de informação para a tomada de decisões em negócios, a integração de informação de diferentes fontes, a gestão de dados etc. Nesse eixo, destaca-se, ainda, o núcleo “Sociedade da Informação”, em que os estudantes aprendem sobre os aspectos éticos e legais de processos de captura de dados, conhecimento sobre fontes de informação, sua qualidade e proteção de dados (FÜHLES-UBACH *et al.*, 2019).

Acredita-se que o curso de “Ciência de Dados e Informação” da Universidade Técnica de Colônia capture a expressão e o potencial da Biblioteconomia de Dados, especialmente em relação às áreas de atuação deste profissional, que atualmente ainda está centralizado em instituições acadêmicas. Um curso de graduação que contemple o mercado tradicional abre espaço para que o bibliotecário se desenvolva e se aproprie de posições que surgem diariamente em decorrência dos impactos do Big Data nas organizações. Isso pois “profissionais da informação precisam compreender esse novo cenário de dilúvio de dados e como podem contribuir para uma gestão eficiente da informação” (COSTA; DA CUNHA, 2014, p. 201). Assim, devem ampliar “[...] sua atuação profissional para além dos muros das instituições, e verifi[car] que na ambiência digital sua expertise e seus saberes são essenciais.” (CONEGLIAN; GONÇALEZ; SANTARÉM SEGUNDO, 2017, p. 131).

7 CONCLUSÃO

As mudanças proporcionadas pelo Big Data produzem efeitos em diversas camadas da sociedade e têm impactado principalmente o modo como as pessoas e instituições trabalham e lidam com dados. Esta partícula elementar da informação vem sendo gerada em grande volume, em diversos formatos, a uma velocidade que está além da habilidade humana de processamento. O fenômeno proporcionou o desenvolvimento de áreas de atuação que se preocupam com o tratamento desse dilúvio de dados, por meio da criação de ferramentas e métodos para traduzi-lo de modo que indivíduos e instituições possam se beneficiar ao máximo desses recursos.

Exemplo de setor que surge neste cenário é a Biblioteconomia de Dados. Ela deve ser compreendida como uma evolução natural da Biblioteconomia para atender às demandas informacionais da sociedade hodierna. Para tanto, ela busca adaptar as técnicas existentes da Biblioteconomia para o contexto atual ao complementá-las com conhecimentos multidisciplinares, em especial, oriundos da Ciência de Dados.

A Biblioteconomia de Dados não tem como objetivo substituir as práticas e fazeres da Biblioteconomia tradicional, mas ampliá-los com os recursos do universo biblioteconômico informacional. Isso não representa uma descaracterização da profissão do bibliotecário, mas da formação de um profissional da informação especializado no tratamento de dados.

Acredita-se que o bibliotecário de dados tem capacidade e proficiência para atuar com dados, independente do domínio, e tal fato destaca que esse profissional tem diversas possibilidades de atuação, tal como bibliotecas de diversas naturezas, instituições de pesquisa e empresas do mercado e indústria, podendo atuar em diferentes frentes de trabalho, tal como atividades de ordem administrativa, gerencial, analítica e educativa, dentre outras, colocando-a como protagonista nos processos de organização e tratamento de dados que são cada vez mais recorrentes e urgentes na era dos dados.

REFERÊNCIAS⁷

AFFELT, A. L. **The accidental data scientist**: big data applications and opportunities for librarians and information professionals. Medford, New Jersey: Information Today, 2015. 194 p.

ALMEIDA, N. B. F.; BAPTISTA, S. G. Breve histórico da Biblioteconomia brasileira: formação do profissional. In: XXV CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, 15., 2013, Florianópolis. **Anais [...]**. Florianópolis: FEBAB, 2013. p. 1–12. Disponível em: <http://repositorio.febab.org.br/files/original/8/2396/1508-1521-1-PB.pdf>. Acesso em: 26 ago. 2020.

AL-SALIM, A. M. *et al.* Greening big data networks: velocity impact. **IET Optoelectronics**, Reino Unido, v. 12, n. 3, p. 126–135, jun. 2018. Disponível em: <https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/iet-opt.2016.0165>. Acesso em: 27 jul. 2020.

ARAÚJO, D. G. *et al.* Contribuições para a gestão de dados científicos: análise comparativa entre modelos de ciclo de vida dos dados. **Liinc em Revista**, Rio de Janeiro, v. 15, n. 2, p. 32–51, 11 dez. 2019. Disponível em: <http://revista.ibict.br/liinc/article/view/4686>. Acesso em: 23 mar. 2020.

ASSIS, C. B. **Governança da informação**: viabilizadores e inibidores para adoção organizacional. 2018. Tese (Doutorado em Engenharia de Produção) - Escola Politécnica, Universidade de São Paulo, São Paulo, 2018. Disponível em: <https://www.teses.usp.br/teses/disponiveis/3/3136/tde-27042018-102121/pt-br.php>. Acesso em: 26 set. 2021.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6023**: informação e documentação: referências: elaboração. Rio de Janeiro, 2018. 68 p.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 10520**: informação e documentação: citações em documentos: apresentação. Rio de Janeiro, 2002. 7 p.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 17424**: informação e documentação: trabalhos acadêmicos: apresentação. Rio de Janeiro, 2011. 11 p.

AUSTIN BOOTH, H.; HENDRIX, D. Libraries and institutional data analytics: challenges and opportunities. **The Journal of Academic Librarianship**, Reino Unido, v. 41, n. 5, p. 695–699, set. 2015. Disponível em: <https://rrpress.utsa.edu/handle/20.500.12588/20>. Acesso em: 13 abr. 2020.

BARATA, A. M. **Governança de dados em organizações brasileiras**: uma avaliação comparativa entre os benefícios previstos na literatura e os obtidos pelas organizações. 2015. Dissertação (Mestrado em Sistemas de Informação) - Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2015. Disponível em: <https://www.teses.usp.br/teses/disponiveis/100/100131/tde-28072015-215618/pt-br.php>. Acesso em: 26 set. 2020.

⁷ De acordo com a Associação Brasileira de Normas Técnicas (ABNT NBR 6023).

BARBARO, A. On the importance of being a data-savvy librarian. **Journal of EAHIL**, [S.I.], v. 12, n. 1, 4 p. 24–27, mar. 2016. Disponível em: <http://ojs.eahil.eu/ojs/index.php/JEAHIL/article/view/100>. Acesso em: 28 fev. 2020.

BARRETO, A. A. A aventura de perceber significados. **DataGramZero - Revista de Informação**, Rio de Janeiro, v. 15, n. 3, p. 1–8, out. 2014. Disponível em: <https://brapci.inf.br/index.php/res/v/8211>. Acesso em: 19 mai. 2020.

BAWDEN, D.; ROBINSON, L. **Introduction to information science**. Londres: Facet publishing, 2012. 351 p.

BIHL, T. J.; YOUNG II, W. A.; WECKMAN, G. R. Defining, understanding, and addressing Big Data. **International Journal of Business Analytics**, Estados Unidos, v. 3, n. 2, p. 1–32, abr. 2016. Disponível em: <https://www.igi-global.com/gateway/article/149153>. Acesso em: 1 jun. 2020.

BOWEN, M.; BARTLEY, A. **The basics of data literacy**: helping your students (and you!) make sense of data. Arlington: NSTA Press, 2014. 166 p.

BRODSKY, M. The Role of Business Librarians in Teaching Data Literacy. **Ticker: the academic business librarianship review**, [S.I.], v. 1, n. 3, p. 1–6, jul. 2019. Disponível em: <https://quod.lib.umich.edu/cgi/p/pod/dod-idx/role-of-business-librarians-in-teaching-data-literacy.pdf?c=ticker;idno=16481003.0001.301;format=pdf>. Acesso em: 23 mar. 2020

CAESARIUS, L. M.; HOHENTHAL, J. Searching for big data. **Scandinavian Journal of Management**, [S.I.], v. 34, n. 2, p. 129–140, jun. 2018. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0956522118301064>. Acesso em: 27 jul. 2020.

CALZADA PRADO, J.; MARZAL, M. Á. Incorporating data literacy into information literacy programs: core competencies and contents. **Libri**, Berlin, v. 63, n. 2, p. 123–134, jan. 2013. Disponível em: <https://www.degruyter.com/document/doi/10.1515/libri-2013-0010/html>. Acesso em: 19 mai. 2020.

CAO, L. **Data science thinking**: the next scientific, technological and economic revolution. Cham: Springer International Publishing, 2018. 390 p.

CAPURRO, R.; HJØRLAND, B. O conceito de informação. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 12, n. 1, p. 148–207, jan./abr. 2007. Disponível em: <https://www.brapci.inf.br/index.php/res/v/33134>. Acesso em: 6 mai. 2020.

CARLSON, J.; JOHNSTON, L. **Data information literacy**: librarians, data, and the education of a new generation of researchers. Indiana: Purdue University Press, 2015. 282 p.

CASTELLS, M. O espaço de fluxos. *In: A sociedade em rede*. São Paulo: Editora Paz e Terra, 1999. p. 403–455.

CAVANILLAS, J. M.; CURRY, E.; WAHLSTER, W. **New horizons for a data-driven economy**. Cham: Springer International Publishing, 2016. 303 p.

CLOUDMOYO. **Sources of Big Data**: where does it come from? Disponível em: <https://www.cloudmoyo.com/blog/data-architecture/what-is-big-data-and-where-it-comes-from>. Acesso em: 17 maio. 2020.

CONEGLIAN, C. S.; GONÇALEZ, P. R. V. A.; SANTARÉM SEGUNDO, J. E. O profissional da informação na era do big data. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 22, n. 50, p. 128–143, set. 2017. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2017v22n50p128>. Acesso em: 19 mar. 2020.

CONWAY, D. The data science Venn diagram. **Drew Conway**, 30 set. 2010. Disponível em: <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>. Acesso em: 20 maio 2020.

CORRALL, S.; KENNAN, M. A.; AFZAL, W. Bibliometrics and research data management services: emerging trends in library support for research. **Library Trends**, Baltimore, v. 61, n. 3, p. 636–674, 2013. Disponível em: <http://d-scholarship.pitt.edu/18948/>. Acesso em: 1 fev. 2020.

COSTA, M. M.; DA CUNHA, M. B. O bibliotecário no tratamento de dados oriundos da e-science: considerações iniciais. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 19, n. 3, p. 189–206, set. 2014. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/1900/1445>. Acesso em: 3 mar. 2020

DALE, N.; WALKER, H. M. A classification of data types. **Computer Science Education**, Austin, v. 3, n. 3, p. 223–232, jun. 1992.

DAVENPORT, T. H.; PATIL, D. J. Data scientist: the sexiest job of the 21st century. **Harvard Business Report**, Massachusetts, v. 90, n. 10, p. 70–76, out. 2012. Disponível em: <https://www.hbs.edu/faculty/Pages/item.aspx?num=43110>. Acesso em: 30 mai. 2020

DEVLIN, B.; ROGERS, S.; MYERS, J. **Big data comes of age**. Cornwall: 9sight Consulting; Colorado: EMA, 2012. 43 p. Disponível em: http://9sight.com/pdfs/Big_Data_Comes_of_Age.pdf. Acesso em: 17 maio 2020.

DOSSANTOS, J. **What librarians can teach us about managing big data**. Disponível em: https://infocus.delltechnologies.com/joe_dossantos/what-librarians-can-teach-us-about-managing-big-data/. Acesso em: 20 maio. 2020.

DST4L. About. **DST4L**, s.l., s.d. Disponível em: <http://www.dst4l.info/about.html>. Acesso em: 17 maio. 2020.

ECLEVIA, M. R. *et al.* What makes a data librarian? : an analysis of job descriptions and specifications for data librarian. **Qualitative and Quantitative Methods in**

Libraries, [S.l.], v. 8, n. 3, p. 273–290, nov. 2019. Disponível em: <http://qqml-journal.net/index.php/qqml/article/view/541>. Acesso em: 4 mar. 2020

EKSTRØM, J. **Theories and practice in data science librarianship**. London: FACET Publishing, 2019. 224 p.

ESPÍNDOLA, P. L. *et al.* Governança de dados aplicada à ciência da informação: análise de um sistema de dados científicos para a área da saúde. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 16, n. 3, p. 274–298, ago. 2018. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8651080>. Acesso em: 13 mar. 2020.

FEDERER, L. Defining data librarianship: a survey of competencies, skills, and training. **Journal of the Medical Library Association**, Pittsburgh, v. 106, n. 3, p. 294–303, jul. 2018. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6013124/>. Acesso em: 10 jun. 2020.

FEDERER, L. *et al.* **Developing the librarian workforce for data science and open science**. [S.l.]: OSF Preprints, 2020. Disponível em: <https://osf.io/uycax>. Acesso em: 18 out. 2020. 44 p.

FONTICHIARO, K.; OEHRLE, J. Why data literacy matters. **Knowledge Quest**, Chicago, v. 44, n. 5, p. 21–27, maio/jun. 2016. Disponível em: <https://files.eric.ed.gov/fulltext/EJ1099487.pdf>. Acesso em: 20 maio 2020.

FRICKÉ, M. The knowledge pyramid: the DIKW hierarchy. **Knowledge Organization**, Baden-Baden, v. 46, n. 1, p. 33–46, 2019. Disponível em: <https://www.nomos-elibrary.de/10.5771/0943-7444-2019-1-1/titelei-inhaltsverzeichnis-volume-46-2019-issue-1>. Acesso em: 6 jul. 2020.

FÜHLES-UBACH, S. *et al.* Data librarian – ein neuer studienswerpunkt für wissenschaftliche bibliotheken und forschungseinrichtungen. **Bibliothek Forschung und Praxis**, Colônia, v. 43, n. 2, p. 255–261, jul. 2019. Disponível em: <https://edoc.hu-berlin.de/handle/18452/20699>. Acesso em: 10 jun. 2020.

FURNER, J. Data: The data. In: KELLY M., BIELBY J. **Information cultures in the digital age: a festschrift in honor of Rafael Capurro**. Wiesbaden: Springer, 2016. p. 287–306.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2004. 176 p.

GOBEN, A.; NELSON, M. S. The data engagement opportunities scaffold: development and implementation. **Journal of eScience Librarianship**, [S.l.], v. 7, n. 2, p. 1–7, mar. 2018. Disponível em: <https://escholarship.umassmed.edu/jeslib/vol7/iss2/1/>. Acesso em: 7 jul. 2020.

GOBEN, A.; ZILINSKI, L.; BRINEY, K. Going beyond the data management plan: services and partnerships. In: FEDERER, L. **The Medical Library Association Guide**

to **Data Management for Librarians**. Lahan: Rowman & Littlefield, 2016. p. 1–19. Disponível em: https://dc.uwm.edu/lib_staffart/9/. Acesso em: 1 jun. 2020

GORDON-MURNANE, L. Big Data: A Big Opportunity for Librarians. **Online**, [S.l.], v. 36, n. 5, p. 30–34, set./out. 2012.

GOYAL, D. *et al.* Emerging trends and challenges in data science and big data analytics. **Anais [...]** In: 2020 INTERNATIONAL CONFERENCE ON EMERGING TRENDS IN INFORMATION TECHNOLOGY AND ENGINEERING, 2020. Vellore: IEEE, 2020. p.1–8 Disponível em: <https://ieeexplore.ieee.org/document/9077786/>. Acesso em: 17 set. 2020

GUSS, S. A studio model for academic data services. In: **Databrarianship: the academic data librarian in theory and practice**. Chicago: Association of College and Research Libraries, 2016. p. 9–24.

HIGGINS, S. The DCC Curation Lifecycle Model. **International Journal of Digital Curation**, [S.l.], v. 3, n. 1, p. 134–140, 2 dez. 2008. Disponível em: <http://www.ijdc.net/article/view/69>. Acesso em: 13 abr. 2020.

HJØRLAND, B. Data (with Big Data and Database Semantics). **Knowledge Organization**, Baden-Baden v. 45, n. 8, p. 685–708, nov. 2018. Disponível em: <https://www.isko.org/cyclo/data>. Acesso em: 9 jun. 2020.

HOY, M. B. Big data: an introduction for librarians. **Medical Reference Services Quarterly**, Estados Unidos, v. 33, n. 3, p. 320–326, jul. 2014. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/02763869.2014.925709>. Acesso em: 13 ago. 2020.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS. **IFLA Statement on Digital Literacy**, 18 ago. 2017. Disponível em: https://www.ifla.org/publications/node/11586#_ftn11. Acesso em: 20 jun. 2020.

JOHNSON, E. O. **Working as a data librarian**: a practical guide. Santa Barbara, California: Libraries Unlimited, 2019. 245 p.

KELLAM, L. M.; PETER, K. **Numeric data services and sources for the general reference librarian**. Oxford: Chandos Publishing, 2011. 248 p.

KELLAM, L. M.; THOMPSON, K. **Databrarianship**: the academic data librarian in theory and practice. Chicago: Association of College and Research Libraries, 2016. 385 p.

KERBAUY, M. T. M. Diversidade cultural e política de informação. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 5, n. 1/2, p. 60–69, jan./dez. 2009. Disponível em: <https://brapci.inf.br/index.php/res/v/4290>. Acesso em: 14 mai. 2020.

KIM, J. Who is teaching data: meeting the demand for data professionals. **Journal of Education for Library and Information Science Online**, Westford, v. 57, n. 2, p. 161–

173, abr. 2016. Disponível em:
<https://www.utpjournals.press/doi/abs/10.3138/jelis.57.2.161>. Acesso em: 28 fev. 2020.

KOLTAY, T. Data governance, data literacy and the management of data quality. **IFLA Journal**, California, v. 42, n. 4, p. 303–312, nov. 2016. Disponível em:
<https://journals.sagepub.com/doi/abs/10.1177/0340035216672238>. Acesso em: 31 jul. 2020.

KOLTAY, T. Data literacy for researchers and data librarians. **Journal of Librarianship and Information Science**, California, v. 49, n. 1, p. 3–14, mar. 2017. Disponível em:
<https://journals.sagepub.com/doi/abs/10.1177/0961000615616450>. Acesso em: 13 ago. 2020.

KOLTAY, T. Data literacy: in search of a name and identity. **Journal of Documentation**, [S.l.], v. 71, n. 2, p. 401–415, mar. 2015. Disponível em:
<https://www.deepdyve.com/lp/emerald-publishing/data-literacy-in-search-of-a-name-and-identity-6qJ1GU8tEc>. Acesso em: 14 mai. 2020.

LE COADIC, Y.F. **A ciência da informação**. Brasília: Briquet de Lemos, 1996. 119 p.

LEE, I. Big data: Dimensions, evolution, impacts, and challenges. **Business Horizons**, v. 60, n. 3, p. 293–303, 2017. Disponível em:
<https://www.sciencedirect.com/science/article/abs/pii/S0007681317300046>. Acesso em: 7 jul. 2020.

LI, Y.; DRESSEL, W.; HERSEY, D. Research Data Management: what can librarians really help?. In: INTERNATIONAL CONFERENCE ON GREY LITERATURE “RESEARCH DATA FUELS AND SUSTAINS GREY LITERATURE”, 20., 2018, New Orleans. **Proceedings [...]**. New Orleans: TextRelease, 2019. p. 67–74. Disponível em:
http://greyguide.isti.cnr.it/attachments/category/30/GL20_Conference_Proceedings.pdf. Acesso em: 19 fev. 2020.

LIBRARY CARPENTRY. About Us. **Library Carpentry**, s.l., s.d. Disponível em:
<https://librarycarpentry.org/about/>. Acesso em: 17 maio. 2020.

LINKEDIN. **Profissões Emergentes 2020**. Disponível em:
https://business.linkedin.com/content/dam/me/business/en-us/talent-solutions/emerging-jobs-report/Emerging_Jobs_Report_Brazil.pdf. Acesso em: 15 jul. 2020.

MAATTA, S. L. The emerging databrarian digital and data management roles reach critical mass among recent LIS graduates. **Library Journal**, Estados Unidos, v. 138, n. 17, p. 26–33, out. 2013.

MARCHIONINI, G. Information science roles in the emerging field of data science. **Journal of Data and Information Science**, China, v. 1, n. 2, p. 1–6, mar. 2016.

MARQUESONE, R. **Big Data: técnicas e tecnologia para extração de valor dos dados**. São Paulo: Casa do Código, 2016. 242 p.

MARTIN, E. What do data services librarians do? **Journal of eScience Librarianship**, Massachusetts, v. 1, n. 3, p. 115–117, mar. 2013. Disponível em: <https://escholarship.umassmed.edu/cgi/viewcontent.cgi?article=1038&context=jeslib>. Acesso em: 6 jul. 2020.

MCCAFFREY, M.; GIESBRECHT, W. Teaching data librarianship to LIS students. In: **Databrarianship: the academic data librarian in theory and practice**. Chicago: Association of College and Research Libraries, 2016. p. 355–371.

MOREIRA, J.; CARVALHO, A. C. P. DE L. F.; HORVÁTH, T. **A general introduction to data analytics**. Hoboken: John Wiley & Sons, 2019. 325 p.

ORTEGA, C. D. Relações históricas entre biblioteconomia, documentação e ciência da informação. **DataGramZero - Revista de Informação**, Rio de Janeiro, p. 13, out 2004. Disponível em: <https://www.brapci.inf.br/index.php/article/download/7649>. Acesso em: 23 abr. 2020.

OUSSOUS, A. *et al.* Big data technologies: a survey. **Journal of King Saud University - Computer and Information Sciences**, [S.l.], v. 30, n. 4, p. 431–448, out. 2018. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1319157817300034?via%3Dihub>. Acesso em: 31 jul. 2020.

PARTLO, K.; SYMONS, D.; CARLSON, J. D. Revolutionary or evolutionary? Making research data management manageable. In: EDEN, B. L. **Creating Research Infrastructures in the 21st-Century Academic Library: Conceiving, Funding, and Building New Facilities and Staff**. Lanham: Rowman & Littlefield, 2015. p. 175–201. Disponível em: https://digitalcommons.csbsju.edu/library_pubs/8/. Acesso em: 9 jun. 2020.

PINTO, A. L.; SEMELER, A. R. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. **Ciência da Informação**, Brasília, v. 48, n. 1, p. 130–129, maio. 2019. Disponível em: <http://revista.ibict.br/ciinf/article/view/4461>. Acesso em: 18 ago. 2020.

PIZARRO, D. C.; DAVOK, D. F. O papel do bibliotecário na gestão da informação empresarial: uma pesquisa bibliográfica em periódicos nacionais de biblioteconomia e ciência da informação. **Revista ACB**, Florianópolis, v. 13, n. 1, p. 37–58, jan./jun. 2008. Disponível em: <https://revista.acbsc.org.br/racb/article/view/546>. Acesso em: 18 ago. 2020.

REINSEL, D.; GANTZ, J.; RYDNING, J. **The digitization of the world: from edge to core**. Framingham: International Data Corporation, 2018. 28 p. Disponível em: <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-data-age-whitepaper.pdf>. Acesso em: 19 ago. 2020.

RICE, R.; SOUTHALL, J. **The data librarian's handbook**. London: Facet Publishing, 2016. 192 p.

RIDSDALE, C. *et al.* Strategies and best practices for data literacy education. **Knowledge Synthesis Report**. [S.l.]: Dalhousie University, 2015. 75 p. Disponível em: <https://dalspace.library.dal.ca/handle/10222/64578>. Acesso em: 19 ago. 2020.

ROSENBERG, D. Early modern information overload. **Journal of the History of Ideas**, Pennsylvania, v. 64, n. 1, p. 1–9, jan. 2003. Disponível em: <https://muse.jhu.edu/article/42178>. Acesso em: 19 fev. 2020.

ROWLEY, J. The wisdom hierarchy: representations of the DIKW hierarchy. **Journal of Information Science**, Reino Unido, v. 33, n. 2, p. 163–180, abr. 2007. Disponível em: <http://www.public.imtbs-tsp.eu/~gibson/Teaching/Teaching-ReadingMaterial/Rowley06.pdf>. Acesso em: 23 abr. 2020.

ROY, L.; SHOLLER, D. What reference librarians should know about data science. **The Reference Librarian**, [S.l.], v. 60, n. 2, p. 109–116, dez. 2018. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/02763877.2018.1560964?needAccess=true&journalCode=wref20>. Acesso em: 1 fev. 2020.

SANTOS, A. P. L. DOS; RODRIGUES, M. E. F. Biblioteconomia: gênese, história e fundamentos. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 9, n. 2, p. 116–131, jul. 2013. Disponível em: <https://brapci.inf.br/index.php/res/v/1186>. Acesso em: 6 mai. 2020.

SANTOS, I. M. F. Dos. **Uma proposta de governança de dados baseada em um método de desenvolvimento de arquitetura empresarial**. 2010. 128 f. Dissertação (Mestrado em Informática) - Centro de Ciências Exatas e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010. Disponível em: <http://www2.uniriotec.br/ppgi/banco-de-dissertacoes-ppgi-unirio/ano-2010/uma-propostade-governanca-de-dados-baseada-em-um-metodo-de-desenvolvimento-de-arquiteturaempresarial/view>. Acesso em: 10 abr. 2020.

SANTOS, J. M. O processo evolutivo das Bibliotecas da Antiguidade ao Renascimento. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 8, n. 2, p. 175–189, jul. 2012. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/237>. Acesso em: 27 mar. 2020.

SANTOS, P. L. V. A. DA C.; SANTANA, R. C. G. Dado e granularidade na perspectiva da informação e tecnologia: uma interpretação pela ciência da informação. **Ciência da Informação**, Brasília, v. 42, n. 2, p. 199–209, jan. 2013. Disponível em: <http://revista.ibict.br/ciinf/article/view/1382>. Acesso em: 13 mar. 2020.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, Londrina, v. 21, n. 2, p. 90–115, dez. 2016. Disponível em: <https://www.uel.br/revistas/uel/index.php/informacao/article/view/27939>. Acesso em: 19 fev. 2020.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. 90 p. Disponível em: <http://www.aben.com.br/Arquivos/420/420.pdf>. Acesso em: 27 mar. 2020.

SEMELER, A. R. **Ciência da informação em contextos de e-science**: bibliotecários de dados em tempos de Data Science. 2017. 165 p. Tese (Doutorado)— Florianópolis. Universidade Federal de Santa Catarina. Centro de Ciências da Educação, Programa de Pós-Graduação em Ciência da Informação, 2017. Disponível em: <https://repositorio.ufsc.br/handle/123456789/185593>. Acesso em: 30 mar. 2020.

SEMELER, A. R.; PINTO, A. L. **Data librarianship venn diagram handbook**: beta version. Rio Grande do Sul: Alexandre Ribas Semeler, 2019. 134 p.

SEMELER, A. R.; PINTO, A. L.; ROZADOS, H. B. F. Data science in data librarianship: core competencies of a data librarian. **Journal of Librarianship and Information Science**, California, v. 51, n. 3, p. 771–780, nov. 2017. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/0961000617742465>. Acesso em: 1 fev. 2020.

SEMIDÃO, R. **Dados, informação e conhecimento enquanto elementos de compreensão do universo conceitual da ciência da informação**: contribuições teóricas. 2014. 198 f. Dissertação (mestrado) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências de Marília, 2014. Disponível em: <http://hdl.handle.net/11449/110783>. Acesso em: 16 ago. 2020.

SONG, I.Y.; ZHU, Y. Big data and data science: what should we teach? **Expert Systems**, [S.l.], v. 33, n. 4, p. 364–373, ago. 2016. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/exsy.12130>. Acesso em: 28 fev. 2020.

SOUZA, C. M. de. Biblioteca: uma trajetória. In: CONGRESSO INTERNACIONAL DE BIBLIOTECONOMIA, 3., 2005, Rio de Janeiro. **Anais [...]**. Rio de Janeiro: S.N., 2005. p. 1-12. Disponível em: <http://www.geocities.ws/csouza952/IIICIB.pdf>. Acesso em: 30 mar. 2020.

THE ECONOMIST. Data is giving rise to a new economy. **The Economist**, Reino Unido, 6 mai. 2017. Disponível em: <https://www.economist.com/briefing/2017/05/06/data-is-giving-rise-to-a-new-economy>. Acesso em: 21 set. 2020.

THE ECONOMIST. The world's most valuable resource is no longer oil, but data. **The Economist**, Reino Unido, 06 mai. 2017. Disponível em: <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>. Acesso em: 21 set. 2020.

THOMAS, C. V. L.; URBAN, R. J. What do data librarians think of the MLIS? Professionals' perceptions of knowledge transfer, trends, and challenges. **College and Research Libraries**, Chicago, v. 79, n. 3, p. 401–423, abr. 2018. Disponível em: <https://crl.acrl.org/index.php/crl/article/view/16726>. Acesso em: 18 maio. 2020.

TYGEL, A. F.; KIRSCH, R. Contributions of Paulo Freire for a critical data literacy: a popular education approach. **Journal of Community Informatics**, Waterloo, v. 12, n. 3, p. 108–121, out. 2015. Disponível em:

<https://openjournals.uwaterloo.ca/index.php/JoCI/article/view/3279>. Acesso em: 4 mar. 2020.

VECHIATO, F. L. *et al.* **Repositórios digitais: teoria e prática**. Curitiba: EDUTFPR, 2017. 271 p. Disponível em: <http://repositorio.utfpr.edu.br/jspui/handle/1/2495>. Acesso em: 3 mar. 2020.

WANG, C. *et al.* Exposing library data with big data technology: a review. In: IEEE/ACI INTERNATIONAL CONFERENCE ON COMPUTER AND INFORMATION SCIENCE (ICIS), 15, 2016, Okayama. **2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)**. Nova Iorque: Ieee, 2016. p. 1–8. Disponível em: <https://ieeexplore.ieee.org/document/9077786>. Acesso em: 25 jun. 2020.

WANG, L. Twinning data science with information science in schools of library and information science. **Journal of Documentation**, [S.l.], v. 74, n. 6, p. 1243–1257, jun. 2018. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/JD-02-2018-0036/full/html>. Acesso em: 22 mai. 2020.

WERTHEIN, J. A sociedade da informação e seus desafios. **Ciência da Informação**, Brasília, v. 29, n. 2, p. 71–77, ago. 2000. Disponível em: <http://revista.ibict.br/ciinf/article/view/889>. Acesso em: 19 mar. 2020.

WOLFF, A. *et al.* Creating an understanding of data literacy for a data-driven society. **Journal of Community Informatics**, Waterloo, v. 12, n. 3, p. 9–26, jun. 2016. Disponível em: <https://openjournals.uwaterloo.ca/index.php/JoCI/article/view/3275>. Acesso em: 3 mar. 2020.

XIA, J.; WANG, M. Competencies and responsibilities of social science data librarians: an analysis of job descriptions. **College and Research Libraries**, Chicago, v. 75, n. 3, p. 362–388, maio. 2014. Disponível em: <https://crl.acrl.org/index.php/crl/article/view/16367>. Acesso em: 22 mai. 2020.

YAMS, E. Data librarianship: a gap in the market. **CILIP Update**, jun. 2008, p. 20–21. Disponível em: <https://era.ed.ac.uk/handle/1842/2499>. Acesso em: 16 ago. 2021.

ZHU, Y.; XIONG, Y. Defining data science. **arXiv**, Cornell University, New York, 21 jan. 2015 (2015b). 8 p. Disponível em: <https://arxiv.org/abs/1501.05039>. Acesso em: 15 maio. 2020.

ZHU, Y.; XIONG, Y. Towards Data Science. **Data Science Journal**, Londres, v. 14, n. 8, p. 1–7, maio 2015 (2015a). Disponível em: <https://datascience.codata.org/articles/10.5334/dsj-2015-008/>. Acesso em: 30 mai. 2020.