

**UNIVERSIDADE FEDERAL DE SERGIPE  
PRÓ- REITORIA DE PÓS-GRADUAÇÃO E PESQUISA  
PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO  
MESTRADO PROFISSIONAL EM GESTÃO DA INFORMAÇÃO E DO  
CONHECIMENTO**

**MAKSON DE JESUS REIS**

**CIÊNCIA DE DADOS E CIÊNCIA DA INFORMAÇÃO: guia de alfabetização de dados  
para bibliotecários.**

**São Cristóvão/SE  
2019**

**MAKSON DE JESUS REIS**

**CIÊNCIA DE DADOS E CIÊNCIA DA INFORMAÇÃO: guia de alfabetização de dados para bibliotecários.**

Dissertação apresentada à Universidade Federal de Sergipe, como parte das exigências do Programa de Pós-Graduação em Ciência da Informação para obtenção do título de Mestre em Gestão da Informação e do Conhecimento.

Orientadora: Profa. Dra. Telma de Carvalho

**São Cristóvão/SE  
2019**

**Dados Internacionais de Catalogação na Publicação (CIP)**

Reis, Makson de Jesus

R375c

Ciência da Informação e Ciência de Dados: guia para alfabetização de dados para bibliotecários / Makson de Jesus Reis. – São Cristóvão-SE, 2019.  
83 f. il. color

Orientadora: Dra. Telma de Carvalho.

Dissertação (Mestrado Profissional em Gestão da Informação e do Conhecimento) – Universidade Federal de Sergipe - UFS, Programa de Pós-Graduação em Gestão da Informação e do Conhecimento, 2019.

1. Ciência de dados 2. Ciência da Informação. 3. Ciência aberta. I. Carvalho, Telma de, orienta. II. Título.

CDU: **02(004.65)**

CDD: **020**

**Bibliotecário responsável: Makson Reis - CRB-5ª Região/1926**

**CIÊNCIA DE DADOS E CIÊNCIA DA INFORMAÇÃO: Guia de alfabetização de dados  
para bibliotecários.**

**MAKSON DE JESUS REIS**

Dissertação apresentada à Universidade Federal de Sergipe, como parte das exigências do Programa de Pós-Graduação em Ciência da Informação para obtenção do título de Mestre em Gestão da Informação e do Conhecimento.

**Avaliação:** \_\_\_\_\_

**Data da defesa:** \_\_\_\_\_

**BANCA EXAMINADORA**

---

**Profa. Dra. Telma de Carvalho**  
**(Orientadora)**

---

**Profa. Dra. Barbara Coelho Neves**  
**(Membro Convidado Externo-UFBA)**

---

**Profa. Dra. Martha Suzana Cabral Nunes**  
**(Membro Convidado Interno- UFS)**

## **AGRADECIMENTOS**

Primeiramente a Deus e ao Universo, alicerce para o meu sustento espiritual!

Aos meus pais Jailton e Acácia, minha irmã Michelly e à toda família que me instruíram e acreditaram em meus sonhos

À minha querida orientadora, amiga e genial Profa. Dra. Telma de Carvalho, sem a qual os projetos realizados, os ensinamentos vividos e o compartilhamento do saber ao longo do desenvolvimento deste trabalho não seriam possíveis.

Aos meus queridos membros da banca (interno, externo e suplentes) pelos ensinamentos conselhos e orientações para o desenvolvimento do projeto: Profa Dra. Martha Suzana Cabral Nunes e Profa. Dra. Janaina Fialho (UFS); Profa Dra. Barbara Coelho (UFBA) e Prof. David Vernon (UFCA).

Aos amigos conquistados e reencontrados da primeira turma do Mestrado Profissional em Gestão da Informação e do Conhecimento (PPGCI/UFS).

Aos amigos da UNIFTC, do município de Feira de Santana-BA

Aos amigos da Biblioteca Julieta Carteador da Universidade Estadual de Feira de Santana, em especial aos (as) bibliotecários (as) e servidores que acreditaram e aceitaram participar do projeto.

*“Nosso papel é desbloquear o uso da ciência de dados, simplificar, moldar, educar e disseminar a cultura analítica, a fim de gerar impacto positivo nas pessoas e organizações”.*

**Ricardo Capra**

## RESUMO

A dissertação apresentada tem como objetivo geral desenvolver habilidades e competências em alfabetização e em gerenciamento de dados para bibliotecários. Assim, apresenta-se em seu referencial teórico, um processo evolutivo baseado nos paradigmas de Thomas Kuhn, a fim de discutir e compreender sobre o quarto paradigma que é baseado em dados. As pesquisas científicas serão transformadas pela grande quantidade de informação desenvolvidas pelas pesquisas e outros meios surgindo o denominado dilúvio de dados. O trabalho aborda a evolução dos dados na ciência, os tipos e ciclo de dados, bem como o processo de armazenamento e gerenciamento desses dados com reflexão sobre os paradigmas anteriores e o surgimento da e-Science. Traz, também, discussão sobre a interdisciplinaridade entre a Ciência da Informação e a Ciência de dados, até o surgimento da biblioteconomia de dados, como campo de atuação para o denominado bibliotecário de dados. Nesse sentido, apresentam-se, as habilidades e competências necessárias para sua formação, com base na matriz de competências de Chantel Ridsdale (2015). Utilizou-se como universo da pesquisa a Biblioteca Central da Universidade Estadual de Feira de Santana (UEFS), como campo de atuação para a promoção da oficina ministrada sobre alfabetização de dados aos bibliotecários desta instituição. Apresenta-se como produto da dissertação um guia que servirá como embasamento teórico para prática de alfabetização de dados e gerenciamentos de dados. Conclui-se que o produto gerado tem caráter inovador e contribuirá com a temática de dados na Ciência da Informação. Além disso poderá ser utilizado pela equipe bibliotecária da BICEN/UEFS em treinamentos e serviços na biblioteca.

**Palavras-chave:** Ciência da Informação. Ciência de Dados. *Data Literacy*. Bibliotecário de Dados. Alfabetização de dados.

## ABSTRACT

A literature dissertation containing developmental skills and skills in literacy and data management for librarians. Thus, an evolutionary process in the paradigms of Thomas Kuhn, a process of discussion and understanding about the paradigm that is based on data, is presented in its theoretical reference. Scientific research is transformed by the vast amount of information by research and other means emerging or called data. The career to the science in the science of the date, the science of the date and the science of the date and the resource of the date and the progressions and the science. It also discusses the interdisciplinarity between data from Science and Science, for the emergence of data librarianship, as the field of action for the so-called data librarian. In this sense, the skills and competences necessary for their formation are presented, based on the competency matrix of Chantel Ridsdale (2015). Used as a research student, at the State University of Feira de Santana (UEFS), as research professor for the data literacy course for librarians of the institution. It is presented as a product of the dissertation, a guide that serves as a theoretical basis for the practice of data literacy and data management. It is concluded that the product presents innovative features and contributes with a series of data in Information Science. BICEN / UFS librarian in training and library services.

**Keywords:** Information Science. Science of Data. Data Literacy. Data Librarian. Literacy of data.

## LISTA DE FIGURAS

<b>Figura 1</b>	Paradigmas da Ciência.....	19
<b>Figura 2</b>	Infográfico que representa a evolução dos dados.....	26
<b>Figura 3</b>	Ciclo de vida dos dados.....	33
<b>Figura 4</b>	Diagrama de Venn.....	36
<b>Figura 5</b>	Relação da Ciência da informação com a Ciência de dados.....	38
<b>Figura 6</b>	Habilidades do cientista de dados.....	40
<b>Figura 7</b>	Matriz de Competências de Alfabetização de dados.....	48
<b>Figura 8</b>	Mantra Research Management Training.....	50
<b>Figura 9</b>	Foster.....	51
<b>Figura 10</b>	Dmptool.....	52
<b>Figura 11</b>	Entrada da UEFS.....	57
<b>Figura 12</b>	Análise SWOT.....	61

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>09</b>
<b>2</b>	<b>A EVOLUÇÃO DOS DADOS NA CIÊNCIA.....</b>	<b>13</b>
<b>2.1</b>	<b>Dilúvio de dados e <i>Big data</i>.....</b>	<b>21</b>
2.1.1	Tipos de Dados.....	29
2.1.2	Ciclo de dados e gerenciamento de dados .....	31
<b>3</b>	<b>RELAÇÕES CONCEITUAIS ENTRE CIÊNCIA DE DADOS E CIÊNCIA DA INFORMAÇÃO.....</b>	<b>35</b>
<b>3.1</b>	<b>Biblioteconomia e o bibliotecário de dados.....</b>	<b>41</b>
<b>3.2</b>	<b><i>Data literacy</i> ou alfabetização de dados para o bibliotecário no Brasil</b>	<b>45</b>
<b>3.3</b>	<b>Web sites para alfabetização em dados.....</b>	<b>49</b>
<b>4</b>	<b>METODOLOGIA.....</b>	<b>54</b>
<b>4.1</b>	<b>A Universidade Estadual de Feira de Santana (UEFS) e a Biblioteca Central Julieta Carteadó.....</b>	<b>56</b>
<b>4.2</b>	<b>Procedimentos metodológicos.....</b>	<b>57</b>
<b>5</b>	<b>DIAGNÓSTICO.....</b>	<b>60</b>
<b>6</b>	<b>RESULTADO DA INTERVENÇÃO.....</b>	<b>63</b>
<b>6.1</b>	<b>Curso de alfabetização em dados.....</b>	<b>63</b>
<b>6.2</b>	<b>Guia para alfabetização em dados.....</b>	<b>63</b>
<b>7</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>68</b>
	<b>REFERÊNCIAS.....</b>	<b>71</b>
	<b>APÊNDICES.....</b>	<b>77</b>

## 1 INTRODUÇÃO

Os dados sempre foram elementos constantes nas pesquisas, especialmente as pesquisas científicas. Com o desenvolvimento das tecnologias de informação e das possibilidades de busca por meio de mecanismos mais aprimorados, surgem estudos que enfatizam a Ciência de Dados, que pode ser considerada como a ciência que reúne múltiplos aspectos da informação a partir de seus dados, contando com uma equipe multidisciplinar de profissionais, a exemplo de estatísticos, programadores, analistas de dados, matemáticos e também bibliotecários. A partir desse entendimento, o grande volume de dados armazenados passa a fazer parte daquilo que se convencionou chamar de *Big Data*. Ou seja, quando se fala em *Big Data* deve-se levar em consideração grande volume de dados armazenados na internet e, nesse sentido encontra-se a possibilidade do armazenamento desses dados em grandes espaços, denominados “nuvens”. As nuvens são entendidas como plataformas digitais que abrigam informações geradas a partir de grande volume de dados, pois a cada dia mais e mais informações e dados de pesquisa são inseridos no ambiente digital e nas nuvens. Com isso, surge o conceito de “dilúvio de dados” que nada mais é do que a grande quantidade de informação gerada diariamente pelas pesquisas e por outros meios, uma vez que mesmo o usuário comum está conectado a uma rede de informação e gera dados e conteúdo, que também são armazenados nessas plataformas.

Diante desse panorama, surge um novo conceito de ciência chamada E-Science que tem por características expandir a compreensão e o uso destes dados, especificamente trabalhando com os dados abertos, configurando-se como o quarto paradigma da ciência.

Tendo em vista estes acontecimentos e suas respectivas mudanças na forma de busca, recuperação, tratamento e disseminação de dados surge também a necessidade de aprofundar conhecimentos nessa nova dinâmica imposta e, conseqüentemente, a necessidade de uma alfabetização em dados. Entende-se por alfabetização em dados os conhecimentos necessários para lidar com esse campo de atuação que envolve múltiplos profissionais de variadas formações. Assim, não poderia deixar de aparecer a figura do bibliotecário de dados, cujas habilidades e competências estão relacionadas à ciência de dados, na busca, tratamento e uso da informação.

Gray (2007) destacou que as pesquisas científicas serão transformadas pelo grande volume de dados e que a e-Science relaciona-se com a tecnologia da informação e cientistas. Portanto, a tecnologia tem seu papel de destaque, quando para fazer ciência.

Conforme Koltay (2017), os desafios tecnológicos viabilizaram o processo de

difusão da informação através do seu processo de organização, busca, recuperação e preservação dos dados extraídos promovendo o tratamento técnico. O ambiente digital está possibilitando a criação de acesso a um registro científico muito expandido, que é cada vez mais, expresso em termos de redes de links e associações entre diversos artefatos de pesquisa.

Para suportar a reutilização e integração de dados exigirá compreensão e documentação da proveniência dos dados, o desenvolvimento de ontologias, anotações de especialistas e análises. Os serviços habilitados por essas atividades de visualização, simulação, mineração e modelagem de dados e outras formas de representação e extração de conhecimento (KOLTAY, 2017)

Conforme Peter Kellam (2013) fortalecer a capacitação de bibliotecários apresenta desafios significativos para a construção de habilidades e de conhecimentos necessários para realizar novos papéis. A especialização de domínio é essencial para trabalhar efetivamente com pesquisadores nas fases iniciais do gerenciamento e planejamento de dados de pesquisa.

O conhecimento de domínio também pode ser necessário para auxiliar em problemas ou criar ferramentas de gerenciamento de dados. Argumenta-se que o treinamento para estes profissionais, com a temática em dados, promoverá uma alfabetização de dados.

Em ambos os casos, os bibliotecários poderão fortalecer suas relações de trabalho, incorporando-se com equipes de docentes, estudantes e pesquisadores. Bertin, Visoli e Ducker (2017) afirmam que, tradicionalmente, as comunidades de informação e dados se desenvolveram paralelamente, de modo convergentes, mas a mudança de atitudes em relação ao acesso aberto dos resultados da pesquisa científica resultou em novas parcerias nas quais bibliotecas e gerentes de informações estão trabalhando com a comunidade de dados sobre novos produtos de informação. Habilidades de gerenciamento de dados como padrões, metadados, gerenciamento de direitos, serviços de descoberta, preservação e particularmente prestação de serviços estão sendo aceitos como uma base vital para o sucesso da capacitação em dados (AMARAL, 2016).

Uma vez que bibliotecas e bibliotecários compreendam as oportunidades de integração da Biblioteconomia na perspectiva de dados em seus serviços e perspectivas e que estejam convencidos do valor que podem ser incluídos nas rotinas e em suas atividades, precisarão investir tempo no desenvolvimento de novas habilidades e cruzar fronteiras

culturais para a perfeita compreensão da necessidade de entendimento em relação à alfabetização de dados.

A expansão de estudos com a temática dados encontra-se de forma exponencial na literatura em função das possibilidades que os mesmos representam para o avanço científico, o que chamou a atenção para esta pesquisa. Afinal, com a crescente importância dada aos dados, surgem diversos movimentos preocupados com a preservação e disponibilização de ferramentas e de serviços que venham promover a coleta, a organização e o gerenciamento desses dados.

Com acesso frequente à Biblioteca Central da Universidade Estadual de Feira de Santana, na Bahia e, em diálogo com os bibliotecários de lá, pôde-se perceber que a temática alfabetização de dados seria interessante naquele contexto. Assim, surgiu a possibilidade de discutir o tema, tendo por objetivo desenvolver habilidades em alfabetização e gerenciamento de dados aos bibliotecários, pois o profissional que irá gerir ou trabalhar com os dados na instituição deve estar capacitado, a fim de obter melhores resultados nestes estudos, chegando assim à problematização deste trabalho: como os bibliotecários da Biblioteca Central (BICEN) da Universidade Estadual de Feira de Santana realizam o gerenciamento de dados em pesquisa?

Desta forma, a presente pesquisa tem por objetivo geral desenvolver habilidades e competências em alfabetização e em gerenciamento de dados para bibliotecários da Biblioteca Central da Universidade Estadual de Feira de Santana e tem por objetivos específicos: promover um curso próprio de alfabetização de dados com demonstração prática das ações necessárias em cada etapa; levantar as ações promovidas pelos bibliotecários relacionadas à inclusão de novos serviços na biblioteca; elaborar um guia com materiais que sirvam de embasamento teórico e prático para alfabetização e gerenciamento de dados para bibliotecários e avaliar o curso ministrado com vistas a identificar o grau de satisfação do mesmo face aos assuntos abordados.

Justifica-se a escolha desse tema pois a biblioteca universitária está em processo constante de evolução e de interação tecnológica e, nestes espaços, possuir um profissional capacitado, com as atualizações do mercado, potencializa as relações entre os usuários e comunidade. Entende-se que a alfabetização de dados, como campo prático para promover novas habilidades e instigar novos conceitos que emergem sobre os dados deve ser compreendida pelos bibliotecários atuantes nas bibliotecas universitárias, referindo-se, especificamente, ao caso apresentado.

Além disso, outro fator de destaque, é oportunizar à biblioteca central da UEFS

atuar como campo de estudo para o aprendizado sobre alfabetização de dados, a fim de possibilitar em seu espaço a prática de atividades e suporte aos dados da pesquisa da instituição.

Assim, a primeira parte desta Dissertação apresenta, na Introdução, o problema, os objetivos e a justificativa da pesquisa. A segunda e terceira parte dizem respeito ao referencial teórico, com capítulos específicos inerentes ao tema proposto. A quarta parte foi destinada à Metodologia com os procedimentos metodológicos empregados na pesquisa para se atingir aos objetivos propostos. O diagnóstico realizado na Instituição é apresentado na quinta parte do trabalho tendo por sexta e sétimas partes os resultados da intervenção e as considerações finais. Por fim, apresentam-se as Referências utilizadas.

## 2 A EVOLUÇÃO DOS DADOS NA CIÊNCIA

A história da ciência possui inúmeros aspectos e pode ser constituída por grandes personagens e filósofos, a partir de eventos ou episódios marcantes realizados por descobertas dos cientistas, onde a evolução da ciência se constituiu na criação da história humana (FOUREZ,1995).

As ciências que têm como objeto a realidade efetiva histórico-social buscam, hoje, de maneira mais intensa do que jamais aconteceu, a sua conexão mútua e a sua fundamentação. Causas, que residem no estado das ciências positivas atuam nessa direção juntamente com os impulsos mais poderosos, que emergiram dos abalos da sociedade desde a Revolução Francesa. O conhecimento das forças que vigoram na sociedade, das causas que produziram os seus abalos, dos meios para um progresso saudável presentes na sociedade, transformou-se em uma questão de vida ou morte para a nossa civilização (DILTHEY, 2010, p. 14).

A notoriedade e conceituação da ciência constroem-se a partir do processo evolutivo da história. Conforme Dilthey (2010), em sua conceituação, dá-se a partir das grandes narrativas desenvolvidas ao longo do tempo, com o progresso da sociedade, das causas e efeitos inerentes ao homem e às revoluções concebidas em grandes descobertas por filósofos.

Segundo Miranda (2016), os filósofos: René Descartes (1596-1650), John Locke (1632-1704), David Hume (1711-1776), Immanuel Kant (1724-1804), em suas publicações, desenvolveram conceitos de ciência e teorias do conhecimento, promovendo a filosofia da ciência, defendida pela notoriedade de cada época. Discutir o significado da ciência remete ao acesso ao conhecimento filosófico.

Uns defendem que o acesso ao conhecimento se dá por intermédio da razão (Platão e Descartes), outros, por intermédio da experiência (Aristóteles, Locke e Hume). Em particular, Kant considera aspectos da razão e da experiência para explicar o acesso que temos ao conhecimento (apriorismo kantiano) (MIRANDA, 2016, p. 35).

Esse acesso ao conhecimento deve-se ao fato de que a ciência está diferenciada no científico ou no tido como científico e sua definição teórica tem como base o seu paradigma. Miranda (2016, p. 33) define como sendo a ciência produtora do conhecimento, preocupando-se interinamente com a densidade da base empírica, ou seja, baseando-se na experiência vivida e na observação das coisas, bem como nos fenômenos estudados.

A ciência não é um sistema de enunciados certos ou bem estabelecidos, nem é um sistema que avance continuamente em direção a um estado de finalidade. Nossa ciência não é conhecimento (espiteme): ela jamais pode reclamar haver atingido a verdade ou um substituto da verdade, como a probabilidade (POPPER, 2007, p. 305).

Isto significa dizer que a ciência não traz, em si, verdades absolutas pois a cada nova descoberta, outras possibilidades podem surgir, e deste modo, outros fatores poderão influenciar o conhecimento já existente.

Demo (2011, p. 10) afirma que “a definição de ciência é naturalmente controversa, embora se possa arrumar algum consenso em torno dela, permitindo que a academia funcione razoavelmente”. Por mais que possam ser divergentes, a partir dos seus paradigmas diversificados, a intersubjetividade de critérios forma a lógica da ciência e os dados mensuráveis, comprováveis e retestáveis. As controvérsias das quais a ciência se constitui dividem-se por múltiplos métodos.

A ciência, mais do que uma instituição, é uma atividade. Podemos mesmo dizer que ‘ciência’ é um conceito abstrato. O que conhecemos concretamente são cientistas que trabalham e os resultados do trabalho destes [...] o processo de investigação, em todos os ramos da Ciência, revelou que os princípios tidos como basilares em certa época tiveram de ser modificados ou substituídos para manter adequação a fatos revelados por novas descobertas (MORAIS, 2013, p. 26-27).

Assim, na concepção do autor, a ciência em si pode ser considerada como algo abstrato, a partir do momento em que o concreto diz respeito ao processo da pesquisa e que as novas descobertas, a cada período, podem modificar ou substituir os conhecimentos anteriores.

Kuhn (2005) assegura que, de fato, a ciência é uma das formas, não a única, do acesso ao conhecimento produzido pelo homem, podendo definir-se como concreta, à medida que se evolui, tendo seu processo de discussão baseado em paradigma. O paradigma define-se como caminho para o processo de evolução da ciência, na qual o homem lê o mundo conforme sua concepção, conseguindo distinguir o que deve ser aceito ou não pela comunidade.

Para o autor, paradigmas são as realizações científicas, que oferecem problemas ou soluções para uma comunidade praticante de uma ciência.

A tentativa de descobrir a fonte dessa diferença levou-me ao reconhecimento do papel desempenhado na pesquisa científica por aquilo que, desde então, chamo de ‘paradigmas’. Considero ‘paradigmas’ as realizações científicas por aquilo

que, durante algum tempo, fornecem problemas e soluções modelares para uma comunidade de praticantes de uma ciência (KUHN, 2005, p. 13)

Kuhn (2005) afirma que o primeiro paradigma é caracterizado pelo período de mitos, cujos fenômenos naturais foram atribuídos a deuses como revelação do divino e que se basearam no dogma da Igreja, através das Escrituras Sagradas, do Século I ao Século XVIII, na Idade Média. Quando um paradigma é aceito pela totalidade de uma comunidade, da sociedade ou de cientista, o autor afirma que este paradigma é capaz de explicar fenômenos extraordinários que seu antecedente era incapaz de explicar, porém, muitos problemas considerados relevantes podem ser seguramente abandonados.

A primeira quebra de paradigma teria ocorrido há algumas centenas de anos, com o surgimento das primeiras tentativas de se explicar fenômenos por meio de modelos teóricos. Modelos como as Leis de Kepler, as Leis de Newton ou a Lei de Boyle-Marriott (dentre tantas outras) permitiram não só um melhor entendimento dos fenômenos observados empiricamente, como também a realização de previsões sobre o comportamento de novos fenômenos (CORDEIRO; BRAGHETTO; GOLDMAN; KON, 2013, p. 75)

No segundo momento, Kuhn (2005) descreve que as ideias de Copérnico e Galileu promoveram revolução na ciência, através da premissa do Heliocentrismo versus Geocentrismo, bem como a Lei da Gravidade e, por fim, a teoria de que o Sol estava no centro do universo e os planetas giravam em sua volta.

Uma das contribuições mais importantes de Galileu é ‘a lei da gravidade’, isto é, como se comporta, em termos da sua velocidade e do tempo, um objeto em queda livre. Galileu enfrentou um problema duplo ao tratar de resolvê-lo [...] e a solução nos legou aspectos fundamentais para a tarefa da ciência. Em primeiro lugar, a queda dos corpos é um fenômeno que ocorre muito rápido, e não existiam relógios suficientemente precisos para documentar o processo. É indispensável contar com instrumentos fiáveis para medir. Demonstrando genialidade, desenvolveu um sistema para medir intervalos muito curtos de tempo de forma reproduzível e fiável (FENTANES, 2014, p. 9)

Gérard (2007) afirma que a ciência, a partir dessa época, foi sustentada na ciência fundamental ou teórica e na ciência aplicada. Na ciência teórica, por volta do século XVIII, tem-se a ciência de Leis de Kepler, de Newton, Equações de Maxwell assumindo o ramo teórico. Na ciência experimental, a ciência acontecia de forma empírica, por meio da observação e da descrição das leis naturais.

É necessário, todavia, esclarecer que a chamada ciência fundamental, voltada para a compreensão e explicação do real, não pode ser dissociada da ciência aplicada, que visa a ação de controle sobre o real. Enquanto a ciência fundamental

pode ser interpretada como uma filosofia da natureza, a ciência aplicada consiste em, por assim dizer, uma forma racional da técnica. Repetimos: tais atividades são simultaneamente complementares, guardando tal independência que, nos dias atuais, só se faz delas distinções teóricas (MORAIS, 2013, p. 49)

Ou seja, ciência fundamental e ciência aplicada têm distinções entre si, estando a primeira voltada para a explicação do real, a partir de uma filosofia da natureza e a segunda para ação de controle sobre o real, em uma abordagem racional e com utilização de técnicas.

Conforme Fentanes (2014), os termos “provem ou experimentem” contextualizam a existência da ciência experimental, apesar de Galileu, há quase quatro séculos, nas ciências naturais, ter comprovado que ao lançar dois objetos diferentes do alto da Torre de Pisa, ambos chegavam ao solo ao mesmo tempo. Esses fatos ajudam a afirmar que a maneira como a ciência é, se modifica ao longo da história, a partir da validação de diversas pesquisas científicas, baseando-se tanto em dados obtidos por observações de experimentos, como nos resultados por simulações ou números. Com o processo evolutivo da ciência, a partir da Pós-Guerra até os anos 1960, começa a surgir um novo conceito de paradigma para a ciência. Um dos primeiros computadores digitais, o ENIAC (Eletronic Numerical Integrator and Computer), surge como simulador de modelos balísticos, como armazenamento de dados, realizando operações matemáticas, a exemplo da soma e da notação binária, permitindo cálculos à distância.

Segundo Cordeiro; Braghetto; Goldman e Kon (2013), a ciência começa a atingir limites de exploração, a partir de desenvolvimento de instrumentos ou coleta de dados, que permitiam ampliar ou aproximar fenômenos estudados, a fim de melhor analisá-los ou mensurá-los, tendo como base o uso da tecnologia, permitindo, assim, acesso ao conhecimento.

Agora, como ‘campo do conhecimento’, a tecnologia vai além da aplicação de um saber-fazer: pergunta-se por um saber teórico, procura-se um aperfeiçoamento. A invenção técnica apresenta caráter deliberado, reforçado pela tecnologia, que supõe um desenho (ou projeto) tecnológico como uma representação antecipada de um artefato e auxiliado por algum conhecimento científico, com o propósito de criar sistemas funcionais que desempenhem com eficiência e eficácia funções úteis para as pessoas (MIRANDA, 2016, p. 256-257).

Assim, considera-se que a tecnologia assume também um caráter teórico a partir do momento que o produto tecnológico, em si, é uma representação do conhecimento

científico.

Velho (2011) afirma que a transformação do conhecimento científico e o processo de descoberta na complexidade das simulações em tecnologia, especialmente no conhecimento científico e na apropriação da sociedade, foi impulsionada com o desenvolvimento computacional, construindo um novo paradigma.

A relação entre ciência e tecnologia ainda era concebida como linear, mas agora com ênfase na demanda. Ou seja, não é mais a ciência que empurra a tecnologia (science push), mas o mercado, as necessidades dos usuários que puxam o desenvolvimento científico (demand pull). Nessa relação entre ciência e tecnologia, as empresas eram tidas como possuidoras de capacitação e habilidades para julgar as demandas do mercado, identificar oportunidades tecnológicas e articular necessidades e demandas (VELHO, 2011, p. 140)

As transformações em curso, que apontam uma crescente ressignificação do dado como uma unidade de análise, chegam a se constituir em um paradigma. Kuhn (2005) conceitua paradigma como uma representação a ser seguida, tendo como pressuposto a filosofia, embasada por teoria e conhecimento, a fim de originar um campo científico como modelo de estudo, sendo demonstrado seu processo evolutivo de paradigma desde as origens até o século XXI. Segundo o autor, o primeiro paradigma foi originado com as contribuições, fenômenos e técnicas utilizadas até o séc. III a.C., onde, na Grécia Antiga, berço da ciência, iniciava-se o princípio dos primeiros pensadores filósofos, como Hipócrates e Pitágoras, sendo despertados os grandes interesses pelo saber e conflito entre o pensamento sagrado e o pensamento racional ou filosófico questionando a repressão religiosa e a busca pelo pensamento livre, sem doutrinação. O segundo paradigma, conforme Kuhn (2005) constituiu-se a partir da Idade Moderna, nos séculos XVI e XVII, quando as ideias e conceitos se voltaram às perspectivas do iluminismo e da reforma protestante, baseadas nas ideias dos filósofos Keppler, Copérnico, Galileu e Leonardo Da Vinci, contrapondo-se ao conhecimento antes religioso, formalizando o método indutivo como solução para descobertas de relações novas causadas na natureza, sendo privilegiadas a observação e a coleta de dados. No terceiro paradigma, na Idade Contemporânea, século XIX, têm-se a ideia evolucionista de Charles Darwin, a teoria da evolução das espécies. Neste paradigma, os experimentos são relativamente importantes e o conceito de probabilidade e de estatística versus tomada de decisão reafirma as teorias estabelecidas, sendo anunciadas através de aplicações, por método hipotético-dedutivo, a fim de incluir e sistematizar como prática os instrumentos de coleta de dados para resolução dos problemas.

Assim surge o novo (quarto) paradigma, abordado pelo cientista Jim

Gray<sup>1</sup>(2007 apud TOLLE; TANSLEY; HEY, 2011, p. xi), na palestra proferida para o Conselho de Telecomunicações e Ciências da Computação, na Califórnia. O cientista destacou o conceito de “ciência” intensiva em dados, sendo estes em dois modos: observado ou simulados, unificando-se em teoria, experimentos e processados por software, utilizando-se de gerenciamento de dados e estatística, para resolução de problemas. Segundo Gray (2007), esse paradigma é construído pelo novo processo científico; antes, o processo de resolução de problemas era construído com base em resoluções analíticas e, com a evolução da ciência, essas resoluções passaram a ser simuladas, gerando uma grande quantidade de dados gerados.

Para construção do quarto paradigma, o autor Gray (2007) utiliza as revoluções científicas apresentadas por Kuhn (2006, p. 125), apontadas como “[...] aqueles episódios de desenvolvimento não-cumulativo, nos quais um paradigma mais antigo é total ou parcialmente substituído por um novo [...]” e afirma que “se a ciência é a reunião de fatos, teorias e métodos reunidos nos textos atuais, então os cientistas são homens que, com ou sem sucesso, empenharam-se em contribuir com um ou outro elemento para essa constelação específica” (KUHN 2006, p. 120). Cabe, assim, ação com intuito de harmonizar os fatores envolvidos nessa dinâmica cooperando com a evolução da ciência.

Gray (2007, p. 20) retrata essa evolução da ciência, para construção do novo paradigma baseado em dados:

**Primeiro paradigma:** mil anos atrás, a ciência era empírica, com a descrição dos fenômenos naturais;

**Segundo paradigma:** há poucos séculos; se desenvolveu a ciência teórica com o uso de modelos e generalizações, mostrando como exemplos as leis de Newton, Kepler e as equações de Maxwell;

**Terceiro paradigma:** nas últimas décadas apareceu o ramo computacional, com a simulação de fenômenos complexos gerando uma grande quantidade de dados mostrando o caminho para o quarto paradigma destacado por Gray como e-Science;

**Quarto paradigma:** a ciência do século XXI é apresentada como exploração de dados, a e-Science, que unifica os três paradigmas anteriores (experimento, teoria e simulação) destacando como características: grandes quantidades de dados capturados por instrumentos ou gerados por simulações e processados por softwares; informação e/ou conhecimento armazenados em computadores; cientista analisa base de dados e arquivos por meio de gerenciamento de dados e estatísticas.

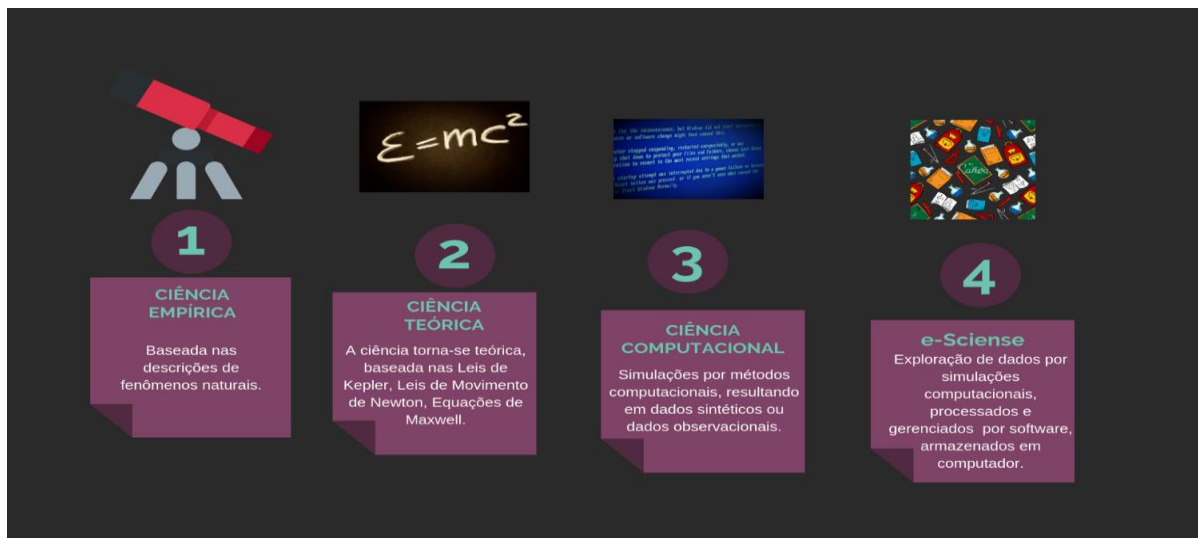
---

<sup>1</sup> Jim Gray é pesquisador e foi gerente do grupo de eScience da Microsoft Research. Seus principais interesses de pesquisa são bancos de dados e sistemas de processamento de transações-com foco particular no uso de computadores para tornar os cientistas mais produtivos. Em 28 de janeiro de 2007, Jim desapareceu no mar. Disponível em site: <https://www.microsoft.com/en-us/research/people/gray/>

Conforme Gray (2007), o novo (quarto) paradigma é centrado em 4 (quatro) campos de paradigmas da ciência, sendo estes: a Ciência Empírica; a Ciência Teórica; a Ciência Computacional e a Ciência Baseada em Dados (e-Science). A ciência empírica é embasada na pesquisa científica puramente de fenômenos, da observação de leis naturais e fenômenos da natureza (GRAY, 2007).

A ciência teórica, segundo Jim Gray (2007), é seguida das Leis de Kepler e Leis de Newton, onde seus modelos teóricos formam uma construção para a ciência computacional, em que as simulações das pesquisas geram dados. A ciência empírica, teórica e computacional (Figura 1), são afetadas pelo dilúvio de dados conforme Bell et.al (2009), estabelecidos pelo padrão intelectual de conhecimento científico, construindo, assim, a nova ciência, a chamada: e-Science. (INTERAGENCY WORKING GROUP ON DIGITAL DATA, 2009). Entende-se por “dilúvio de dados” a grande quantidade de dados gerados a partir das técnicas e das tecnologias utilizadas para simulações e experimentos (BELL et al., 2009). Esse dilúvio de dados ocorre devido à grande participação dos cientistas que produzem cada vez mais dados.

Figura 1 - Paradigmas da Ciência



Fonte: GRAY, (2007) adaptado pelo autor (2019)

Dentre seus paradigmas, a ciência tem sido afetada pelo “dilúvio de dados” coletados, gerados ou simulados em grandes laboratórios e observatórios e delinea um novo paradigma, o paradigma do uso intensivo de dados, denominado por alguns pesquisadores como e-Science, também conhecida como Ciência orientada a dados, Computação fortemente orientada a dados – Cyber infraestrutura - e dos dados ao

conhecimento. (HEY; T.; HEY; J., 2006).

A interferência de Jim Gray era que a ciência experimental e a ciência teórica estiveram desde Newton. E sobre os últimos 50 anos, a ciência computacional amadureceu como uma metodologia para a pesquisa científica. Jim pensou que agora estamos vivendo o surgimento de um quarto paradigma para a pesquisa científica, ou seja, a ciência intensiva de dados. Para isso, os pesquisadores precisam de um conjunto diferente de habilidade, daqueles exigidos para a ciência experimental, teórica e computacional. (HOFFMANN, 2010, p. 112, tradução nossa).

O termo e-Science, segundo Yang (2009), foi iniciado pelo diretor geral do gabinete de Ciência e Tecnologia do Reino Unido, John Taylor, em 1999. Taylor significou o termo E-Science como desafios científicos na utilização de tecnologia para investigações baseadas em dados.

O princípio fundamental na e-Science baseia-se na tendência de que os procedimentos e práticas de forma tradicional em que a ciência é conduzida estão passando por mudanças radicais. Esta mudança é baseada na inclusão da moderna infraestrutura cibernética como parte da ciência ou ambiente experimental que inclui não só os computadores de ponta já onipresente, armazenamento e infraestrutura de rede, mas também tecnologias emergentes na Web. (YANG, LIZHE; LASZEWSKY, 2010, p. 353, tradução nossa).

A e-Science, conforme Schroeder e Fry (2007, p. 564, tradução nossa), pode ainda ser definida por diferentes maneiras, “utilizamos a e-Science no singular para indicar as ciências naturais e sociais (as humanidades só serão tratadas de passagem), e adotamos uma definição restrita que inclui apenas os esforços que aprimoram a pesquisa científica”.

Schroeder e Fry (2007) confirmam que a e-Science possui várias vertentes, podendo ser abrangentes nos compartilhamentos de acesso a serviços de uma rede, como recursos computacionais da Web Semântica, ou utilizando-se de recursos de pesquisa (instrumentos ou dados). Definida como elemento científico e social, o seu compartilhamento permite o uso de recursos computacionais ou recursos de pesquisa, sendo as ferramentas desenvolvidas como um modo em rede, que podem ser usadas e acessadas em diferentes locais. Entende-se esta nova ciência baseada em dados (e-Science) como a representação da informação, tecnologia e dados, dando um significado às práticas de teorias científicas contemporâneas nas diversas áreas do conhecimento. Martin Frické (2014) afirma que a e-Science tem base em outras teorias, além de produzir descobertas, tornando-se paradigmática e interdisciplinar.

Hoffman (2010) afirma que a nova ciência baseada em dados utiliza-se dos

campos teóricos e computacionais para compreender, conforme Jim Gray (2007), a chamada e-Science como um caminho da evolução da ciência empírica para a teórica, da teórica para computacional e da computacional para a ciência baseada em dados, em seus diversos termos. Apesar de ser um termo ascendente, existem estudos de caso em que a formulação teórico-conceitual sobre e-Science na qual a área é eminentemente oriunda da inovação das tecnologias nas pesquisas científicas.

A e-Science é interdisciplinar e reúne pesquisadores da ciência e informática. Isso significa que o desenvolvimento de aplicativos de e-Science não é trivial como deveria ser. As competências exigidas incluem a aplicação de computação paralela e computação distribuída, bem como a compreensão do domínio do conhecimento específico (por exemplo, Ciências da terra, bioinformática), que é significativo (YANG; LIZHE; LASZEWSKY, 2010, p. 355, tradução nossa).

Com a e-Science considerada essa nova ciência orientada em dados, vale abordar a crescente expansão dos dados, na próxima seção.

## 2.1 Dilúvio de dados e *Big Data*

Os dados, segundo Setzer (1999, p.1), são “[...] símbolos quantificados ou quantificáveis. Portanto, um texto é um dado. De fato, as letras são símbolos quantificados, já que, por si só, constituem uma base numérica”. Os dados podem ser representados por números, categoriais, textos físicos ou digitais, imagens, áudio, locais ou conexões entre elementos de rede. Dados são como matéria-prima na tomada de decisão e informação.

Segundo Hjørland (2018), os dados possuem seu conceito utilizado em diversos campos como: banco de dados, arquivo de dados, mineração de dados, dados descritivos, metadados, dados vinculados e Big Data.

Gleick (2011) afirma que os primeiros indícios de dados surgiram quando o homem desenhou o primeiro elemento neandertal, em sua caverna, dando início para o processo de escrita.

O povo paleolítico começou, há pelo menos 30.000 anos, emaranhar e pintar formas que recordavam as imagens dos olhos de cavalos, peixes e caçadores. Estes sinais na argila e nas paredes da caverna serviram como finalidades da arte ou da mágica, sendo artefatos iniciais da escrita, e começaram a gravação de estados mentais e visualizações do cotidiano em meios externos. De outra forma, nós em cordões e entalhes em paus, serviram como auxílios à memória e dados (GLEICK, 2011, p. 33)

Gleick (2011), afirma que as pinturas marcadas demonstravam significados de

posse, abordando um entendimento de representação do cotidiano, como, por exemplo, as tábuas de barro da Suméria, que além de possuírem caracteres diferentes eram usados em maneiras diferentes, podendo denotar números e pesos.

Logo após o sistema de escrita e pinturas, surgiu, conforme Gleick (2011), o alfabeto, um símbolo para relação da fala através do domínio da escrita.

Ao Leste, estava a grande civilização da Mesopotâmia, com o seu roteiro cuneiforme, descendo a costa para o sudoeste do Egito, onde hieróglifos eram desenvolvidos simultaneamente e de forma independente. Os comerciantes viajaram, também, a Chipre e Creta, trazendo seus próprios sistemas incompatíveis. As classes sacerdotais dominantes investiram em seus sistemas de escrita. Quem possuía o domínio, possuía as leis e os ritos. (GLEICK, 2011 p. 35)

Com a evolução da escrita, Gleick (2011) afirma que os monges escribas, tiveram um papel importante no resguardo cultural e artístico através da produção de manuscritos. Além da escrita, os escribas tinham a função de registrar dados numéricos, escrever leis, copiar ou arquivar informações.

Com o avanço da tecnologia, a reprodução de cópias destas publicações, antes restrita a limites de cópias, como papiros, ideogramas e escrita cuneiforme, foram propagadas com a invenção da imprensa de Gutenberg, no século VX, com o surgimento da máquina de réplica, aumentando a reprodução de livros e ampliando acesso ao conhecimento. (GLEICK,2011).

Guizzo (2003), afirma que, após esse período, foi inventada a máquina de calcular com base em engrenagens mecânicas em que os números eram representados pelas engrenagens das rodas. Em 1830, o inglês Charles Babbage criou o que chamou de mecanismo analítico, que tinha características iniciais de um computador moderno, possuindo um dispositivo de entrada, armazenamento para números, aguardando para serem processados, um processador e um dispositivo de saída. O autor menciona o surgimento do primeiro dispositivo acionado eletricamente, em 1890, que se utilizava de cartões perfurados e hastes de metal que passavam pelos furos para fechar um circuito elétrico e, assim, fazer com que um contador avançasse, elaborado pelo matemático Claude Shannon, conhecido como pai da Teoria da Informação, expôs que um dado, independentemente do seu formato, deve ser transmitido e arquivados para outro lugar, de forma computacional. Baseou-se em aplicação da lógica booleana de dois valores a circuitos, assim fundando o primeiro computador digital, tornando possível partilhar uma gama de informações (sons ou imagens), através de fios elétricos.

Em seu trabalho, Shannon mostrou que cada canal tem uma taxa máxima de transmissão de dados eletrônicos de forma confiável, que ele chamou a capacidade do canal. Tente enviar informações a uma taxa maior que esse limite e você sempre perderá parte da sua mensagem. Este limite máximo, medido em bits por segundo, tornou-se um referencial essencial para os engenheiros de comunicação (GUIZZO, 2003, p. 8, tradução nossa)

O surgimento do computador revolucionou a forma de tratar e armazenar dados. Segundo Hjørland (2018) os computadores começaram a representar dados incluindo vídeo, imagens, sons e texto, através dos valores binários, usando padrões de apenas dois números: 1 e 0. Kelleher e John (2018) consideram que um bit é a menor unidade de dados representadas por apenas um único valor. Um byte tem oito dígitos binários, seu armazenamento e memória são medidos, em megabytes e gigabytes, representando grandes quantidades de dados. Manovich (2015) contextualiza que os formatos de dígitos binários estão subjacentes aos computadores, memórias, unidades de disco, bem como dispositivos periféricos comuns na computação.

Tipicamente, os recursos são numéricos. Mas eles podem também tomar outra forma. Por exemplo, um computador pode analisar uma imagem e gerar algumas palavras descrevendo conteúdo da imagem. Em geral, ambos metadados e recursos podem usar vários tipos de dados: números, categorias, texto livre, relações de rede, coordenadas espaciais, datas, horas e assim por diante (MANOVICH, 2005, p. 17).

Conforme Hjørland (2018), a entrada inicial do computador para controle de dados tomava forma a partir de cartões perfurados seguido pela fita magnética e pelo disco rígido. Logo após, o processamento eletrônico de dados passou a abranger toda a gama do que conhecemos por tecnologia da informação. A Internet marcou o começo dos computadores, compartilhando informação, formando uma grande explosão de compartilhamento de dados (Big Data) no mundo.

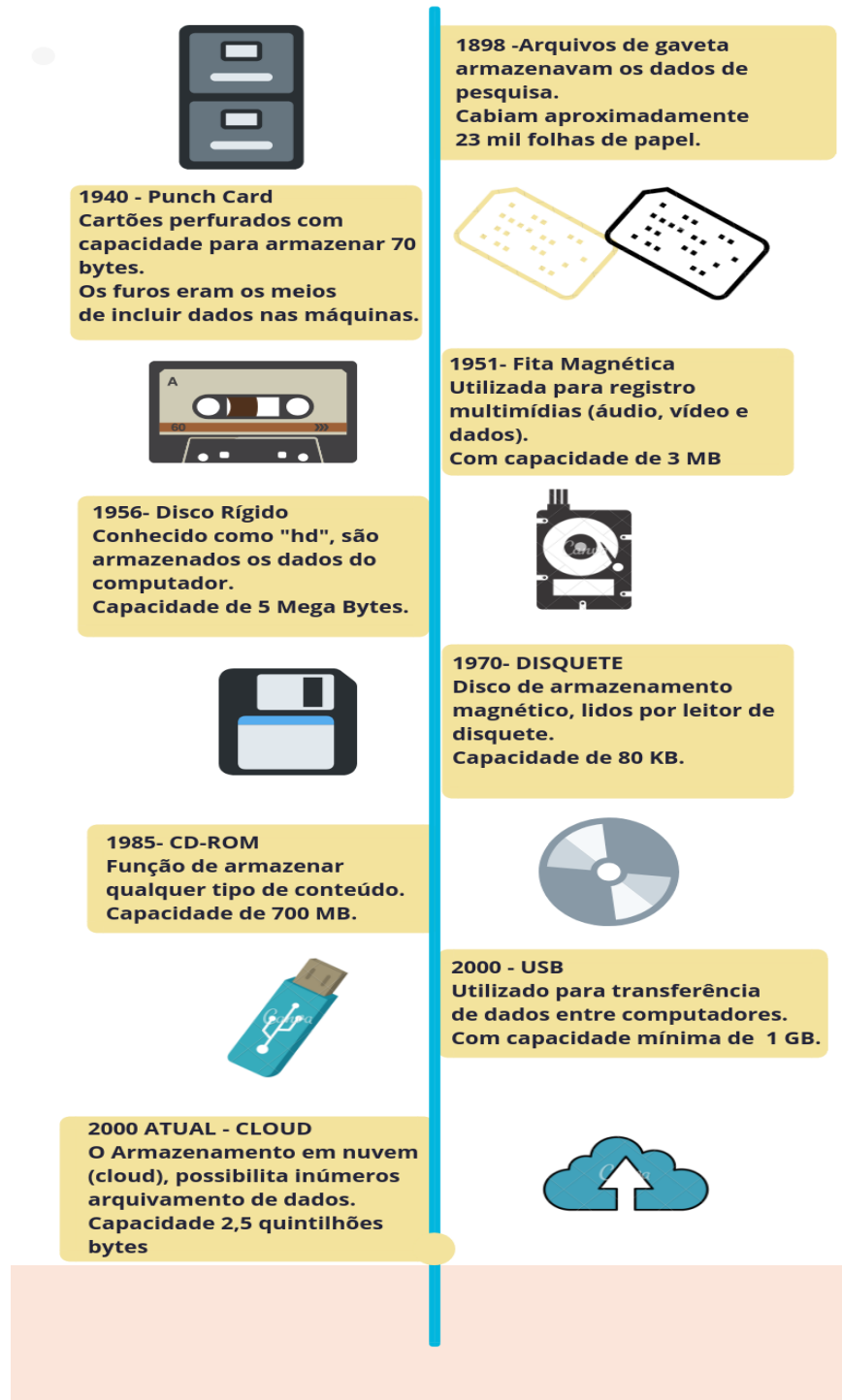
Com o desenvolvimento dos computadores, os suportes de armazenamento dos dados foram expandidos, conforme Figura 2. A evolução dos sistemas de armazenamento de arquivos levou a uma complexa combinação de uso de sistemas de nuvem. Com o processo evolutivo e a grande massa de produção em dados, novas formas de processamento e reorganização surgiram, possibilitando, com o computador, a ressignificação do dado como unidade de análise. Em paralelo com a lei de Moore, criada em 1965, conceituada por Goldon Earl Moore, vislumbrava-se que conforme a evolução da indústria avançava na tecnologia, o poder de capacidade de armazenamento dos computadores dobraria a cada um

ano e meio bem como a sua velocidade de compartilhamento dos dados duplicaria, pela evolução dos transistores facilitando, assim, a troca de sinais elétricos e de potência elétrica (VAN DER AALST, 2014). Kelleher e John (2018) afirmam que outros armazenamentos de dados começaram a surgir devido à grande capacidade de projeções de objetos conectados à internet, gerando quantidades maciças de dados, o então chamado “dilúvio de dados”. Bell et al. (2009) afirmam que o dilúvio de dados é originado pela grande massa de dados produzidas pela ciência ao longo do século. Esses dados foram gerados e capturados em laboratórios ou mediados por linguagem computacional. Consequentemente, mecanismos de acesso e crescimento de dados estão surgindo com tecnologias que aceleram a importância em como utilizá-los e gerenciá-los, sendo necessário automatizar o processo de descoberta de dados e da informação para conhecimento. Gray (2007) afirma que a crescente explosão dos dados, ou dilúvio de dados, tende a ser resolvido através do armazenamento em nuvem e de novas estruturas de dados.

Chen, Mao e Liu (2014) afirmam, ainda, que com o crescimento e o aumento explosivos da tecnologia nos últimos 20 anos, os dados serão multiplicados em quase dois anos no futuro próximo. Assim o Big Data surge como descritor para conjunto de dados enormes, sendo incluídas massas de dados que possuem e necessitam de análise em tempo real. Além disso, o Big Data possibilita oportunidades para descobrir novos valores e obter conhecimento, com novos desafios em como organizar e gerenciar tais conjuntos de dados. Assim, um computador tem uma série de limites para utilização e gerenciamento de dados como: memória, tamanho do disco, velocidade do processador. Com isso, a expansão tecnológica utiliza-se de nova escala de armazenamento de dados, a fim de fornecer capacidade com a mesma rapidez de acesso. De acordo com o relatório de Gartner, até 2020, em torno de 230 bilhões de dispositivos estarão conectados à internet utilizando-se de medidores inteligentes e sensores, atuando continuamente com uma quantidade enorme de dados analisados e armazenados. Definido pelos 3Vs (Volume, Velocidade e Variedade), o Big Data, possui desafios e oportunidades em relação à quantidade e capacidade de armazenamento de dados. O primeiro V referente ao volume, atribuído aos múltiplos datasets (conjuntos de dados), sendo definidos por escala de armazenamentos: um 1 gigabyte de armazenamento é o equivalente a cerca de um vídeo de 7 minutos em alta-definição; um pentabyte de armazenamento equivale à 13,3 anos de vídeos de alta-definição e um brontobyte equivale a 10 vezes o armazenamento de 1 gigabyte (CHEN; MAO; LIU, 2014).

Desta forma, com o advento da tecnologia, cada vez mais surgem inovações que carregam em si uma grande quantidade de dados. Um exemplo para ilustrar essa dimensão poderia ser os carros autônomos, que gerarão enormes quantidades de dados de sensores e poderia trazer os brontobytes como uma unidade para a esfera da mídia de armazenamento.

Figura 2 - Infográfico que representa a evolução dos dados



Fonte: Adaptado de Manovich (2015), elaborado pelo autor (2019)

Os estudos de exploração de dados para extrair informação evoluem conforme a necessidade de armazenamento em nuvem utilizando-se de recurso computacional, com processador no servidor ou armazenamento de dados na rede, sendo produzidos velozmente e amplamente difundidos nos dias atuais em plataformas digitais, sem a necessidade de utilização de recursos físicos para salvaguardar dados, como as mídias eletrônicas: CDs, USB ou DVDs. O armazenamento em nuvem possibilita o acesso aos dados de forma simples e remota permitindo seu acesso sem necessidade de conhecimento sobre a tecnologia utilizada (VAN DER AALST, 2014).

O uso de tecnologias digitais, incluindo computação para modelos e simulações cada vez mais complexas, grandes matrizes de sensores, poderosos equipamentos de imagem e detectores, e ferramentas de acesso, interação e disseminação em rede, transformou a ciência. Os dados que são ‘nascidos-digitais’ — disponíveis apenas em formato digital e preservados apenas eletronicamente — estão cada vez mais sendo a principal saída da ciência e o ponto de partida para novas pesquisas. A taxa em que esses dados digitais são produzidos está aumentando a cada ano, produzindo fluxos de dados massivos e exponencialmente crescentes no que foi descrito como um ‘dilúvio de dados’ (INTERAGENCY WORKING GROUP ON DIGITAL DATA, 2009, p. 6, tradução nossa).

As organizações e pesquisadores começam a perceber a importância da validação, limpeza, tipologia e integração dos dados, a fim de afunilar os conjuntos extraídos de sites, dos dados próprios gerados e, especialmente, na governança de dados.

Diferentes sociedades, culturas, comunidades e setores (estaduais e municipais, bancos, igrejas, hospitais, Ciências etc.) estabelecem instituições, regras e processos para produzir dados e documentação que servem, como anteriormente explicado, como suficientes fatos para as suas principais atividades. Na ciência os dados servem como medições e experimentos, incluindo dados computacionais, são bons exemplos. Nas humanidades podem ter dados sobre documentos (dados bibliográficos), sobre obras de artistas, sobre indivíduos, sobre datas, eventos e períodos históricos. (HJØRLAND, 2018 p.698, tradução nossa)

Conforme Sagioglu e Sinanc (2013), à medida que o comércio eletrônico, baseado na Web, se espalhou, evoluíram modelos de negócios orientados a dados, que tratam os dados como um ativo em si. Tais tendências também geraram maior preocupação com os usos sociais de dados e privacidade de dados. A maioria das empresas, com as evoluções tecnológicas que marcaram o início do século XXI, percebeu que precisava começar a promover o tratamento de dados, seja para aproveitar novas oportunidades para suas empresas (os primeiros a adotar) ou, simplesmente, para se manter à tona em um mundo onde as empresas promovem utilizações possíveis desses dados.

Com isso surge o Big Data e, conforme Hjørland (2018), o termo foi cunhado pelo estatístico Francis Diegbold, em 2000, caracterizando-se como um conjunto de dados considerado grande. O Big Data é composto por 3 tipos de Vs (volume, variedade e velocidade). O volume refere-se à quantidade de dados gerados e armazenados; a variedade refere-se à sua composição, com os diferentes tipos de dados, sendo extraídos, por exemplo, de texto, imagens e áudio e a velocidade refere-se ao ritmo com que são processados e gerados em tempo real. O Big Data repousa na tecnologia e na análise de grande quantidade de dados, favorecendo percepções que não eram conhecidas anteriormente e isso requer estudos comparativos e precisos.

Definimos o Big Data como um fenômeno cultural, tecnológico e acadêmico que repousa na interação de: (1) Tecnologia: maximizar o poder de computação e precisão algorítmica para coletar, analisar, vincular e comparar grandes conjuntos de dados. (2) Análise: desenho em grandes conjuntos de dados para identificar padrões a fim de fazer reivindicações econômicas, sociais, técnicas e legais. (3) Mitologia: a crença generalizada de que grandes conjuntos de dados oferecem uma forma superior de inteligência e conhecimento que pode gerar percepções que antes eram impossíveis, com a aura de verdade, objetividade e precisão (BOYD; CRAWFORD, 2012, p. 663)

De acordo com Davenport (2014), o Big Data revoluciona não por estar nas máquinas que calculam os dados, porém nos dados e na maneira como são utilizados. Apesar do conceito revolucionário sobre Big Data, sua característica está no grande volume de dados desestruturados, provenientes de diversas fontes, com capacidade de análise contínua e nas diferentes áreas de aplicação.

O entendimento dos variados tipos de dados se deu após o surgimento do Big Data e que, nas palavras de Amaral (2016, p. 7), “é um fenômeno em que dados são produzidos em vários formatos e armazenados por uma grande quantidade de dispositivos e equipamentos”.

A história do ENIAC, [ ], mostra que a evolução tecnológica, em termos de miniaturização, capacidade de processamento e barateamento é algo surpreendente. Os insumos de tecnologia, como CPUS, memórias e unidade de armazenamento vêm se tornando cada vez mais baratos. O barateamento, miniaturização e aumento da capacidade de processamento levam à disseminação de equipamentos, dispositivos e processos capazes de produzir e armazenar dados, virtualização, computação na nuvem, internet. Temos, então, o Big Data. (AMARAL, 2016, p. 7).

As relações entre os dados e organizações que fundamentam relações essenciais para o Big Data corroboram para o surgimento de novos tipos pesquisa, políticas, negócios

e tomada de decisão, orientados à dados, bem como novos campos profissionais de atuação, gerando a necessidade de um novo perfil profissional para a gestão e organização desses dados.

Segundo Davenport (2014), com o Big Data, as organizações permitem a análise dos dados através de tecnologia eficiente, com significado estatístico e probabilidades maiores de resultados. O Big Data é envolvido pelo uso de diversos tipos de conceitos e tecnologias, como computação em nuvens, Internet, processamento, armazenamento e gestão de projetos. Amaral (2016) afirma que esse é um fenômeno da intensa produção e armazenamento de dados, estabelecida por seus tipos e formas de gerenciamento. Devendo levar em conta, o processo de preservação dos dados, tendo em vista, a necessidade de uso na extração de informações e conhecimento.

### 2.1.1 Tipos de dados

Conforme Kelleher e Tierney (2018), os dados se tornaram essenciais em diferentes vertentes de pesquisa. Com o crescimento da ciência de dados impulsionado pelo surgimento do *Big Data*, veio a explosão de dados com o desafio de desenvolver métodos mais assertivos de análise e modelagem de dados, bem como conhecer os fatores que levam as organizações a coletar, armazenar e processar dados. Com a proliferação de dados nas organizações deu-se ênfase à garantia da qualidade dos dados para registros mais precisos e atuais. As diversas etapas envolvidas no gerenciamento de dados modernos incluem a limpeza e os processos de extração, transformação e carregamento para a integração desses dados.

Existem diferentes tipos de dados sendo estes: numérico, nominal e ordinal e suas distinções podem ser úteis em relação aos dados estruturados e não estruturados. Os dados estruturados devem ser armazenados em uma tabela, possuindo um conjunto de atributos, tendo como exemplificação os dados demográficos de uma população na qual cada linha pode descrever um conjunto de atributos demográficos. Os dados estruturados possibilitam melhor armazenamento e organização dos dados ao serem pesquisados, realizando mesclagem com outros dados. Os dados não estruturados são dados que possuem estrutura interna própria, ou seja, não têm relação com outros dados, sem armazenados de forma independente, como um conjunto de páginas da web, com cada página tendo uma estrutura (KELLEHER; TIERNEY, 2018).

Em relação à dados estruturados e dados não estruturados os autores apontam alguns exemplos de suas finalidades e de sua extração.

[...] Dados não estruturados são muito mais comuns do que os dados estruturados. Por exemplo, as coleções de texto humano (e-mails, tweets, mensagens de texto, postagens, romances, etc.) podem ser consideradas dados não estruturados, assim como coleções de arquivos de som, imagem, música, vídeo e multimídia. [ ] Muitas vezes podemos extrair dados estruturados de dados não estruturados usando técnicas de inteligência artificial (como processamento de linguagem natural e ML), processamento de sinal digital e visão computacional. (KELLEHER; TIERNEY, 2018, p. 49, tradução nossa).

Os dados estruturados são bem definidos nas palavras de Inmon e Linstedt (2015). Segundo os autores, eles são, em sua maioria, gerenciados por um sistema de banco de dados e consistem em: registros, atributos, chaves e índices. Os dados estruturados são bem definidos e previsíveis possibilitando uma localização muito rápida e fácil. Os dados não estruturados são inversos, possuindo dados imprevisíveis e sem nenhuma estrutura reconhecível para um computador, seu acesso é feito por sequências de dados, pesquisados em ordem para encontrar uma determinada unidade de dados. Os dados científicos incluem informações de texto, áudio e vídeo, bem como registros de atividades de registro e da web.

Dados in loco das diversas áreas do conhecimento, são utilizados de maneira adequada por pesquisadores, quando se têm atrelado, aos dados o uso de vocabulários controlados. Dessa forma, o uso de ontologias na construção ou reestruturação de repositórios, ou das bases de dados que disponibilizam dados primários, propiciam uma melhor recuperação e utilização desses dados (MONTEIRO; SANT'ANA; SEGUNDO; 2016, p. 11).

Bertin, Visoli e Drucker (2017) afirmam que, com os avanços em tecnologia da informação, a produção desses dados estruturados e não estruturados aumentou e trouxe expansão global. Porém, com isso, veio a dificuldade de organização e preservação, inclusive para pesquisadores, no início do século XX, quando se utilizavam de dados de suas pesquisas para exploração de potencial avanço científico e tecnológico.

No entanto, tais dados estavam escritos em cadernos científicos ou armazenados em suportes de mídia, tornando-se, ao longo do tempo, ilegíveis ou descartados em pequenos laboratórios, acabando, por fim, sendo resguardados na biblioteca, até seu processo de descarte (MANTRA, 2012). Os dados de pesquisa, podem ser constituídos conforme o quadro 1:

Quadro 1 -Tipos de dados da pesquisa

Formato	Suporte
Texto	MS Word, Portable Document Format (PDF), Rich Text Format (RTF), Hyper-Text Markup Language (HTML) e Extensible Markup Language (XML)
Numérico	SPSS, Stata, MS Excel, SAS e arquivos hierárquicos.
Multimídia	JPEG, TIFF, GIF, MPEG, Quicktime, Bitmap,PNG
Modelo	3D, estatístico, macroeconômico, causal e de similaridade
Software	JAVA, C, Perl, Python e PHP

Fonte: adaptado de Bertin, Visoli e Drucker (2017, p. 37)

Rice e Southall (2016) afirmam que os dados da pesquisa ou dados científicos são registros, arquivos ou conteúdos em formato impresso ou digital que contêm resultados de observações de pesquisa podendo ser compartilhado entre a comunidade acadêmico, podendo ser contidos em documentos; atas de laboratório; questionários e relatórios; fitas; CDs; DVDs; fotografias; Slides; arquivos de dados estatísticos; conteúdos de banco de dados (vídeo, áudio, texto, imagens); scripts e mapas.

Com o aumento da produção de dados, surge a necessidade de iniciar um processo de gerenciamento e de formação de análise dos dados, que abordaremos a seguir.

### 2.1.2 Ciclo de dados e gerenciamento de dados

Os elementos da ciência têm como pilares: a observação, o experimento, a teoria e a modelagem, sendo transformados pelo ciclo contínuo de geração, acesso e uso de alcance cada vez maiores de dados. Assim, um ciclo de dados determina: o nascimento, a agregação dos dados a serem substituídos com o crescimento, a análise durante sua reprodução e a exclusão dos dados, referente à sua inutilidade. O ciclo de dados representa

as fases de uso, gerenciamento e armazenamento desses dados levando em conta no primeiro processo, a criação dos dados - produção em diferentes tamanhos dependendo da aplicação e do propósito do dado - podendo compor grandes volumes de dados; o segundo processo refere-se à agregação dos dados relacionados com a semelhança ou correlação das coletas e armazenamentos, sendo distribuídos de forma centralizada; o terceiro processo é a análise dos dados gerados e agregados a fim de lidar com dois fatores: o volume e a complexidade dos tipos de dados existentes ou criados, conforme explica a *Interagency Working Group On Digital Data* (2009)

Durante o ciclo de dados, vale destacar que o armazenamento e gerenciamento são fatores importante para a preservação, tendo como pilares elementares para o uso de dados: a observação, experiência, teoria e modelagem, sendo verificados as formas de acesso e uso, para um alcance e volume cada vez maior das informações disponíveis, levando em percurso que um bom gerenciamento e política desenvolvida, facilita na recuperação em caso de perda ou desconfiguração desses dados. Conforme *Interagency Working Group On Digital Data* (2009), o gerenciamento efetivo de cada etapa do ciclo, junto com coordenação entre as etapas é essencial para garantir que os dados sejam preservados e possam ter seu acesso e uso de forma confiável e eficiente. Para um bom uso de dados de pesquisa é fundamental elaborar um ciclo de vida de dados no seu processo criacional, no processo de análise de dados, bem como gerenciamento, levantando como pontos de vista as principais características para uso e desuso.

Nas palavras de Bertin, Visoli e Drucker (2017, p. 37), “[...] para alcançarem sustentabilidade e competitividade no sistema científico moderno, portanto, instituições e pesquisadores devem garantir o apropriado gerenciamento e preservação dos dados de pesquisa, de modo a possibilitar a verificação futura de resultados [...]” . Vale afirmar que os dados da pesquisa possuem tempo de vida maior que o projeto de pesquisa desenvolvido.

Bertin, Visoli e Drucker (2017, p.43) afirmam que o ciclo de dados e seu gerenciamento são compostos em quatro etapas:

**Implementação, Definição de parâmetros, Exploração de dados:** possibilidade de preservação de dados científicos, desenvolvimento de políticas de gerenciamentos de ciclo de dados para projetos e atividades relevantes; **Ingestão de dados, Obtenção de dados:** estratégias que preveem a preservação e o acesso a longo prazo e rentáveis à qualidade adequada, garantindo proteção de alta confiança e confidencialidade ;**Tomada de decisão:** aplicações para os requisitos legais e regulamentados para toda a gama de tipo de dados e **Utilização do modelo:** recuperação dos dados de pesquisa, tendo em vista a sua implementação.

A Figura 3, a seguir, demonstra a vinculação dessas etapas, para assegurar a descrição, controle adequados dos dados a longo prazo, sendo necessários para sua coleta e preservação.

Figura 3 – Ciclo de vida dos dados



Fonte: adaptado de Bertin; Visoli e Drucker (2017), elaborado pelo autor (2019)

Amaral (2016) afirma que os dados são inseridos em processos, modelos e tecnologias desde o início da produção até o seu descarte. O sistema que permitirá que esses dados de pesquisa sejam prontamente descobertos, avaliados e usados devem estar devidamente baseados em um ciclo de dados em que estabelecem os princípios de gestão contínua e eficaz de novas tecnologias e métodos de identificação e adoção, sem pôr em

risco a preservação e o acesso. Conforme Bertin, Visoli e Drucker (2017, p. 39), a importância de se elaborar um ciclo de vida dos dados, envolvem aspectos de “planejamento, aquisição, organização, estruturação, definição de fluxos analíticos e ferramenta apropriada para o armazenamento de dados”. Além disso deve-se atentar também à questões relativas à preservação, à organização, ao compartilhamento e à proteção. Esses itens são necessários por representarem a estruturação e organização, afim de promover o uso e reuso dos dados, de forma segura e adequada.

Bugnion, Manivannan e Nicolas (2017) afirmam que o ciclo de dados facilita a boa gestão de dados quando disponibilizados em formatos como textos: MS Word; PDF; RTF, em formato numérico: Excel, em multimídia: JPEG, GIF, MPEG, etc.; em software; e quando disponibilizados em variedade de suportes: fitas, CDs, slides, modelos, mapas e arquivos de dados, etc.

O gerenciamento de dados e informações é um problema que vem sendo cada vez mais comum à medida que mais recursos computacionais são disponibilizados e mais dados são produzidos como parte de atividades colaborativas de grande escala. Uma das motivações para o gerenciamento de dados é baseada no fato de que, embora os recursos de computação são cada vez mais poderosos, eles não são, no entanto, baratos. Portanto, novas soluções para facilitar a execução de simulações devem ter dados de arquivamento, descoberta e capacidades de acesso construído desde o início (YANG; LIZHE; LASZEWSKY, 2010, p. 354, tradução nossa).

O ciclo de vida dos dados é uma ferramenta conceitual, que possibilita entender os diferentes passos que os dados seguem até o seu gerenciamento. O compartilhamento dos dados começa com uma prática de ciclo de vida dos dados e é promovido desde o planejamento até sua análise e publicação.

Portanto, ao elaborar um ciclo de vida de dados deve-se considerar a gama de informações e tipologias existentes para o desenvolvimento da pesquisa. Sendo assim, vale discutir o surgimento da ciência de dados pelos pesquisadores, que possibilitou ferramentas e técnicas para o gerenciamento de dados.

Com isso, a próxima seção apresenta uma breve história da ciência de dados e sua relação com a Ciência da Informação.

### 3 RELAÇÕES CONCEITUAIS ENTRE CIÊNCIA DE DADOS E CIÊNCIA DA INFORMAÇÃO

Zhu e Xiong (2015) afirmam que os dados armazenados em computadores são utilizados e mapeados para que possamos usá-los. No entanto, o método de uso de dados mudou de acesso simples para análise de Big Data, especialmente no campo da ciência. A necessidade da utilização dos dados foi modificada mediante seu crescimento, fenômenos e regras.

Nesse contexto, os dados tornam-se um desafio técnico-científico, potencializando estudos para a extração máxima de acesso ao conhecimento, produzindo um desafio em diversas áreas no eixo global. Assim, surge a chamada “Ciência de Dados”, emergindo como uma nova área interdisciplinar, com aplicações em áreas tão distintas, com o desafio de identificar os métodos e técnicas fundamentais para o gerenciamento e análise de grandes volumes de dados.

A chamada Ciência de Dados ou Data Science, baseada em técnica e teorias, incorpora elementos de campos básicos em engenharia e ciência básica, além de relacionar-se com a ciência natural. Segundo Zhu e Xiong (2015), a ciência de dados é conhecida também como ciência intensiva em dados.

Embora a expressão “Data Science” venha dos anos 1960, a ciência de dados é uma ciência nova. Uma ciência trata de obter conhecimento e informação, de forma sistemática, bem como normalizar e organizar esse conhecimento. Da mesma forma, a ciência de dados trata de estudar o dado em todo o seu ciclo de vida, da produção ao descarte (AMARAL, 2016, p. 4).

Conforme Zhu e Xiong (2015) a ciência de dados tem seu termo usado, em 1996, pelo pesquisador Peter Naur, que sugeriu que a ciência da computação deveria ser chamada de “datalogia”, a partir dos anos 1990, na reunião do Committee on Data of the International Council for Science (CODATA2) no Conselho Internacional para a Ciência (ICSU). A CODATA foi criada para promover a colaboração global para melhorar a disponibilidade e usabilidade de dados para todas as áreas de pesquisa estabelecendo, em 2015, princípios e estratégias para a ciência de dados em três áreas prioritárias: promoção princípios, políticas e práticas para dados em ciência; avanço das fronteiras da ciência de

---

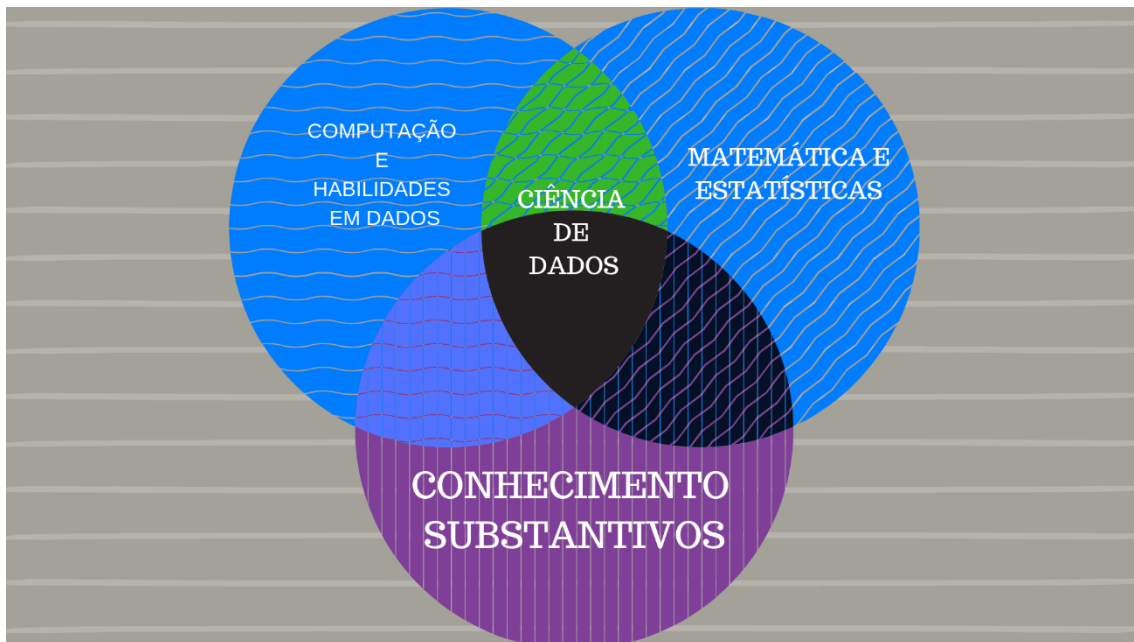
<sup>2</sup> Disponível em CODATA: <http://www.codata.org/about-codata>. Acesso em 11 jan. 2019

dados e desenvolvimento da capacidade do acesso aberto, ou open access, melhorando competências em dados e funções para apoiar os dados abertos.

Finzer (2013) apresenta o diagrama de Venn (figura 4) como mecanismo que constitui a ciência de dados. O diagrama está formado por três círculos que identificam o pensamento, referindo-se, o primeiro aos princípios básicos e experiência requeridos em matemática e estatística para compreensão da variabilidade e para interpretar e diferenciar os tipos de dados; o segundo exemplifica a compreensão do contexto disciplinar para a escolha de uma metodologia de análise válida para dados (conhecimentos substantivos) e, o terceiro refere-se à computação e habilidades em dados para resolução de problemas, permitindo que se visualize a estrutura de dados o que requer habilidades para programar, extrair e estruturar dados.

Sobre o diagrama de Venn, Amaral (2016, p. 5) afirma que a ciência de dados é “composta por várias outras ciências, modelos, tecnologia, processos e procedimentos relacionados ao dado” criando relações interdisciplinares na área.

Figura 4 – Diagrama de Venn (2010)



Fonte: adaptado de Amaral (2016)

Conforme Smith (2006, p. 163, tradução nossa) "A ciência de dados se desenvolveu desde então para incluir o estudo da captura de dados, sua análise, metadados, recuperação, arquivamento, troca, mineração para encontrar conhecimento inesperado e relações de dados."

Desta maneira, a Ciência de dados é definida como método para extração da informação útil a partir de complexas e dinâmicas bases de dados (BUGNION; MANIVANNAN; NICOLAS, 2017).

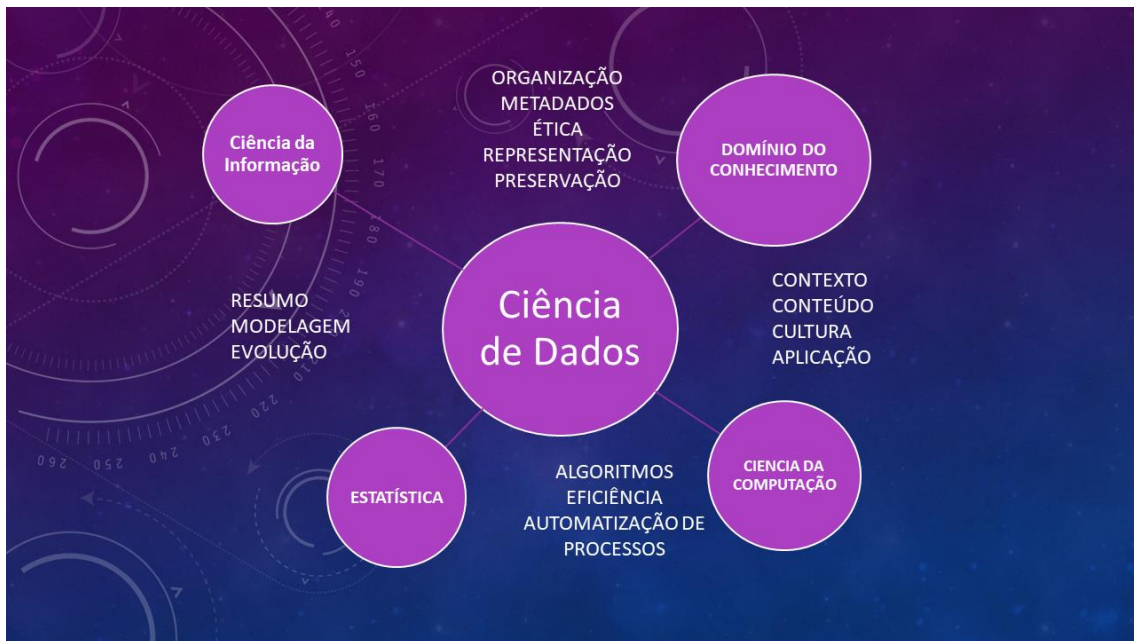
Nessa perspectiva, a ciência de dados é a descoberta de conhecimento ou de informações acessíveis de dados e a ciência da informação relaciona-se nesse sentido com suas práticas para armazenar e recuperar informações. Embora sejam áreas distintas, estas complementam-se na preservação e expansão do conhecimento. Gray (2007) afirma que as ciências tradicionais se juntam à ciência de dados através da técnica e prática para uma variedade de campos do conhecimento.

Assim, entende-se que a Ciência da Informação está diretamente ligada à ciência de dados. As diversas epistemologia da Ciência da Informação permeiam a coleta, a curadoria, o processamento, o armazenamento, a análise e a visualização dos dados. Seu campo multidisciplinar se relaciona com a análise, classificação, armazenamento, coleta, disseminação e proteção da informação, com a interação entre pessoas, organizações ou sistemas de informação existentes (MARCHIONINI, 2016).

Setzer (1999, p. 20) afirma que os temas basilares para a Ciência da Informação relacionam-se a dados, informação e conhecimento, onde "o compartilhamento de dados deve estabelecer pesquisas e políticas de compartilhamento de dados científicos na ciência da informação". O autor enfatiza a necessidade de "adoção de política de compartilhamento de dados em periódicos da ciência da informação". Além disso, Ferreira e Santos (2013) abordam que, na web, o acesso à informação distribuída globalmente se dá através dos dados estruturados, que melhoram a descoberta e o acesso a tais informações, por meio de adoção de políticas de dados em instituições ou periódicos.

Segundo Marchionini (2016, p. 3) "a ciência da informação é como termo geral que realiza interdisciplinaridade com a informática, concentrando em distinções e semelhanças para a ciência de dados". O pesquisador, em seu artigo intitulado: *Information Science Roles in the Emerging Field of Data Science*, ressalta que a ciência da informação está atribuída à ciência de dados, como pode-se perceber na figura 5.

Figura 5 - Relação da Ciência da Informação com a Ciência de dados



Fonte: adaptado de Marchionni (2016)

Marchionini (2016) argumenta que a ciência da informação é interdisciplinar e a Ciência de Dados um subconjunto desta área, sendo que os programas de formação em biblioteconomia devem utilizar-se dessas estratégias para assuntos relacionados aos dados. Além disso, os departamentos da ciência da informação devem promover, em seus currículos, habilidades técnicas para gerenciamento de dados distribuídos, para o processo de mineração de dados, programação, relatórios e visualizações para dados e saída, desenvolvendo a interdisciplinaridade da área, surgindo novas oportunidades para biblioteconomia e serviços de informação.

A ciência de dados provocou mudanças, pois a partir da complexidade existente na manipulação e extração de dados surge a necessidade de profissionais habilitados no tratamento de dados, como o desenvolvedor de algoritmos e ferramentas eficientes para manipulação de dados e comportamentos, e o profissional da informação, com especialidade em dados, para promoverem eficiência na interação com os usuários e ofertarem as necessidades informacionais da sociedade.

A ciência de dados, além de incitar o surgimento do profissional de dados, deu espaço para o surgimento de outras profissões, como o cientista de dados (MARCHIONINI, 2016).

À medida que a pesquisa se torna mais intensiva em dados, pesquisadores enfrentam novos desafios no gerenciamento e compartilhamento de informações. Com isso, as bibliotecas são redutos de profissionais que começam a se especializar, a fim de promoverem variedades de serviços de suporte em dados, incluindo: instrução, treinamento, gerenciamento, orientação de planejamento, curadoria e visualização.

Com a grande quantidade de dados, o processo de organização e reuso deve ser feito por um profissional que tenha informações relevantes para atuar na recuperação de grandes volumes de dados.

Uma nova profissão surgiu nos últimos anos, o "cientista de dados", que combina as habilidades do estatístico, do programador de software e do designer gráfico para o gerenciamento de dados. Na literatura sobre este assunto, o cientista de dados é capaz de minerar massas desarrumadas dos dados e de produzir conhecimento que tanto Estados quanto corporações anseiam, sendo profissionais valiosos porque são capazes de discernir o valor de grandes conjuntos de dados (MAYER-SCHÖNBERGER; CUBIN, 2013).

Finzer (2003) afirma que o profissional do século, terá formação em ciência de dados, tendo como característica a expansão das áreas técnicas no campo da estatística, programação e computação, sendo intitulado como um: Data Analyst ou Analista de Dados, porém, o autor descreve poucas características desse novo profissional.

Swan e Brown (2008) afirmam que o relatório publicado pela Joint Information Systems Committee (JISC) descreve o profissional em ciência de dados reconhecendo que seu perfil precisa ser definido, porém, o relatório aponta o cientista de dados como aquele que realiza e desenvolve pesquisa, que está envolvido na análise e estudos dos dados. Van Der Aalts (2014) defende que o profissional em ciência de dados deve ter conhecimento em negócios, estatísticas, ciência da computação e programação, entre outros, tendo domínios computacionais, habilidades com matemática e conhecimentos na aplicação de dados. Com base em Van Der Aalts (2014), elaborou-se um infográfico (Figura 6), com as principais habilidades e conhecimentos de um cientista de dados.

Figura 6 – Habilidades do cientista de dados



Fonte: extraído de Van Der Aalts (2014), adaptado pelo autor (2019).

A ecologia da ciência de dados é entendida como o domínio de técnicas relacionadas à formação do profissional interessado em qualificar-se na área por meio de participação em programas de capacitação ou formação, com duração curta ou longa, para tornar-se um especialista em determinado assunto. Assim surgem diferentes perfis de profissionais em campos de atuação dentro da área de dados com especificidades diferentes e com funções relacionados aos dados existentes. Através dessa multidisciplinaridade, Swan e Brown (2008), com base na ecologia da ciência de dados - entendida como campo de variações das habilidades e competências desenvolvidas pelos cientistas de dados -

definem quatro funções voltadas ao profissional qualificado para trabalhar com dados: Data Manager ou Data Creator (gerenciadores de dados ou criadores de dados): pesquisadores com conhecimento de domínio de produção e criação de dados devendo ter alto nível de especialização no manuseio, manipulação e uso de dados; Data Scientis (Cientista de Dados): profissional com qualificação, e que trabalha com dados digitais e com desenvolvimento de banco de dados possuindo conhecimento e qualificação em estatística e programação; Data Manager (Gestor de Dados): cientistas da computação ou tecnólogos da informação responsáveis pelo acesso contínuo e preservação de dados e Data Librarians (bibliotecários de dados): profissional bibliotecário com especialização em curadoria, preservação, gestão e arquivamento de dados, que será melhor abordado na próxima seção (VAN DER AALTS, 2014).

### **3.1 Biblioteconomia e o bibliotecário de dados**

É sabido que as bibliotecas têm uma longa tradição em fornecer acesso e efetuar o gerenciamento de documentos e de serviços que auxiliam o usuário na identificação e obtenção de informações para pesquisa. Com o crescente interesse em dados e o movimento de acesso aberto as bibliotecas adquiriram um papel fundamental. Hey T.e Hey J. (2006) afirmam que grande parte do material produzido por pesquisadores, como conjunto de dados, resultados experimentais e dados de mídia avançada, bem como outros documentos, são depositados e armazenados na biblioteca. Com a e-Science, além do papel tradicional na organização e tratamento da informação (catalogação e indexação), as bibliotecas vêm promovendo uma ressignificação de gerenciamento dos dados, a fim de possibilitar sua conservação.

Na década de 1990, além dos serviços de referência nas bibliotecas, os dados de pesquisa começaram a ser adquiridos, armazenados e apresentados em formato digital. Jacobs teoriza os níveis de dados como serviços em um ambiente de biblioteca acadêmica, propando serviços de computação e de dados. Com o surgimento e fundação da IASSIST3 (International Association Science Information Services & Technology), na década de 2000, o pesquisador Jacobs junto com Humphrey, deram início a estudos relacionados ao surgimento da Biblioteconomia de dados, como uma ramificação da área da

---

<sup>3</sup>Web site da international organization of professionals working in and with information technology and data services : <https://iassistdata.org/>

Biblioteconomia, identificando novas formas de serviços para o profissional bibliotecário no auxílio a pesquisadores no processo de utilização dos dados, bem como em sua preservação e reutilização a longo prazo.

Em 2007 o pesquisador Gold publica artigos enfatizando o ciclo de pesquisa e o ciclo de vida de dados onde descreve como os bibliotecários de dados poderiam selecionar, adquirir e licenciar dados de pesquisa, bem como organizar e realizar serviços de curadoria, além de trabalhar com pesquisadores para auxiliarem na descoberta de ferramentas que possibilitem outros recursos para dados criados ou existentes. Assim surge a Biblioteconomia de Dados, com um campo emergente da Ciência de Dados, compreendida como uma formação educativa promissora para auxílio em pesquisa, em e-Science e e em Big Data. (XIA; WANG 2014).

A Biblioteconomia de dados é um ramo da ciência que aborda novas práticas da Biblioteconomia na geração e preservação de dados de pesquisa. Rice e Southall (2016) definem a Biblioteconomia de dados como evolução da Biblioteconomia tradicional para a Biblioteconomia do e-science, da descoberta de recursos em formatos de dados ou digital. Segundo Ferreira (2016, p. 79), os bibliotecários estão trabalhando nas bibliotecas com “captura, arquivamento, curadoria, acesso, interoperabilidade, análise, serviços, ciclo de vida, instrumentos políticos e modelos de negócios relacionados aos dados de pesquisa.”

Rice e Southall (2016) afirmam ainda que os bibliotecários, dentro de seus espaços de trabalho, estão adotando essa perspectiva, ampliando e iniciando uma série de atividades de apoio à pesquisa, fortalecendo e conscientizando-se sobre questões de dados dentro das instituições e seus benefícios de gerenciar ativamente os dados de pesquisa. Assim, eles promovem o desenvolvimento de políticas sobre gerenciamento e preservação de dados prestando consultoria e assessoria aos pesquisadores sobre o ciclo de vida da pesquisa, influenciando a forma como os pesquisadores irão utilizar seus dados e criando ou instigando o uso de repositórios para publicação ou preservação de seus dados. Isso possibilita também ao bibliotecário de dados estabelecer uma comunicação com profissionais de serviço de dados ou profissionais da Tecnologia da Informação (TI) para desenvolver ou modificar softwares, que tenham capacidade de armazenamento e recuperação de dados e promover em conjunto, para seus usuários, cursos que possibilitem maior oferta de treinamento e introdução aos conceitos de gerenciamento de dados e curadoria.

Swan e Brown (2008) afirmam que a comunidade de biblioteconomia e ciência da informação tem um papel importante a desempenhar na área de ciência de dados, particularmente na compreensão de uma boa prática de gestão de dados e curadoria de dados, além de habilidade como manipulação e gerenciamento de dados, podem ser parte de treinamento de habilidade de pesquisa em uma instituição. Além disso, outras funções podem ser ensinadas diante da especialização do assunto, como: treinar os pesquisadores para serem mais conscientes dos usos de dados; adoção de medidas para arquivamento e preservação de dados e treinamento para formação e plano de gerenciamento de dados.

Segundo Christensen-Dalsgaard et al. (2012), surgiu em 2010 um grupo de trabalho abordando a temática E-science na Liber (Association of European Research Libraries), uma rede de biblioteca de pesquisa da Europa que têm como missão em suas mais de 450 bibliotecas afiliadas promover e discutir assuntos como: Humanidade Digitais, Acesso Aberto, Métricas e Gerenciamentos de dados de pesquisa. Este grupo lançou dez recomendações para que as bibliotecas começassem a tratar o gerenciamento de dados de pesquisa. Essas recomendações incitaram as bibliotecas a pensarem na preservação e no gerenciamento de dados da pesquisa e abriram caminho para a Biblioteconomia de dados. Abaixo, as recomendações:

Oferecer suporte ao gerenciamento de dados de pesquisa, incluindo planos de gerenciamento de dados para pedidos de subsídios, consultoria de direitos de propriedade intelectual e materiais de informação. Auxiliar o corpo docente com planos de gerenciamento de dados e a integração do gerenciamento de dados no currículo.

Envolver-se no desenvolvimento de padrões de metadados de dados e fornecer serviços de metadados para dados de pesquisa.

Criar publicações para bibliotecários de dados e desenvolver habilidades profissionais para a biblioteconomia de dados.

Participar ativamente do desenvolvimento de políticas de dados institucionais de pesquisa, incluindo planos de recursos. Incentivar e adotar políticas de dados abertos, quando apropriado, no ciclo de vida dos dados de pesquisa.

Fazer parcerias com pesquisadores, grupos de pesquisa, arquivos e centros de dados para promover uma infraestrutura interoperável para acesso, descoberta e compartilhamento de dados.

Conhecer o ciclo de vida para dados de pesquisa, fornecendo serviços para armazenamento, descoberta e acesso permanente.

Promover a citação de dados de pesquisa aplicando identificadores persistentes aos dados de pesquisa.

Fornecer um Catálogo de Dados ou um Repositório de Dados institucional, dependendo da infraestrutura disponível.

Evolver-se na prática específica de gerenciamento de dados.

Oferecer ou intermediar o armazenamento seguro para dados de pesquisa dinâmicos e estáticos em cooperação com unidades de TI institucionais e/ou buscar a exploração de serviços em nuvem apropriados. (CHRISTENSEN-DALSGAARD et al., 2012, p. 20, tradução nossa).

A Biblioteconomia de dados é um campo crescente onde os bibliotecários estão sendo chamados para apoiar as tarefas de análise de dados. Bibliotecário de dados é um

termo novo e sua aplicação se encontra nos princípios de práticas tradicionais do bibliotecário aos atuais recursos de dados, e emergiu conforme demandaram as tecnologias digitais e a internet, na década de 1990. O periódico americano *Library Journal*, fundado em 1876, por Melvil Dewey - que tem como assuntos abordados em suas publicações as bibliotecas e as práticas profissionais - publicou, em 2013, um artigo da autora Stephanie Smith, professora do Departamento de Ciência da Informação da Universidade do Sul da Flórida sobre sua pesquisa com 1.900 graduados em Biblioteconomia e Ciência da Informação com foco para os empregos emergentes e as novas responsabilidades desse profissional e, dentre estes, apareceu bibliotecários de dados, como novo campo de atuação. (MAATTA, 2013).

Nas palavras de Martínez-Urbe e Macdonald (2008, p. 276), “os dados trabalham de mãos dadas com pesquisadores aconselhando sistemas de busca de dados, formulários de acesso e auxiliando em sua gestão”.

Xia e Wiang (2014) afirmam que o perfil do bibliotecário de dados está relacionado à: competência em informática, conhecimento de disciplinas que abordem práticas, investigação e fluxos de trabalhos, gestão de dados: na aquisição (desenvolvimento de coleções), na organização (catalogação e metadados), na preservação e na implementação de serviços adequados aos usuários.

Alvaro et al. (2011) afirmam que, além da habilidade tradicional do profissional bibliotecário, três categorias relacionam-se à ciência de dados: a primeira inclui responsabilidades relacionadas à coleta, ao gerenciamento e ao armazenamento de dados; a segunda, refere-se à referência, ao desenvolvimento de coleções e à comunicação acadêmica; a terceira, a e-Science, na forma de funções de curadoria de dados e acesso aberto.

Segundo Federer (2018), os bibliotecários estão assumindo novas funções em serviços de dados de pesquisa e, neste campo exponencial, o novo profissional deve fornecer uma ampla gama de serviços.

A autora classifica esta nova função de bibliotecário de dados considerando as habilidades básicas de dados e a capacidade de desenvolver relacionamentos com pesquisadores: o generalista de dados e o especialista de dados, sugerindo que a ciência da biblioteca de dados, não pode ser única.

Os generalistas de dados, com amplo conhecimento de como os dados são usados em várias áreas ou habilidades, podem ser adequados para trabalhar em ambientes acadêmicos, onde terão oportunidades de se envolver com alunos,

professores e pesquisadores de várias maneiras. Por outro lado, os especialistas em assuntos desenvolvem algumas habilidades especializadas e têm experiência prática trabalhando com grupos de usuários específicos e podem ser uma boa opção para uma instituição na qual eles possam se concentrar mais especificamente no tipo de usuários que podem se beneficiar de sua experiência (FEDERER, 2018, p. 300, tradução nossa)

Dada a emergente natureza do campo, a prática da biblioteconomia de dados pode acarretar diversas formas de especificidades de profissionais, com perfis diferentes e atuações construídas por habilidades e competências, através de qualificações e formação baseadas em dados, promovendo, assim, a popularização de um novo profissional do século XXI.

### **3.2 *Data literacy* ou alfabetização de dados para o bibliotecário no Brasil**

Uribe-Martínez e Macdonald (2008) afirmam que as comunidades científicas estão sendo transformadas pelas maneiras de trabalhar e pelos serviços oferecidos pela biblioteca. Os profissionais da informação, desde os anos 60, quando os formatos digitais mais primitivos emergiram, trabalham com acesso à informação, quer seja no processo técnico: (classificação, catalogação e indexação) ou no resguardo e preservação documental.

Sua principal tarefa é fornecer acesso à informação aos pesquisadores que desejam utilizar esses recursos em suas atividades de pesquisa. Os avanços tecnológicos influenciaram a evolução e a automação da biblioteca, bem como os serviços e as atividades executadas pelos bibliotecários. Hoje, as bibliotecas de dados são a combinação das próprias coleções em formato eletrônico e os respectivos serviços correspondentes para ajudar a encontrá-los e utilizá-los (URIBE-MARTÍNEZ; MACDONALD, 2008).

Assim, o bibliotecário precisa aprender e alfabetizar-se sobre os dados e ciência de dados, tendo um pensamento crítico para resolução de problemas, bem como aprender a interpretar e conhecer dados, desenvolvendo habilidades para esses tipos serviços e para pesquisa orientada à dados. Além disso, o termo *data literacy* ou alfabetização é um fator crescente para a atualmente chamada Ciência de Dados e para a Ciência da Informação.

Conforme Koltay (2017), as bibliotecas tendem a ser um campo para a prática da alfabetização de dados e, sendo um espaço de treinamento e gerenciamento, são disponibilizados pelas orientações sobre questões relacionadas à gestão de dados, acesso aberto, comunicação acadêmica, direitos autorais e propriedades de arquivamento, reuso e

preservação de dados. A capacitação para a alfabetização de dados pode ser dada pelo profissional bibliotecário, na biblioteca.

O desenvolvimento de habilidades pelo bibliotecário, relacionadas a dados, relacionam-se, no nível inicial e intermediário, conforme demonstra a Quadro 1:

Quadro 1 – Níveis de habilidade em dados

Nível Inicial	Na formação dos profissionais da informação, tendo como especialidade a capacidade de realizar a gestão de dados e curadoria dos dados de pesquisa, mediante sua educação formal em instituições de ensino e participação de eventos na temática de dados.
Nível Intermediário	Após o nível inicial, o bibliotecário em sua <i>Literacy Data</i> , pode desenvolver treinamento a pesquisadores e estudantes, possibilitando compartilhamentos de competências, cujo termo é <i>data literacy</i>
Nível Avançado	Para o Nível Avançado, o bibliotecário de dados deve possuir conhecimentos relativos à estatística e à programação; neste nível as competências e qualificações exigem maior experiência e disponibilidade de serviços aos seus usuários.

Fonte: Adaptado de Prado; Marzal (2013) e Koltay (2017),

Assim, no nível inicial, o bibliotecário deve adquirir na sua formação e pela participação em eventos da área, a capacidade de realizar a gestão e curadoria de dados. No nível intermediário ele adquire a habilidade de ensinar e de desenvolver treinamentos junto aos pesquisadores e estudantes. No nível avançado, a partir das qualificações e competências desenvolvidas, o bibliotecário deverá possuir conhecimento em estatística e programação, a fim de ampliar gamas de serviços, bem como ter conhecimento de: pacotes de software; metadados baseados em padrões computacional, de análise quantitativa e qualitativa de dados e pensamento crítico para resolução de problemas, voltados aos dados. O sucesso na realização dessas atividades dependerá de uma formação de qualidade requerendo das faculdades de Biblioteconomia e de Ciência da Informação, bem como de seus docentes, o empenho para um bom planejamento no preparo de profissionais adequadamente habilitados para atenderem às demandas da e-Science e exercerem seus papéis na gestão e na curadoria de dados da pesquisa.

Conforme Koltay (2017), a data literacy tem como definição a capacidade de processar, classificar e filtrar grandes quantidade de informações, devendo o profissional estar habilitado a pesquisar, filtrar e processar, bem como produzir e sintetizar dados.

Esses atributos serão amplamente potencializados com a competência informacional em dados discutindo-se atividades como dados, gerenciamento de dados e

curadoria de dados tornando os profissionais qualificados a trabalharem com dados e alfabetização de dados.

Embora a alfabetização em dados esteja passando por um período de gestação, começa a ser amplamente aceita como uma habilidade crucial para profissionais da informação envolvidos no apoio à pesquisa intensiva de dados. Por outro lado, a terminologia no campo de dados ainda não está padronizada. Há literacia em dados e literacia em dados de investigação. Carlson et al. (2011) defendem a alfabetização informacional de dados, porque, de acordo com sua abordagem, esta difere de um significado mais restrito de alfabetização, ou seja, a capacidade de ler gráficos e tabelas apropriadamente, tirando conclusões corretas dos dados e reconhecendo quando os dados estão sendo usados de maneira enganosa ou inadequada (KOLTAY, 2017, p. 306, tradução nossa)

Além de alfabetização de dados, a alfabetização estatística também é essencial e, nas palavras de Peter e Kellam (2013) consiste na capacidade de ler e interpretar informações numéricas, conforme são apresentadas em tabelas, gráficos e textos. O bibliotecário exerce a função de mediador da informação direcionando informações aos seus usuários, onde podem ser encontradas fontes estatísticas, de sites governamentais a banco de dados baseados em assinaturas.

Apesar de os bibliotecários não terem os conhecimentos estatísticos, eles têm a capacidade de interpretar e recuperar informações no tocante à alfabetização em dados, auxiliando os pesquisadores no gerenciamento de seus dados de pesquisa.

Koltay (2017) afirma que dentro dessa nova ocupação deve-se ter como fundamental na gestão de dados: apoio a pesquisadores, manutenção dos dados, vulnerabilidade dos dados disponíveis, gerenciamento e compartilhamento e a orientação no treinamento prático para elaboração de plano de gestão de dados, além da elaboração de citações de dados e na vinculação dos dados à publicação.

As competências básicas e avançadas dependem da finalidade da educação em alfabetização de dados, sendo sintetizadas em um conjunto que compreendem vários níveis representados na matriz de competências, exemplificado pela Figura 7 e adaptada de

Ridsale et al. (2017); essa matriz está organizada pelos cinco aspectos fundamentais da definição de alfabetização em dados (Estrutura Conceitual, Coleta de dados, Gerenciamento de dados, Avaliação de dados e Aplicação de dados). A matriz é composta por 23 competências e 64 tarefas com habilidades relacionadas, sendo as competências fator evolutivo para formar e instigar o pensamento crítico do bibliotecário. Segundo Ridsale et al. (2017, p. 52), a matriz de competência possibilita “pesquisar, filtrar, produzir e sintetizar dados”.

Assim, esta matriz destina-se a formar a base dos entendimentos em curso sobre avaliação dos níveis de alfabetização em dados, bem como atua como suporte para possibilitar ao bibliotecário formação ampla em diversos campos da alfabetização de dados.

Figura 7- Matriz de Competências de Alfabetização de dados

Legenda: Competências Conceituais Competências Essenciais Competências Avançadas

Área do Conhecimento	Competência	Conhecimento /Tarefas				
Estrutura conceitual	Introdução aos Dados	Conhecimento e compreensão dos dados	Conhecimento e compreensão dos usos e aplicações de dados			
	Coleta de dados	Descoberta de dados e Coleta	Realiza a exploração de dados	Identifica dados úteis	Coleta de dados	
	Avaliando e assegurando dados e fontes	Avalia criticamente fontes de dados para Confiabilidade	Avalia criticamente a qualidade de conjuntos de Dados para erros ou problemas			
Gerenciamento de dados	Organização de dados	Conhecimento da organização básica de dados métodos e ferramentas	Acesso e requisitos de organização de dados	Organizar dados		
	Manipulação de dados	Acesso e métodos para limpar dados	Identifica os valores e as anomalias	Limpa dados		
	Conversão de dados (de formato para formatar)	Conhecimento de diferentes tipos de dados e métodos de conversão	Atribui os descritores de metadados apropriados para conjuntos de dados originais			
	Criação e uso de metadados	Cria descritores de metadados	Converter dados de um formato ou tipo de arquivo para outro			
	Curadoria de Dados	Avalia os requisitos de Curadoria de dados (por exemplo, cronograma de retenção, armazenamento, acessibilidade, requisitos de partilha, etc.)	Avalia os requisitos de segurança de dados (por exemplo, acesso restrito, unidades protegidas, etc.)	Curadoria de dados		
	Preservação de dados	Avalia os requisitos de preservação	Métodos e ferramentas de acesso para preservação de dados	Preserva os dados		
Avaliação de dados	Ferramentas de dados	Conhecimento de ferramentas de análise de dados e técnicas	Seleciona a ferramenta de análise de dados apropriada ou Técnica	Aplica ferramentas de análise de dados e técnica		
	Análise básica de dados	Desenvolve planos de análise	Aplica métodos e ferramentas de análise	Realiza análise exploratória	Avalia os resultados da análise	Compara os resultados da análise com outros achados
	Interpretação de dados (Noções básicas sobre dados)	Lê e compreende gráficos, tabelas	Identifica os principais pontos e integra isso com outras informações importantes	Identifica discordância dentro os dados		
	Identificando problemas usando dados	Usa dados para identificar os problemas em situações práticas (por exemplo, eficiência no local de trabalho)	Usa dados para identificar problemas de nível superior (por exemplo, política, ambiente, experimentação científica, marketing, economia)			
	Visualização de dados	Cria tabelas significativas para organizar e apresentar visualmente dados	Cria representações gráficas significativas de dados	Avalia eficácia de representações gráficas	Avalia criticamente representações gráficas para precisão e deturpação de dados	
	Apresentando dados (verbalmente)	O resultado pretendido (s) para apresentar os dados	Avalia as necessidades do público e a familiaridade com sujeito (s)	Planeja a reunião apropriada ou tipo de apresentação	utiliza tabelas significativas e visualizações para comunicar dados	Apresenta argumentos e/ou resultados claros e coerentemente
	Decisões orientada a dados Tomada de decisões em dados	Prioriza a informação obtida a partir de dados	Converter dados em informações	Pesa o mérito e os impactos de possíveis soluções/decisões	Implementa decisões/soluções	

Fonte: Adaptado de RIDSDALE; CHANTEL et al. (2015).

A matriz de competência em alfabetização de dados é um ponto de partida para discussões de padrões na promoção de conteúdo didático para conhecimentos disciplinares ou de domínio específico, nos cursos de Biblioteconomia. Incentiva também a forma de aplicações permitindo ao graduando sair do viés tradicional potencializando seu o pensamento crítico.

Nesse processo, a biblioteconomia, no Brasil, deve indagar se dentro do seu currículo acadêmico há inclusão de disciplinas que tornem os bibliotecários aptos a atuarem no contexto e no gerenciamento dos dados da pesquisa. A função de bibliotecário de dados deve entrar na Classificação Brasileira de Ocupações, haja vista seu caminho inovador de colaboração e cooperação.

Desenvolver o sólido conhecimento da literacia de dados é essencial para a construção de conhecimentos específicos de domínio disciplinar e para garantir que os cidadãos possam utilizar e aplicar essas habilidades de forma adequada e diversamente ao longo de suas vidas pessoais e profissionais. O melhor lugar para começar esta iniciativa é o currículo de graduação em instituições pós-secundárias, devido ao seu objetivo de produzir algo globalmente competitivo, com pensamento crítico e bem equipado. Adicionando a alfabetização de dados, explicitamente, os currículos de graduação ajudarão a garantir que os graduados estejam melhor equipados para atender às habilidades de dados (RIDS DALE; ROTHWELL; SMIT, 2015, p. 2, tradução nossa)

Portanto, os autores consideram que as instituições de ensino devem prever a alfabetização em dados em seus currículos uma vez que existem lacunas neste sentido e que esta alteração auxilia os alunos de Biblioteconomia a reconhecerem a alfabetização de dados e seus elementos constituintes.

A seguir serão apresentados alguns web sites que tratam da temática da alfabetização de dados.

### **3.3 Web sites para alfabetização em dados**

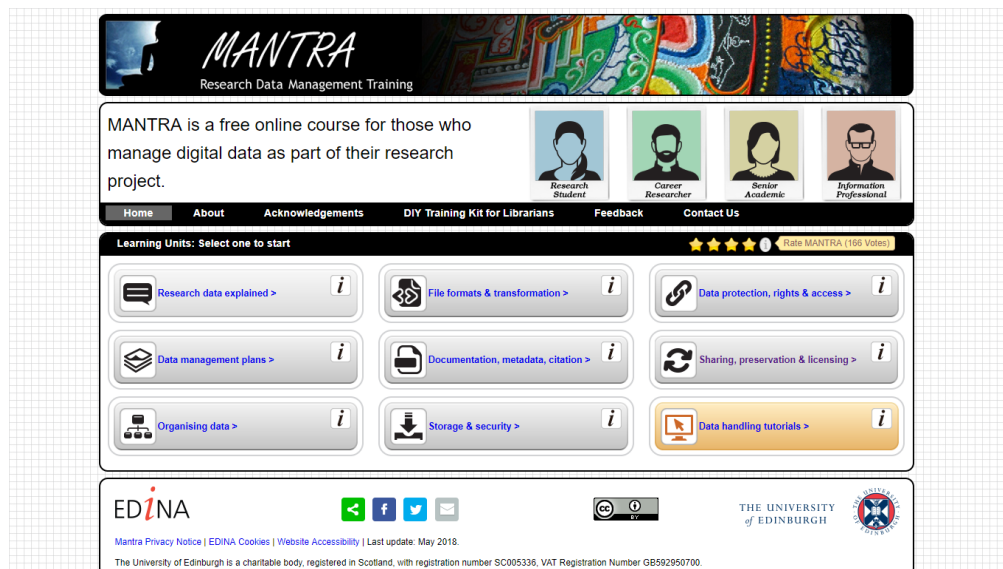
Os programas de alfabetização informacional, bem como recursos digitais disponibilizados, devem estar alinhados com as atuais práticas e culturas disciplinares. Têm surgido, através da expansão de publicações e frente às necessidades, portais de conteúdos interdisciplinares que possibilitam a apropriação do conhecimento em dados.

Porém, alguns fatores ainda podem impedir o acesso do profissional da informação a estes conteúdos como o conhecimento prático e o domínio idiomático, tendo em vista que a maioria dos sites de alfabetização em dados e ciência de dados não estão traduzidos para o português (KOLTAY, 2017).

O uso desses sites pode melhorar a eficácia dos processos e dos serviços a serem disponibilizados pelo bibliotecário para possibilitar amplo conhecimento em dados. Por isso, serão apresentadas algumas ferramentas disponibilizadas por esses sites para formação complementar e ampliação de competências e de habilidades para os bibliotecários que se interessam pela temática em alfabetização de dados. Esses sites foram localizados durante o desenvolvimento da pesquisa a partir de buscas no buscador do Google utilizando-se de termo em inglês: “data librarian training”; “data management software” e “data literacy training”, onde foram recuperados inúmeros sites em idioma inglês voltados à temática de gerenciamento e de alfabetização de dados. Dentre eles destacaram-se três que se relacionavam com o propósito deste trabalho: o site Mantra, voltado especificamente para alfabetização de dados e as ferramentas de gerenciamento e criação de plano de dados: Foster e DMPTool, que serão apresentados a seguir.

Em 2012, na Universidade de Edimburgo, no Reino Unido, surgiu a plataforma Mantra<sup>4</sup> (figura 8). O site Mantra é um canal de comunicação desenvolvido por bibliotecários desta Universidade e disponibiliza, por meio de unidades interativas, um curso on-line para gerenciamento e gestão de dados da pesquisa a estudantes de pós-graduação, pesquisadores e profissionais da informação; além disso oferece um kit de treinamento em alfabetização de dados para bibliotecários.

Figura 8- Mantra Research Management Training- recursos para alfabetização de dados



Disponível em: <https://mantra.edina.ac.uk>

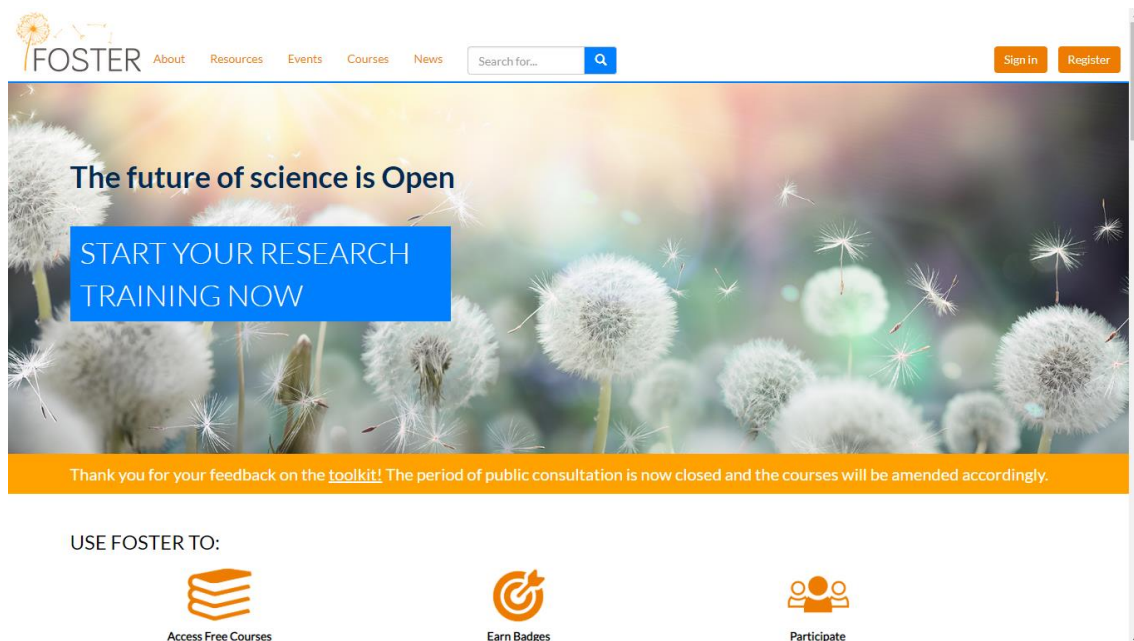
<sup>4</sup> Site do Mantra. Disponível em: <https://mantra.edina.ac.uk/> Acesso em: 25 jan. 2019.

Com este recurso o bibliotecário pode criar sua biblioteca de dados e utilizar as ferramentas indicadas para a gestão de dados e proporcionar serviços aos pesquisadores. O site possibilita realizar atividades como: criar um plano de dados para a instituição; criar formas de nomeações para arquivos de dados; identificar tipos de dados de pesquisa; elaborar ferramentas para armazenamento e backup dos dados - sendo disponibilizado um Kit de Treinamento para download nos formatos: PDF; PPT; Word e Zip, com explicações em cada aba sobre a etapa selecionada - além de fornecer atividades para fixação de conteúdo. O site Mantra é de conteúdo aberto e possibilita ao bibliotecário realizar cursos e também modificar seu conteúdo para utilizar de forma customizada a alfabetização de dados em seu espaço de trabalho.

O site Foster (Figura 9) foi criado pela União Europeia com o objetivo de contribuir com os pesquisadores e com os bibliotecários em recursos voltados à ciência de dados e à ciência aberta. É uma plataforma que reúne recursos de treinamento dirigidos àqueles que precisam saber mais sobre ciência de dados, ou que precisam desenvolver estratégias e habilidades para implementarem as práticas de gerenciamento de dados em seus fluxos de trabalhos diários.

Muitos usuários podem beneficiar-se do portal, como bibliotecários, gerentes do dados, administradores e pesquisadores. Os materiais e os conteúdos de treinamento são disponibilizados de forma gratuita, no idioma inglês (FOSTER, 2019).

Figura 9- FOSTER- recursos para treinamento em gestão de dados



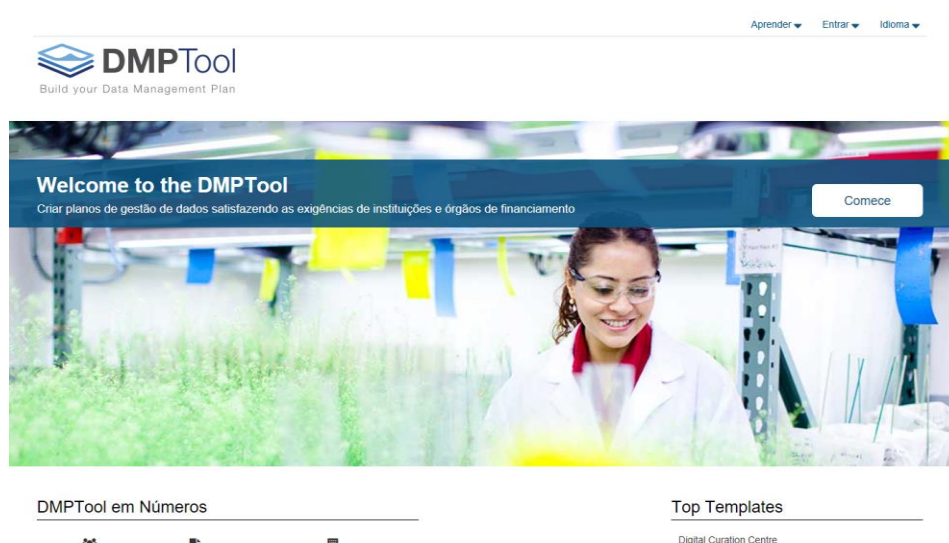
Disponível em: <https://www.fosteropenscience.eu/> Acesso em: 25 jan. 2019

O site oferece habilidades que podem e devem ser desenvolvidas pelo bibliotecário de dados desde o planejamento, tratamento, análise, preservação e reutilização dos dados, como consultoria, treinamento e infraestrutura para armazenamentos e curadoria dos dados. Além disso, cursos como: acesso aberto aos dados, melhores práticas de pesquisas em dados educacionais, formação em gestão e curadoria de dados e gestão de dados.

O site Dmptool.Org (Figura 10), apresentado a seguir, é utilizado no gerenciamento de dados e originou-se a partir da necessidade dos pesquisadores de criarem planos de gerenciamentos de dados, que atendessem aos requisitos institucionais e de financiamento. O site auxilia a consolidação do conhecimento às necessidades de gerenciamento de dados, auxiliando pesquisadores nos seus projetos e permite às instituições de pesquisa realizarem parcerias, bem como criarem seus planos de dados, tendo como vantagem a possibilidade de personalizar o plano de gerenciamento de dados aos usuários da instituição.

O DMPTool é um aplicativo on-line gratuito e de código aberto que ajuda os pesquisadores a criarem planos de gerenciamento de dados. Esses planos, ou DMPs, são agora exigidos por muitas agências de financiamento como parte do processo de submissão de propostas de subsídios. O DMPTool fornece um campo para o pesquisador criar seu plano de gerenciamento de dados, além de possibilitar o cadastro da instituição com os requisitos inseridos pelo usuário sobre tipos de dados; formas de dados; quantos dados o projeto de pesquisa poderá coletar e criar; formatos de arquivo e diretrizes para organização dos dados. A Figura 10 traz uma visão do site.

Figura 10- DMPTOOL- site para elaborar plano de gestão de dados



Disponível em: <https://dmptool.org/> Acesso em: 25 jan.2019

Tendo em vista as habilidades necessárias a um cientista de dados, propõe-se interconectar, através desses sites, ampla difusão dos serviços e produtos voltados à alfabetização e gerenciamento de dados, sendo uma possibilidade de uso e formação complementar para os bibliotecários de dados, tendo em vista, que essas plataformas possibilitam realizar além de capacitações no âmbito da biblioteca, promover e criar seu gerenciamento de dados, utilizando-se de políticas e diretrizes internacionais com acesso aberto.

Na próxima seção apresentamos os procedimentos metodológicos utilizados para a realização da pesquisa ora apresentada.

## 4 METODOLOGIA

A metodologia, conforme Barros e Lehfeld (2007, p.2) é “o estudo da melhor maneira de abordar determinados problemas no estado atual de nossos conhecimentos”, sendo aplicados processos e técnicas. Assim, serão apresentados os métodos e procedimentos utilizados para a construção deste trabalho durante o desenvolvimento desta pesquisa.

A pesquisa, via de regra, fundamenta-se na descrição e análise de métodos a serem empregados considerando-se os aspectos de teorias, hipóteses, variáveis, recursos de pessoal e de equipamentos ressaltando a relação entre teoria e prática.

A parte teórica abordaria o problema da natureza do conhecimento e método científico, que estariam referenciando e direcionando modelos analíticos de explicação da realidade em questão. A parte prática compreenderia a produção científica referente a técnicas e métodos operacionais necessários para o estudo e compreensão da realidade ou do objeto de análise (BARROS; LEHFELD, 2007, p. 16).

A observação sistemática, segundo Barros e Lehfeld (2007, p. 74), “caracteriza-se por ser estruturada e realizada em condições controladas, tendo em vista objetivos e propósitos pré-definidos”. Desta forma a pesquisa ora apresentada tem como abordagem metodológica de estudo a pesquisa exploratória e a descritiva. A primeira diz respeito à questões que envolvam maior familiaridade com o assunto (LAKATOS; MARCONI, 2001) e a segunda analisam as características de um fenômeno para descobrir sua natureza (LAKATOS; MARCONI, 2001).

Definido por Diehl (2004, p. 50), o método fenomenológico, aborda “a realidade, construída socialmente, é entendida como o compreendido, o interpretado, o comunicado, o sujeito/ator é reconhecidamente importante no processo de construção do conhecimento”.

Esta pesquisa caracteriza-se por uma abordagem qualitativa, tendo como efeito, conforme Diehl (2004, p. 52), “compreender e classificar os processos dinâmicos vividos por grupos sociais, contribuir no processo de mudança de dados grupo”.

Neste tipo de abordagem, o pesquisador e pesquisados interagem por meio de aperfeiçoamentos técnicos e conjectura teórica-metodológica, com liberdade para realizar seu estudo.

Considera-se, desta forma, que esta abordagem se adequa ao que foi realizado na Biblioteca Central da Universidade Estadual de Feira de Santana, onde foi possível a realização de uma oficina para sete bibliotecários da instituição, sob a temática de alfabetização de dados, a fim de levantar informações sobre o conhecimento da equipe no assunto e de fornecer

embasamento teórico e prático sobre o tema. Posteriormente foi aplicado aos bibliotecários participantes da oficina um questionário que foi respondido por intermédio do Google Forms para verificar o grau de satisfação com o mesmo.

Segundo Diehl (2007, p. 73-74), “o procedimento mais adequado para conhecer o universo de publicações acerca de um assunto é consultar os especialistas na área, as bibliotecas especializadas, a internet, as indicações de bibliografias”. Neste trabalho utiliza-se para o arcabouço teórico, a pesquisa bibliográfica.

A pesquisa bibliográfica, conforme Gil (2008), estimula a discussão dos assuntos através de consultas em teses, periódicos, livros, artigos, obras de referências e bases de dados, como o Portal de Periódicos da Capes, que provê o acesso gratuito e pago às revistas de cunho nacional e internacional, Banco de Teses e Dissertações do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e Google Acadêmico.

Nesse sentido, deve-se considerar, ainda, os Repositórios Institucionais como fontes de informação que podem ser pesquisadas no momento da revisão da literatura, além das Bibliotecas de Teses e Dissertações das próprias universidades e outros bancos de dados que elas provêm.

Os dispositivos de busca permitem refinar a pesquisa para direcioná-la de forma mais precisa, evitando a obtenção de resultados muitos amplos ou restritos; para isso, recomenda-se a utilização de palavras-chave ou frase que descrevam a informação procurada por meios de operadores booleanos (DIEHL, 2007, p.75)

No levantamento bibliográfico realizado para este estudo, as palavras-chaves utilizadas foram: data literacy, ciência de dados, ciência da informação, data Science, e-Science, data librarian, bibliotecário de dados, data management of librarian, alfabetização de dados, história dos dados, history of data Science, data curation. As bases de dados utilizadas foram: Portal de Periódicos da Capes, EBSCO, Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) e nas palavras-chave foram inseridas entre elas o operador booleano OR, uma vez que a intenção era recuperar o maior número de informações possíveis. Assim, foram recuperados 56 artigos, que posteriormente passaram por uma seleção com a leitura dos resumos e, em alguns casos, de partes do texto. Desta forma, o conjunto de artigos selecionados foram de trinta e seis. Em se tratando de livros e outros materiais, a recuperação se deu por meio do banco de dados bibliográficos da UFS, o Pergamum.

#### **4.1 A Universidade Estadual de Feira de Santana (UEFS) e a Biblioteca Central Julieta Carteado**

A UEFS é uma instituição de ensino superior, de pesquisa e estudo, autorizada pelo Decreto Federal nº. 77.496, de 27 de abril de 1976. Com a portaria Ministerial nº: 847/86 de 1972 tem sua instalação em 31 de maio de 1976, com os seguintes cursos: Licenciatura de 1º e 2º graus Plena em Letras- Inglês Francês, Licenciatura em Ciências, Licenciatura em Estudos Sociais e cursos de Enfermagem, Engenharia de Operações, Administração, Economia e Ciências Contábeis (UEFS)

A UEFS possui um campus central, localizado na Avenida Transnordestina, S/No., no bairro Novo Horizonte, no município de Feira de Santana, contendo: sete pátios de aula, creche, biblioteca central, restaurante universitário e prédio da Reitoria. Além disso, possui outras unidades extra campus: Horto Florestal, Observatório Antares, o Centro Universitário de Cultura e Arte (CUCA) e um campus avançado da Chapada Diamantina. Possui atualmente 28 cursos de graduação, sendo 14 cursos de bacharelado, 11 cursos de licenciatura e 03 cursos de modalidade, licenciatura e bacharelado. Contém 18 cursos Stricto Sensu, sendo 4 de mestrado profissional e 14 de mestrado acadêmico (UEFS, 2019).

O Sistema Integrado de Bibliotecas (SISBI) é composto por sete bibliotecas setoriais: a Biblioteca Setorial Monteiro Lobato; Biblioteca Setorial Monsenhor Renato Galvão (Localizada no Museu Casa do Sertão); a Biblioteca Setorial do Observatório Astronômico Antares; a Biblioteca do Centro de Educação Básica; a Biblioteca Setorial Pierre Klose, Biblioteca Setorial Ernesto Simões Filho e a Biblioteca Setorial Chapada Diamantina, sendo gerenciado pela Biblioteca Central Julieta Carteado.

A Biblioteca Central (BICEN), localizada no campus universitário foi implantada em 1976, junto com a criação da UEFS, com capacidade para abrigar aproximadamente 100.000 títulos e mais de 350.000 exemplares de acervo básico. Os serviços da biblioteca são informatizados e está interligada à internet e outras redes de comunicação nacionais e internacionais. Seu funcionamento é de segunda a sexta feira, das 7:30h às 22:45h e aos sábados das 9:00h às 17:00h.

Na figura 11, a seguir, tem-se a vista da entrada da Universidade.

Figura 11- Entrada da UEFS



Disponível em: <http://www.uefs.br> Acesso em: 28 abr.2019

#### 4.2 Procedimentos metodológicos

O universo de pesquisa, constitui-se, segundo Gil (2008, p. 89) de um “conjunto de elementos que possuem determinadas características”. Nesta perspectiva, a amostra selecionada para participação nesta pesquisa foi caracterizada pelos bibliotecários da Biblioteca Central (BICEN), da Universidade Estadual de Feira de Santana (UEFS) que possuíam alguma relação com gerenciamento de dados totalizando, desta forma sete servidores que atuam diretamente com o Repositório Institucional na UEFS.

Portanto, para se atingir objetivo geral do trabalho “desenvolver habilidades e competências em alfabetização e em gerenciamento de dados para bibliotecários da Biblioteca Central da Universidade Estadual de Feira de Santana” buscou-se conhecer as atividades executadas pelos bibliotecários e, com base na literatura levantada identificar as habilidades e competências requeridas. Para isso, utilizou-se a Matriz de Competências de Ridsdale; Chantel et al. (2015) e também o infográfico elaborado na plataforma Canva para elaboração de apresentações de design e templates em diversos formatos como: cards para rede social, materiais em formato A4 e apresentações, sendo utilizada a versão gratuita para construção do produto desta dissertação.

Em relação aos objetivos específicos, o primeiro trata de “promover um curso próprio de alfabetização de dados com demonstração prática das ações necessárias em cada etapa” e, para isso, realizou-se, em 07/06/2019, no auditório da Biblioteca Central da UEFS,

uma oficina sobre alfabetização de dados com carga horária de 06 horas dividida em cinco módulos: 1º: Planejamento e gerenciamento de dados; 2º: Organizando e documentando dados; 3º: Armazenamento de dados e segurança; 4º: Ética e Direitos Autorais e 5º. Compartilhamento de dados para bibliotecários, que buscou capacitar a equipe sobre o gerenciamento de dados. A Matriz de competências de Ridsdale; Chantel et al. (2015) foi a base para a construção dos conteúdos.

Para se atingir objetivo de “levantar as ações promovidas pelos bibliotecários relacionadas à inclusão de novos serviços na biblioteca” foi a interação do pesquisador com os bibliotecários durante a realização do curso acima mencionado. Assim, constatar que os serviços atualmente desenvolvidos são compartilhamento e criação de dados, através da pasta do servidor da Instituição, possibilitando o armazenamento e backup diário desses dados, juntamente com o suporte da TI. Os bibliotecários participantes da capacitação em gerenciamento de dados informaram que novos serviços serão discutidos e incorporados na biblioteca, como a elaboração do plano de gerenciamento de dados da Instituição, novas formas padronizadas para nomeação e preservação de dados e curadoria dos dados, especialmente agora, que poderão efetuar ações mais concretas direcionadas ao tema de “dados”.

Assim, os participantes elaboraram e conheceram, durante o curso, atividades principalmente voltadas para o plano de gerenciamento de dados destacando a necessidade de possuírem habilidades e competências para os próximos anos. Além disso, pontuaram que a oficina possibilitou pensar em modificações em seu repositório institucional para a disponibilização dos dados de pesquisa de seus pesquisadores, com o intuito de fornecerem novas formas de depósito dos trabalhos acadêmicos dos cursos de graduação e pós-graduação. Além disso incitou a promoção e incorporação em seus serviços de treinamentos à comunidade acadêmica da UEFS e ao público externo, voltadas para a temática de alfabetização de dados.

O terceiro objetivo específico “elaborar um guia com materiais que sirvam de embasamento teórico e prático para alfabetização e gerenciamento de dados para bibliotecários”, torna-se o produto deste Mestrado Profissional. Sua concepção se deu a partir da necessidade de oferecer materiais para alfabetização de dados, utilizando-se, especialmente, do site Mantra, que fornece informações específicas sobre conteúdos de dados, com programas de treinamento já elaborados e de livre-acesso.

Em contato com a equipe do site Mantra obteve-se a informação de que o conteúdo deles poderia ser adaptado para a elaboração do guia em questão que também se baseou nas recomendações verificadas na Matriz de Competências de Ridsdale; Chantel et al. (2015). Assim, o guia foi elaborado com a finalidade de capacitar bibliotecários em alfabetização de

dados, sendo os bibliotecários da BICEN, da UEFS, os protagonistas para seu desenvolvimento. O mesmo está hospedado no link <http://sites.uefs.br/portal/sites/bibuefs/menus/treinamentos>.

O Apêndice 1 traz a impressão de algumas telas do guia. E, por fim, o quarto objetivo, destinado a “avaliar o curso ministrado com vistas a identificar o grau de satisfação do mesmo, face aos assuntos abordados” foi realizado por meio de questionário elaborado pelo Google Forms e respondido por cinco dos sete participantes do curso. O Apêndice 2 traz as questões levantadas no questionário.

A seguir apresenta-se o Quadro 2 com informações sobre os instrumentos de coletas de dados e as fontes de coletas de dados utilizadas nos objetivos da pesquisa.

Quadro 2- Informações sobre os instrumentos de coletas de dados e as fontes de coletas de dados utilizadas nos objetivos da pesquisa

OBJETIVOS	PROCEDIMENTOS ADOTADOS	FONTE DE COLETA DE DADOS
promover um curso próprio de alfabetização de dados com demonstração prática das ações necessárias em cada etapa	Oficina ministrada na Biblioteca Central Julieta Carteadó, da UEFS, para sete bibliotecários	Informações do site Mantra, referências bibliográficas, Matriz de Competência
levantar as ações promovidas pelos bibliotecários relacionadas à inclusão de novos serviços na biblioteca	Levantamento de informações durante a realização do curso	Bibliotecários participantes do curso
elaborar um guia com materiais que sirvam de embasamento teórico e prático para alfabetização e gerenciamento de dados para bibliotecários	Localização de informações e adaptação do conteúdo do curso para elaboração do guia	Levantamento bibliográfico sobre os tópicos ministrados; utilização do site Mantra e adaptação de conteúdo já preparados Utilização da Matriz de Competências para definição dos conteúdos do guia
avaliar o curso ministrado com vistas a identificar o grau de satisfação do mesmo face aos assuntos abordados	Questionário elaborado por meio do Google Forms	Bibliotecários participantes do curso

Fone: elaborado pelo autor (2019).

Para tanto, foi aplicado um questionário para a equipe que participou da oficina para a capacitação em dados, sendo o mesmo respondido por meio do Google Forms. As perguntas avaliativas visaram obter informações sobre a satisfação no conteúdo abordado, que enfocou: Planejamento de dados, Organização de dados, Armazenamento de dados e Ética, além de buscar verificar quais foram os pontos mais importantes, a maior dificuldade encontrada durante a realização da oficina, além de verificar a aplicação das informações obtidas nas atividades do trabalho e de um espaço para comentários adicionais. Os resultados serão apresentados no capítulo Resultados da Intervenção.

A seguir apresenta-se o Diagnóstico realizado na BICEN/UEFS.

## 5 DIAGNÓSTICO

O diagnóstico refere-se à etapa do projeto de pesquisa que visa fornecer informações sobre o local onde a pesquisa será desenvolvida, direcionando aspectos que promovam conhecer o ambiente onde a instituição se enquadra, enfocando o ambiente externo e o interno, assim como os pontos fortes e fracos.

Desta forma tem-se como caracterização do objeto de pesquisa:

- a) **nome e natureza:** Biblioteca Central Julieta Carteado da Universidade Estadual de Feira de Santana;
- b) **histórico:** Instituída em 1976, a partir da criação da UEFS;
- c) **descrição dos principais serviços:** A biblioteca apresenta serviços e produtos para a comunidade acadêmica, composta por discentes docentes e comunidade externa. Dentre os principais serviços oferecidos estão: empréstimos automatizados, através do Sistema Pergamum; reserva e renovação eletrônica; levantamento bibliográfico; acesso à bases de dados; acesso à internet para a comunidade interna (docentes, discentes, servidores e prestadores de serviço), bem como comunidade externa; acesso ao Portal de Periódicos da Capes (local e remoto via VPN) e acesso ao acervo físico; normalização de publicações; orientação à normalização de trabalhos acadêmicos (ABNT e Vancouver); elaboração de ficha catalográfica; orientação ao usuários, com visitas e treinamentos e serviços de alerta: BC Informa, murais e boletim eletrônico, através do site: <http://sites.uefs.br/portal/sites/bibuefs>
- d) **porte, instalação e tipo:** a biblioteca possui prédio próprio; acervo aproximadamente de 100.000 títulos e mais de 350.000 exemplares de acervo físico, sendo compostos por áreas do conhecimento, classificados conforme a CDU, além de acervo com conteúdo regionais: jornais e revistas, bem como assinatura e acervo físico de revistas científicas.
- e) **Principal foco:** Regional
- f) **Tipos de usuários:** Discentes, Docentes, Servidores, Terceirizados e Comunidade externa.
- g) **Recursos Humanos da Biblioteca Central:** Possui 09 bibliotecários, 05 técnicos, e 8 auxiliares de bibliotecas, sendo 4 terceirizados.

Estas informações permitem observar o desempenho organizacional da instituição e elaborar a análise SWOT, que é um modelo que tem sido utilizado para observar o ambiente da organização oferecendo informações sobre as ameaças e as oportunidades advindas do

ambiente externo e dos pontos fortes e fracos advindos do ambiente interno sendo utilizada dentro da metodologia do planejamento estratégico.

Com base nas informações coletadas, segue-se a análise de SWOT da BICEN/UEFS, exibida na Figura 12.

Figura 12 – Análise SWOT da Biblioteca Central da UEFS

<b>ASPECTOS INTERNOS</b>	<b>Forças</b>	<b>Fraquezas</b>
	<ul style="list-style-type: none"> <li>• Equipe com habilidade técnica</li> <li>• Apoio da direção</li> <li>• Suporte tecnológico acessível</li> <li>• Profissional engajado em desenvolver habilidades voltadas a temática em “dados”</li> <li>• Atendimento especializado ao pesquisador e discente</li> </ul>	<ul style="list-style-type: none"> <li>• Limite de habilidade na temática de dados</li> <li>• Falta de software/site específico em gerenciamento de dados</li> <li>• Falta de conhecimento técnico em gestão de dados</li> <li>• Falta de domínio de idioma em língua estrangeira</li> <li>• Falta de um guia para o profissional, direcionando-o sobre gerenciamento de dados</li> <li>• Ausência de política institucional sobre dados</li> </ul>
<b>ASPECTOS EXTERNOS</b>	<b>Oportunidades</b>	<b>Ameaças</b>
	<ul style="list-style-type: none"> <li>• Possibilidades de cursos e treinamentos externos para conhecimento e capacitação na área de gestão de dados</li> <li>• Existência política de gestão de dados a nível nacional e internacional, para gerenciamento em pesquisas de dados</li> <li>• Internacionalização das pesquisas em dados nas pós-graduações brasileiras</li> </ul>	<ul style="list-style-type: none"> <li>• Falta de acompanhamento da tendência em gestão de dados</li> <li>• Falta de recursos financeiros</li> <li>• Falta de investimento em Tecnologias</li> </ul>

Fonte: elaborado pelo autor (2019)

Assim, considera-se abordar características da instituição de maneira organizada e, conforme Chiavenato e Sapiro (2003), a análise do ambiente permite identificar suas oportunidades e ameaças e verificar os seus pontos fortes e fracos. Sendo assim a análise do ambiente interno, conforme Ferrel e Hertline (2009) baseia-se na existência de forças e fraquezas, como os recursos possuídos ou não na Instituição, bem como o relacionamento com sua comunidade acadêmica. Percebe-se que no campo da:

- **Forças:** a biblioteca da UEFS, possui vantagens em relação a sua equipe, que além de possuir habilidade técnica, possuem interesse em adquirir competências para formação em

relação a temática “dados”, tendo como apoio, a direção da instituição e suporte tecnológico acessível;

- Fraquezas: percebe-se limitação e falta de conhecimento sobre a temática dados, inexistência de uma política institucional de gerenciamento de dados, falta de conhecimento ou uso de software específico, bem como a falta de domínio de idioma em língua estrangeira, tendo em vista que a maioria das referências em “dados” são disponibilizadas em literatura inglesa, percebe-se, também, que a falta de um guia para o profissional, em idioma português, dificulta as atividades profissionais relativas à gestão de dados.

Para o estudo da análise do ambiente externo, Cobra (2003) afirma que devem ser analisadas as variáveis macro como: demografia, economia, tecnologia, políticas institucionais, sociedade e cultura e, além disso, o seu consumo e os canais de distribuição. Sendo assim, o ambiente externo caracteriza-se como:

- Oportunidades: a existência de cursos e treinamentos externos voltados a temática “dados” que auxiliarão a construção do gerenciamento de dados, com base nas políticas publicadas em nível nacional e internacional. Além disso é importante para a instituição promover o processo de internacionalização de suas pesquisas, a partir da capacitação de seus servidores na área de gestão de dados.

- Ameaças: A partir do momento que ocorrer a falta de acompanhamento das tendências, voltadas a temática “dados”, pode gerar estagnação dos novos serviços a serem disponibilizados na biblioteca da UEFS. Outro fator como ameaça é a falta de recursos financeiros e investimentos em tecnologias, tendo em vista a crescentes evolução no mercado.

A análise da matriz SWOT permitiu identificar as oportunidades e ameaças, as forças e fraquezas mais significativas da Biblioteca Central da UEFS. A maioria dos pontos fortes da biblioteca estão relacionados à habilidade dos profissionais bibliotecários em adquirir competências sobre a temática de gestão de dados, sendo necessário adotar e priorizar as tarefas, procedimentos ou possibilidades de ação, baseando-se na Matriz de Competências de Ridsdale; Chantel et al. (2015).

Assim, foi possível desenvolver e fortalecer as competências em alfabetização de dados para a equipe bibliotecária, tendo em vista que o interesse da temática foi fator principal para o desenvolvimento da pesquisa.

A seguir apresentam-se os Resultados da Intervenção, a partir da oficina desenvolvida na BICEN/UEFS.

## **6 RESULTADO DA INTERVENÇÃO**

Tendo em vista os passos executados para a elaboração desta pesquisa que teve por objetivo principal promover a alfabetização de dados para a equipe de bibliotecários da Biblioteca Central da UEFS que atuam no Repositório institucional da universidade, consideram-se como resultados da intervenção:

### **6.1 Curso de alfabetização em dados**

O curso foi ministrado no auditório da Biblioteca Central da UEFS no dia 07/06/2019, com carga horária de 06 (seis) horas, abordando os seguintes tópicos:

- Planejamento de gerenciamento de dados
- Organizando e documentando dados
- Armazenamento de dados e segurança
- Compartilhamento de dados

Pôde-se observar, durante a realização do curso, o interesse da equipe na temática, uma vez que as informações passadas seriam utilizadas na prática, em atividades rotineiras da equipe. Sob este aspecto, analisando-se a necessidade de melhor compreensão sobre vários itens que envolvem os dados e levando em consideração os objetivos desta pesquisa em prover capacitação para a equipe da Biblioteca Central da UEFS, pode-se elaborar, a partir do curso oferecido o Guia para alfabetização de dados para bibliotecários que será tratado a seguir.

### **6.2 Guia para alfabetização em dados**

Para sua elaboração, utilizou-se a Matriz de Competências de Ridsdale, Chantel et al. (2015), desenvolvida para formação e capacitação em alfabetização de dados, distribuídas por assuntos como os tipos de dados, a gestão de dados, tratamento de dados e curadoria de dados

O guia é produto desta Dissertação está constituído em capítulos que auxiliam os bibliotecários a encontrarem as informações necessárias para adquirir competências e habilidades afim de desenvolverem a alfabetização de dados nos espaços da biblioteca. Além disso, o guia está hospedado no site da Biblioteca Central, na aba de treinamento, no link: <http://sites.uefs.br/portal/sites/bibuefs/menus/treinamentos>

O guia foi construído utilizando-se da oficina e do feedback após o encerramento, enfatizando os assuntos mais relevantes relacionados no tocante à alfabetização de dados. Utilizou-se a plataforma Canva para a sua confecção, em formato A4, sendo disponibilizado online e no formato PDF. Além disso foi efetuada a licença do guia no site da Creative Commons, permitindo a distribuição gratuita da obra, sendo protegida por direitos autorais. O guia está constituído por unidades e, ao final de alguns capítulos, são disponibilizadas atividades como auxílio ao processo de memorização do conteúdo disponibilizado. Segue a seguinte estrutura de conteúdo:

- Unidade 1- Dados da Pesquisa e Tipologia
- Unidade 2-Planejando e Gerenciando Dados
- Unidade 3- Organizando e Documentando Dados
- Unidade 4- Armazenamento, Compartilhamento e Segurança de Dados
- Unidade 5- Ética e Direitos Autorais

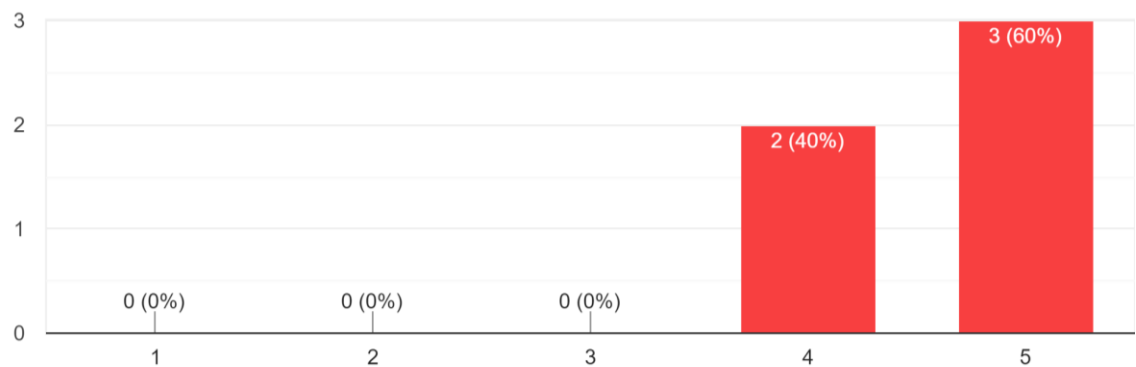
Por fim, foi possível verificar que os profissionais participantes da capacitação em alfabetização de dados estavam representados na divisão de classes de Prado, Marzal (2013) e Koltay (2017), no Nível Iniciante, ou seja, com pouco conhecimento em dados. Este dado pôde ser obtido pelas discussões ocorridas durante o treinamento, onde falas como “não tínhamos conhecimento sobre a temática de alfabetização em dados” ou “na nossa biblioteca, não desenvolvemos atividades de gerenciamento de dados” nortearam a classificação. Entretanto, observou-se, também, a possibilidade de avançarem para o nível Intermediário, a partir da constatação de que com a oferta da oficina, esses bibliotecários obtiverem a experiência de vivenciar a formação complementar, a partir das atividades desenvolvidas e dos conteúdos ministrados. Dentre esses destacam-se as atividades de Plano de Gerenciamento de dados e Identificação de tipos de dados da pesquisa, observando a interação entre os participantes e a aplicação prática do conteúdo ministrado para assim suscitar e desenvolver discussões inerentes à temática de dados. Essas atividades foram realizadas após a aplicação do conteúdo teórico tendo o ministrante realizado a mediação e a distribuição das atividades, sempre instigando os participantes a questionarem-se sobre inserir em suas atividades profissionais a gestão de dados, pois poderão, futuramente elaborar para a instituição uma Política de Gestão de Dados da Pesquisa, bem como repensarem a forma de disponibilização dos documentos no Repositório Institucional da UEFS.

Outro ponto que merece destaque nesta pesquisa do Mestrado Profissional foi a avaliação realizada após o curso de alfabetização de dados que foi ministrado no dia 07/06/2019. Os gráficos a seguir demonstram os resultados do questionário elaborado no Google Forms, considerando os cinco respondentes do total dos sete bibliotecários que estiveram presentes. Portanto, para essa análise, considera-se como 100% os cinco respondentes. O questionário foi elaborado utilizando uma escala de um a cinco para aferir o grau de satisfação com variações de pouco satisfeito a muito satisfeito.

O primeiro item buscou conhecer o grau de satisfação com a oficina ministrada, onde pode-se considerar que houve aproveitamento, devido que os resultados estão dentro das pontuações quatro e cinco, com 40% e 60% respectivamente, conforme demonstra o Gráfico 1

Gráfico 1 - Satisfação com a oficina ministrada

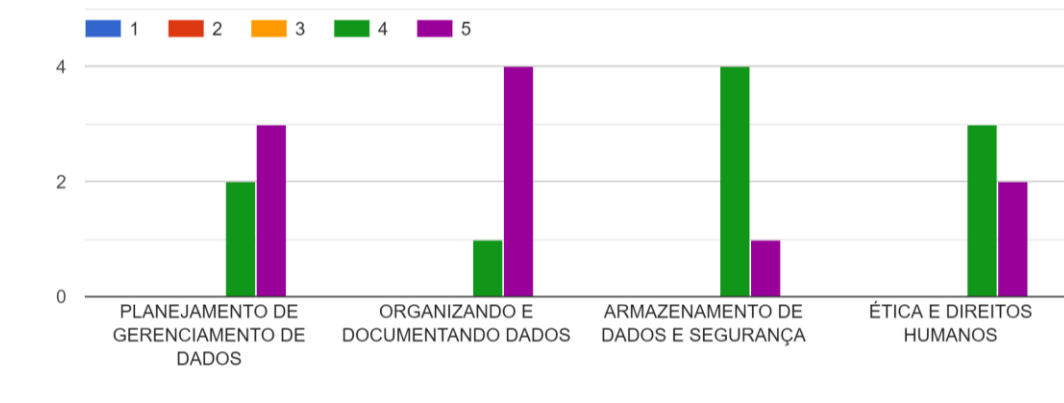
Fonte: Google Forms (2019).



O segundo item buscou conhecer o quanto os bibliotecários adquiriram conhecimento em cada item dos assuntos ministrados considerando-se: o planejamento de gerenciamento de dados, a organização e documentação dos dados, o armazenamento e segurança de dados e a ética e direitos humanos, onde pode-se constatar, em relação ao planejamento de gerenciamento de dados, que três (60%) se consideraram muito satisfeitos e que 40% satisfeitos. Em relação ao item organizando e documentando dados, os resultados demonstraram que 80% se consideram muito satisfeitos e 20% satisfeitos. Por sua vez, o item que teve menor grau de satisfação foi em relação ao armazenamento e segurança de dados, talvez por ser um item que requeira melhor preparo e conhecimento. Os resultados demonstraram que 20% se sentiram muito satisfeitos, enquanto 80% se sentiram satisfeitos. O

último item abordado, ética e direitos humanos obteve 40% de muito satisfeito e 60% em satisfeito. Considera-se, assim, que de maneira geral, a oficina proporcionada satisfaz os bibliotecários da UEFS. O Gráfico 2 demonstra esses resultados.

Gráfico 2 - Grau de conhecimento adquirido nos temas abordados na oficina



Fonte: Google Forms (2019).

O terceiro item buscou conhecer quais foram os pontos mais importantes que os bibliotecários consideraram durante a oficina, sendo mencionadas as atividades práticas que foram utilizadas durante o curso e que foram proveitosas para o entendimento dos temas abordados, assim como o item relacionado à organização e documentação de dados. Também houve destaque para o tópico que envolveu a ética e direitos humanos na organização de dados, além das explicações sobre a relevância do gerenciamento de dados, destacando-se, ainda, a clareza e desenvoltura do ministrante. Isso representa, em suma, que os bibliotecários da UEFS participantes da oficina, destacaram pontos diferentes entre si no julgamento de importância, embora observa-se que foram mencionados todos os tópicos abordados, o que sugere concluir que a temática foi importante e assimilada por todos, com especificidades inerentes ao âmbito de atuação de cada um.

O quarto item analisado em relação à satisfação da oficina recaiu sobre os conteúdos que apresentaram maior dificuldade para os participantes, sendo considerados: armazenamento e segurança de dados, a terminologia atual utilizada para os elementos da tecnologia de informação, a segurança de dados e a compreensão da dimensão sobre o que abarca o conceito de dados. Isso demonstra que a ciência de dados, por ser uma área emergente, ainda não está totalmente integrada nas atividades dos bibliotecários da UEFS, embora estes tenham

assimilado suas diretrizes e atuem com o repositório institucional, fato que possibilita maior entendimento do assunto.

O quinto item da avaliação buscou verificar se os bibliotecários a UEFS conseguiriam aplicar as informações obtidas nas atividades que executavam, tendo como respostas que sim e que parte do aprendizado seria utilizada para o gerenciamento de dados do novo site institucional, para a organização de dados no Repositório Institucional, bem como no trabalho diário.

A sexta parte da avaliação foi destinada aos comentários sobre a oficina, tendo sido considerado que a carga horária poderia ser maior, em função do conteúdo, que os vídeos apresentados poderiam ser mais breves, que se pensasse em práticas para o armazenamento de dados, assim como houve para a nomeação dos arquivos, que a oficina deve ser disseminada para outros participantes em diversos ambientes e, por fim, parabenizaram o ministrante pelo trabalho, muito importante dentro da Ciência da Informação.

## 7 CONSIDERAÇÕES FINAIS

A tecnologia tem inovado as formas de trabalho no mundo, modificadas pelas transformações da ciência ao longo do século. Essas transformações se deram no processo de descoberta, possibilitando e criando relações impulsionadas pelo desenvolvimento da computação. A partir da teoria do quarto paradigma apresenta-se o surgimento e explosão dos dados trazendo uma nova representação da informação, tecnologia e dados para a descoberta do conhecimento. Percebe-se que estas representações ocorrem mediante a evolução de suportes e ressignificação do dado como uma unidade de análise formando assim o Big Data.

Respondendo ao objetivo geral da pesquisa: desenvolver habilidades e competências em alfabetização e em gerenciamento de dados para bibliotecários da Biblioteca Central da Universidade Estadual de Feira de Santana, elaborou-se uma oficina para tratar da temática de dados e foi possível a confecção do guia.

Assim com o intuito de revisar e possibilitar um referencial teórico na temática de ciência de dados, a Seção 2: A evolução dos dados da ciência, discute as relações teóricas e evolutivas dos paradigmas, como processo de evolução da ciência. Nesta mesma seção aborda-se o quarto paradigma e o surgimento do termo e-Science, como tendências e práticas para discussões sobre a nova ciência orientada em dados. Assim, com o intuito de relacionar e conceituar a explosão de dados, a Seção 2.1: Dilúvio de dados e Big Data abordou o surgimento dos dados caracterizando seus diversos conceitos, a fim de esclarecer quais são os dados relacionados ao quarto paradigma, realizando-se um recorte histórico desde o período neorodental até o controle de dados, por computadores, memórias, unidade de disco e outros dispositivos.

Em continuidade, a Seção 2.1.1: Tipos de dados, apresenta os tipos de dados originados por computadores, sendo estes estruturados e não-estruturados, reproduzidos e originados por meio diversos meios eletrônicos e armazenados em diversos suportes; neste capítulo mencionam-se os dados da pesquisa como registro em formato impresso ou digital, que contenham como assunto principal resultados de observações de pesquisa no meio acadêmico. Na Seção 2.1.2: Ciclo de dados e gerenciamento de dados apresentam-se as fases para gerenciamento e armazenamento desses tipos de dados sendo estes fatores importantes para o processo de armazenamento e desenvolvimento de política para uso e desuso; assim apresenta-se o ciclo de vida dos dados onde devem ser tidos como processos e modelos estabelecidos por um princípio de gestão contínua e mediante fatores da tecnologia. Responde-se, assim, ao

objetivo específico: elaborar um guia com materiais que sirvam de embasamento teórico e prático para alfabetização e gerenciamento de dados para bibliotecários.

Com o sentido de abordar o objetivo específico: *levantar as ações promovidas pelos bibliotecários relacionadas à inclusão de novos serviços na biblioteca*, na seção 3: **Relações conceituais entre Ciência de Dados e Ciência da Informação e a seção 3.1: Biblioteconomia e o bibliotecários de dados** discorrem sobre a interdisciplinaridade entre as áreas considerando-se que a Ciência da Informação é um subconjunto da Ciência de Dados, ou seja, deve estar inserida e utilizar-se dos referenciais teóricos para promoção de estudos aos dados sendo a biblioteconomia de dados um ramo para ciência, a fim de abordar o novo profissional: o bibliotecário de dados, que surge a partir da necessidade de reinvenção das bibliotecas e de novas formas de serviços e produtos, como: captura, curadoria, acesso, análise e serviços relacionados aos dados de pesquisa.

Utilizando-se da matriz de competência, na seção: 3.2 *Data literacy* ou **alfabetização de dados para o bibliotecário no Brasil** na seção: 3.3 **Web sites para alfabetização em dados** nortearam, a partir do arcabouço teórico, a elaboração da oficina de alfabetização em dados tendo em vista promover competências básicas e avançadas no tocante à educação em alfabetização de dados, assim respondendo ao objetivo específico: *promover um curso próprio de alfabetização de dados com demonstração prática das ações necessárias em cada etapa*.

Em relação à problematização da pesquisa: **Como os bibliotecários da Biblioteca Central (BICEN) da Universidade Estadual de Feira de Santana realizam o gerenciamento de dados em pesquisa?** apresentam-se na seção 4.2 **procedimentos metodológicos e análise SWOT**, a realização da oficina com informações do site Mantra; referências bibliográficas e a matriz de competências. Após a oficina e com vistas a levantar e perceber o conhecimento adquirido aplicou-se um questionário elaborado no Google Forms, considerando os conteúdos ministrados e o nível de habilidades em dados sendo apresentado no **gráfico 1: Satisfação com a oficina ministrada**, localizada na seção 6.2 **Guia para alfabetização em dados**, nível satisfatório de conhecimento e de aproveitamento absorvidos.

A avaliação da oficina possibilitou a confecção do Guia de Alfabetização de dados para Bibliotecários que, utilizando-se da plataforma Canva desenhou suas seções com base no grau de necessidade de conhecimento dos bibliotecários. Além disso levou em conta também os conteúdos abordados durante a oficina. O guia está estruturado em cinco unidades enfocando temas relacionados à matriz de competências com vistas a possibilitar a estes profissionais a

aquisição de conhecimentos sobre o tema, tanto para auxiliar em suas atividades diárias, como para realizar a promoção de alfabetização de dados nos espaços da biblioteca.

O guia apresenta-se como um produto final desta dissertação e, tendo em vista seu caráter inovador poderá suscitar discussões, tanto teórica como práticas, na temática de alfabetização de dados além de possibilitar ao profissional bibliotecário desenvolver novos serviços e produtos inerentes aos dados.

Nesse sentido, conclui-se que esta dissertação abre caminho para promover na Ciência da Informação novas de discussões e, devido à interdisciplinaridade com a Ciência de Dados, apresenta o Bibliotecário de Dados, que pode atuar eficazmente no ciclo de dados de pesquisas junto à outros profissionais, bem como atuar com treinamentos e capacitações para a alfabetização em dados.

## REFERÊNCIAS

- ALVARO, E., BROOKS, H., HAM, M., POEGEL, S., ROSENCRANS, S. E-science librarianship: Field undefined. **Issues in Science and Technology Librarianship**, v. 66, 2011. Disponível em: <<http://www.istl.org/11-summer/article1.html>>. Acesso em: 20 jan. 2019.
- AMARAL, Fernando. **Introdução à ciência de dados: mineração de dados e big**. Rio de Janeiro: ALTA Books, 2016.
- BARROS, Aidil Jesus da Silveira; LEHFELD, Neide Aparecida de Souza. **Fundamentos de metodologia científica**. 3 ed. São Paulo: Pearson Prentice Hall, 2007.
- BELL, Gordon. Prefácio. In: HEY, Tony; TRANSLEY, Stewart; TOLLE, Kristin (orgs). **O quarto paradigma: descobertas científicas na era da e-Science**. São Paulo, Oficina de Textos, 2011.
- BELL, Gordon; HEY, Tony; SZALAY, Alex. Beyond the data deluge. **Science**, v. 323, n. 5919, p. 1297-1298, 2009.
- BERTIN, P. R. B.; VISOLI, M. C.; DRUCKER, D. P. A gestão de dados de pesquisa no contexto da e-science: benefícios, desafios e oportunidades para organizações de p&d. **Ponto de Acesso**, v. 11, n. 2, p. 34-48, 2017. Disponível em: <<https://portalseer.ufba.br/index.php/revistaici/article/view/21449>>. Acesso em: 03 jan. 2019.
- BOYD, Danah; CRAWFORD, Kate. “Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon”. In: **Information, Communication & Society** 15, no. 5: 662–7, 2012.
- BUGNION, P.; MANIVANNAN, A.; NICOLAS, P.R. **Scala: Guide for Data Science Professionals**. Birmingham: Packt Publishing, 2017, 1077 p.
- CHEN, Min; MAO, Shiwen; LIU, Yunhao. Big Data: A survey. **Mobile networks and applications**, v. 19, n. 2, p. 171-209, 2014.
- CHIAVENATO, Idalberto; SAPIRO, Arão. **Planejamento Estratégico: fundamentos e aplicações**. 1. ed. 13° tiragem. Rio de Janeiro: Elsevier, 2003.
- CHRISTENSEN-DALSGAARD, Birte et al. Ten recommendations for libraries to get started with research data management. **Final report of the LIBER working group on E-Science/Research Data Management**. Retrieved January, v. 30, p. 2014, 2012. Disponível em: <https://libereurope.eu/wpcontent/uploads/The%20research%20data%20group%202012%20v7%20final.pdf>. Acesso em: 19 jan. 2019.
- COBRA, Marcos. **Consultoria em Marketing Manual do Consultor**. 1. ed. São Paulo: Cobra Editora e Marketing, 2003.
- CORDEIRO, D.; BRAGHETTO, K.; GOLDMAN, A.; KON, F. Da ciência à e-ciência: paradigmas da descoberta do conhecimento. **Revista USP**, n. 97, p. 71-81, 30 maio 2013.

Disponível em: < <http://www.revistas.usp.br/revusp/article/view/61867> > Acesso em: 19 de jan. 2019.

DAVENPORT, T. H. **Big Data at Work**: dispelling the myths, uncovering the opportunities. Boston: Harvard Business Review Press, 2014.

DEMO, Pedro. **Praticar ciência**: metodologias do conhecimento científico. 1. ed. São Paulo: Saraiva, 2011. 208 p.

DIEHL, Astor Antônio. **Pesquisa em ciências sociais aplicadas**: métodos e técnicas. São Paulo: Prentice Hall, 2004.

DILTHEY, Wilhelm. **Introdução às ciências humanas** – tentativa de uma fundamentação para o estudo da sociedade e da história. Trad. de Marco Antônio Casanova. Rio de Janeiro: Forense Universitária, 2001

Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878>. > Acesso em: 19 jan. 2019.

FEDERER, Lisa “**Defining Data Librarianship: A Survey of Competencies, Skills, and Training**”. *Journal of Medical Librarianship (JMLA)*. Vol 106, No 3 (2018)

FENTANES, E.G. **A tarefa da ciência experimental**: um guia prático para pesquisar e informar resultados nas ciências naturais. Rio de Janeiro: LTC, 2014. 187p.

FERREIRA Valdinéia Barreto. **e-Science e políticas públicas para ciência, tecnologia e inovação no brasil**: colaboração, infraestrutura e repercussão nos institutos nacionais de ciência e tecnologia da área de nanotecnologia. Salvador, 2016. 400f. Tese (Doutorado) – Programa de Pós-Graduação em Ciência da Informação, Universidade Federal da Bahia, Salvador, 2016. Disponível em: <<https://repositorio.ufba.br/.../Valdineia%20Barreto%20Ferreira%20-%20Tese%20de%20...>>. Acesso em: 24 junho. 2019.

FERREIRA, Jaider Andrade; SANTOS, Plácida L. V. A. da Costa. O modelo de dados Resource Description Framework (RDF) e o seu papel na descrição de recursos. **Informação & Sociedade: estudos**, v. 23, n. 2, 2013. Disponível em: <<http://www.brapci.inf.br/index.php/res/v/92565>> Acesso em: 21 jan. 2019.

FERRELL, O. C.; HERTLINE, Michael D.. **Estratégia de Marketing**. Tradução All Tasks e Marlene Cohen. 4. ed. São Paulo: Cengage Learning, 2009.

FINZER, W. The data science education dilemma. **Technology Innovations in Statistics Education, California**, v. 7, n.2, 2013. Disponível em: <<https://escholarship.org/uc/item/7gv0q9dc> >. Acesso em: 22 dezembro. 2018

FOUREZ, Gérard. **A construção das ciências**: introdução à filosofia e à ética das ciências. São Paulo: UNESP, 1995. 319 p.

FRICKÉ, Martin. Big Data and its Epistemology. **Journal Association for Information Science and Technology**, V. 66, Issue 4. 615-661, 2014. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/asi.23212>>. Acesso em: 10 jan.2018.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6. Ed. – São Paulo: Atlas, 2008.

GLEICK, J. **The information: A history, a theory, a flood**. New York, NY, US: Pantheon Books, 2011, p. 535.

GRAY, Jim. Jim Gray on eScience: A Transformed Scientific Method. Based on the transcript of a talk given by Jim Gray to the NRC-CSTB1 in Mountain View, CA, on January 11, 2007. In: HEY, Tony; TRANSLEY, Stewart; TOLLE, Kristin (orgs). **The Fourth Paradigm. Data-Intensive Scientific Discovery**. Redmond, WA: Microsoft Research, 2009. 284pp.

GUIZZO, Erico Marui. **The Essential Message: Claude Shannon and the Making of Information Theory**. [on line] Doctoral dissertation. (Master of Science in Science writing at the Massachusetts Institute of Technology), 2003. Disponível em: <<https://dspace.mit.edu/handle/1721.1/39429>>. Acesso em: 21 jan. 2019.

HEY ; Tony, HEY, Jessie "e-Science and its implications for the library community", **Library Hi Tech**, Vol. 24 Issue: 4, 2016, pp.515-528. Disponível em: <<https://doi-org.ez20.periodicos.capes.gov.br/10.1108/07378830610715383>>. Acesso em: 28 dez. 2018.

HJØRLAND, Birger.. “Data (with dig data and database semantics)”. **Knowledge Organization** 45, no. 8: 685-708. Also available in *ISKO Encyclopedia of Knowledge Organization*, ed. Birger Hjørland, coed. Claudio Gnol, 2018, Disponível em: <<http://www.isko.org/cyclo/data>>. Acesso em: 20 jan. 2019

HOFFMANN, L. Q&A Gray’s Paradigm. **Communications of the ACM**, V. v. 53, n. 10, p. 112–111, 2010. Disponível em: <<https://dl.acm.org/citation.cfm?id=1831407>>. Acesso em: 10 jan. 2019.

INMON, W. H.; LINSTEDT. **Data Architecture: a primer for the data scientist- Big Data, Data Warehouse and Data Vault**. USA: Elsevier, 2015.

INTERAGENCY WORKING GROUP ON DIGITAL DATA. **Harnessing the Power of Digital Data for Science and Society**, 2009. Disponível em: <[https://www.nitrd.gov/about/harnessing\\_power\\_web.pdf](https://www.nitrd.gov/about/harnessing_power_web.pdf)>. Acesso em: 19 jan. 2019

KELLEHER, John D.; TIERNEY, Brendan. **Data Science**. Cambridge, MA: The MIT Press, 2018, p.282.

KOLTAY, Tibor. Data literacy for researchers and data librarians. **Journal of Librarianship and Information Science**, v. 49, n. 1, 2017. Disponível em: <<http://journals.sagepub.com/doi/abs/10.1177/0961000615616450>>. Acesso em: 15 dez. 2018.

KUHN, T. S. **A Estrutura das Revoluções Científicas**. 9. ed. São Paulo: Perspectiva, 2005.

LAKATOS, Eva Maria; MARCONI, Marina de Andrade. Fundamentos de metodologia científica. 4. ed. rev. e ampl. São Paulo: Atlas, 2001.

MAATTA,S. Placements & salaries 2013: **The emerging databrarian. Library Journal.** Disponível em: < <http://lj.libraryjournal.com/2013/10/placements-and-salaries/2013survey/theemerging-databrarian/#> >

MANOVICH, Lev. Data Science and Digital Art History. **International Journal for Digital Art History**, [S.l.], n. 1, june 2015. ISSN 2363-5401. Disponível em: <<https://journals.ub.uni-heidelberg.de/index.php/dah/article/view/21631/15404>>. Acesso em: 15 dez. 2018.

MARCHIONINI, Gary. Information Science Roles in the Emerging Field of Data Science. **Journal of Data and Information Science**. Vol. 1 N2, 2016 pp 1-6. Disponível em: <[http://manu47.magtech.com.cn/Jwk3\\_jdis/Y2016/V1/I2/1](http://manu47.magtech.com.cn/Jwk3_jdis/Y2016/V1/I2/1)> Acesso em: 15 dez. 2018.

MARTÍNEZ-URIBE, Luis, MACDONALD, Stuart. Un nuevo cometido para los bibliotecarios académicos: data curation. **El profesional de la información**, vol. 17, n. 3, 2008, 273-280 p. Disponível em: <<http://www.elprofesionaldeinformacion.com/contenidos/2008/mayo/03.pdf>. > Acesso em: 15 dez. 2018.

MAYER-SCHÖNBERGER Viktor, CUKIER, K. **Big Data: A Revolution That Will Transform How We Live, Work, and Think**, Boston, M: Houghton Mifflin Harcourt, 2013.  
MIRANDA, Luis Felipe Sigwait de. **Introdução histórica à filosofia das ciências**. Curitiba: InterSaberes, 2016.

MONTEIRO, Elizabete Cristina de Souza de Aguiar; SANT'ANA, Ricardo Cesar Gonçalves; SANTAREM SEGUNDO, José Eduardo. e-Science Semântica: integração dos dados na comunicação científica. **Informação em Pauta**, Fortaleza, v. 1, n. 1, p. 9-29, june 2016. ISSN 2525-3468. Disponível em: <<http://periodicos.ufc.br/informacaoempauta/article/view/2942>>. Acesso em: 20 jan. 2019.

MORAIS, Regis de. **Filosofia da ciência e da tecnologia: introdução metodológica e crítica**. Campinas, SP: Papirus, 2013.

PENTLAND, A.S. (2013). The data----driven society. *Scientific American*, 309(4), 78----83. Retrieved from Peter, Katharin and Kellam, Lynda. “Incorporating Statistical Literacy into Information Literacy Instruction.” **LOEX Quarterly**, 40: 1 (2013), 2-3, 10, 2013. Disponível em: <[http://www.ala.org/acrl/publications/keeping\\_up\\_with/statistical\\_literacy](http://www.ala.org/acrl/publications/keeping_up_with/statistical_literacy)> Acesso em: 20 jan. 2019.

POPPER. Karl Raiumund. **A Lógica da pesquisa científica**. 8 ed. São Paulo: Cultrix, 2007. 567 p.

PRADO, J., MARZAL, M. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. **Libri**, 63(2), 123–134. Disponível em: <<https://www.degruyter.com/view/j/libr.2013.63.issue-2/libri-2013-0010/libri-2013-0010.xml>> . Acesso em: 15 jan. 2019.

RICE, R.; SOUTHALL, S. **The data librarian’s handbook**. London: Facet Publishing, 2016. Ridsdale, Chantel; ROTHWELL, James; SMIT, Mike; HOSSAM, Ali-Hassan, BLIEMEL, Michael; IRVINE Dean, KELLEY, Daniel; MATWIN ;Stan; WUETHERICK. Brad

**Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report**, Dal housie University, Halifax, NS, 2015. Disponível em: <[http://www.mikesmit.com/wp-content/papercite-data/pdf/data\\_literacy.pdf](http://www.mikesmit.com/wp-content/papercite-data/pdf/data_literacy.pdf). > Acesso em: 15 jan. 2019.

SAGIROGLU, S; SINANC, D. Big Data: A review. In: Collaboration Technologies and Systems (CTS), 2013 **International Conference on. IEEE**, 2013. P. 42-47. Disponível em: <[https://www.scirp.org/\(S\(351jmbntvnsjt1aadkposzje\)\)/reference/ReferencesPapers.aspx?ReferenceID=2163642](https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/ReferencesPapers.aspx?ReferenceID=2163642)> Acesso em: 15 jan. 2019

SCHROEDER, Ralph; FRY, Jenny. Social Science Approaches to e-Science: Framing and Agenda. **Journal of Computer- Mediated Communication**, V. 12, Issue 2, p. 563-582, Jan. 2007. Disponível em: <<https://academic.oup.com/jcmc/article/12/2/563/4583047>>. Acesso em: 20 jan. 2019.

SETZER, V. W. *Dado*, informação, conhecimento e competência. **DataGramZero**, v. 0, n. 0, p. A01, 1999. Disponível em: <<http://www.brapci.inf.br/index.php/article/download/14562>> .Acesso em: 27 jan. 2019.

SMITH, F. Jack Data Science as an academic discipline. **Data Science Journal** , v.5, (2006) p. 163–164. Disponível em: <<https://datascience.codata.org/articles/abstract/10.2481/dsj.5.163/>> Acesso em: 27 jan. 2019.

SWAN, A.; BROWN, S. The skills, role and care restructure of data scientists and curators: an assessment of current practice and future needs. **Report to the Joint Information Systems Committee (JISC)**. Truro: Key Perspectives for JISC, 2008. p. 34 Disponível em: <[https://eprints.soton.ac.uk/266675/1/Data\\_skills\\_report\\_final\\_draft.doc](https://eprints.soton.ac.uk/266675/1/Data_skills_report_final_draft.doc)> Acesso em: 27 jan. 2019.

VAN DER AALST, W. M. P. Data scientist: the engineer of the future. In: Enterprise Interoperability VI: interoperability for agility, resilience and plasticity of collaborations. Springer: New York, 2014. Disponível em: <<https://research.tue.nl/en/publications/data-scientist-the-engineer-of-the-future>> Acesso em: 28 dez. 2018.

VELHO, Léa. Conceitos de Ciência e a Política Científica, Tecnológica e de Inovação. **Sociologias**, Porto Alegre, v. 13, n. 26, p. 128-153, 2011.

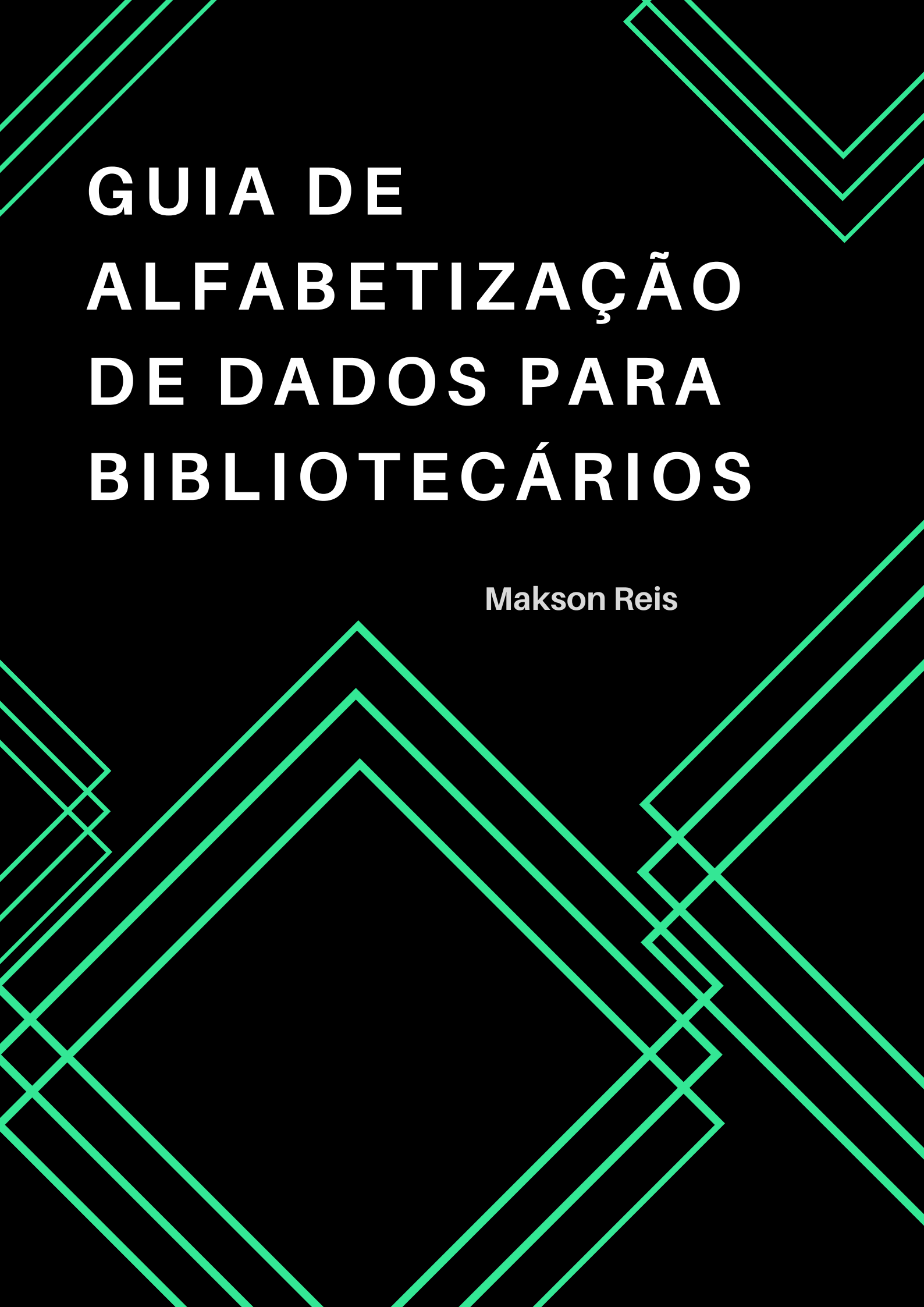
XIA, Jingfeng; WANG, Minglu. Competencies and responsibilities of Social Science Data Librarians: an analysis of job dDescriptions. **College & Research Libraries**, [S.l.], v. 75, n. 3, p. 362-388, may 2014. Disponível em: <<https://crl.acrl.org/index.php/crl/article/view/16367>>. Acesso em: 24 jan. 2019.

YANG X, WANG L, LASZEWSKI, G. Recent research advances in e-science. **Cluster Computing**, v.12, n. 4, 2009, p. 353–356. Disponível em: <<https://dl.acm.org/citation.cfm?id=1666191.1666210>. > Acesso em: 10 jan. 2019.

ZHU, Y. and XIONG, Y. Towards data science. **Data Science Journal**, v.14, 2015 p.8. Disponível em: <<https://datascience.codata.org/article/10.5334/dsj-2015-008/>> Acesso em: 27 jan. 2019.

ZHU, Y.; XIONG, Y. Towards Data Science. **Data Science Journal**, 14, p. 8, 2015.  
Disponível em:<<http://doi.org/10.5334/dsj-2015-008>>. Acesso em: 06 jan. 2019.

## **APÊNDICE A – GUIA DE ALFABETIZAÇÃO DIGITAL**



# GUIA DE ALFABETIZAÇÃO DE DADOS PARA BIBLIOTECÁRIOS

Makson Reis



**GUIA DE ALFABETIZAÇÃO DE DADOS PARA BIBLIOTECÁRIOS**

**Makson Reis**



Elaboração: Makson de Jesus Reis

Orientação: Profa Dra. Telma de Carvalho

Projeto oriundo da Dissertação: "Ciência de Dados e Ciência da Informação: Guia de alfabetização de dados para bibliotecários."

Programa de Pós Graduação em Ciência da Informação

Mestrado Profissional em Gestão da Informação e do Conhecimento

Universidade Federal de Sergipe

Todos os Direitos reservados aos autores.



Este trabalho está licenciado sob a Licença Creative Commons  
Atribuição 3.0 Não Adaptada.

Para ver uma cópia desta licença, visite

<http://creativecommons.org/licenses/by/3.0/> ou envie uma carta para  
Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

## ELABORAÇÃO

Makson Reis

## ORIENTAÇÃO

Dra. Telma de Carvalho

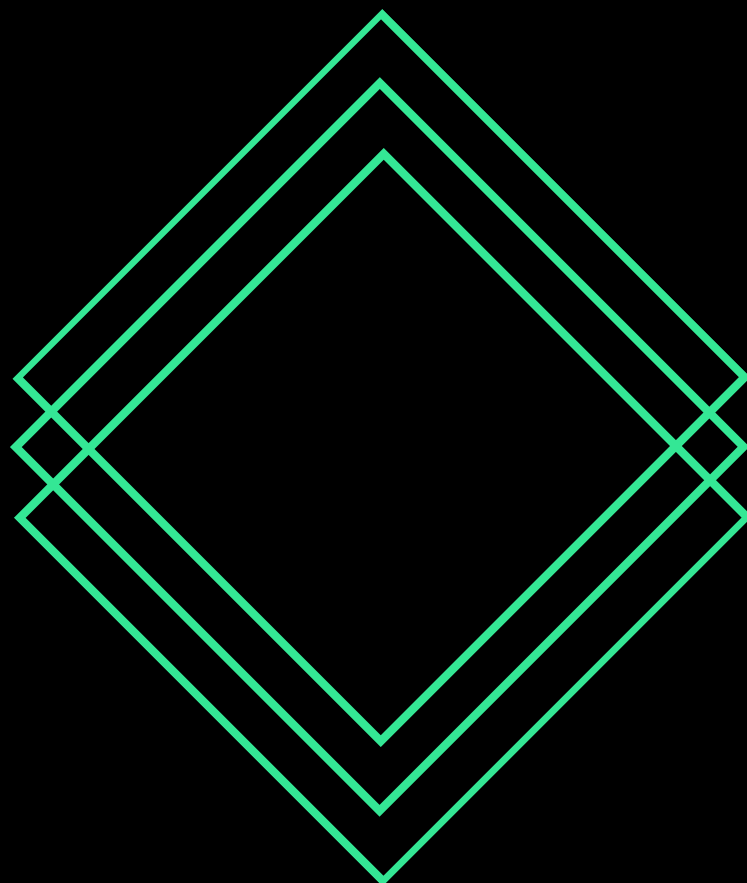
## DIAGRAMAÇÃO

Makson Reis

## Contatos:

maksonacademico@gmail.com

carvalhotel@gmail.com



### Dados Internacionais de Catalogação na Publicação (CIP)

	Reis, Makson de Jesus
R375g	<p>Guia de Alfabetização de Dados para Bibliotecários / Makson de Jesus Reis. – São Cristóvão-SE, 2019. 62 p. il. color</p> <p>Orientadora: Dra. Telma de Carvalho. Guia apresentado como produto da Dissertação intitulada: Ciência de Dados e Ciência da Informação: guia de alfabetização de dados para bibliotecários (Mestrado Profissional em Gestão da Informação e do Conhecimento) – Universidade Federal de Sergipe - UFS, Programa de Pós-Graduação em Gestão da Informação e do Conhecimento, 2019.</p> <p>1. Ciência de dados 2. Ciência da Informação. 3. Ciência aberta. I. Carvalho, Telma de, orienta. II. Título.</p> <p>CDU: 030(031) CDD: 361.02</p>
Bibliotecário responsável: Makson Reis - CRB-5ª Região/1926	

# SUMÁRIO

APRESENTAÇÃO	6
PREFÁCIO	8
A NOVA CIÊNCIA BASEADA EM DADOS DA PESQUISA	10
DADOS DA PESQUISA E TIPOLOGIA	16
PLANEJANDO E GERENCIANDO DADOS DE PESQUISA	24
ORGANIZANDO E DOCUMENTANDO DADOS DE PESQUISA	38
ARMAZENAMENTO, COMPARTILHAMENTO E SEGURANÇA	43
ÉTICA E DIREITOS AUTORAIS	49
ANEXO	54
REFERÊNCIAS	57

# APRESENTAÇÃO



6

O propósito deste guia é ser um meio de divulgação e de promoção de informações no âmbito dos dados.

Trata-se de um produto desenvolvido a partir da relação entre a ciência da informação e a ciência de dados oriundo do Mestrado Profissional em Gestão da Informação e do Conhecimento da Universidade Federal de Sergipe (PPGCi/UFS).

Diante dos estudos analisados e da mineração de informações realizadas sobre a temática, deparei-me com autores que difundem o assunto no tocante à capacitação e ao gerenciamento de dados para bibliotecários. Isso interessou-me muito, pois precisamos que os bibliotecários dominem esse aspecto da informação.

# APRESENTAÇÃO

Foram utilizados para construção das atividades deste guia o site Mantra Data Librarian com adaptação do seu Kit de Treinamento para Bibliotecários de Dados, a fim de serem incluídas algumas partes neste produto possibilitando ampla disseminação sobre a alfabetização de dados.

Por fim, agradeço de forma carinhosa e singela, aos bibliotecários (as) e aos colaboradores da Biblioteca Central Julieta Carteadó, da Universidade Estadual do Município de Feira de Santana BA, que auxiliaram para o andamento e aplicação da pesquisa, sendo acolhedores e participativos.

E a minha orientadora Telma de Carvalho, que acolheu, acreditou e sonhou com o projeto.

"Os dados eram enigmas do futuro, hoje são estudos do presente."

(MAKSON REIS, 2019)

# PREFÁCIO



8

DRA. TELMA DE CARVALHO,  
UNIVERSIDADE FEDERAL DE SERGIPE

É com grande alegria que elaboro o prefácio deste guia, produto final da Dissertação de Mestrado do aluno Makson de Jesus Reis, do PPGCI/UFS.

Desde nossos primeiros contatos percebia-se no aluno uma ânsia em trabalhar com a gestão de dados, no âmbito da ciência aberta, especialmente voltada para a capacitação de equipe bibliotecária neste tema tão atual e importante.

Assim, após várias leituras e desenvolvimento do trabalho, optou-se por utilizar a Matriz de Competências de Ridsdale e Chantel et al. (2015) para a formulação do Guia de Alfabetização de Dados, utilizando-se a plataforma de edição CANVA para a sua elaboração.

## PREFÁCIO POR

.....  
DRA. TELMA DE CARVALHO,  
UNIVERSIDADE FEDERAL DE SERGIPE

No decorrer da elaboração da dissertação ocorreu uma oficina voltada para os bibliotecários da Biblioteca Central da Universidade Estadual de Feira de Santana, que culminou na possibilidade de construção deste guia, como apoio aos ensinamentos ministrados no curso e adquiridos com os autores utilizados para a construção desta pesquisa.

Esperamos que sua repercussão seja promissora para que tenhamos a certeza de que o guia cumpriu o seu papel.

Boa leitura e bons estudos para todos que se interessam pela temática!

A NOVA CIÊNCIA  
BASEADA EM DADOS  
DE PESQUISA

Gray (2009) destacou que as pesquisas científicas e a ciência serão transformadas pelo grande volume de dados tendo como novo método de formulação de hipóteses e análise dos resultados: a coleta, manipulação e exibição dos dados.

Este novo tipo de ciência baseada em dados (E-science), possibilita estudos que viabilizam novos desafios científicos na utilização de tecnologia para pesquisas baseadas em dados. Conforme Gray (2009), a ciência vem sendo afetada pela grande quantidade de dados coletados, gerados ou simulados em pesquisas ou laboratórios.

Assim surge o termo Big Data como um conjunto de dados em grande escala sendo sua composição por tipos de dados: em formatos de textos, imagens e áudio, ou por sua velocidade: processamento, armazenamento e criação desses dados (HJØRLAND, 2018).

Conforme Boyd e Crawford (2012), o big data tem se caracterizado como uma ação cultural, tecnológica e acadêmica, por maximizar o poder dos computadores para coletar, analisar e armazenar conjuntos de dados na identificação de padrões, a fim de possibilitar acessos econômico, social e técnico para fornecerem um conjunto de conhecimentos e informações sobre a pesquisa.

## ***E-SCIENCE***

Realizado por simulações de fenômenos muito grandes ou pequenos, rápidos ou lentos ou muito complexos para serem explorados em um laboratório convencional;  
um conglomerado (estático ou orgânico) ou uma coleção de conjuntos de dados menores (revisados por pares), provavelmente publicados e curados, por exemplo: bancos de dados de genes, bancos de dados de cristalografia

## ***PESQUISA ORIENTADA POR DADOS***

Pesquisa na qual a análise de grandes quantidades de dados experimentais é uma etapa do projeto geral de pesquisa. Ainda mais distinguível por: dados capturados por instrumentos digitais de alta produtividade e dispositivos de gravação; dados que requerem envolvimento humano em algum momento na captura ou preparação de dados para análise.

## **PESQUISA INTENSIVA EM DADOS**



Uma ampla gama de atividades acadêmicas realizadas por pesquisadores e acadêmicos em um amplo leque de disciplinas. A pesquisa com uso intensivo de dados também revela novas áreas para bolsas de estudo e pesquisa adicional, mais conhecidas por: conjuntos de dados híbridos baseados na Internet que consistem em dados nascidos digitais e dados resultantes da digitalização de grandes coleções e informações analógicas; dados associados ao aprimoramento digital e/ou representação digital de entidades e espaços físicos, por exemplo, artefatos culturais, manuscritos antigos; explorações digitalizadas de coleções de bibliotecas e museus.

Neste contexto, surge a necessidade de alfabetização em dados, para possibilitar competências e habilidades nos bibliotecários. Segundo Koltay (2007), as tecnologias viabilizaram o processo de ampliação e difusão da informação em dados, através do seu processo de organização, busca, recuperação e preservação.

Com base na matriz de competências em Alfabetização de Dados (RIDSDALE; CHANTEL et al., 2015), este guia apresentará os seguintes passos para sua elaboração:

- UNIDADE 1: Dados da Pesquisa e Tipologia
- UNIDADE 2: Planejando e Gerenciando Dados
- UNIDADE 3: Organizando e Documentando Dados
- UNIDADE 4: Armazenamento, Compartilhamentos e Segurança de Dados
- UNIDADE 5: Ética e Direitos Autorais

# Matriz de Competências de Alfabetização de dados

15

Legenda:

Competências Conceituais	Competências Essenciais	Competências Avançadas
--------------------------	-------------------------	------------------------

Área do Conhecimento	Competência	Conhecimento /Tarefas
----------------------	-------------	-----------------------

Estrutura Conceitual	Introdução nos Dados	Conhecimento e compreensão dos dados	Conhecimento e compreensão dos usos e aplicações de dados			
Coleta de dados	Descoberta de dados e Coleção	Realiza a exploração de dados	Identifica dados úteis	Coleta de dados		
	Avaliando e assegurando dados e fontes	Avalia criticamente fontes de dados para Confiabilidade	Avalia criticamente a qualidade de conjuntos de Dados para erros ou problemas			
Gerenciamento de dados	Organização de dados	Conhecimento da organização básica de dados métodos e ferramentas	Acesso e requisitos de organização de dados	Organizar dados		
	Manipulação de dados	Acesso e métodos para limpar dados	Identifica os valores e as anomalias	Limpa dados		
	Conversão de dados (de formato para formatar)	Conhecimento de diferentes tipos de dados e métodos de conversão	Atribui os descritores de metadados apropriados para conjuntos de dados originais			
	Criação e uso de metadados	Cria descritores de metadados	Converter dados de um formato ou tipo de arquivo para outro			
	Curadoria de Dados	Avalia os requisitos de Curadoria de dados (por exemplo, cronograma de retenção, armazenamento, acessibilidade, requisitos de partilha, etc.)	Avalia os requisitos de segurança de dados (por exemplo, acesso restrito, unidades protegidas, etc.)	Curadoria de dados		
	Preservação de dados	Avalia os requisitos de preservação	Métodos e ferramentas de acesso para preservação de dados	Preserva os dados		
Ferramentas de dados	Conhecimento de ferramentas de análise de dados e técnicas	Seleciona a ferramenta de análise de dados apropriada ou Técnica	Aplica ferramentas de análise de dados e técnica			
	Análise básica de dados	Desenvolve planos de análise	Aplica métodos e ferramentas de análise	Realiza análise exploratória	Avalia os resultados da análise	Compara os resultados da análise com outros achados
	Interpretação de dados (Noções básica sobre dados)	Lê e compreende gráficos, tabelas	Identifica os principais pontos e integra isso com outras informações importantes	Identifica discordância dentro os dados		
Avaliação de dados	Identificando problemas usando dados	Usa dados para identificar os problemas em situações práticas (por exemplo eficiência no local de trabalho)	Usa dados para identificar problemas de nível superior (por exemplo, política, ambiente, experimentação científica, marketing, economia)			
	Visualização de dados	Cria tabelas significativas para organizar e apresentar visualmente dados	Cria representações gráficas significativas de dados	Avalia eficácia de representações gráficas	Avalia criticamente representações gráficas para precisão e deturpação de dados	
	Apresentando dados (verbalmente)	O resultado pretendido (s) para apresentar os dados	Avalia as necessidades do público e a familiaridade com sujeito (s)	Planeja a reunião apropriada ou tipo de apresentação	utiliza tabelas significativas e visualizações para comunicar dados	Apresenta argumentos e/ou resultados claros e coerentemente
	Decisões orientada a dados Tomada de decisões em dados	Prioriza a informação obtida a partir de dados	Converter dados em informações	Pesa o mérito e os impactos de possíveis soluções/decisões	Implementa decisões/soluções	

DADOS DA PESQUISA  
E  
TIPOLOGIA

# UNIDADE 1

Esta Unidade aborda como se constituem os dados da pesquisa e quais são os seus formatos.

Depois de completar esta unidade, você irá:

- Conceituar os dados de pesquisa;
- Ser capaz de conhecer os dados de pesquisa;
- Classificar os tipos de dados de pesquisa e seus formatos.

# O QUE SÃO DADOS DE PESQUISA?



18

Conforme o site Mantra (2013) os dados de pesquisa são coletados, observados ou criados para fins de análise, a fim de produzirem e validarem os resultados originais das pesquisas.

Os dados são vistos onde a informação e o conhecimento da pesquisa possam ser extraídos.

Conforme Setzer (1999), os dados são símbolos quantificados, portanto, podem ser representados por números, categorias, textos físicos ou digitais, imagens ou áudio. São como matéria-prima na tomada de decisão e informação.

Entretanto, os dados de palavras significam coisas diferentes para diferentes pessoas em diferentes contextos. Diferentes disciplinas têm e usam linguagem específica em torno dos dados da pesquisa.

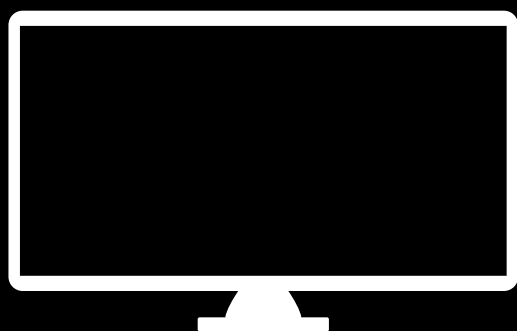
Algumas pessoas se referem a tudo que é digital como dados; outros a materiais analógicos e digitais. Para os propósitos deste módulo, os dados de pesquisa devem ser considerados como aqueles criados em uma forma digital (nascidos digital) ou convertidos em um formulário digital (digitalizado).

Assim, os dados de pesquisa referem-se a quando eles são usados, o que constituem e o propósito para o qual devem ser usados. Os dados também podem ser criados para uma finalidade e usados, em uma data posterior, para uma agenda de pesquisa completamente diferente. (MANTRA,2013; SETZER 1999)

## EXEMPLOS:



Uma imagem fotográfica de um antigo prédio municipal em um arquivo histórico é uma imagem arquivada em um banco de imagens, mas quando usada por um pesquisador para estudar a história de uma cidade, a imagem fotográfica se torna um dado para aquele pesquisador. (MANTRA,2013,tradução nossa)



Imagens de TV podem ser arquivadas (ou destruídas) por uma empresa de segurança. No entanto, quando usado por um pesquisador para estudar o comportamento humano ou os métodos de vigilância do século XXI, as imagens de vídeo se tornam dados para esse pesquisador. (MANTRA,2013,tradução nossa)

# O QUE SÃO DADOS DE PESQUISA?

20

## **OBSERVACIONAL**

Dados capturados em tempo real, geralmente únicos e insubstituíveis, por exemplo, imagens do cérebro

## **EXPERIMENTAL**

Dados de resultados experimentais, por exemplo, de equipamentos de laboratório, muitas vezes reproduzíveis.

## **SIMULAÇÃO**

Dados gerados a partir de modelos de teste em que o modelo e os metadados podem ser mais importantes do que os dados de saída, por exemplo, modelos econômicos ou climáticos

## **DERIVADO OU COMPILADO**

Resultante do processamento ou combinação de dados "brutos", geralmente reproduzíveis, mas caros, por exemplo, bancos de dados compilados, mineração de texto, dados censitários agregados.

## **CANÔNICOS**

um conglomerado (estático ou orgânico) ou uma coleção de conjuntos de dados menores (revisados por pares), provavelmente publicados e curados, por exemplo, bancos de dados de genes, bancos de dados de cristalografia

Os dados de pesquisa vêm em formato de arquivos físicos e tipos de conteúdo variados, alguns dos quais estão listados aqui:

## **NÚMERICOS**

SPSS, Stata , MS Excel, SAS. Arquivos simples: arquivos de formato de campo fixo, arquivos delimitados e arquivos hierárquicos.

## **MULTÍMÍDIA**

JPEG, TIFF, GIF, Dicom, MPEG, Quicktime, Bitmap e PNG.

## **TEXTO**

Arquivos de texto simples como: EMBL, MS Word, Portable Document Format (PDF), Rich Text Format (RTF), Hyper-Text Markup Language (HTML) e Extensible Markup Language (XM).

## **SOFTWARE**

Arquivos binários e de código escritos em uma variedade de linguagens de programação/script: Java, C, Perl, Python, Rubi e PHP

## **ESPECÍFICOS**

Sistema flexível de transporte de imagens (FITS), astronomia. Arquivo de Informação Cristalográfica (CIF), química Binary GRIdded (GRIB), meteorologia

## **INSTRUMENTOS**

Formato de dados do microscópio confocal da Olympus. Formato de Imagem Microscópica ou de laboratório.

## OS DADOS DE PESQUISA (TRADICIONAL E ELETRÔNICA) PODEM INCLUIR OS SEGUINTE OBJETOS:

- Documentos (texto, Word), planilhas;
- Cadernos de laboratório, cadernos de campo, agendas;
- Questionários, transcrições, livros de códigos;
- Audiocassetes, fitas de vídeo;
- Fotografias, filmes;
- Respostas de teste;
- Slides, artefatos, amostras;
- Coleta de objetos digitais adquiridos e gerados durante o processo de pesquisa
- Arquivos estatísticos ou outros arquivos de dados;
- Conteúdos da base de dados (vídeo, áudio, texto, imagens);
- Modelos, algoritmos, scripts;
- Conteúdo de uma aplicação (entrada, saída, arquivos de log para software de análise, software de simulação, esquemas);
- Metodologias e Fluxos de Trabalho e
- Procedimentos operacionais e protocolos padrão

# ATIVIDADE

## UNIDADE 1

23

VOCÊ FOI APRESENTADO AO CONCEITO DE DADOS DE PESQUISA, O QUE CONSTITUI DADOS DE PESQUISA E COMO ELES DIFEREM DE OUTROS TIPOS DE INFORMAÇÕES.

COMO TAREFA DE RESUMO, ARRASTE E SOLTE AS PALAVRAS NA PARTE INFERIOR DA PÁGINA NAS LACUNAS CINZAS DEIXADAS NO TEXTO.

**DICA:** SE A PALAVRA QUE VOCÊ SELECIONAR NÃO FOR A PALAVRA CORRETA, VOCÊ NÃO PODERÁ SOLTÁ-LA NA LACUNA

1) Ao contrário de outros tipos de informação, os dados de pesquisa são coletados, observados ou criados, para fins de análise, para produzir e [ ] resultados originais da pesquisa.

2) Dados de pesquisa também podem ser considerados [ ] na medida em que as mesmas informações ou materiais digitais podem ser dados de pesquisa para algumas questões de pesquisa, mas não para outras.

validar

importante

3) Os dados da pesquisa podem ser classificados da seguinte forma:

- \_\_\_\_\_: dados capturados em tempo real, geralmente exclusivos.
- \_\_\_\_\_: dados de resultados experimentais, por exemplo, de equipamentos de laboratório, muitas vezes reproduzíveis, mas podem ser caros.
- \_\_\_\_\_: dados gerados para reproduzir o comportamento do sistema por modelos de teste.
- \_\_\_\_\_: dados resultantes do processamento de dados, geralmente em um nível mais refinado de granularidade.
- \_\_\_\_\_: uma coleção de conjuntos de dados menores (revisados por pares), mais provavelmente publicados e com curadoria.

Simulação - Derivado - Canônicos- Experimental - Observacional

**PLANEJANDO  
E GERENCIANDO  
DADOS DE PESQUISA**

# UNIDADE 2

Esta Unidade apresenta os conceitos de gerenciamentos de dados e de planos de gerenciamentos de dados

Depois de completar esta unidade, você irá:

- Compreender as regras gerais de gestão de dados;
- Estar ciente de boa prática no gerenciamento de dados de pesquisa em geral e aplicar em seu próprio contexto de trabalho;
- Usar as informações apresentadas na unidade para elaborar um plano de gerenciamento de dados e mantê-lo durante o curso de sua pesquisa.

# PLANEJANDO E GERENCIANDO DADOS DE PESQUISA

26

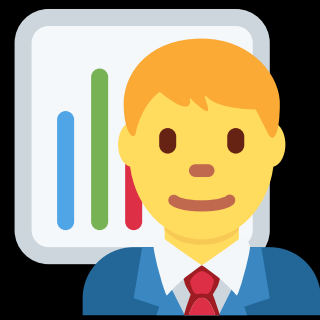
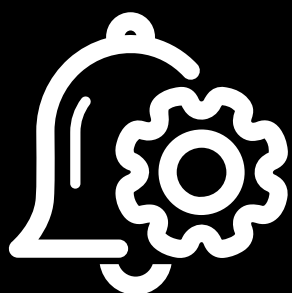
O gerenciamento de dados se refere a todos os aspectos da criação, hospedagem, entrega, manutenção, arquivamento e preservação de dados. É uma das áreas essenciais para uma conduta responsável na pesquisa (AMARAL, 2016).

Conforme o site Mantra (2013), cada vez mais as universidades, agora, encorajam todos os pesquisadores (inclusive estudantes de pós-graduação) e bibliotecários a realizarem o planejamento do gerenciamento de dados (PGD) no início de seu projeto de pesquisa ou na inserção em bases de dados ou repositórios institucionais.

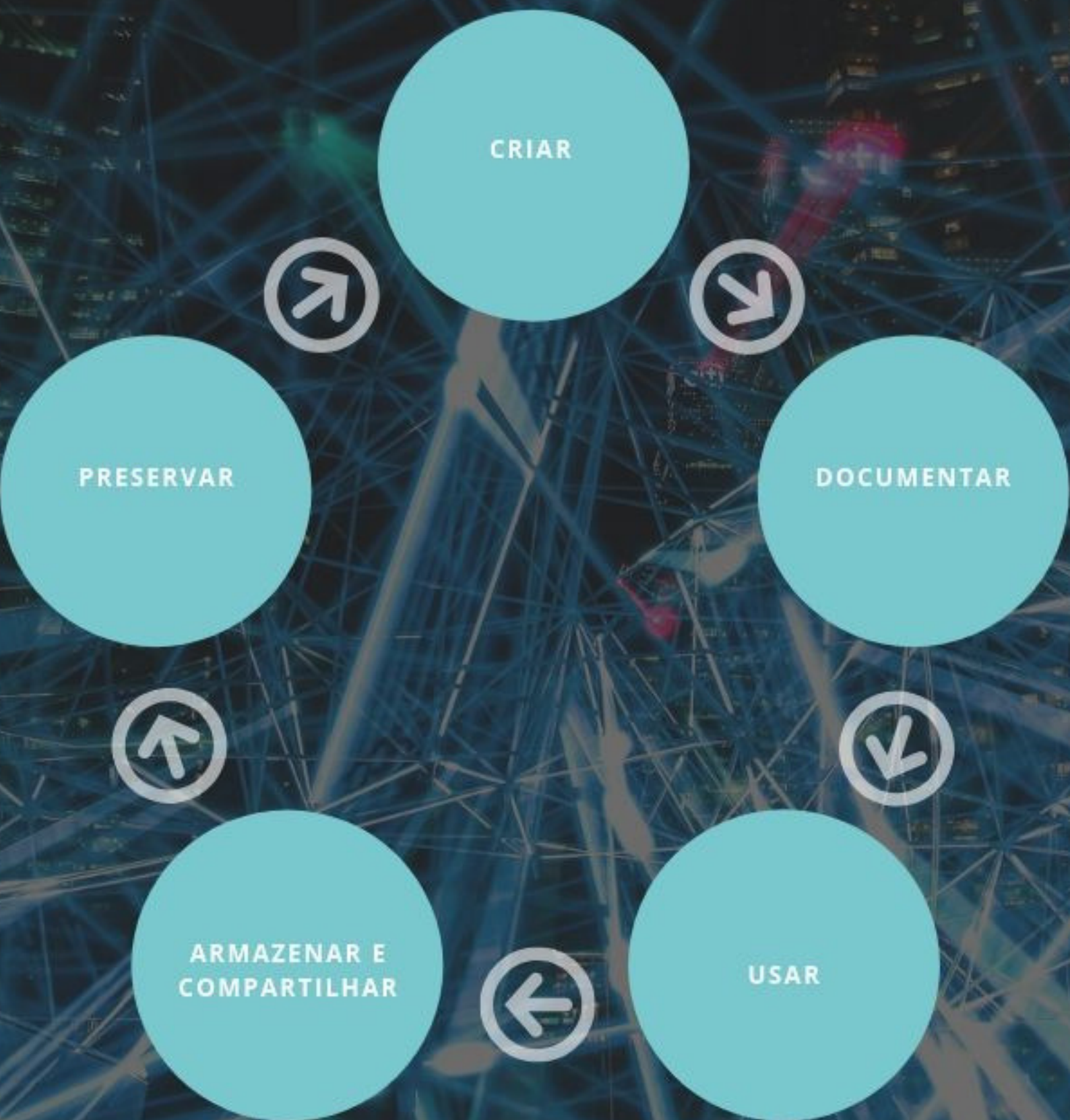
Um plano de gerenciamento de dados é um documento formal que se desenvolve no início do projeto de pesquisa e que descreve todos os aspectos dos dados (ou seja, o que será feito com os dados durante e após o projeto de pesquisa).

O gerenciamento desses dados deve ser de forma eficaz, a partir de um ciclo de vida dos dados, sendo fator importante para preservação e compartilhamento.

Na próxima página, será apresentado o ciclo de vida do gerenciamento de dados, como modelo inicial para construção de um plano de gerenciamento de dados da pesquisa.



# PLANEJANDO E GERENCIANDO DADOS DE PESQUISA



# CICLO DE VIDA DE DADOS DE PESQUISA



28

Conforme Bertin, Visoli e Drucker (2017), os dados possuem tempo de vida maior que o projeto de pesquisa desenvolvido e é fundamental perceber qual ciclo de vida será importante para o processo de análise de dados, bem como seu gerenciamento, tendo em vista seu uso e disponibilidade. Assim, observam-se as seguintes etapas:

**1º CRIAR DADOS:** verificar a escolha do tipo e do formato de dados que será criado e como os dados serão gerados.

**2º DOCUMENTAR:** analisar se a estrutura/nomenclatura de arquivos será compreensível para os outros, quais dados serão mantidos e quais podem ser descartados.

**3º USAR:** verificar como se dará a organização e uso dos dados de pesquisa durante a vida do projeto

# CICLO DE VIDA DE DADOS DE PESQUISA

**4º ARMAZENAR E COMPARTILHAR:** planejar como fará para armazenar e salvar os dados com segurança durante projeto e disponibilizá-los publicamente ao final. Deve-se pensar se a pessoa que está elaborando o projeto tem permissão para compartilhar seus dados e também em quais fatores poderiam restringir a capacidade de compartilhamento dos dados.

**5º PRESERVAR:** Verificar como se dará preservação dos dados após o final do projeto

## O PLANO DE DADOS DEVE DESCREVER:

- Quais dados de pesquisa será criado ou coletado.
- Quem será responsável pelos aspectos do plano de gerenciamento que será desenvolvido
  - Quais políticas (financeiras, institucionais e legais) serão aplicadas aos dados do projeto em execução.
  - Como os dados serão organizados (estruturas de pastas, convenções de nomenclatura de arquivos, controle de versão de arquivos).
  - Como os dados serão documentados durante a fase de coleta e análise de pesquisa.
  - Quais práticas de gerenciamento de dados (backups, armazenamento, controle de acesso, arquivamento) serão utilizadas para armazenar e proteger os dados.
  - Quais instalações e equipamentos serão necessários (espaço no disco rígido, servidor de backup, repositório).
- Quem terá direitos de propriedade e acesso aos dados do projeto.
- Como os dados serão preservados e disponibilizados a longo prazo quando a pesquisa estiver concluída.

## 1. INFORMAÇÕES ADMINISTRATIVAS

- Nome e identificação do projeto
- Descrição do projeto
- Órgãos de financiamento
- Nome do pesquisador principal e identificação
- Contato de dados do projeto
- Políticas relacionadas
- Data da primeira versão
- Data da última atualização

## 2. COLETA DE DADOS

- Descrição dos dados, incluindo tipo, formato e volume
- Conjuntos de dados existentes a serem reutilizados
- Métodos pelos quais os dados serão coletados ou criados
- Estruturas, sistema de nomenclatura e versão para pastas e arquivos
- Processos de garantia de qualidade

## 3. DOCUMENTAÇÃO E METADADOS

- Lista de informações esperadas que sejam necessárias para que os dados sejam lidos e interpretados no futuro
- Informar sobre o planejamento para coleta ou criação da documentação e de metadados
- Indicar os padrões de metadados que será utilizado

## EXEMPLOS DE DOCUMENTAÇÃO DE DADOS

OBS: ADICIONAR DESCRIÇÕES DETALHADAS PARA COLEÇÕES OU ARQUIVOS, POR EXEMPLO: ARQUIVO, DE ONDE VEIO, COMO PODERIA SER RECUPERADO SE NECESSÁRIO; QUAISQUER PROBLEMAS EXISTENTES ETC.

- Cadernos de laboratório e protocolos experimentais
- Questionários, livros de códigos, dicionários de dados
- Sintaxe de software e arquivos de saída
- Informações sobre configurações de equipamentos
- Esquema de banco de dados
- Relatórios de metodologia
- Informações sobre fontes de dados de pesquisa

## 4. ÉTICA E CONFORMIDADE LEGAL

- Detalhes do consentimento necessário para preservação e compartilhamento de dados
- Passos a serem tomados, se necessário, para proteger a identidade de qualquer participante
- Passos a serem tomados, se necessário, para garantir que os dados confidenciais sejam armazenados e transferidos com segurança

## 5. COPYRIGHT E DIREITOS DE PROPRIEDADE INTELECTUAL

- Nome (s) do (s) proprietário (s) dos dados
- Licença (s) para reutilização que será aplicada (por exemplo, uma das licenças disponíveis na Creative Commons ou Open Data Commons)
- Restrições ao uso de terceiros
- Qualquer atraso esperado para o compartilhamento de dados, por exemplo, pendência de um pedido de patente ou embargo relacionado à publicação em um periódico

## 6. ARMAZENAMENTO E BACKUP

- Onde (fisicamente) os dados serão armazenados
- Fornecimento de backup
- Pessoa ou equipe responsável pelo backup
  - Procedimentos de recuperação

## 7. SEGURANÇA

- Riscos envolvidos e como eles serão gerenciados
- Acordos de acesso
- Qualquer acordo, se necessário, para transferência segura de dados coletados no campo

## 8. SEGURANÇA

- Detalhes de quais dados devem ser retidos, compartilhados e / ou preservados, com particular referência a requisitos contratuais, legais ou regulatórios.
- Pesquisas previsíveis para os dados.
- Tempo pelo qual os dados serão (ou deverão) ser mantidos além da vida útil dos dados. o projeto
- O repositório ou arquivo onde os dados serão mantidos e quaisquer cobranças associadas
- Tempo e esforço necessários para preparar dados para preservação e compartilhamento de dados

## 9. COMPARTILHAMENTO DE DADOS

- Passos a tomar para maximizar a descoberta dos dados
- Quaisquer condições ou restrições à divulgação de dados, e se estas serão definidas num acordo de partilha de dados
- Mecanismo de partilha (ou seja, através de um repositório, correspondência direta ou outro arranjo)
- Momento de publicação de dados
- Arranjos, se houver, para obter um identificador persistente para os dados

## 10. RESPONSABILIDADE E RECURSOS

- Pessoa designada responsável pela implementação do Plano de Gerenciamento de Dados
- Pessoa designada responsável pelas atividades de gerenciamento de dados
- Hardware e software necessários (qualquer que seja adicional à provisão institucional existente)
- Experiência especializada adicional ou treinamento necessário
- Cobranças a serem aplicadas pelos repositórios de dados

# ATIVIDADE 1

## UNIDADE 2

36

UM PLANO DE GERENCIAMENTO DE DADOS PODE INCLUIR INFORMAÇÕES SOBRE:

- Contexto
- Dados [redacted] por exemplo, nomeação de pastas e arquivos
- Documentação e [redacted] (descritivo, administrativo, estrutural)
- Armazenamento [redacted] (incluindo métodos de backup e uso de criptografia)
- Dados [redacted], direitos e acesso ( para ajudá-lo a garantir altos padrões éticos e conformidade legal e regulamentar
- Preservação, [redacted] e licenciamento (para ajudá-lo a maximizar o alcance e o impacto de sua pesquisa)

metadados

diferir

preciso

organização

Claro

proteção

compartilhamento

segurança

Atualizada

# ATIVIDADE 2

## UNIDADE 2

37

VOCÊ FOI APRESENTADO AOS CONCEITOS DE GERENCIAMENTO DE DADOS E PLANOS DE GERENCIAMENTO DE DADOS.

CONFORME CONTEÚDO APRENDIDO, INSIRA AS PALAVRAS CORRETAS NO TEXTO.

1-Garantir que seus dados e registros de pesquisa sejam , completo, autêntico e confiável ajudando você a garantir que a rotulagem seja significativa e completa).

2-Salve  e recursos a longo prazo.

3-Melhorar os dados  e assim minimizar o  de perdas de dados.

4- Garantir a integridade da pesquisa e  por outros.

2-Evita  de esforço, permitindo que outros usem seus dados

Tempo

risco

preciso

reprodutibilidade

segurança

duplicação



**ORGANIZANDO E  
DOCUMENTANDO  
DADOS DE PESQUISA**

# UNIDADE 3

Esta Unidade apresenta habilidades de organização e documentação para os dados de pesquisa

Depois de completar esta unidade, você irá:

- Entender a organização dos dados de pesquisa;
- Facilitar o processo de recuperação da informação de dados de pesquisa;
- Nomear os arquivos e seus formatos de maneira eficiente e eficaz.

# ORGANIZANDO E DOCUMENTANDO DADOS DE PESQUISA

Os arquivos e suas pastas devem estar organizados de maneira eficaz e eficiente, sendo identificados para utilização por outros usuários. Levando em consideração alguns procedimentos disponibilizados pelo site Mantra (2015), elencamos:

- Os arquivos de dados devem ser distintos entre si, dentro de sua pasta;
- A nomenclatura a ser estabelecida facilitará na localização e navegação;
- A classificação dos arquivos deve obedecer uma sequência lógica
- Se os arquivos forem movidos para outras pastas, mantêm-se seus nomes e sua classificação.

# MÉTODOS E TÉCNICAS PARA ORGANIZAÇÃO E DOCUMENTAÇÃO

Conforme Mantra (2015), elencamos algumas informações para a organização e documentação de dados:

- O nome do arquivo deve fornecer conteúdo principal, distinguindo-se de arquivos semelhantes ou de versões diferentes.
- O arquivo deve ser identificado mas sugere-se evitar o uso de caracteres; usar underline, em vez de espaços.
- Utilizar data de criação ou modificação (Ex.: ANO\_MÊS\_DIA ou ANOMÊSDIA).
- Outros elementos podem auxiliar a nomear o arquivo, como: número da versão; data de criação; nome do criador; descrição de conteúdo; nome da equipe ou departamento e data da Publicação.



## NOMEANDO E ORGANIZANDO DADOS DA PESQUISA



### Nomeando por tempo:

Se estiver usando uma data, use o formato Ano-Mês-dia:

AAAA-MM-DD OU AAAA-AAAA

Isto manterá a ordem de tempo de seus arquivos



### Nomeando por assunto:

Mantenha os assuntos do arquivo curto e utilize de caracteres dos arquivos para capturar informações relevantes:

AAAA-MM-DD\_assunto principal\_versão.formato

Data é fácil de distinguir de arquivos. O nome do arquivo deve ser o conteúdo principal. Usar uma número de versão facilita a distinção do arquivo.



**ARMAZENAMENTO,  
COMPARTILHAMENTO  
E SEGURANÇA**

# UNIDADE 4

Esta Unidade apresenta os conceitos de armazenamento, de compartilhamento e de segurança de dados.

**Depois de completar esta unidade, você irá:**

- Conhecer os tipos de Driver e software utilizados para armazenamentos de dados;
- Entender os benefícios e desafios para compartilhamentos de dados;
- Resolver problemas envolvendo a proteção e backup dos dados.

Conforme Kelleher e John (2018) os armazenamentos de dados surgiram em função do grande número de objetos conectados à internet que geraram quantidades maciças de dados.

Com isso, no gerenciamento de dados deve-se levar em conta, independentemente do formato, o armazenamento em backup de dados, sendo estes mantidos regularmente e armazenados em:

- **Unidades de Rede:** serviços de armazenamentos de dados em nuvens, disponíveis sob as formas gratuita ou paga e que minimizam as perdas e riscos;
- **Computadores:** convenientes para armazenar dados durante o uso; porém podem falhar ou corromper em função de vírus digital;
- **Mídia Externa** (Discos rígidos, USB, CDs, DVDs): podem ser uma opção de armazenamento temporário, com baixo custo, porém, se não forem armazenados e acondicionados em locais físicos apropriados, podem correr o risco de dano e de perda de dados.

O compartilhamento de dados, deve levar em conta o acesso, uso e disponibilidade no futuro.

Para o seu compartilhamento, deve ter em mente a necessidade de: integridade do dados da pesquisa; a reutilização desses dados da pesquisa por outra pessoa; podendo levar a melhorias e correções na documentação ou comparação com outros dados de pesquisa. Além de maximizar a transparência e a responsabilidade.

Dicas:

- No Plano de Gerenciamento de Dados deve ser fornecido ao autor um termo de depósito e aceite para compartilhamento do dados;
- o autor deve ter ciência que seus dados podem ser modificados ou alterados, a partir do uso público;
- os dados de pesquisa podem servir como propósito de reinterpretação ou reinvenção para pesquisas futuras, promovendo a inovação por possíveis novos usuários de dados, principalmente em ambientes web, como nos repositórios institucionais.

Recomenda-se para maior segurança dos dados e proteção criptografada os serviços de armazenamentos online, que permitem maior sincronização de arquivos de dados entre pastas e computadores, bem como seu acesso, ponderando-se sobre alguns questionamentos importantes:

- Como será o armazenamento destes dados?
- Possui algum serviço ou assinatura de armazenamento online?
- Por quanto tempo os dados serão armazenados?
- Qual o espaço de memória no disco rígido, ou quantas mídias serão disponíveis para manter os dados?
- Se os dados forem confidenciais como armazená-los de forma segura e apropriada?

O site Mantra (2015) apresenta alguns exemplos de provedores utilizados:



DropBox

Serviço de hospedagem baseado em Web, utiliza computação em nuvem para permitir que usuários armazenem e compartilhem arquivos e pastas.



GoogleDrive

Serviço on-line que oferece 15 GB de armazenamento gratuito, permitindo o acesso aos arquivos por qualquer lugar, sendo necessário possuir uma conta Google para criação e edição dos arquivos em diversos formatos.



OneDrive

Serviço de hospedagem que possibilita sincronizar arquivos a partir de um navegador da Web ou de dispositivo local, sendo necessário obter conta da Microsoft.

ÉTICA  
E  
DIREITOS  
AUTORAIS

# UNIDADE 5

Esta Unidade apresenta os conceitos de Ética e de Direitos Autorais para uso e gerenciamentos de dados de pesquisa

Depois de completar esta unidade, você irá:

- Estar ciente dos requisitos de ética ao gerenciamento de dados da pesquisa;
- Entender as privacidades, confidencialidade e suas consequências e
- Identificar os direitos autorais dos dados.

Os dados da pesquisa devem cumprir – além dos seu gerenciamento e segurança –, os requisitos éticos que aplicam-se à sua natureza, ou seja, aos dados de pesquisa envolvendo estudos de seres humanos.

No Plano de Gerenciamento de Dados (PGD) da pesquisa deve constar um Código de Ética de compartilhamento e disponibilidade desses dados, sendo referenciados conforme o consenso do pesquisador, as normas e publicações.


No documento PGD deve conter um termo de direito à privacidade, a fim de o autor dos dados da pesquisa estar ciente da proteção de privacidade dos sujeitos, incluindo se dará permissão ou não ao acesso aos participantes.

Vale destacar que existem maneiras de obter consentimento dos participantes da pesquisa sendo apresentando um termo de livre de consentimento em que o pesquisador deve descrever o uso desses dados, seu acesso e sua publicação.

A fim de se manter a confidencialidade sobre os dados da pesquisa, deve considerar os seguintes fatores:

- Coletar os dados sem usar informações de identificação pessoal;
- Em casos de necessidade de coleta de fontes pessoais, após a análise, esses dados devem ser retirados o mais rápido possível;
- Evitar transmitir ou compartilhar dados sem o consentimento informado.

# ANEXO

A decorative horizontal dotted line in a light teal color spans the width of the page. Below it, three concentric diamond shapes are formed by teal outlines, centered on the page. The text is centered within the innermost diamond.

Modelo de Plano de  
Gerenciamento  
de Dados da pesquisa

# #Dicas de como começar a criar seu plano de gerenciamento de dados



Use a Ferramenta DMPonline do Centro de Curadoria Digital:  
<https://dmponline.dcc.ac.uk/>



Obtenha uma compreensão da terminologia e dos problemas de gerenciamentos de dados.



Adquira uma compreensão do seu projeto.



Converse com seu supervisor/colegas/ coordenadores de pesquisa/ Diretor de TI da sua instituição para obter informações.



Use os cabeçalhos fornecidos nesta unidade para desenvolver seu próprio plano de gerenciamentos de dados.



Mantenha-o prático e simples, não gaste muito tempo.



Não hesite em pedir ajuda quando precisar!



Verifique a disponibilidade na sua equipe da biblioteca universitária, para ajudar no projeto.



1

Que dados são produzidos?



2

Como os dados serão documentados e descritos?



3

Como seus dados serão estruturados e armazenados?



4

Quais são os planos para compartilhamento de dados e acesso após a submissão de um dado?



5

Existe requisitos "Especiais" para seus dados?



6

Quais são os planos para arquivamento a longo prazo dos dados digitais que apoiam a pesquisa?

# PASSOS PARA FORMAÇÃO: HABILIDADES E COMPETÊNCIAS PARA BIBLIOTECÁRIO DE DADOS

I  
N  
F  
O  
R  
M  
Á  
T  
I  
C  
O



# REFERÊNCIAS: PLANO DE GERENCIAMENTO DE DADOS

57

## TEXTUAIS

AMARAL, Fernando. Introdução à ciência de dados: mineração de dados e big. Rio de Janeiro: ALTA Books, 2016.

BERTIN, P. R. B.; VISOLI, M. C.; DRUCKER, D. P. A gestão de dados de pesquisa no contexto da e-science: benefícios, desafios e oportunidades para organizações de p&d. Ponto de Acesso, v. 11, n. 2, p. 34-48, 2017. Disponível em: <<https://portalseer.ufba.br/index.php/revistaici/articloe/view/21449>> Acesso em: 03 jan. 2019.

BOYD, Danah; CRAWFORD, Kate. “Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon”. In: Information, Communication & Society 15, no. 5: 662–79. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878>. > Acesso em: 19 jan. 2019.

HJØRLAND, Birger.. “Data (with dig data and database semantics)”. Knowledge Organization 45, no. 8: 685-708. Also available in ISKO Encyclopedia of Knowledge Organization, ed. Birger Hjørland, coed. Claudio Gnoi, 2018, Disponível em: <<http://www.isko.org/cyclo/data>. > Acesso em: 20 jan. 2019

## TEXTUAIS

GRAY, Jim. Jim Gray on eScience: A Transformed Scientific Method. Based on the transcript of a talk given by Jim Gray to the NRC-CSTB1 in Mountain View, CA, on January 11, 2007. In: HEY, Tony; TRANSLEY, Stewart; TOLLE, Kristin (orgs). The Fourth Paradigm. Data-Intensive Scientific Discovery. Redmond, WA: Microsoft Research, 2009. 284pp.

KOLTAY, Tibor. Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*, v. 49, n. 1, 2017. Disponível em: <<http://journals.sagepub.com/doi/abs/10.1177/0961000615616450>>. Acesso em: 15 dez. 2018.

Ridsdale, Chantel; ROTHWELL, James; SMIT, Mike; HOSSAM, Ali-Hassan, BLIEMEL, Michael; IRVINE Dean, KELLEY, Daniel; MATWIN ;Stan; WUETHERICK, Brad Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report, Dal housie University, Halifax, NS, 2015. Disponível em: <[http://www.mi-kesmit.com/wp-content/papercite-data/pdf/data\\_literacy.pdf](http://www.mi-kesmit.com/wp-content/papercite-data/pdf/data_literacy.pdf)>. Acesso em: 15 jan. 2019.

SETZER, V. W. Dado, informação, conhecimento e competência. *DataGramZero*, v. 0, n. 0, p. A01, 1999. Disponível em: <<http://www.brapci.inf.br/index.php/article/download/14562>>. Acesso em: 27 jan. 2019.

## VIDEOS

1- Bibliotecário Jeff Haywood- Gerenciamento de Dados para estudantes de doutorado e pesquisador:  
<https://youtu.be/PXr14Urf268>

2- Professora Lynn Jamieson (Importância do Gerenciamento de Dados):  
<https://www.youtube.com/watch?v=YQNadL5t8hg>

3- Instituição com serviços, recurso e orientações disponíveis, para acessar, armazenar, preservar e compartilhar seus dados de pesquisa:  
<https://www.youtube.com/watch?v=cuaAVum9mZk>

4- Preparação de planos de gerenciamento de dados:  
<https://www.youtube.com/watch?v=Lc82pxxRkMo>

5- Assessorando estudantes de doutorado no planejamento de gerenciamento de dados:  
<https://www.youtube.com/watch?v=rIW369PRfu0>

6- Plano de gerenciamento de dados durante um doutorado:  
<https://www.youtube.com/watch?v=tBRKM-XO12E>

# REFERÊNCIAS



60

## VIDEOS

7- Dados Espaciais:

<https://www.youtube.com/watch?v=FLL74eXcl8c>

8- Dados Secundário:

<https://www.youtube.com/watch?v=G8lonzDwYsQ>

**APÊNDICE B - QUESTIONÁRIO**



## Feedback da Oficina: Alfabetização de Dados

Agradecemos sua participação na oficina de capacitação: Alfabetização de Dados, ministrado pelo Mestrando Makson de Jesus Reis, sobre a orientação da Profa. Telma de Carvalho, do Mestrado Profissional em Gestão da Informação e do conhecimento do PPGCI/UFS.

Gostaríamos de sua contribuição na avaliação da oficina para fornecer subsídios na construção do Guia para Alfabetização de dados para Bibliotecários, que será o produto da minha dissertação de mestrado.

Por gentileza, se aceitar participar deste feedback, clique em preencher formulário, assinalando abaixo.

As informações coletadas são apenas para fins didáticos.

- Aceito participar  
 Não aceito participar

Obrigado!

**\*Obrigatório**

Você ficou satisfeito com a oficina ministrada? \*

	1	2	3	4	5	
Pouco satisfeito	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Muito satisfeito

Em uma escala de 1 a 5 indique o grau de conhecimento sobre o assunto obtido com a oficina. (1 = Muito insatisfeito 5 = Muito satisfeito) \*

	1	2	3	4	5
PLANEJAMENTO DE GERENCIAMENTO DE DADOS	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ORGANIZANDO E DOCUMENTANDO DADOS	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ARMAZENAMENTO DE DADOS E SEGURANÇA	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ÉTICA E DIREITOS HUMANOS	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Indique o seu grau de satisfação com as atividades desenvolvidas na oficina

	1	2	3	4	5	
Fraco	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Excelente

Quais foram os pontos mais importantes da oficina? \*

Sua resposta

---

Quais foram os conteúdos que você teve maior dificuldade com a oficina? \*

Sua resposta

---

Você conseguirá aplicar as informações obtidas nas atividades de seu trabalho? \*

Sua resposta

---

Deixe seus comentários adicionais sobre a oficina

Sua resposta

---

 Página 1 de 1

ENVIAR

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. [Denunciar abuso](#) - [Termos de Serviço](#)

Google Formulários

