


CIÊNCIA DE DADOS E CIÊNCIA DA INFORMAÇÃO: evolução e paradigmas da ciência

DATA SCIENCE AND INFORMATION SCIENCE: evolution and paradigms of science

 Makson de Jesus Reis ¹

 Telma de Carvalho ²

¹ Doutorando em Ciência da Informação pela Universidade Federal da Bahia (UFBA). Mestre em Gestão da Informação e do Conhecimento pela Universidade Federal de Sergipe (UFS).


E-mail: maksonacademico@gmail.com

² Doutora em Ciência da Informação pela Universidade de São Paulo (USP). Docente do Curso de Biblioteconomia e Documentação da Universidade Federal de Sergipe (UFS).

E-mail: carvalhotel@gmail.com



ACESSO ABERTO

Copyright: Esta obra está licenciada com uma Licença Creative Commons Atribuição 4.0 Internacional. 

Conflito de interesses: Os autores declaram que não há conflito de interesses.

Financiamento: Não há.

Declaração de Disponibilidade dos dados: Todos os dados relevantes estão disponíveis neste artigo.

Recebido em: 26 fev. 2022.

Aceito em: 30 ago. 2022.

Publicado em: 28 dez. 2022.

Como citar este artigo:

REIS, Makson de Jesus; CARVALHO, Telma de. Ciência de Dados e Ciência da Informação: evolução e paradigmas da ciência. **Informação em Pauta**, Fortaleza, v. 7, p. 1-21, 2022. DOI: 10.36517/2525-3468.ip.v7i00.2022.78490.1-21.

RESUMO

A evolução da Ciência possibilita percorrer o entendimento do processo evolutivo da sociedade, sendo embasada pelos paradigmas científicos e pelas pesquisas. Os conhecimentos no campo científico têm sua definição pautada na densidade da base empírica, sendo construída por seus paradigmas. Neste alicerce, pelo processo de desenvolvimento tecnológico, destaca-se que quatro paradigmas compõem a ciência, sendo o primeiro pela descrição dos fenômenos naturais, o segundo pelo uso de modelos, o terceiro pelo ramo computacional, nas simulações desses fenômenos e o quarto dessa quantidade de dados. Por meio desse avanço, este grande compartilhamento de dados e seu armazenamento, propiciaram o surgimento da e-science, formada pelo dilúvio de dados, sendo estes coletados, gerados ou simulados pela computação. O artigo tem como objetivo discutir a relação da ciência de dados na ciência da informação. Tem como procedimento metodológico a utilização de revisão de literatura com bases em pesquisas nas bases de dados científicas. Deste modo, conclui-se que compreender os paradigmas da ciência até chegar-se à e-science são essenciais para estudar o comportamento da atual era a partir da grande quantidade de dados emergentes da web, chegando-se, atualmente, ao quarto paradigma.

Palavras-chave: Ciência de Dados; Ciência da Informação; e-science; quarto paradigma; evolução da ciência.

ABSTRACT

The evolution of Science makes it possible to go through the understanding of the evolutionary process of society, being based on scientific paradigms and research. Knowledge in the scientific field has its definition based on the density of the empirical base, being built by its paradigms. In this foundation, through the process of technological development, which four paradigms make up science, being first for the description of natural phenomena, the second for the use of models, third for the computational branch, in the simulations of these phenomena and the fourth for the great amount of data, which through this advance, this great sharing of data

and its storage, propitiate the emergence of e-science, formed by the flood of data, which are collected, generated or simulated by computing. The article aims to discuss the relationship between data science and information science. Using as methodological study the use of literature review. Thus, it is concluded, that seeking to understand the paradigms of e-science are essential to study the behavior of the current era of the large amount of data emerging from the web.

Keywords: Data Science; Information Science; e-science; fourth paradigm; evolution of science.

1 INTRODUÇÃO

Os dados sempre foram elementos constantes nas pesquisas, especialmente nas pesquisas científicas. Com o desenvolvimento das tecnologias de informação e das possibilidades de busca por meio de mecanismos mais aprimorados, surgiram estudos que enfatizaram a Ciência de Dados, que reúne múltiplos aspectos da informação e estudos aprofundados sobre a temática dos dados, tendo uma equipe multidisciplinar de profissionais, a exemplo de estatísticos, programadores, analistas de dados, matemáticos e, também, bibliotecários para a extração de informações, interessados na temática. Desta forma os dados podem ser compreendidos como um grande volume de dados armazenados na internet, em grandes espaços, denominados “nuvens” e passam a fazer parte daquilo que se convencionou chamar de *Big Data*.

As nuvens são entendidas como plataformas digitais que abrigam informações geradas pelo grande volume de dados, além das pesquisas científicas inseridas no ambiente digital. Com isso, surgiu também o conceito de “dilúvio de dados”, interpretado pela grande quantidade de informação gerada diariamente pelas pesquisas e por outros meios, considerando-se também o usuário comum que está conectado a uma rede de informação e que gera dados e conteúdo, que também são armazenados nessas plataformas.

Diante desse panorama, surge a 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 (GRAY, 2007).

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 que envolve os conhecimentos necessários para lidar com esse campo de atuação que abriga 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) destaca que as pesquisas científicas serão transformadas pelo grande volume de dados e que a *e-science* se relaciona com a tecnologia da informação e cientistas. Portanto, a tecnologia tem papel de destaque no 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.

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 de dados serão desempenhados pelo processo de visualização, simulação, mineração e modelagem de dados e outras formas de representação e extração de conhecimento, sendo construídos diante de habilidades e conhecimentos adquiridos por capacitação (KOLTAY, 2017).

Conforme comenta Peter Kellam (2013) fortalecer a capacitação de bibliotecários apresenta desafios significativos para a construção de habilidades e de conhecimentos necessários para a realização de novos papéis. O conhecimento de domínio específico 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 fortalecerão suas relações de trabalho, incorporando-se às 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 convergente, 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 informação 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 de 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 Ciência da Informação na perspectiva de dados em seus serviços, e que estejam convencidos que podem ser incluídos nas rotinas dessas atividades, precisarão investir tempo no desenvolvimento de novas habilidades para a perfeita alfabetização de dados. A expansão de estudos com a temática dados encontra-se de forma exponencial na literatura em função das possibilidades que representam para o avanço científico, o que chamou a atenção para esta pesquisa. Afinal, com a crescente importância relacionada aos dados, surgem diversos movimentos preocupados com a preservação e disponibilização de ferramentas e de serviços que promovam a coleta, a organização e o gerenciamento desses dados.

A partir do exposto levantou-se o seguinte problema de pesquisa para este estudo: Quais foram os paradigmas percorridos para o processo evolutivo da chamada ciência de dados e de que maneira a Ciência da informação se comporta frente a essa situação?

Este questionamento deu origem ao seguinte objetivo geral: discorrer sobre os paradigmas percorridos para a evolução da ciência de dados e situar a Ciência da Informação, a Ciência de Dados e o bibliotecário nesta nova era.

Como objetivos específicos o trabalho busca: 1) discutir como são constituídos os paradigmas na ciência; 2) identificar a tipologia dos dados da pesquisa a fim de mapear os formatos dos arquivos utilizado pelo bibliotecário de dados; 3) apontar o relacionamento Ciência da Informação com a Ciência de Dados e 4) compreender o conceito de bibliotecário de dados. Entende-se que a Ciência de Dados e a Ciência da Informação possuem relação interdisciplinar no processo de entendimento entre os paradigmas.

Em termos metodológicos trata-se de pesquisa bibliográfica, utilizando-se as palavras-chaves: ciência de dados, ciência da informação, e-Science, bibliotecário de dados e alfabetização de dados. As bases de dados utilizadas foram: Portal de Periódicos da Capes, BRAPCI, EBSCO, Biblioteca Digital Brasileira de Teses e Dissertações, considerando-se nos campos de buscas os termos português e inglês e o uso dos operadores booleanos AND e OR, como demonstrado a seguir: (ciência de dados OR e-Science) AND (bibliotecário de dados OR alfabetização de dados) AND ciência da informação AND paradigmas.

2 OS PARADIGMAS DA CIÊNCIA

A notoriedade e conceituação da ciência constroem-se a partir do processo evolutivo da história. Conforme Dilthey (2010), ocorre 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.

O conhecimento no campo 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”. Demo (2011, p. 10), por sua vez, 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 de paradigmas diversificados, a adoção de critérios forma a lógica da ciência e os dados mensuráveis, comprováveis e testáveis. A ciência se constitui por múltiplos métodos.

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 evolui, tendo seu processo de discussão baseado em paradigma. O paradigma se define 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. Para o autor, quando um paradigma é aceito pela totalidade de uma comunidade, da sociedade ou de cientistas, ele é capaz de explicar fenômenos extraordinários que seu antecedente era incapaz, porém, muitos problemas considerados relevantes podem ser abandonados.

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.

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

Para Fentanes (2014,) contextualiza a existência da ciência experimental, apesar de Galileu, há quase quatro séculos, nas ciências naturais, ter comprovado, ao lançar dois objetos diferentes do alto da Torre de Pisa, que ambos chegavam ao solo ao mesmo tempo. Esses fatos ajudam a entender 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 paradigma. Um dos primeiros computadores digitais, o ENIAC (*Electronic 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 *et al.*, (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.

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.

Kuhn (2005, p.15), afirma que

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 Kepler, 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, privilegiando-se a observação e a coleta de dados. No terceiro paradigma, na Idade Contemporânea, no 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, conforme aponta Gray (2007).

Assim, surge o novo (quarto) paradigma, abordado pelo cientista Jim Gray¹ (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, provocando uma grande quantidade de dados gerados. Para construção do quarto paradigma, 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 [...]”. Em continuidade à sua linha de raciocínio o autor 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).

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

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;

¹ Jim Gray é pesquisador e foi gerente do grupo de e-Science 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/>

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.

Conforme Gray (2007), o novo (quarto) paradigma é centrado em 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 é afetada 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.

Dentre seus paradigmas, a ciência tem sido afetada pelo “dilúvio de dados” coletados, gerados ou simulados em grandes laboratórios e observatórios e delineia 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).

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 descreve o termo *e-science* como desafios científicos na utilização de tecnologia para investigações baseadas em dados. 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.

Para Hoffman (2010) a nova ciência baseada em dados utiliza-se dos teóricos e da computação 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.

3 TIPOLOGIA DOS DADOS DE PESQUISAS

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. Os dados estruturados possibilitam melhor armazenamento e organização ao serem pesquisados, realizando mesclagem com outros. Os dados não estruturados são aqueles que possuem estrutura interna própria, ou seja, não têm relação com outros dados, são 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 e 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; SANTARÉM 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 também 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.

Rice e Southall (2016) afirmam que os dados da pesquisa ou dados científicos são registros, arquivos ou conteúdo, em formato impresso ou digital, que contêm resultados de observações de pesquisa podendo ser compartilhado entre a comunidade acadêmica e estarem contidos em documentos como: 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.

3.1 Relações da Ciência da Informação e Ciência de Dados

Zhu e Xiong (2015) afirmam que os dados armazenados em computadores são utilizados e mapeados para que possam ser utilizados. No entanto, o método de uso de dados para análise de *Big Data*, modificou-se 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, a Ciência de Dados, emerge 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 e Amaral (2016, p. 4) esclarece em relação à *data science* que:

Embora a expressão 'Data Science' venha dos anos 1960, a ciência de dados é uma ciência nova. Uma ciência que 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.

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 (CODATA)* 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. Estabeleceu, em 2015, princípios e estratégias para a ciência de dados em três áreas prioritárias: promoção de princípios, políticas e práticas para dados em ciência; avanço das fronteiras da ciência de dados; 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, ver figura 1, 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 1 – Diagrama de Venn



Fonte: Adaptado de Amaral (2016)

Conforme Smith (2006, p. 163, tradução nossa), entende-se que a ciência de dados “se desenvolveu 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.

Alguns estudos que envolvem os dados, dentro 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”. 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 dessas estratégias para assuntos relacionados aos dados.

Além disso, os departamentos e/ou Faculdades de cursos de Biblioteconomia e/ou 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. Com isso, deve ampliar seus campos epistemológicos, possibilitando, a partir de estudos sobre manipulação de dados e extração de dados, por exemplo, entendê-lo como um campo novo para os bibliotecários, desenvolvendo habilidades no tocante aos dados.

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, da mesma forma, 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).

Desta forma, considera-se que à medida em 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 armazenados atualmente, o processo de organização e reuso deve ser feito por um profissional que tenha competência para atuar na recuperação de grandes volumes de dados. Assim, 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). Em meio às demandas requeridas para essa nova especialização.

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, mas deixa evidente as habilidades requeridas para tratar os dados de forma precisa.

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 Aalst (2014) pontua 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. Vê-se, portanto, que a ciência de dados envolve, além de equipe multidisciplinar, habilidades e competências voltadas para dados estatísticos, de programação, computação e, também, de tratamento técnico da informação, onde os bibliotecários atuam fortemente.

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 Scientist* (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 (VAN DER AALTS, 2014).

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 bibliotecários de 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 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 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.

4 CONCLUSÃO

Para identificar como os paradigmas são percorridos pela ciência de dados buscou-se compreender a evolução da ciência, que é um processo desafiador, especialmente quando se procura elencar os paradigmas pelos quais as novas tendências os modificaram e/ou contemplaram, alterando-se, inclusive, os métodos científicos adotados. A contribuição da ciência para o desenvolvimento da sociedade ocorre ao longo do tempo e, à medida que evolui o processo de pesquisa e novas descobertas são conhecidas e reconhecidas os paradigmas estabelecidos se mantêm ou surgem novos, para explicar os fenômenos observados.

Os paradigmas foram construídos, em seu primeiro momento, pelo pensamento racional e filosófico, na busca para entender o conflito entre o sagrado e o pensamento filosófico. No segundo momento, pelo método indutivo, com a observação e coleta de dados para resolução de problemas. No terceiro paradigma, com aplicações que envolveram probabilidade e estatística. O quarto paradigma surge pautado na grande quantidade de dados capturados por instrumentos ou gerados por simulações com uso do computador. Para identificar a tipologia dos dados da pesquisa, faz-se necessário explorar os diferentes tipos de dados em que são constituídos, como oriundos de arquivos de computador. As coleções de texto humano (e-mails, Tweets, mensagens de texto, postagens etc.) podem ser consideradas dados não estruturados, assim como coleções de

arquivos de som, imagem, música, vídeo e multimídia, sendo que estes podem ser compostos a partir da proliferação no ambiente web.

No momento atual, a grande quantidade de análises desses dados é crescente, devido, principalmente, à produção intelectual no campo científico construída em uma ciência computacional, chamada de e-science, baseada em vertentes de armazenamento e tecnologias emergentes na web.

Considera-se, assim, a ciência de dados, como a representação da informação, tecnologia e dados para as práticas científicas em diversas áreas do conhecimento e sua importância caracteriza-se pela interdisciplinaridade e ressignificação do dado como ciência, efetivando-se um relacionamento entre Ciência da Informação e Ciência de Dados. Assim, nesse contexto, surge o bibliotecário de dados, profissional voltado a habilidade com dados, sendo construído por variedades de suporte em dados, gerenciamento, planejamento, curadoria e visualização.

Por fim, este profissional é responsável pela coleta, armazenamento e gerenciamento de dados, com ações de curadoria e acesso aberto.

REFERÊNCIAS

ALVARO, E.; BROOKS, H.; HAM, M.; POEGEL, S.; ROSENCRANS, S. E-science librarianship: field undefined. *Issues in Science and Technology Librarianship*, v. 66, Summer 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.

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.

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**, Salvador, v. 11, n. 2, p. 34-48, 2017. Disponível em: <https://portalseer.ufba.br/index.php/revistaici/article/view/21449>. Acesso em: 3 jan. 2019.

BUGNION, P.; MANIVANNAN, A.; NICOLAS, P.R. Scala: **Guide for data science professionals**. Birmingham: Packt Publishing, 2017, 1077 p.

CORDEIRO, D.; BRAGHETTO, K.; GOLDMAN, A.; KON, F. Da ciência à e-ciência: paradigmas da descoberta do conhecimento. *Revista USP*, São Paulo, n. 97, p. 71-81, maio 2013. Disponível em: <http://www.revistas.usp.br/revusp/article/view/61867>. Acesso em: 19 de jan. 2019.

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

DILTHEY, W. **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, 2010. 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)**. v. 106, n. 3, July 2018. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6013124/pdf/jmla-106-294.pdf>. Acesso em: 28 dez. 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, 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.

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 dez. 2018

FRICKÉ, Martin. Big data and its epistemology. **Journal Association for Information Science and Technology**, New Jersey, v. 66, issue 4, p. 651-661, May 2014. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/asi.23212>. Acesso em: 10 jan.2018.

LEBRUN, Gérard. A filosofia e sua história. Org. de Carlos Alberto Ribeiro de Moura, Maria Lúcia M. O. Cacciola e Marta Kawano. São Paulo: Cosac Naify, 2007.

GRAY, Jim. Jim Gray on e-Science: A transformed scientific method based on the transcript of a talk given by Jim Gray to the NRC-CSTB1 in Mountain View, CA, (orgs). **The fourth paradigm. Data-intensive scientific discovery**. Redmond, WA: Microsoft Research, 2007. 284p.

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

HOFFMANN, L. Q& A: Gray's paradigm. **Communications of the ACM**, v. 53, n. 10, p. 112. 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. Acesso em: 19 jan. 2019

KELLAM, PENTLAND, A.S. The data-driven society. **Scientific American**, v. 309, n. 4, p. 78-83, Oct. 2013. Disponível em: https://connection.mit.edu/sites/default/files/publication-pdfs/data%20driven%20society%20sci%20amer_0.pdf. Acesso em: 22 dez. 2018

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

KOLTAY, T. Data literacy for researchers and data librarians. **Journal of Librarianship and Information Science**, Newbury Park, 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.

MAATTA, S. Placements & salaries: The emerging databrarian. London: **Library Journal**, New York, Oct. 2013. Disponível em: <https://www.libraryjournal.com/story/the-emerging-databrarian>. Acesso em: 15 dez. 2018

MARCHIONINI, G. Information science roles in the emerging field of data science. **Journal of Data and Information Science**, Beijing, v. 1, n. 2, p. 1-6, 2016 Disponível em: 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**, v. 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, V.; CUKIER, K. **Big data: a revolution that will transform how we live, work and think**, Boston, M: Houghton Mifflin Harcourt, 2013.

MIRANDA, L. F. S. **Introdução histórica à filosofia das ciências**. Curitiba: InterSaber, 2016.

MONTEIRO, Elizabete Cristina de Souza de Aguiar; SANT'ANA, Ricardo Cesar Gonçalves; SANTARÉM 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. ISSN2525-3468. Disponível em: <http://periodicos.ufc.br/informacaoempauta/article/view/2942>. Acesso em: 20 jan. 2019.

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

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**, Rio de Janeiro, 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. J. Data science as an academic discipline. *Data Science Journal*, v. 5, p. 163-164, 2006. 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 Acesso em: 27 jan. 2019.

HEY, T, TANSLEY, S, TOLLE, K. **O Quarto paradigma**: descobertas científicas na era da eScience. São Paulo: Oficina de Textos; 2011.

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 descriptions. **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, p. 353-356, 2009. Disponível em: <https://dl.acm.org/citation.cfm?id=1666191.1666210>. Acesso em: 10 jan. 2019.

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

¹ O presente artigo baseia-se em elementos extraídos da dissertação de mestrado intitulada "Ciência de dados e ciência da informação: guia para alfabetização de dados para bibliotecários", defendida no ano 2019 junto ao Programa de Pós-Graduação de Pós-Graduação em Ciência da Informação, do Mestrado Profissional em Gestão da Informação e do Conhecimento, da Universidade Federal de Sergipe.