

CONJECTURAS DA CIÊNCIA ABERTA NA CONTEMPORANEIDADE DO BIG DATA

NIVALDO CALIXTO RIBEIRO^{*}
DALGIZA ANDRADE OLIVEIRA^{**}
RONALDO FERREIRA DE ARAÚJO^{***}

RESUMO

Objetivou-se neste artigo discutir a Ciência Aberta, por meio dos dados científicos abertos, na contemporaneidade da revolução do *big data*. A gestão adequada de dados abertos pode apresentar-se como potencializadores para gerar o conhecimento científico, e como consequência favorecer avanços para a pesquisa científica e para a sociedade. O aporte para a discussão deste artigo centra-se na Escola de Pensamento Democrático e na Escola de Pensamento da Infraestrutura da Ciência Aberta e sua abordagem nos dados abertos de pesquisa (*open data*). Trata-se de uma pesquisa bibliográfica baseada em artigos de periódicos científicos sobre o *big data*, a Ciência Aberta e o relatório do acordo internacional sobre dados abertos produzido na *Science International Conference*. Concluiu-se que, apesar de iniciativas para melhor proveito do *big data*, de organização e

^{*} <https://orcid.org/0000-0003-0650-0121>. Doutorando em Gestão e Organização do Conhecimento pela Universidade Federal de Minas Gerais. Mestre em Administração, linha de pesquisa em Gestão Estratégica, Marketing e Inovações. Graduação em Biblioteconomia (2002) e Especialista em Gestão do Conhecimento e Tecnologia da Informação pelo Centro Universitário de Formiga (2006).

^{**} <https://orcid.org/0000-0002-0814-6325>. Docente da Escola de Ciência da Informação/Programa de Pós-Graduação em Gestão da Organização e do Conhecimento da Universidade Federal de Minas Gerais. Doutora em Ciência da Informação pela Universidade Federal de Minas Gerais (PPGCI/UFMG-2011). Mestre em Ciência da Informação (PPGCI/UFMG-2005). Bacharel em Biblioteconomia (Escola de Biblioteconomia/UFMG-1991). Coordenadora do Programa Carro-Biblioteca/ECI (10/2013-10/2015).

^{***} <http://orcid.org/0000-0003-0778-9561>. Doutor em Ciência da Informação (2015) e Mestre em Ciência da Informação (2009) pelo Programa de Pós-Graduação em Ciência da Informação da Escola de Ciência da Informação da Universidade Federal de Minas Gerais (UFMG). Possui graduação em Ciência da Informação (2006) pela Pontifícia Universidade Católica de Minas Gerais (PUC Minas). Foi professor visitante no Departamento de Jornalismo e Ciências da Comunicação da Universidade do Porto (2012). É Professor Adjunto do Curso de Biblioteconomia do Instituto de Ciências Humanas, Comunicação e Artes (ICHCA), Universidade Federal de Alagoas (UFAL).

disponibilização dos dados abertos de pesquisa, ainda há um *locus* para avanço da ciência nesse contexto.

PALAVRAS-CHAVE: Ciência Aberta; *Big data*; Dados científicos abertos.

OPEN SCIENCE CONJECTURES
IN THE CONTEMPORANEITY OF BIG DATA

ABSTRACT

The purpose of this article is to establish a Open Science, through open scientific data, the contemporaneity of the big data revolution. Adequate management of open data can be used to enhance scientific knowledge and, as a consequence, favor advances in scientific research and society. The contribution to the discussion of this article focuses on the school of democratic thought and the infrastructure school of thought of the science open and their approach in the open data. This work is a bibliographic research based on articles from scientific journals on the big data, Open Science and the report on the international open data agreement produced at the Science International Conference. In conclusion, despite initiatives to better benefit from big data, the organization and availability of open research data, there is still a way forward in this context.

KEYWORDS: Open Science; Big data; Open data.

1 INTRODUÇÃO

A revolução digital das últimas décadas pode ser considerada como um fenômeno histórico mundial tão profundo e difundido quanto a introdução da imprensa. (BOULTON *et al.*, 2015). Segundo Magalhães, Hartz e Martins (2016), o infográfico dinâmico *The Internet in Real Time* mostra que gigantes como o *Google*, a *Apple*, a *Microsoft*, a *Pandora*, o *Facebook*, a *Netflix*, o *Linkedin* e outros lucram cerca de 142 mil dólares por minuto, em razão da geração de aproximadamente um milhão de *gigabytes* de dados a cada um minuto na Internet. Nesse contexto de multiplicação e grande quantidade de dados sendo gerados a todo instante, encontra-se o fenômeno do *big data*.

O *big data* pode proporcionar maneiras de estudos e análises mais eficientes, já que é extremamente relevante às organizações públicas e privadas, no sentido de dar suporte para a elaboração, definição e planejamento de suas estratégias e ações, pois permite observar o crescimento, a disponibilidade e o uso de informações estruturadas e/ou não estruturadas, de acordo com o seu escopo de atuação.

Possivelmente associado a esse fenômeno e à necessidade de informação para acesso ao conhecimento, emerge junto o movimento da Ciência Aberta. A ciência deve ser aberta e visar ao bem-comum da sociedade. Ela representa o foco na pesquisa colaborativa, transparente e acessível, envolvendo diferentes significados, tipos de práticas e iniciativas, bem como distintas perspectivas, pressupostos e implicações. Um dos pressupostos da Ciência Aberta é que, devidamente apoiada por uma ciberinfraestrutura tecnológica e metodológica, se procura possibilitar que o conhecimento científico possa ser livre para ser usado, reutilizado e distribuído, evitando-se ou reduzindo-se as restrições legais, as tecnológicas e as sociais. (ALBAGLI; CLINIO; RAYCHTOCK, 2014).

Nesse cenário, o *big data* pressupõe uma excelente oportunidade para o avanço da ciência, ao passo que a ideia dos "dados abertos" aumenta a eficiência, a produtividade e a criatividade das pesquisas, podendo neutralizar também as tendências da privatização do conhecimento. Entretanto, a exploração do *big data* dependerá das condições dos sistemas nacionais de ciência para desenvolver recursos para usá-lo e decidir quais dados poderão ser abertos para o uso e a reutilização. Em consonância com Xavier (2018), observa-se que iniciativas e movimentos de Ciência Aberta estão mais comuns na atualidade, como reação à rápida evolução tecnológica da sociedade do conhecimento.

Uma conjectura pode ser considerada como uma hipótese, um pressuposto, uma fórmula ou uma ideia, a qual ainda precisa ser confirmada. Assim, este artigo debate-se sobre as premissas e contribuições que a Ciência Aberta pode proporcionar para a gestão do desenfreado surgimento de diversos dados gerados a cada instante, o *big data*.

É relevante mencionar que foram observadas, para a análise deste estudo, as Escolas de Pensamento da Ciência Aberta, a Democrática e a da Infraestrutura, em razão de sua relevância para o contexto da gestão dos dados abertos e da necessidade de melhoria da ciberinfraestrutura para melhor uso do *big data*.

No intuito de ampliar o debate sobre o tema, objetivou-se neste artigo discutir a relação entre a Ciência Aberta, por meio dos dados científicos abertos, na contemporaneidade da revolução do *big data*. Embora não seja o propósito neste estudo esgotar o assunto, busca-se contribuir com as discussões a respeito dessas duas expressões, que são relevantes para diversas áreas do conhecimento na atualidade.

Com relação aos procedimentos metodológicos, neste artigo não é apresentada uma sessão específica. Não houve estudo de campo, pois trata-se de uma pesquisa bibliográfica baseada em artigos de periódicos científicos e documentos sobre os temas *big data* e ciência aberta e suas contribuições para a ciência. Complementa o estudo uma análise do relatório do acordo internacional sobre dados abertos, produzido na Conferência *Science International*, de 2015.

2 A REVOLUÇÃO DO *BIG DATA*

Nos últimos anos, observa-se o desenvolvimento do *big data* de forma vertiginosa, atraindo a atenção da academia, da indústria e até mesmo dos governantes em diversos países, como Estados Unidos, Reino Unido, entre outros da União Europeia e demais regiões. Não há uma definição universal e totalmente aceita ou acabada; entretanto, é consenso de que se trata de conjuntos de dados que necessitam ser processados, tratados e armazenados. Pode ser considerado como um vínculo que sutilmente conecta e integra o mundo físico e a sociedade ao ciberespaço. (JIN *et al.*, 2015). Geralmente são obtidos dados por meio de sensores, experimentos científicos e observações, tais como: dados biológicos, dados neurais, dados astronômicos, dados de sensoriamento remoto e outros, oriundos de redes sociais, Internet, programas de saúde, de finanças, de economia, de transporte e de diversos recursos ou aparatos tecnológicos.

O volume surpreendente de dados requer que ele seja organizado e estruturado para possível subsídio dos gestores em suas estratégias de negócios e nichos de atuação. Fiorini *et al.* (2018) expõem que o *big data* auxilia na previsão, monitoramento, avaliação e adaptação de processos, estruturas internas e nas tomadas de decisões.

Em função do aumento de informação adicionada diariamente aos bancos de dados, certificar-se do estado da arte científica, tecnológica e extrair a informação certa para decisões corretas torna-se tarefa quase impossível, ao se empregarem os meios tradicionais. Diante de todo esse imbróglio de dados, a tecnologia da informação tem contribuído com ferramentas que auxiliam nesse processo nos mais diversos ramos de atuação. (MAGALHÃES; HARTZ; MARTINS, 2016).

Para Ferrinho e Fortunato (2016), o século XXI impulsionou inquietações e circunstâncias oportunas devido ao surgimento de

muitos dados adicionados a cada instante na *web*. No campo da ciência e do desenvolvimento tecnológico não é diferente. Dessa forma, é destaque a necessidade de metodologias mais avançadas para identificar, extrair e tratar os dados. Portanto, “minerar” o “*big data*” é fundamental, pois proporciona agilidade nos processos decisórios e na evolução da sociedade.

Disponibilizar uma abordagem consistente no tratamento do constante crescimento e da complexidade dos dados é uma tentativa de *big data*, considerando para a ocasião os cinco Vs que anteriormente, estruturavam-se em três: Volume, Velocidade, Variedade, Veracidade e Valor. (PRÁ NETTO; MORO; FERREIRA, 2015; MASHINGAIDZE; BACKHOUSE, 2017). Complementando, Brands (2014), ao se referir ao *big data*, acrescenta mais dois termos aos cinco Vs: Validade e Vocabulário, que se referem, pela ordem, à precisão e ao contexto dos dados.

Devido ao seu grande valor, o *big data* está transformando a maneira como se vive, se trabalha e se pensa; o mesmo ocorre com a ciência, levando a comunidade científica a reexaminar sua metodologia de pesquisa, desencadeando uma revolução no pensamento e nos métodos científicos, gerando um novo paradigma de pesquisa, isto é, com muitos dados, os pesquisadores só precisam encontrar ou extrair as informações necessárias, conhecimento e inteligência. (JIN *et al.*, 2015). Emerge, então, em decorrência do exposto, uma disciplina interdisciplinar chamada Ciência de Dados.

Ciência de dados diz respeito às técnicas para extrair conhecimento de dados diversos, com um foco particular em muitos dados, caracterizando o atributo dos Vs: volume, velocidade, variedade, valor, veracidade (GUTIERREZ, 2014), e também com a validade e o vocabulário. O campo da Ciência de Dados está se tornando cada vez mais influente nos setores público e privado, com seu objetivo abrangente de aumentar a compreensão de prestação de serviços, produtos e partes interessadas em todas as áreas.

Técnicas de Ciência de Dados estão sendo desenvolvidas e usadas em pesquisas aplicadas e interdisciplinares. No panorama da Ciência Aberta, muitas ações buscam a abertura de dados de pesquisa, que antes eram tratados como fechados, para seu uso e reuso por toda comunidade, convergindo para o estado similar ao identificado para o *big data*.

3 CIÊNCIA ABERTA: SUA TAXONOMIA E SUAS ESCOLAS DE PENSAMENTO

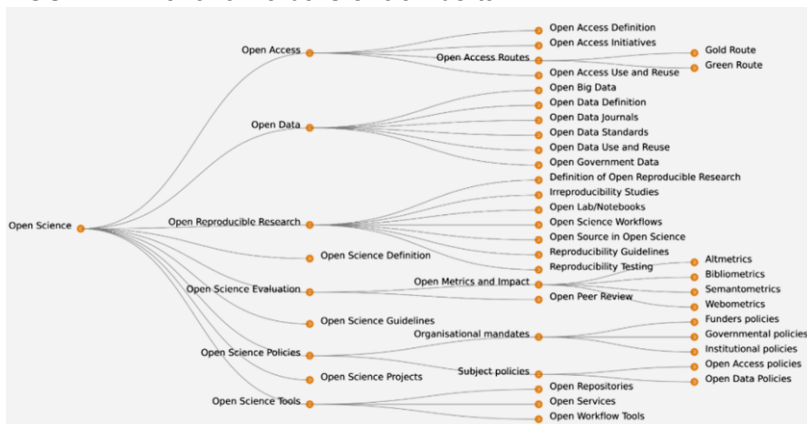
Em uma perspectiva histórica, o que se tem observado é que o desenvolvimento do conhecimento científico fundamentou-se, significativamente, por meio do acesso aberto, embora tem se constatado um avanço não ocorrido antes, voltado para interesse privado e comercial sobre a pesquisa, em diversas conjecturas da ciência. (GARCIA, 2014). Defende-se que os resultados de pesquisa deveriam ser de acesso aberto para a sociedade, pois uma grande parte das pesquisas científicas na atualidade são desenvolvidas em universidades e instituições públicas de pesquisa financiadas pelo Estado e/ou agências governamentais de fomento. (DENG, 2008). Em decorrência do predadorismo científico de algumas empresas, nem sempre é o que se tem testemunhado.

Segundo Santos (2017), a exigência do planejamento de gerenciamento de dados, como parte dos requisitos obrigatórios ao financiamento de projetos de pesquisa, pode ser considerado um dos marcos da abertura de dados científicos, liderada por agências de fomento, o que resultou na maioria das políticas relacionadas a essa questão. Associada à estratégia para o desenvolvimento científico, econômico e social em alguns países, na seara da administração pública, as políticas de abertura de dados coletados e/ou armazenados por governos visam à transparência pública, ao acesso à informação e à participação do cidadão.

A Ciência Aberta é um fenômeno disruptivo com uma amplitude que abrange práticas científicas, envolvendo desde o acesso aberto à literatura científica e aos dados abertos de pesquisa, perpassando pela ciência cidadã, por códigos, *software* e *hardware* abertos, entre outras. Ela é conduzida por uma metodologia adequada à filosofia da cultura digital aberta, valorizando a colaboração e o compartilhamento das *praxis* e conteúdos científicos. (SANTOS, 2017).

No estudo de Vicente-Saez e Martinez-Fuentes (2017), os autores buscaram construir uma definição rigorosa, integrada e atualizada do movimento da Ciência Aberta, por meio de uma revisão sistemática da literatura. Com base em seus estudos, os autores chegaram à conclusão de que a Ciência Aberta é o conhecimento transparente e acessível, compartilhado e desenvolvido por meio de redes colaborativas. Os pilares e as práticas da Ciência Aberta podem ser observadas em sua taxonomia, Figura 1, elaborada por Pontika e Knoth (2015).

FIGURA 1 – Taxonomia da Ciência Aberta



FONTE: Pontika e Knoth (2015).

Como pode ser verificado, o movimento da Ciência Aberta está estruturado em diversos pilares que visam o conhecimento, possibilitando o avanço tecnológico, científico e social em uma esteira que parece caminhar na mesma direção – a socialização do conhecimento.

Mirowski (2018) tece críticas à Ciência Aberta e afirma que a forma como vem sendo implementada não solucionará aquilo que se julga problemático, o antigo regime da ciência. Para o autor, o pano de fundo de suas discussões pode ser creditado à quebra do monopólio histórico da produção de conhecimento, aparentemente sem a participação do mercado. Ainda descreve que a Ciência Aberta é um dispositivo do formato atual da ciência, emplacada pelos defensores do estado mínimo, estruturando-se na “ciência radicalmente colaborativa” e no “capitalismo de plataformas”, associando-se à natureza de que a geração de conhecimento deve adequar-se às exigências do mercado. Entretanto, apesar da crítica do autor, neste artigo, o que é percebida é uma inclinação pró-Ciência Aberta. Autores como Fecher e Frieseke (2013, 2014), em uma ótica mais positivista, reconhecem a existência de cinco correntes ou escolas de pensamento que representam perspectivas diferentes e complementares da Ciência Aberta.

a) A Escola Pública (*Public School*) preza por pesquisas científicas que engloba e se dialogue com um público em geral e não somente com especialistas. São duas vertentes - acessibilidade para o processo de pesquisa (produção) e a compreensibilidade dos

resultados da pesquisa (o produto). Ambas as perspectivas envolvem a relação entre os cientistas e o público, para definir a abertura como uma forma de aproximação do último. (FECHER; FRIESEKE, 2013, ALBAGLI; CLINIO; RAYCHTOK, 2014).;

b) A Escola Pragmática (*Pragmatic School*) atua com o entendimento de “aberto” voltado para a inovação aberta e vê que o processo científico pode ser mais eficiente integrando o conhecimento externo e a colaboração por meio de ferramentas on-line. (FECHER; FRIESEKE, 2013, ALBAGLI; CLINIO; RAYCHTOK, 2014).;

c) A Escola da Infraestrutura (*Infrastructure School*) trabalhar nas alternativas e nas adversidades da Ciência Aberta, dedicando atenção às questões relacionadas às suas práticas e à infraestrutura necessária para o seu desenvolvimento. (FECHER; FRIESEKE, 2013, ALBAGLI; CLINIO; RAYCHTOK, 2014).;

d) A Escola das Métricas (*Measurement School*) tem como propósito a avaliação da produção científica, observando sua diversidade, visto que a migração para os ambientes digitais gerou a possibilidade de novos formatos de publicação, cujas métricas tradicionais não atribuíam nenhum tipo de impacto. (FECHER; FRIESEKE, 2013, ALBAGLI; CLINIO; RAYCHTOK, 2014).;

e) A Escola Democrática (*Democratic School*) preconiza o acesso ao conhecimento como um direito da humanidade; portanto, vislumbra que é ainda mais necessária a abertura dos resultados de pesquisa, ao se tratar de uma pesquisa científica financiada com recursos públicos. (FECHER; FRIESEKE, 2013, ALBAGLI; CLINIO; RAYCHTOK, 2014).

Neste estudo, os esforços direcionam sua atenção à duas escolas: a primeira, relacionada à Infraestrutura, pois com o surgimento do *big data*, alteraram-se representativamente os padrões tecnológicos e a maneira de se fazer ciência, considerando nessa perspectiva o uso da *web* e a rede de computadores, o que possibilitou o compartilhamento e a realização de atividades comuns. Além disso, entende-se que as soluções que permitiram a constituição de comunidades e redes sociais relacionadas aos pesquisadores e à comunidade acadêmica em ações do ensino, da pesquisa e da extensão são alvos de muita atenção. A segunda e mais relevante para esta pesquisa é a Escola Democrática, pois se apodera de duas vias estratégicas: 1. Acesso aberto – foca na abertura dos resultados de pesquisa, comumente divulgados por meio de artigos publicados em anais e revistas científicas e, 2. Dados abertos – busca garantir a disponibilização dos dados

primários, coletados no andamento da pesquisa, em formatos que possibilitem a sua consulta, mas também suas análises e a sua reutilização, de forma conveniente, em pesquisas posteriores, via essa que se passa a discorrer nas próximas linhas.

4 DADOS ABERTOS DE PESQUISA

Os dados de pesquisa sempre foram fundamentais nas mais diversas áreas de investigação e, recentemente, tornaram-se vitais, crescendo substancialmente em escala e complexidade. (SILVA, 2017). Dados de pesquisa são elementos protagonistas da dinâmica da pesquisa. São registros científicos que amparam os resultados de investigações publicadas no formato de dissertações, teses, artigos, patentes e trabalhos científicos. (DUDZIAK, 2016).

De acordo com Sayão e Sales (2015), os dados de pesquisa podem ser classificados como dados observacionais, dados computacionais, dados experimentais ou podem ser categorizados como brutos ou derivados, simulações, coleções físicas, modelos, *software*, imagens, vídeos e muito mais. Incluem documentos textuais, planilhas, estatísticas, cadernos de laboratório, cadernos de campo, diários, questionários, transcrições, arquivos de áudio, vídeo, fotografias, sequências de proteínas ou genéticos, artefatos, amostras, modelos, algoritmos, *scripts*, arquivos de *log*, *software* de simulação, metodologias e fluxos de trabalho, procedimentos operacionais, padrões, protocolos, entre outros. Os dados coletados podem variar conforme a área de conhecimento. (DUDZIAK, 2016).

Segundo Boulton *et al.* (2015), a respeito dos dados abertos de pesquisa e do *big data*, há uma crescente chamada de vários atores, dentro e fora da comunidade científica, e de órgãos intergovernamentais, como o grupo dos 8 países mais ricos e influentes do mundo (G8), a Organização de Cooperação e de Desenvolvimento Econômico (OCDE) e a Organização das Nações Unidas (ONU), para o acesso aberto a dados científicos financiados com recursos públicos, especialmente em relação a dados de particular importância para grandes desafios globais. A exploração do *big data* dependerá de como os sistemas de ciência desenvolverão sua capacidade para usá-lo e para decidir quais dados poderão se tornar abertos para uso e reuso.

O relatório do acordo internacional *Open data in a big data world*, apresentado como resultado da *Science International Conference 2015*, propõe 12 princípios para orientar cientistas, editores, bibliotecas, outras partes interessadas e em requisitos

técnicos para dados abertos. (BOULTON *et al.*, 2015). Os princípios apontados no acordo são apresentados a seguir, de forma sintetizada:

I. Cientistas financiados com recursos públicos têm a obrigação de contribuir por meio da criação e da comunicação de novos conhecimentos. Os dados produzidos em suas pesquisas devem ser disponibilizados o mais rápido possível, para que outros pesquisadores possam utilizá-los e/ou reaproveitá-los.

II. Os dados oriundos de evidências das produções científicas devem ser publicados simultaneamente e estar disponíveis de forma inteligível em formato aberto. Na medida do possível, os dados devem ser depositados em repositórios confiáveis e bem gerenciados.

III. As instituições de pesquisa e universidades têm a responsabilidade de criar um ambiente e suporte para dados abertos. Isso inclui o fornecimento de capacitação em gerenciamento de dados, preservação e análise, e de suporte técnico relevante, incluindo biblioteca e serviços de gerenciamento de dados. Instituições que empregam cientistas, juntamente com as agências de fomento à pesquisa, devem desenvolver políticas de incentivos e critérios para o avanço das suas carreiras.

IV. As editoras têm a responsabilidade de disponibilizar os dados aos revisores durante o processo de revisão, para exigir acesso aberto inteligível aos dados, simultaneamente com a publicação que os utilizam e para exigir a completa referência e citação desses dados. Os editores têm responsabilidade de disponibilizar o registro científico para posterior análise, por meio do fornecimento de metadados e acesso aberto para texto e mineração de dados.

V. As agências de financiamento devem considerar os custos dos processos de gestão de dados abertos como parte de um projeto de pesquisa, além de fornecerem recursos e políticas adequadas de sustentabilidade a longo prazo para infraestrutura e para repositórios. As avaliações do impacto da pesquisa devem considerar a contribuição dos criadores de dados.

VI. As associações profissionais, sociedades acadêmicas e academias devem desenvolver diretrizes e políticas para dados abertos, promover meios para que eles reflitam as normas epistêmicas e fomento de práticas e engajamento dos seus membros quanto aos dados abertos.

VII. As bibliotecas, arquivos e repositórios têm a

responsabilidade pelo desenvolvimento, prestação de serviços e padrões técnicos de dados, para garantir que estejam disponíveis e acessíveis a longo prazo.

VIII. Dados abertos devem ser exigidos para financiamento público em ciência. As exceções devem ser limitadas a questões de privacidade, de segurança e para uso comercial no interesse público, devidamente justificadas, caso a caso, e não de forma aleatória ou abrangente.

IX. Quando, em publicações acadêmicas, os pesquisadores usam dados criados por outros, eles devem mencionar seu autor ou produtor, sua origem e um identificador digital permanente.

X. Tanto os dados da pesquisa quanto os metadados, que permitem que os dados sejam identificados, avaliados, usados e reutilizados, devem ser o mais interoperáveis possível.

XI. Se os dados de pesquisa já não estiverem no domínio público, eles devem ser descritos como reutilizável, por meio de uma isenção de direitos ou licença não restritiva que deixa evidente que os dados podem ser reutilizados, sem mais exigência do que a de reconhecer o seu autor ou seu produtor.

XII. Os dados abertos devem, sempre que possível, ser vinculados a outros dados com base em seu conteúdo e contexto, a fim de maximizar seu valor semântico. (BOULTON *et al.*, 2015).

Para atender ao máximo os princípios estabelecidos neste acordo, mencionam-se as lições de Dudziak (2016), por meio das quais recomenda-se que todo pesquisador gerencie os dados, pensando quais dados coletar, como descrevê-los, como garantir a qualidade dos dados, o que deverá ser armazenado, qual o período e qual será a maneira de deixá-los acessíveis a longo prazo. É recomendado ainda elaborar um Plano de Gestão de Dados de Pesquisa e escolher um repositório para manter esses dados.

No entendimento de Banks *et al.* (2019), pode haver os mais variados motivos legítimos para que pesquisadores não compartilhem dados de pesquisa. Entre eles, aqueles que poderiam comprometer as identidades de certos participantes da pesquisa. Os dados que se compreende serem anonimizados podem, às vezes, ser comparados com outros dados existentes para revelar a identidade dos participantes. Em tempos de tratamento de muitos dados, com vistas ao *big data*, é possível que isso ocorra com mais intensidade. Além disso, há preocupações legais, éticas, econômicas e estratégicas que dificultam a abertura dos dados de determinadas pesquisas. Outra barreira pode estar na preocupação

dos pesquisadores quanto ao receio de outros apontarem falhas em suas pesquisas, podendo ameaçar a reputação ou credibilidade de um pesquisador. (NOSEK; BAR-ANAN, 2012).

Contudo, entende-se que a gestão de dados de pesquisa em tempos de Ciência Aberta e do *big data* se apresenta em emulação, contribuindo com passos largos para o desenvolvimento da ciência e da sociedade. Ambos, Ciência Aberta e o *big data*, estão passíveis a críticas e vieses que induzem as suas aplicações ou não. Caberá contar com *feeling* de governantes, empresários, gestores, pesquisadores e diversos outros atores, para que sejam estabelecidas políticas e diretrizes éticas, eficazes e sérias voltadas para o interesse do público em geral.

8 CONSIDERAÇÕES FINAIS

Tudo e nada ao mesmo tempo. Talvez essa seja a melhor forma para representar o *big data*, diante da imensidão de dados e informações que surgem a cada segundo. Localizar o que se precisa na *web* pode ser considerada uma tarefa tão árdua como encontrar uma “agulha no palheiro”. O mesmo pode ocorrer com os dados científicos abertos, quando não estão organizados, de forma prevista, nas linhas que discorrem sobre Ciência Aberta.

Na perspectiva da ciência, o que se percebe é um crescimento de ações voltadas para a disponibilização de dados científicos abertos, colaborando ainda mais para o aumento de dados na *web*. Entretanto, uma luta tem sido travada vislumbrando a gestão de tais dados, de forma inteligível, tanto por máquina quanto por humanos, especialistas ou membros da sociedade em geral. É inegável que as tecnologias influenciaram essa emblemática ascensão e o excesso de dados. Tal contexto parece ser irreversível na atual realidade mundial.

Muito há para se avançar, o locus para atuação é iminente e requer ações de diversos atores para melhor proveito dos aspectos gerados pelo *big data*. Compreende-se que qualquer área que ignorá-lo está fadada ao fracasso, visto que seu potencial é devastador e sem precedentes na sociedade contemporânea.

Quanto às conjecturas da Ciência Aberta para o fenômeno do *big data*, é possível identificar que a comunidade científica tem demonstrado grande preocupação com os dados científicos abertos, buscando estabelecer diretrizes e políticas que preveem melhor uso e reuso, em especial, quando se trata de pesquisas financiadas com recursos públicos.

O entendimento é que, ao se disponibilizar os dados de pesquisa aberta, há mais possibilidades de se evoluir tecnologicamente. Esse pressuposto é visível nos princípios estabelecidos pelo acordo internacional sobre dados abertos produzido na *Conferência Science* em 2015. Além disso, desdobramentos da Ciência Aberta têm buscado maneiras de apresentar à sociedade as soluções geradas pelas pesquisas científicas.

Quanto às questões iniciais, observou-se que a revista *Ciência da Informação* tem dado mais atenção à temática *big data* e que se identificou uma tendência de ampliar o aceite de artigos com a temática *big data* para os próximos anos. No estudo, constatou-se que os autores têm apresentado diversas abordagens para os temas em estudos, reforçando a interdisciplinaridade da área. Os textos tratam de experimentos que, da área da saúde à agrária, demonstram o potencial do *big data* e do ecossistema da Ciência Aberta. Tão grande foi a revolução dos dados que houve o surgimento de uma nova disciplina, a Ciência de Dados, e a necessidade de formação de um novo profissional, o cientista de dados. A Ciência de dados bem como o cientista de dados continuam em evolução e vêm ganhando cada vez mais espaço no mercado e na pesquisa.

Por fim, pressupõe-se que o movimento de Ciência Aberta e o *big data* devam ser mais explorados, tanto por empresas, quanto por instituições de pesquisa, para melhor compreensão das oportunidades de evolução e dos desafios a serem enfrentados pela humanidade na divulgação da sociedade do conhecimento.

REFERÊNCIAS

ALBAGLI, S.; CLINIO, A.; RAYCHTOK, S. Ciência aberta: correntes interpretativas e tipos de ação. **Liinc em Revista**, Rio de Janeiro, v. 10, n. 2, p. 434-450, nov. 2014. Disponível em: <http://revista.ibict.br/liinc/article/view/3593>. Acesso em: 14 maio 2019.

ARAÚJO JÚNIOR, R. H. de; SOUSA, R. T. B. Estudo do ecossistema de *Big Data* para conciliação das demandas de acesso, por meio da representação e organização da informação. **Ciência da Informação**, Brasília, DF, v. 45, n. 3, fev. 2018. Disponível em: <http://revista.ibict.br/ciinf/article/view/4057>. Acesso em: 17 jan. 2020.

BANKS, G. *et al.* Answers to 18 questions about open science practices. **Journal of Business and Psychology**, [S.l.], v. 34, n. 3, p. 257-270, 2019. Disponível em: <https://link.springer.com.ez27.periodicos.capes.gov.br/article/10.1007/s10869-018-9547-8>. Acesso em: 10 jan. 2020.

BERTIN, P. R. B. *et al.* A política de governança de dados, informação e conhecimento da empresa como mecanismo para a gestão de dados de pesquisa agropecuários. **Liinc em Revista**, Rio de Janeiro, v. 15, n. 2, p. 194-204, nov. 2019. Disponível em: <http://revista.ibict.br/liinc/article/view/4798>. Acesso em: 20 jan. 2020.

BOULTON, G. *et al.* **Open data in a big data world: an international accord**. Paris: ICSU-IAP-ISSC-TWAS, 2015. Disponível em: <http://www.science-international.org/>. Acesso em: 10 jan. 2020.

BRANDS, K. *Big data* and business intelligence for management accountants. **Strategic Finance**, Montvale, p. 56-58, June. 2014. Disponível em: <https://sfmagazine.com/wp-content/uploads/sfarchive/2014/06/TECH-PRACTICES-Big-Data-and-Business-Intelligence-for-Management-Accountants.pdf>. Acesso em: 20 jun. 2019.

BUFREM, L. S. *et al.* Produção internacional sobre ciência orientada a dados: análise dos termos *data science* e *e-science* na *Scopus* e na *Web of Science*. **Informação & Informação**, Londrina, v. 21, n. 2, p. 40-67, dez. 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/26543/20114>. Acesso em: 17 jan. 2020.

CALDAS, C. O. L.; CALDAS, P. N. L. Estado, democracia e tecnologia: conflitos políticos e vulnerabilidade no contexto do *big-data*, das *fake news* e das *shitstorms*. **Perspectiva Ciência da Informação**, Belo Horizonte, v. 24, n. 2, p. 196-220, jun. 2019. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362019000200196&lng=en&nrm=iso. Acesso em: 17 Jan. 2020.

CARVALHO, J. Saiba o que é Qualis Capes e quais os periódicos da Ciência da Informação com essa classificação. **Biblioo Cultural**, [S.l.], 3 mar. 2017. Disponível em: <https://biblioo.cartacapital.com.br/saiba-o-que-e-qualis-capes/>. Acesso em: 20 jan. 2020.

COSTA, M. M.; CUNHA, M. B. da. A literatura internacional sobre e-science nas bases de dados LISA e LISTA. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 20, n. 44, p. 127-144, nov. 2015. ISSN 1518-2924. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2015v20n44p127>. Acesso em: 17 jan. 2020.

CURTY, R. G.; SERAFIM, J. S. A formação em ciência de dados: uma análise preliminar do panorama estadunidense. **Informação & Informação**, Londrina, v. 21, n. 2, p. 307-331, dez. 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27945>. Acesso em: 17 jan. 2020.

DENG, F. **What is “open”?: an economic analysis of open institutions**. Munich Personal RePEc Archive nº 888. [s.l.], 2008. Disponível em: <https://mpa.ub.uni-muenchen.de/8888/>. Acesso em: 10 jan. 2020.

DUDZIAK, E. **Dados de pesquisa agora devem ser armazenados e citados**. [S.l.: s. n.], 2016. Disponível em: <http://www.sibi.usp.br/?p=6189>. Acesso em: 10 jan. 2020.

FACILITATE OPEN SCIENCE TRAINING FOR EUROPEAN RESEARCH (FOSTER). **Open data: open science taxonomy**. [S.l.: s. n.], 2015. Disponível em: <https://www.fosteropenscience.eu/foster-taxonomy/open-data>. Acesso em: 20 jun. 2019.

FECHER, B.; FRIESIKE, S. Open science: one term, five schools of thought. **The RatSWD Working Papers**, [S.l.], 30 maio 2013. Disponível em: <https://ssrn.com/abstract=2272036> ou <http://dx.doi.org/10.2139/ssrn.2272036>. Acesso em: 20 jan. 2020.

FECHER, B.; FRIESIKE, S. Open science: one term, five schools of thought. In: BARTLING, S.; FRIESIKE, S. **Opening science: the evolving guide on how the internet is changing research, collaboration and scholarly publishing**. New York: Springer, 2014.

FERRINHO, P.; FORTUNATO, P. Gestão do conhecimento, meta-avaliação, políticas de saúde, *big data* e ciência aberta. **Anais do Instituto de Higiene e Medicina Tropical**, Lisboa, v. 15,, p. 5-6, 2016. Suplemento 2. DOI: <https://doi.org/10.25761/anaisihmt.111>. Disponível em: <https://anaisihmt.com/index.php/ihmt/article/view/111/95>. Acesso em: 20 jan. 2020.

FIORINI, P. C. *et al.* Management theory and *big data* literature: From a review to a research agenda. **International Journal of Information Management**, Guildford, v. 43, p. 112-129, 2018.

GARCIA, M. Ciência, futuro em aberto. **Ciência Hoje**, Rio de Janeiro, 1 out. 2014. Disponível em: <http://cienciahoje.org.br/ciencia-futuro-em-aberto/>. Acesso em: 22 jan. 2020.

GUTIERREZ, S. **Data scientists at work**. New York: Amazon Digital, 2014. 366p.

HARRIS, C. G. **Applying human computation methods to information science**. 2013. 209 f. Thesis (Doctor of Philosophy) - University of Iowa, 2013. Disponível em: <https://doi.org/10.17077/etd.6kebet6l>. Acesso em: 20 jan. 2020.

HEY, T.; TREFETHEN, A. E. Cyberinfrastructure for e-Science. **Science**, London, v. 308, n. 5723, p.. 817-821, May 2005. Disponível em: <https://science.sciencemag.org/content/308/5723/817/tab-article-info>. Acesso em: 12 fev. 2020.

JIN, X. *et al.* Significance and challenges of *big data* research. **Big Data Research**, Paris, v. 2, p. 59-64, 2015. DOI: <https://doi.org/10.1016/j.bdr.2015.01.006>. Acesso em: 10 jan. 2020.

MAGALHÃES, J.; HARTZ, Z.; MARTINS, M. R. O. *Big data* para a investigação em saúde e a Ciência Aberta: um contributo para a gestão do conhecimento. **Anais do Instituto de Higiene e Medicina Tropical**, Lisboa,

v. 15, p. 75- 82, 2016. Suplemento 2.

MANSELL, R. Open collaboration for social problem solving: converging or diverging norms of governance authority?: colaboração aberta para a solução de problemas sociais: normas de autoridade de governança convergentes ou divergentes?. **Liinc em revista**, Rio de Janeiro, v. 10, n. 2, p. 451-459, nov. 2014. Disponível em: <http://revista.ibict.br/liinc/article/view/3594/3073>. Acesso em: 17 jan. 2020.

MASHINGAIDZE, K.; BACKHOUSE, J. The relationships between definitions of big data, business intelligence and business analytics: a literature review. **International Journal of Business Information Systems**, Bakersfield, v. 26, n. 4, p. 488, 2017. Disponível em: <https://dl.acm.org/doi/10.1504/IJBIS.2017.087749>. Acesso em: 12 fev. 2020.

MIROWSKI, P. The future(s) of open science. **Social Studies of Science**, London, v. 48, n. 2, p. 171-203, 2018. DOI: 10.1177/0306312718772086.

NOSEK, B. A.; BAR-ANAN, Y. Scientific utopia: I. Opening scientific communication. **Psychological Inquiry**, [S.l.], v. 23, n. 3, p. 217-243, 2012. Disponível em: <http://dx.doi.org/10.1080/1047840X.2012.692215>. Acesso em: 18 jan. 2020.

PONTIKA, N.; KNOTH, P. **Open science taxonomy**. 2015. Disponível em: <http://oro.open.ac.uk/47806/>. 10 jan. 2020.

PRÁ NETTO, A. S. D.I; MORO, E. P.; FERREIRA, F. F. *Big Data* e suas influências sobre a estratégia das empresas. Rio de Janeiro: UFRJ, [2015]. Disponível em: https://www.gta.ufrj.br/grad/15_1/bigdata/vs.html. Acesso em: 10 jan. 2020.

RAUTENBERG, S.; CARMO, P. R. *Big data* e ciência de dados. **Brazilian Journal of Information Science: research trends**, Marília, v. 13, n. 1, p. 56-67, 29 mar. 2019. Disponível em: <http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/8315>. Acesso em: 20 jan. 2020.

RODRIGUES, F.; SANT'ANA, R. C.; FERNEDA, E. Análise do processo de recuperação de conjuntos de dados em repositórios governamentais. **INCID: Revista de Ciência da Informação e Documentação**, São Paulo, v. 6, n. 1, p. 38-56, 10 abr. 2015. Disponível em: <http://www.revistas.usp.br/incid/article/view/73496>. Acesso em: 20 jan. 2020.

SANT'ANA, R. C. G. Ciclo de vida dos dados e o papel da ciência da informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013. Apresentações. **Anais...** Florianópolis: UFSC, 2013. ISBN 978-85-65044-06-6. Disponível em: <http://enancib2013.ufsc.br/index.php/enancib2013/XIVenancib/paper/viewFile/284/319>. Acesso em: 10 jan. 2020.

SANTOS, P. X. (Org.). **Livro verde ciência aberta e dados abertos: mapeamento e análise de políticas, infraestruturas e estratégias em perspectiva**

nacional e internacional. Rio de Janeiro: Fiocruz, 2017. 140 p. Disponível em: <https://www.arca.fiocruz.br/handle/icict/24117>. Acesso em: 14 jul. 2019.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/IN, 2015. Disponível em: http://www.cnen.gov.br/images/CIN/PDFs/GUIA_DE_DADOS_DE_PESQUISA_SA.pdf. Acesso em: 22 jan. 2020.

SHINTAKU, M.; DUQUE, C.; SUAIDEN, E. Análise da adesão às tendências da Ciência pelos repositórios institucionais brasileiros. **InCID: Revista de Ciência da Informação e Documentação**, São Paulo, v. 6, n. 2, p. 148-169, 2 out. 2015. Disponível em: <http://www.revistas.usp.br/incid/article/view/89191>. Acesso em: 20 jan. 2020.

SILVA, P. F. P. **As políticas de open data em Portugal**: análise da sua implementação e impacto. 2017. 138p. Dissertação (Mestrado em Ciência da Informação) - Universidade de Coimbra, Coimbra, 2017. Disponível em: <https://estudogeral.sib.uc.pt/handle/10316/47025>. Acesso em: 10 jan. 2020.

SOUZA, R. R.; ALMEIDA, M. B.; BARACHO, R. M. A. Ciência da Informação em transformação: *big data*, nuvens, redes sociais e web semântica. **Ciência da Informação**, Brasília, DF, v. 42 n. 2, p.159-173, maio/ago. 2013. Disponível em: <http://revista.ibict.br/ciinf/article/view/1379>. Acesso em: 17 jan. 2020.

VAN REENEN, J. Acesso aberto e conectividade: estimulando inovações inesperadas com o uso de arquivos abertos institucionais. **Ciência da Informação**, Brasília, DF, v. 35, n. 2, p. 17-26, 2006. Disponível em: <http://dx.doi.org/10.1590/S0100-19652006000200003>. Acesso em: 20 jan. 2020.

VICENTE-SAEZ, R.; MARTINEZ-FUENTES, C. Open science now: a systematic literature review for an integrated definition. **Journal of Business Research**, Athens, v. 88, p. 428-436, July, 2018. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0148296317305441>. Acesso em: 20 jan. 2020.

VICTORINO, M. C. *et al.* Uma proposta de ecossistema de big data para a análise de dados abertos governamentais conectados. **Informação & Sociedade: Estudos**, João Pessoa, v. 27, n. 1, 2017. Disponível em: <https://doi.org/10.22478/ufpb.1809-4783.2017v27n1.29299>. Acesso em: 20 jan. 2020.

XAVIER, P. A Fiocruz frente ao desafio da ciência aberta em prol do Desenvolvimento e da Saúde Pública. **SciELO em Perspectiva**, [S.l.], maio 2018. Disponível em: <https://blog.scielo.org/blog/2018/05/21/a-fiocruz-frente-ao-desafio-da-ciencia-aberta-em-prol-do-desenvolvimento-e-da-saude-publica/#.XRJ-otJKjIU>. Acesso em: 26 jun. 2019.

Recebido em: 11/ 01/2019

Aceito em: 20/11/2019