

CIÊNCIA ABERTA: DIMENSÕES PARA UM NOVO FAZER CIENTÍFICO

CIENCIA ABIERTA: DIMENSIONES A UNA NUEVA PRACTICA CIENTÍFICA

Adriana Carla Silva de Oliveira*
Edilene Maria da Silva**

RESUMO:

Introdução: As práticas da *e-Science* e uso e reuso dos dados científicos se constituem um novo fazer científico que conduz à reflexão sobre novos arcabouços normativos, legais, institucionais e tecnológicos voltados para uma ciência aberta.

Objetivo: Esse estudo apresenta como questão de pesquisa: quais dimensões dão sustentabilidade à formulação de uma política orientada à ciência aberta e suas práticas no contexto brasileiro? O objetivo deste estudo é discutir as dimensões que apoiam transversalmente a formulação de uma política direcionada à ciência aberta e suas práticas científicas.

Metodologia: Teoricamente, o estudo está orientado pelo quarto paradigma científico baseado na *e-Science*. A metodologia está apoiada no estudo de Bufrem (2013) que propõe um modelo alternativo e multidimensional para análise e discussão de pesquisas científicas. Tecnicamente, empregaram-se os métodos de pesquisa bibliográfica e levantamento documental acerca do modelo científico *Data Lifecycle*, legislações e acordos internacionais. Para fins desse estudo, propõem-se cinco dimensões, a saber: epistemológica, política, ético-legal-cultural, morfológica e tecnológica.

Resultados: As dimensões substanciam uma política de informação ou a elaboração de diretrizes mínimas para agenda da ciência aberta no Brasil.

Conclusões: As dimensões afastaram o olhar reducionista sobre dados de pesquisa e conduziram o estudo para a visão multidimensional e multirrelacional da ciência aberta.

Palavras-chave: Ciência aberta. *e-Science*. Políticas de informação. Colaboração. Dimensões.

* Doutora em Ciência da Informação pela Universidade Federal da Paraíba (UFPB). Visiting Scholar pela University of Tennessee, Knoxville, USA. Analista Judiciário do TJRN. E-mail: adrianacarla.a@gmail.com

** Doutora em Ciência da Informação pela Universidade Federal da Paraíba (UFPB). Visiting Scholar pela University of Wisconsin-Milwaukee, USA. Professora do Departamento Ciência da Informação da Universidade Federal de Pernambuco (UFPE). E-mail: dilemabr@gmail.com

1 INTRODUÇÃO

A ciência aberta é o fio condutor de investigações científicas apoiadas por uma ciberinfraestrutura tecnológica e metodológica que permite o uso, reuso e reprodutibilidade de dados de pesquisa. Destarte, figura o surgimento do novo paradigma da ciência nomeado de quarto paradigma no cenário contemporâneo mundial da comunicação científica.

Esse paradigma fomenta uma metodologia científica orientada aos dados de pesquisa. Assim, o dado é o elemento central e seu recurso principal. As mudanças nas práticas científicas têm gerado alterações no *status quo* da ciência tradicional. O novo fazer científico denominado *e-Science* prioriza a colaboração, compartilhamento e gerenciamento dos dados de pesquisa como aspectos principais de uma ciência colaborativa.

No contexto internacional, a realidade de uma ciência orientada aos dados de pesquisa está em amadurecimento e consolidação. Já envolve agências de fomento, instituições de pesquisa e universidades, norteia a submissão dos projetos de pesquisa à apreciação de financiamentos, desenvolvimento de políticas, diretrizes, modelos e infraestrutura tecnológica para apoiar essa nova realidade. Além disso, parcerias institucionais e processo de educação continuada são recomendados para apoiar pesquisadores, grupos de pesquisa e profissionais envolvidos com a nova sistemática da pesquisa aberta.

Para o desenvolvimento da ciência aberta devem ser considerados, além dos aspectos técnicos e tecnológicos, as questões de cunho ético, legal, cultural, político e institucional (formais e informais) que mais interferem no caráter aberto ou proprietário dessas práticas (ALBAGLI, 2015).

A realidade brasileira voltada para a ciência aberta e dados de pesquisa encontra-se em um estágio incipiente. Tal afirmação baseia-se no reduzido número de produções acadêmico-científicas acerca do assunto, das poucas iniciativas em andamento e da ausência de políticas e diretrizes que guiem estas ações. Observando esse fato, percebe-se que as atuais iniciativas precisam ser conduzidas por uma política de informação ou, no mínimo, por

diretrizes que apoiem as práticas da *e-Science* para o uso e reuso de dados de pesquisa.

Assim, o estudo questiona: quais dimensões dão sustentabilidade à formulação de uma política orientada à ciência aberta e suas práticas no contexto brasileiro? O objetivo deste estudo é discutir as dimensões que apoiam transversalmente a formulação de uma política direcionada à ciência aberta e suas práticas científicas.

A ideia preliminar não se reduz à prescrição de uma política de informação, mas substanciar a elaboração de diretrizes mínimas para a ciência aberta a partir das dimensões epistemológica, política, ético-legal-cultural, morfológica e tecnológica. Essas dimensões são estratégicas para a redução de problemas atuais e futuros relativos à gestão dos dados de pesquisa. Ambiciosamente, este estudo antecipa proposições para a consolidação da ciência aberta brasileira.

A metodologia está apoiada nos estudos de Bufrem (2013) que propõem um modelo alternativo e multidimensional para análise e discussão de pesquisas científicas. Teoricamente, o estudo está orientado pelo quarto paradigma científico baseado na *e-Science*. Tecnicamente, empregaram-se os métodos de levantamento documental composto pelo modelo científico *Data Lifecycle*, por mecanismos regulatórios internacionais e pesquisa bibliográfica sobre as temáticas *e-Science*, políticas, ética, cultura, epistemologia e morfologia.

2 METODOLOGIA

A metodologia do estudo está apoiada nas seis dimensões da pesquisa de Bufrem (2013), as quais têm como embasamento o método quadripolar. O método quadripolar é dinâmico e flexível. Foi inicialmente idealizado em 1974 por Paul de Bruyne, J. Herman e M. de Schoutheete (1977) para ser aplicado como um instrumento operacional em pesquisas de naturezas quantitativa e qualitativa. A dinâmica de investigação quadripolar resulta de uma interação entre os polos epistemológicos, teóricos, técnicos e morfológicos. Esses polos

se desdobram em quatro eixos compostos pelos enfoques, quadros de referência, quadros de análise e modelos de investigação.

A trajetória de estudos e exploração do método quadripolar permitiu que Bufrem (2013) elaborasse um modelo alternativo e multidimensional para análise e discussão de pesquisas científicas, ressaltando natureza, gêneros e dimensões. A autora assume a posição epistemológica, na qual “os modos de construir e organizar conhecimento se concretizam em diferentes dimensões e níveis de complexidade”, e isso justifica uma pesquisa com o “propósito de defender a possibilidade de construção de um modelo não reducionista, mas multidimensional, para analisar e compreender os aspectos marcantes da pesquisa na área da Ciência da Informação.” (BUFREM, 2013, p. 2). A referida autora expandiu o método quadripolar ao inserir as dimensões políticas e éticas que envolvem a pesquisa científica. Foram essas seis dimensões do estudo que permitiram ampliar o pensar sobre a ciência aberta e suas questões.

Para fins desse estudo, adotaram-se cinco dimensões por entender que elas dão conta de substanciar uma política ou a elaboração de diretrizes mínimas para a ciência aberta no Brasil. A dimensão teórica foi incorporada à epistemológica ao juntar os pressupostos, conceitos e enfoques. Já a dimensão técnica foi discutida por meio da tecnológica.

Na dimensão epistemológica apresentam-se os pressupostos e a discussão conceitual sobre a ciência aberta e os dados de pesquisa. A ciência aberta é algo em construção que mobiliza distintos e conflitantes interesses e pontos de vista e que admite múltiplas interpretações. Assim, é necessário informar em qual ponto este estudo se situa.

A dimensão política expõe os pressupostos que permitem pensar políticas de informação ou diretrizes mínimas para a ciência aberta. Consideram-se o lado concreto da política relativo às normas, regulamentos, orçamentos e legislações e o lado abstrato, isto é, não formalizado em documentos, tais como, conflitos de interesses dos atores envolvidos, interpretações distintas acerca da ciência aberta, processo negociação e exercício de poder.

Na dimensão ético-legal-cultural reflete-se sobre o novo fazer científico que pressupõe o trabalho colaborativo, a preocupação com o compartilhamento dos dados de pesquisa, sensibilização para tal fazer, processo de educação e postura ética dos pesquisadores e outros profissionais diante das novas práticas científicas.

Na ciência aberta é primordial a ciberinfraestrutura tecnológica que apoia o processo científico. Optou-se por desmembrar a dimensão morfológica e criar uma dimensão específica para tratar dos aspectos tecnológicos. A dimensão morfológica apresenta a estrutura da comunicação científica baseada no quarto paradigma, a qual pressupõe a cultura de compartilhamento. O *modus operandi* da ciência modificou-se para acomodar os dados de pesquisa como ponto central. O modelo *Data Lifecycle* é usado para exemplificar as novas práticas de gestão de dados de pesquisa.

Na dimensão tecnológica apresentam-se os pressupostos que dão sustentabilidade ao desenvolvimento da ciberinfraestrutura para sistematizar a gestão dos dados em longo prazo. Exemplificará essa dimensão por meio das soluções tecnológicas que acompanham e sustentam o compartilhamento, a gestão e a preservação obedecendo à sistemática dos estágios do modelo *Data Lifecycle*.

3 (R)EVOLUÇÃO DOS PARADIGMAS DA CIÊNCIA TRADICIONAL À E-SCIENCE

A evolução da ciência é pautada por quatro revoluções que marcam a transição entre os seus paradigmas epistemológicos. As discussões acerca dos paradigmas destacam que quando há uma grande mudança no *status quo* e no *modus operandi* do fazer científico ocorre uma transição ou rompimento com práticas anteriores.

A mudança de paradigma resulta dos esforços dos cientistas que trabalham incessantemente para que práticas científicas e teorias evoluam e que venham a emergir para novas hipóteses, dinâmicas e modos de fazer e comunicar a ciência. Como afirma Bird (2013) acerca da mudança de paradigma baseada no estudo de Kuhn (2015): o fenômeno da perda, no seu

ponto de vista, elimina a imagem cumulativa tradicional do progresso. A pesquisa é revolucionada quando o paradigma existente não soluciona as anomalias importantes. A transição está na mudança *modus operandi* das práticas e validações científicas.

Nessa direção, estudiosos afirmam que a ciência moderna está experimentando o quarto paradigma científico. O primeiro paradigma ocorreu há mais de um mil anos e remonta à ciência experimental, pautada na descrição dos fenômenos naturais. O segundo paradigma ocorreu nos últimos 100 (cem) anos e tem a ciência teórica baseada nas principais descobertas teóricas, modelos e generalizações. O terceiro paradigma pautou, nas últimas décadas, a ciência computacional com base na simulação de fenômenos complexos e sustentados pelas Tecnologias de Informação e Comunicação (TICs). Atualmente, a ciência se volta para a valorização do compartilhamento dos dados de pesquisa (HEY; TANSLEY; TOLLE, 2009).

Pondera-se, assim, que não há um rompimento de paradigma, mas uma confluência e aprimoramento de teorias, métodos, modelos, práticas e funcionalidades diferentes e revalidadas. A ciência é dinâmica e resultante de um acúmulo histórico, social, cultural, teórico e científico.

O quarto paradigma surge como uma nova perspectiva para a ciência aberta. A *e-Science* se propõe a ser aberta, gratuita e de livre acesso. Tem em seu arcabouço um volume intensivo de dados de pesquisa produzidos em grande escala, apoiado pelas TICs e ferramentas abertas de *altaperformance*, além de apresentar características aprimoradas da ciência tradicional, tais como: compartilhamento, colaboração, preservação e a atribuição autoral.

Isto posto, é mister a busca por dimensões que darão sustentabilidade à formulação de políticas e desenvolvimento de práticas da ciência aberta. Tais dimensões são discutidas nas seções seguintes.

4 DIMENSÕES PARA UMA CIÊNCIA ABERTA

A ciência aberta é uma nova abordagem para investigações no contexto da comunicação científica e da ciência-cidadã. Mediante múltiplas

interpretações, diferentes nomenclaturas são usadas para o termo ciência aberta, tais como: *Open Science*, *e-Science*, *Open Research*, *Research Science* e *Data Science*. Neste artigo, por uma conveniência conceitual, serão adotados os termos *e-Science* e ciência aberta, como sinônimos, num entendimento que a ciência é permeada por um ambiente colaborativo, eletrônico, aberto e digital. A *e-Science* será estudada sob a abordagem científica, pois esta é a nova configuração das práticas científicas.

A ciência aberta é um meio e não um fim. Trata-se de um espectro mais amplo sobre comunicação científica, gerenciamento, análise, representação, curadoria, preservação de dados e colaboração entre pares. Assim, a ciência aberta passa a constituir um termo guarda-chuva, que inclui acesso livre a publicações científicas, “dados científicos abertos, ferramentas científicas abertas, *hardware* científico aberto, cadernos científicos abertos, *wikipesquisa*, ciência cidadã, educação aberta.” (ALBAGLI, 2015, p. 15).

É uma ciência voltada para grandes conjuntos de dados motivados pela crescente sofisticação de bases de dados responsáveis por um amplo conjunto de simulações, metodologias e modelos efetivados em diversas áreas do conhecimento. Por conseguinte, esta ciência conduz a novas pesquisas, investigações e hipóteses a partir do uso, reuso e reprodutibilidade dos dados de pesquisa, sob a égide do quarto paradigma, o qual fomenta o surgimento de um fazer científico baseado no compartilhamento e na colaboração.

Esse novo fazer científico pressupõe que novas regras e práticas sejam ensejadas pela comunidade científica, instituições de fomento, empresas e governo. Esse fato demanda que se pense amplamente nas dimensões epistemológica, política, ético-legal-cultural, morfológica e tecnológica que sustentam a efetivação da ciência aberta.

4.1 Dimensão epistemológica

O quarto paradigma não está substituindo os paradigmas anteriores, mas exige um conjunto diferente de habilidades e práticas para um novo fazer científico. A incapacidade de aproveitar as ferramentas sofisticadas de computadores que manipulam dados impedirá até mesmo o perito mais

altamente treinado a descobrir os *insights* que agora estão começando a entrar em foco (HEY, 2010).

As iniciativas voltadas para as práticas de gerenciamento de dados de pesquisa não são realizadas individualmente por pesquisadores, mas conduzidas por um empreendimento global e institucional promovido por agências de fomento, instituições científicas, acadêmicas, governamentais e não governamentais.

Assim, o entendimento epistemológico de um novo paradigma científico está centrado na mudança do *modus operandi* e no reuso de dados de pesquisa para novas postulações investigativas. Na *e-Science*, “os dados são a infraestrutura da ciência [...]. Além disso, a ciência está se tornando uma ciência de dados intensivos e colaborativos.” (TENOPIR et al., 2011, p.1).

Outro ponto de análise dessa dimensão se concentra na conceituação do termo ‘dados’ e suas tipificações. Desse modo, há o intuito de esclarecer conceitualmente o termo que se apresenta com diferentes terminologias, tais como: dados de pesquisa, dados científicos e dados abertos.

É sabido que a comunicação científica tradicional apresenta-se validada a partir de resultados consolidados ou em processo de investigação científica. Esses resultados são divulgados nos canais formais de publicação, encerrando ou antecipando parcialmente o ciclo tradicional do processo científico. Os dados gerados durante o processo investigativo científico são descartados ou perdidos após a conclusão da pesquisa e da consequente publicação dos artigos científicos resultantes (BORGMAN, 2015). Na *e-Science*, ao contrário, os dados são insumos primários de investigações e resultados de pesquisas. Nesse quadro, os canais formais de comunicação também merecem atenção já que podem divulgar dados de pesquisa utilizados como matéria-prima no processo científico.

O entendimento sobre o termo dado pode variar consideravelmente entre pesquisadores e áreas do conhecimento. A constatação é que os dados são gerados para diferentes propósitos, por distintas comunidades acadêmicas e científicas e por meio de diversos processos (SAYÃO; SALES, 2013).

O termo 'dado' remonta aos meados do século XVII, do idioma latim e da Filosofia. No singular, o termo é denominado de 'data' que significa 'dado' e no plural com origem latina '*datum*'. Na acepção geral, o termo 'dado' pode ser constituído de fatos representados por uma unidade de análise, a qual agrupada constitui um significado (OXFORD ENGLISH DICTIONARY, 2015).

Na comunicação científica contemporânea o termo 'dado' vem tendo uma nova acepção. Nessa direção, a *National Science Foundation* (NSF, 2005) usa o termo dados para se referir a qualquer informação que pode ser armazenada em forma digital, incluindo texto, números, imagens, vídeos, áudio, *software*, algoritmos, equações, animações, modelos, simulações geradas por vários meios da observação, computação ou experimentos.

A Comissão Europeia também considera que os dados de pesquisa se referem à informação, em fatos ou números específicos, coletados, examinados e considerados como base para o raciocínio, discussão ou cálculo. Os dados incluem “estatísticas, resultados de experimentos, medidas e observações resultantes do trabalho de campo, gravações de entrevistas e imagens.” (EUROPEAN COMMISSION, 2016, p.3). Ressalta-se que a Comissão tem como foco dados de pesquisa disponíveis no formato digital.

Já de acordo com os princípios e diretrizes estabelecidos no relatório da Organização para Cooperação e Desenvolvimento Econômico (OECD, 2007), os dados científicos de pesquisa são registros factuais utilizados como fontes primárias da investigação científica, comumente aceitos na comunidade científica como necessários para validar resultados investigativos. “Um conjunto de dados de pesquisa constitui uma representação sistemática e parcial do assunto que está sendo investigado.” (OECD, 2007, p.13). Reside nessas afirmações que os dados de pesquisa são resultantes do processo investigativo que obedeceu aos rigores científico e metodológico requisitados pela ciência.

Dados abertos são dados de pesquisa que “estão disponíveis livremente para reuso e republicação sem restrições de *copyright*, patentes ou outros mecanismos de controle de propriedade intelectual.” (SAYÃO; SALES, 2013). Podem ser originados de naturezas distintas (científicos, governamentais,

institucionais, privados e públicos), desde que estejam disponíveis gratuitamente, com permissão e licenciados para uso, cópia, *download*, análise, reuso e acessíveis também via *software*.

Na concepção da Comissão Europeia, o acesso aberto aos dados da investigação refere-se ao direito de acesso e reutilização de dados de pesquisa digitais sob os termos e condições estabelecidas no contrato de concessão (EUROPEAN COMMISSION, 2016, p.3). Contudo é necessária uma política que descreva a tipologia dos dados, norteie as práticas científicas e normas acadêmico-científicas e institucionais.

4.2 Dimensão política

O foco da comunicação científica, antes concentrado na divulgação dos resultados validados, atualmente também se preocupa com a divulgação dos dados usados na investigação científica. Contudo, qualquer mudança desse porte estimula pensar nos impactos sociais, econômicos e políticos.

A produção científica foi cooptada pela lógica capitalista com diferentes formas de apropriação da produção intelectual ao longo da construção da ciência. De acordo com Albagli (2015, p.10), a preocupação obsessiva pela propriedade intelectual, especialmente nas últimas décadas do século XX, conduziu ao “alargamento dos mecanismos de apropriação privada da produção intelectual e cultural, ampliando e aprofundando relações capitalistas de mercado para áreas que até então constituíam uma reserva social.”

Assim, qualquer proposição de mudança de paradigma enfrentará contradições, confluências e divergências de interpretação acerca dos benefícios e prejuízos. O Movimento do *Open Access* conseguiu adesão de significativa porção da comunidade científica, contudo ainda enfrenta desafios para a sua plena adoção nos âmbitos institucionais, legais e acadêmicos, visto que a cultura e a política científica de cada país podem representar um fator restritivo ou estimulador da adoção de tais ideais. Toma-se como exemplo, o lento avanço da criação do arcabouço normativo-legal brasileiro referente aos repositórios institucionais, bem como a criação de periódicos abertos e gratuitos.

Os fatores restritivos vão para além das questões tecnológicas e técnicas, mas se dão principalmente nos âmbitos políticos, legais, econômicos e culturais. As TICs e a *internet* criam a infraestrutura e dão sustentabilidade à disseminação da comunicação científica. Todavia, as TICs não resolvem as assimetrias no acesso à informação ou dados se não forem respaldadas por políticas que prezem pelas condições e regulações relativas à manutenção do ciclo de produção e distribuição de informação e dados.

Então, os pressupostos de uma ciência aberta, gratuita e de livre acesso são fontes para pensar também os pressupostos de políticas de informação para esta ciência, assim:

1. As políticas devem permitir a compreensão do paradigma científico vigente e de suas práticas. Pois o movimento em prol do compartilhamento dos dados de pesquisa representa um somatório ao processo de produção e comunicação científica estabelecido.
2. As políticas ulteriores devem ser consideradas na concepção de novas políticas e no desenho de arcabouços normativos, legais e institucionais. Políticas científicas, tecnológicas e econômicas são transversais às políticas de informação. Esta última não existe desatrelada das demais.
3. Considerar também as peculiaridades das distintas áreas de conhecimento, suas prioridades e práticas investigativas.
4. Estimular a iniciativa de uso, reuso e preservação dos dados de pesquisa como empreendimento institucional e social encorajando assim a pactuação entre pesquisadores, instituições científicas, acadêmicas, governamentais e não governamentais, setor privado, bibliotecas e demais profissionais envolvidos nesse cenário.
5. Fomentar as condições materiais, humanas e legais necessárias ao compartilhamento, uso, reuso e reprodutibilidade de dados de pesquisa como estratégia para alavancar o desenvolvimento científico, tecnológico e social do país.

Esses pressupostos servem como guias para nortear a reformulação ou elaboração de legislações, orçamentos, financiamento de projetos de pesquisa e reformas institucionais. Essas reformulações são permeadas por conflitos de interesses dos atores envolvidos bem como pela lógica do capitalismo cognitivo que representa, paradoxalmente, um impeditivo às atualizações legais e institucionais que dão suporte à ciência aberta, e por outro lado, um estimulador de práticas de compartilhamento como estratégia de barateamento e agilização da produção científica.

Assim, a ciência aberta se constitui um *locus* de poder e negociação, contradições, ideologias, culturas, saberes e conhecimentos. Os esforços da ciência aberta envolvem instâncias de “ação e decisão diferenciadas, internas e externas à ciência, que vão desde o pesquisador individual e equipes de pesquisa até o nível macro das políticas públicas e das regulações internacionais, passando pelo nível meso das instituições científicas e agências de fomento.” (ALBAGLI, 2015, p. 21).

Este trabalho compreende que a informação científica e tecnológica é um bem público e, por conseguinte, se constitui em um elemento de interesse da coletividade e merecedor de formulações normativas e legais. Isto posto, é mister compreender o arcabouço político, normativo e legal voltado à regulação da informação e de dados.

A política de informação pode ser entendida como conjunto de leis, regulações e posições doutrinárias e outras tomadas de decisões e práticas com efeitos constitutivos em toda sociedade envolvendo a produção, processamento, fluxo, acesso e uso (BRAMAN, 2009). Analogamente, uma política orientada aos dados inclui todos esses elementos e acrescenta os aspectos do reuso e reprodutibilidade.

Contudo, quando esta política é legitimada por autoridade governamental, ela passa a ser compreendida e nomeada como uma política pública de informação. Esta política compreende um conjunto de premissas, decisões e ações produzidas pelo Estado e por múltiplos agentes sociais em nome do interesse coletivo (JARDIM, 2004). O processo de elaboração de tal política se constitui tarefa hercúlea devido à abrangência de objetivos, atores e

instituições envolvidos e a própria área a qual toma a informação como insumo ou produto/serviço, tais como: ciência e tecnologia, saúde, comércio, por exemplo.

Devido à multiplicidade de atores, é intrínseco o conflito de interesses que gera uma arena, na qual o poder de cada um influencia na formulação, implementação e avaliação da política. Por esse motivo, González de Gómez (1999) compreende a política de informação sob o domínio coletivo de ação, no qual os conflitos entre formulações de objetivos, planos, orçamentos, prioridades e metas não são equacionados ou resolvidos apenas tecnicamente, mas requerem a reformulação deliberativa de princípios e regras que dependem da relação de força dos atores envolvidos.

As ações governamentais no âmbito do Brasil para o setor de informação científica e tecnológica são marcadas pela descontinuidade, modismo e dispersão de ações de caráter emergencial para situações pontuais (SILVA; GARCIA, 2014). Esse setor enfrenta problemas relativos aos processos de curadoria digital, infraestrutura tecnológica e suporte legal referente às produções científicas de acesso aberto, haja vista os problemas não resolvidos dos repositórios institucionais.

Enquanto a Comunidade Europeia e os Estados Unidos substanciam empreendimentos que sustentam infraestruturas organicamente integradas que dão suporte aos processos de curadoria digital de dados de pesquisa, no Brasil, ainda são poucas e fragmentadas as ações, agravadas pela incompreensão de suas potencialidades e pela falta de visão de futuro (SAYÃO; SALES, 2013).

Assim, ainda é incipiente o interesse político em torno dos benefícios da ciência aberta e do compartilhamento de dados. Esta falta de visão política tanto dos políticos quanto pela própria comunidade científica impede ou, no mínimo, retarda o aproveitamento dos benefícios do compartilhamento de dados de pesquisa.

No âmbito internacional, as agências financiadoras e instituições de pesquisa, nos últimos anos, emitiram declarações e políticas ressaltando a necessidade de arquivamento de dados provenientes de pesquisa, e algumas

agências de financiamento começaram a exigir que os dados de projetos financiados por eles sejam depositados em um arquivo público (ICPSR, 2012).

A prática mandatória no âmbito internacional torna os dados de pesquisa, acessíveis e públicos, e o seu processo de gestão é uma realidade ainda em construção e desafiante, porque depende de outras variáveis referentes às práticas culturais da comunidade científica e às peculiaridades das áreas de conhecimento. Todavia existe uma compreensão por parte dos pesquisadores, das agências de fomento à pesquisa e das instituições acadêmicas acerca do valor dos dados, que estes se devidamente tratados, gerenciados e preservados, podem ser potencial fonte de recursos informacionais para a pesquisa científica e para o ensino da ciência (SAYÃO; SALES, 2013).

O compartilhamento e o gerenciamento de dados de pesquisa oriundos de investigações científicas são considerados um empreendimento global. Muitas iniciativas internacionais foram ou estão em desenvolvimento para a *Science*. Como por exemplo, a adoção de práticas mandatórias e rotineiras na investigação científica em instituições internacionais, governamentais, acadêmicas e de pesquisas, em especial nos Estados Unidos, Holanda, Inglaterra e Austrália.

Nos Estados Unidos, a *National Science Foundation* (NSF, 2005) financia a iniciativa *Data Observation Network for Earth (DataONE)*, a qual objetiva abordar aspectos essenciais vinculados ao tratamento, curadoria e compartilhamento de dados científicos para uso e reuso.

Um documento endereçado aos chefes de departamentos e agências executivas do governo americano, pelo Departamento Executivo de Política de Tecnologia e Ciência da Casa Branca, denominado de *Memorandum for the Heads of Executive Departments and Agencies*, afirma que a investigação científica apoiada pelo Governo Federal catalisa avanços inovadores que impulsionam a economia. Os resultados dessa pesquisa se tornam o combustível para novas percepções e são ativos para o progresso em áreas como saúde, energia, ambiente, agricultura e segurança nacional.

Este estudo identificou as iniciativas internacionais acerca das políticas e diretrizes para a gestão e compartilhamento de dados de pesquisa a fim de entender como demais países estão formulando-as. A primeira observação diz respeito ao fato que algumas iniciativas se encaixam na noção de *soft law*, a qual se refere a regras que não são estritamente obrigatórias e nem contêm uma significação legal, mas servem como princípios para nortear comportamentos, tais como os princípios elaborados pela OECD para seus membros, os da Comissão Europeia e os Netherlands Organisation for Scientific Research (NWO, 2016).

Iniciando pela OECD (2007) *Principles and Guidelines for Access to Research Data from Public Funding*, pode-se verificar que um dos principais objetivos é promover uma cultura de abertura e compartilhamento de dados de pesquisa entre as comunidades científicas dos países membros. Apesar de os países filiados à OECD não terem a obrigação legal de seguir suas diretrizes, mas por sua adesão aos princípios da instituição e devido a interesses mútuos, acabam por aplicar em diferentes graus tais diretrizes ou princípios em suas realidades. Os princípios se constituem: abertura, flexibilidade, transparência, conformidade legal, proteção da propriedade intelectual, profissionalismo, interoperabilidade, qualidade, segurança, eficiência, responsabilização e sustentabilidade.

A Comissão Europeia por meio de sua *Directorate General for Research & Innovation* também publicou o *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*. Utilizou das declarações: *Budapest Declaration* (2002) e *Berlin Declaration* (2003) para definir o acesso como o direito de ler, baixar, imprimir, copiar, distribuir, buscar, rastrear e minerar informação e dados. E assim, propuseram o *Model Grant Agreement*, o qual estabelece no Artigo 29.2 requisitos legais sobre o acesso aberto às publicações científicas, a exemplo do depósito em repositórios de todos os artigos que passaram pelo processo *peer-review*. Também está sendo proposto nesse documento, o *Open Research Data Pilot*, como forma de estimular pesquisadores a aderirem ao depósito e ao compartilhamento de dados de pesquisa.

Na Holanda, a NWO defende que os resultados de pesquisa financiados por fundos públicos devem ser de livre acesso em todo o mundo. Por isso, um conjunto de regras mais rígidas foi elaborado para guiar a publicação científica dessas pesquisas e dos seus respectivos dados em periódicos abertos e repositórios de dados que, de alguma forma, dão acesso gratuito aos resultados de pesquisa.

Já nos Estados Unidos, há o *Memorandum for the Heads of Executive Departments and Agencies* que estabelece a política de dados abertos como estratégia de gestão da informação como ativo. Isso se constitui em uma política mandatória, na qual estabelece critérios mínimos para a interoperabilidade, plataformas abertas, investimentos necessários para adequar os sistemas existentes, responsabilidades dos atores envolvidos em todo o ciclo de vida da informação e dos dados e mecanismos de responsabilização.

No caso brasileiro, dadas as suas características territoriais continentais e suas assimetrias regionais, pensar em uma política nacional nesse momento se torna retórico. Pode-se, contudo, pensar em diretrizes mínimas que guiem a comunidade científica, agências de fomento, universidades, organismos governamentais e privados acerca da gestão de dados de pesquisa. Essas diretrizes devem considerar as dimensões epistemológicas e morfológicas da ciência aberta e ético-legal-cultural das práticas da comunidade científica brasileira.

No cenário nacional, a prática e adoção de modelos científicos voltados para a ciência aberta ocorrem por meio de algumas iniciativas nas áreas da Biodiversidade, Agropecuária e dos dados abertos governamentais. Potencialmente, a coleção de dados pode “capacitar os pesquisadores a formular novos tipos de indagações, hipóteses e a usar métodos analíticos inovadores no estudo de questões críticas para a ciência e para a sociedade.” (SAYÃO; SALES, 2013).

A segunda observação refere-se ao fato da defesa da abertura dos dados como uma forma de acelerar a produção científica, diminuir custos e possibilitar o desenvolvimento social. Diante disso, este estudo entende que é

imperioso o engajamento do governo e da comunidade científica brasileira na elaboração de diretrizes e princípios para a ciência aberta e a gestão de dados de pesquisa como estratégia de impulsionar o desenvolvimento social e econômico do país. Consideram-se, portanto, para tal empreitada as conjugações ético-legal-culturais do Brasil.

4.3 Dimensão ético-legal-cultural

As questões relacionadas com o trabalho colaborativo, a preocupação com o comportamento ético-legal-cultural, além dos direitos de propriedade intelectual são algumas das inquietações que emergem nesta dimensão. Refletir acerca de como pode ser vista a postura de pesquisadores e autores na condução de uma ciência aberta orientada aos dados de pesquisa é crucial.

Culturalmente, a pesquisa científica foi apoiada por sistemas de propriedade intelectual, de natureza privada que conduziu os mecanismos regulatórios e as legislações autorais de cada país. A recompensa autoral ainda é realizada por meio de reconhecimento científico e de transferência dos direitos de titularidade para sua exploração comercial.

Na *e-Science*, o questionamento está centrado na recompensa autoral baseada na atribuição e citação. Numa acepção generalista, os dados de pesquisa são encapsulados como bem intelectual público e comum a todos. É a ideia preliminar de uma ciência do 'comum' que reflete a noção do *commons law*, a teoria do *commons* e a economia criativa. Desenvolvida a partir da década de 1990, a corrente denominada *commons paradigms* teorizada por Elinor Claire Lin Ostrom (1990) é o marco teórico para o contexto da governança econômica e dos questionamentos acerca de que bens intelectuais comuns e que não privatizados.

Lafuente (2015) enfatiza também essa ideia afirmando ser a 'ciência do comum' e sua própria condição. Defende que essa derivação deve ser um dos imperativos da nova gestão da ciência que rejeitam as regulações impostas pelos mecanismos econômicos da propriedade intelectual que transformam o conhecimento em informação codificada. A dinâmica tradicional, segundo o

autor, gera consequências devastadoras, quando há o incentivo de vários processos de privatização do conhecimento. (LAFUENTE; ESTALELLA, 2015).

As discussões acerca dos mecanismos regulatórios de propriedade privada já são questionáveis, notadamente quando se trata de recursos comuns para toda a sociedade. Muito embora o entendimento de ciência aberta não seja novo no que se refere aos direitos de propriedade intelectual, “a revolução pós-renascentista *Open Science* foi corroída pela proprietização implacável de direitos intelectuais.” (FROSIO, 2014, p.7).

Tal proprietização mantém-se até os dias atuais, quando a indústria editorial e a iniciativa privada, que regem os mercados de conhecimento e informação, pressionam por um reforço dos códigos jurídicos que garantam o direito à propriedade intelectual. Na contramão avança o movimento da ciência aberta com o surgimento de

[...] mecanismos regulatórios visando reverter assimetrias resultantes da apropriação e distribuição privada dos conhecimentos científicos, sobretudo em áreas sensíveis e com forte apelo social, como saúde, agricultura, alimentação e meio ambiente. (ALBAGLI, 2015, p. 18).

É mister que a expropriação de uma propriedade privada e suas consequências econômicas e jurídicas são opostas à cultura do compartilhamento, colaboração, cocriação, inovação e criatividade, contrastando com a abundância e a capacidade de renovação de recursos (culturais e intelectuais) baseados no entendimento da Economia Criativa (CARVALKO; MORRIS, 2014). Pois, os recursos culturais e intelectuais, nessa perspectiva, apresentam natureza pública e universal.

Infere-se que os direitos de propriedade intelectual não contemplam apenas a esfera científica, esses também regem os instrumentos legais regulatórios para o desenvolvimento dos âmbitos sociais, culturais, técnicas, tecnológicas e industriais. As implicações no tocante aos direitos de propriedade intelectual constituem um dos gargalos da ciência aberta. Grande desafio ético-legal está em conduzir uma agenda global para o entendimento

de que a ciência aberta rompe com paradigmas de práticas de apropriação existentes.

A sua promoção vai além do espectro científico-tecnológico e técnico, mas atinge direitos sociais e difusos no sentido de uma sustentabilidade mais ampla, movendo-se para áreas de essenciais, como saúde, meio ambiente, água, alimentação, agricultura, segurança nacional, entre outras.

Assim, ainda é necessária a promoção de uma mudança cultural, através de programas de sensibilização, educação e postura ética do autor-pesquisador em compartilhar seus dados de pesquisa. Aspectos relacionados com a atribuição e a autorialidade são conduzidos em uma dinâmica diferente na ciência aberta.

A ciência aberta tem dois espectros ético-cultural-legal: 1) a transição de um paradigma econômico proprietário que reforça os direitos privados da propriedade intelectual para o paradigma do compartilhamento e, 2) a ênfase do novo paradigma científico baseado nos dados de pesquisa e na cultura do *commons*.

O acesso aos bens intelectuais e culturais, procedentes da própria ecologia social, deve caminhar para uma nova postura assentada na cultura colaborativa e do compartilhamento entre os envolvidos. A economia criativa e o *commons* visam que novas práticas e espaços interativos e colaborativos promovam a criatividade, inovação e saberes científicos, artísticos e culturais a favor e para a sociedade.

Albagli (2015) e Krokosz (2015) reforçam esse pensamento na ideia de que o paradigma subjacente do *commons* e a cultura de bens comunssãoalicerçados por práticas relacionadas com a autoria colaborativa e coletiva por meio das cocriações e cocolaborações que estão representadas na *e-science*, ciência aberta, inovação aberta, produções *peer-to-peer*, *wiki*, *crowdsourcing*, entre outras, reforçando a noção emergente da cultura colaborativa e do compartilhamento.

Tradicionalmente, o *modus operandi* da pesquisa científica era direcionado por um comportamento individual, intrínseco do pesquisador-autor e restrito ao ambiente local e nacional entre pares. A comunicação formal é o

principal meio de publicização de resultados de pesquisa. A intensificação das TICs e da *internet* proporcionou a propagação de publicações eletrônicas, bases de dados e repositórios institucionais. Entretanto o comportamento do pesquisador-autor mantém-se na postura intrínseca e individualista. A concepção tradicional de que a pesquisa é uma propriedade do autor, oriunda de seu labor intelectual, merece ser reconhecida como um mérito acadêmico, de ordem individual e proprietária.

Desse modo, a dinâmica da ciência aberta busca também superar essas relações de autoridade, poder e reconhecimento individual que estava centrada na figura do autor e que ainda prevalece no seio da ciência e reforça a privatização dos saberes. O que é 'autor' e o que é 'autoria' constituem a indagação do contexto da autoralidade na contemporaneidade.

Etimologicamente, o termo 'autor' tem sua origem do latim (*auctor*), derivado do verbo *augeo*, que significa 'aumentar'; o autor é o que produz, gera e inventa. É aquele que cria, causa ou dá origem a alguma coisa, especialmente, obra literária, artística ou científica (HOUAISS, 2001). Do ponto de vista morfológico, o termo é utilizado para explicar o seu significado e as diferentes funções. O autor é o sujeito responsável pela ação que pode ser representada por um objeto, uma ideia ou um fato (KROKOSZ, 2015). Tal ação pode ser empreendida em diferentes propósitos e contextos.

O entendimento doutrinário sobre as acepções do que é autor e autoria, conforme Aulete Valente [2012] e Krokosz (2015), fundamenta que 'autor' é o criador da obra literária, artística e científica. Pessoa responsável por uma invenção ou descoberta; descobridor, inventor, pessoa que faz, realiza, comete um ato ou fato. E 'autoria' é a condição de autor; o trabalho ou produção do autor em diversas formas, tais como: obra literária, música, teoria, simulação, escultura, peça teatral, marca, desenho, dado, *code*; ou seja, o resultado da criação científica, inventiva ou artística fruto da mente humana.

Na *e-Science*, optou-se denominar de pesquisador-autor como o gerador, criador ou inventor de pesquisas que produzem dados originários primários. E a condição de autoria será determinada pelo contexto (científico, governamental, cultural, artística) de onde os dados de pesquisa são oriundos.

O compartilhamento oportuniza ao pesquisador-autor a condição de autoria primária, derivada, coletiva ou colaborativa, aliada ao exercício da sua função autoral. O dueto condição de autor e função autoral determina a autoralidade.

Na cultura do compartilhamento, a visão é que o autor não é o proprietário absoluto de seus textos, escritos ou obras, mas um compilador de ideias, textos e agrupamentos de conhecimentos que geram um novo sentido e, por conseguinte, uma nova obra. Nesse sentido, Foucault (2001, p.265) afirma que “o autor não é exatamente nem o proprietário, nem o responsável por seus textos; não é nem o produtor, nem o inventor deles.”

A visão de Foucault reflete a cognição contemporânea do autor não proprietário dos seus dados de pesquisa, mas, sobretudo, um interlocutor em um espaço coletivo e colaborativo de criação. A autoria perde o *status* de propriedade e ganha o de colaboração. Dois atributos autorais estão correlacionados a essa nova dinâmica: autoria originária e autoria colaborativa. É uma mudança da cultura do *status* da autoralidade.

A autoria original ou primária está relacionada aos direitos morais e à sua paternidade. O pesquisador-autor é o responsável pela geração dos dados primários gozando do prestígio da autoria primária. Na cultura da colaboração e do compartilhamento, por sua vez, a autoria colaborativa enfatiza a troca de saberes com diferentes autores e propósitos, não configurando uma propriedade autoral.

Estimula-se a competitividade sob uma dinâmica ético-cultural respeitando a originalidade dos dados de pesquisa que estão disponíveis para reuso e reprodutibilidade. Também prescreve a garantia da paternidade dos dados primários através da atribuição ao pesquisador-autor original. Promove os princípios da originalidade (quem gerou os dados primários) e da proveniência (de onde eles foram originados).

Além dos pesquisadores-autores, neste cenário, configuram-se também outros participantes, tais como: curadores, bibliotecários, indexadores, especialistas tecnológicos, dentre outros profissionais envolvidos com os meios e fins do processo, que devem receber a sua atribuição mediante a sua responsabilidade.

É mister que o cerne das atuais discussões concentre-se na atribuição autoral. O reconhecimento do direito do autor na cultura colaborativa e do compartilhamento é realçado pela identificação, atribuição e crédito autorais.

Seringhaus e Gerstein (2007) e Borgman (2012) afirmam que a atribuição e o reconhecimento devem ser estabelecidos. Atribuição é feita a parte responsável. Atribuição pode, portanto, ser atribuída a um investigador individual, a uma equipe de pesquisa, a uma universidade, agência de financiamento, repositório de dados, biblioteca ou a outra parte responsável pela coleta, montagem e curadoria.

O importante é estabelecer o crédito corretamente. Realizar a completa identificação do autor e do conteúdo pesquisável facilitará a atribuição e o reconhecimento autoral dos envolvidos no processo. Tal atribuição é mais estreitamente associada à noção de contribuição ou colaboração do que de autoria. A atribuição e o crédito autoral devem ser fixados dentro de sua proporcionalidade (BORGMAN, 2012; SERINGHAUS; GERSTEIN, 2007).

Entende-se que os instrumentos de propriedade intelectual, em seu atual formato, já não cabem no novo paradigma, muito embora os mecanismos norteadores do capital intelectual reforcem que a atribuição e citação são consideradas frágeis em ambientes eletrônicos e digitais. No cenário brasileiro, prevalecem ainda aspectos autorais e culturais arraigados aos processos vigentes, representados em certa medida, pela desonestidade ética, plágio, cultura do produtivismo intelectual desenfreado e da aplicação de mecanismos eficazes de proteção.

A dimensão ético-legal-cultural anda junta nesse processo de compreender que as posturas arraigadas em um paradigma proprietário ainda são presentes na ciência e que o redesenho em favor de uma cultura do *commons* e colaborativa presente na essência da ciência aberta ainda tem um longo caminho a percorrer. Essa nova dinâmica não visa à destituição do pesquisador-autor, mas de seu caráter proprietário e individualista, em prol do empoderamento da sua postura colaborativa.

4.4 Dimensão morfológica

A dimensão morfológica busca apresentar a dinâmica da *e-Science* e a sua relação com novas práticas, técnicas, modelos e tecnologias que representam a ciberinfraestrutura da comunicação científica baseada no quarto paradigma.

Como mencionado anteriormente, o *modus operandi* da pesquisa científica está em mudança e a sua estrutura requer novas dinâmicas. Exige uma representação que otimize a cultura da colaboração, compartilhamento, proporcionando a interatividade e a convergência dos dados de pesquisa para novas hipóteses. O modelo científico contemporâneo baseado na *e-Science* é o foco desta dimensão, utilizando como exemplo o modelo inovador para os fluxos de coleta, gestão e preservação de dados de pesquisa.

Os aspectos metodológicos e tecnológicos são inter-relacionados e intrínsecos ao modelo de ciclo de vida dos dados e ocorrem simultaneamente. As soluções tecnológicas são adotadas e executadas, à medida que os estágios do modelo de ciclo de vida avançam sistematicamente. A adoção de um modelo para o gerenciamento dos dados possibilita que os mesmos sejam armazenados e organizados permitindo pesquisadores ter acesso, participação, representação e análise (TENOPIR et al., 2011).

O modelo *Data Lifecycle* original se adensou por meio da introdução de diferentes desenhos que surgiram a partir de práticas institucionais. Cada realidade institucional desenvolveu um modelo baseado em suas dinâmicas internas, fluxo de trabalho e na natureza dos seus dados de pesquisa. O esforço colaborativo no desenvolvimento das práticas, além de gerar modelos, frutificou tecnologias, ferramentas abertas e guias.

Assim os modelos objetivam adotar um fluxo de procedimentos sistematizados para que os dados de pesquisa sejam tratados desde o estágio inicial da coleta até a sua disponibilidade para uso e reuso. A quantidade de estágios e o fluxo podem ser diferentes de acordo com cada modelo.

Para efeito deste artigo, foi adotado o modelo *Data Lifecycle* do *DataONE* como uma representação dessa estrutura morfológica. A escolha

desse modelo se deu por entender que o mesmo é acessível e pode atender a diversas dinâmicas institucionais e tipologias de dados.

O *DataONE* é um projeto americano que comporta uma rede de observação de dados para a terra fundamentada na nova ciência ambiental inovadora. Esta rede é representada por uma estrutura distribuída em nós e pela ciberinfraestrutura sustentável que atende às necessidades da ciência e da sociedade para o acesso aberto, persistente, robusto, seguro e de fácil descoberta aos dados observacionais da terra (DATAONE, 2005).

É representado por oito estágios do ciclo de vida dos dados. Para cada estágio há um conjunto de procedimentos sistemáticos, os quais devem ser adotados (Quadro 1).

Quadro 1- Estágios do ciclo de vida dos dados

ESTÁGIOS	DESCRIÇÕES DO CICLO
Primeiro estágio: Planejar (<i>Plan</i>)	Está relacionado ao plano de pesquisa que será desenvolvido pelo pesquisador que optou pelo gerenciamento dos seus dados de pesquisa. Todas as etapas e procedimentos metodológicos sobre o projeto de pesquisa devem ser pensados neste estágio.
Segundo estágio: Coletar (<i>Collect</i>)	Coleta dos dados primários. Apoiar todo o processo do ciclo dos dados desde a sua origem. Os dados primários são oriundos de investigações de campo, observacional ou experimental desenvolvida por pesquisadores ou grupos de pesquisa.
Terceiro estágio: Assegurar (<i>Assure</i>)	É a adoção de critérios e procedimentos que asseguram a qualidade e validação dos dados em todos os estágios e o gerenciamento em longo prazo.
Quarto estágio: Descrever (<i>Describe</i>)	Contempla a adoção de padrões e ferramentas de metadados que proveem a descrição minuciosa dos dados de pesquisa. Acrescenta-se a infraestrutura tecnológica com a adoção de padrões, protocolos e repositórios visando ao processamento, armazenagem, curadoria, preservação e disponibilidade dos dados em curto e longo prazos. Esse estágio é um dos mais relevantes ao permitir a interoperabilidade futura dos metadados.
Quinto estágio: Preservar (<i>Preserve</i>)	Deve ser adotado o plano de preservação dos dados em curto prazo com a finalidade de estocar e preservar a coleta primária, e em longo prazo para proporcionar a reutilização dos dados no futuro para outros pesquisadores, parceiros e instituições de pesquisa. A infraestrutura tecnológica e os procedimentos apoiam a preservação de longo prazo.

Sexto estágio do ciclo: Descobrir (<i>Discover</i>)	A identificação de outros <i>datasets</i> e repositórios que podem complementar e agregar valor ao projeto de dados. Os dados e seus metadados disponíveis e localizáveis são potencialmente capazes de serem usados. Algumas estratégias e parcerias institucionais, tal como registrar o projeto de dados em diretórios e repositórios abertos são alternativas para dar maior visibilidade e otimizar o projeto.
Sétimo estágio: Integrar (<i>Integrate</i>)	A integração do projeto de dados a outros recursos de dados que possam ser capazes de proporcionar novas análises e investigações cumpre dois propósitos: 1) agregação interna dos diferentes dados ou conjunto de dados gerados na pesquisa científica em um mesmo ambiente; 2) projeto de dados ser integrado com outras fontes externas.
Oitavo estágio: Analisar (<i>Analyze</i>)	Análise do projeto de dados. Existem ferramentas e <i>software</i> que suportam a exploração, análise e visualização. Neste estágio, todo o processo de pesquisa já está completo, e o pesquisador pode fazer derivações dos seus dados através de publicações em jornais, repositórios, base de dados e câmaras de compensação de dados e portais de dados.

Fonte: Oliveira et al. (2015).

Para que toda a cadeia do ciclo ocorra de forma eficaz e dentro do planejado, o compartilhamento requer uma cuidadosa reflexão durante cada estágio do processo de desenvolvimento, incluindo: descrição e documentação dos dados; depósito e curadoria; preservação dos dados em longo prazo; e a descoberta, que consiste na publicação de dados acessíveis em câmaras de compensação de dados e portais de distribuição (DATA MANAGEMENT FACILITATES SHARING AND REUSE, 2012). Além dos procedimentos metodológicos adotados, as soluções tecnológicas apoiam cada estágio do ciclo de vida.

4.5 Dimensão tecnológica

Na *e-Science*, os pressupostos para uma ciberinfraestrutura tecnológica sustentável exige a adoção de modelos e tecnologias abertas. As soluções tecnológicas devem prover a sustentabilidade de uma ciência orientada aos dados e proporcionar a coleta, gestão, curadoria, preservação para o acesso, uso e reuso em longo prazo.

A ciberinfraestrutura tecnológica é extensa e completa. Exige um agrupamento de plataformas abertas, aplicações, padrões de metadados, protocolos de interoperabilidade, identificadores, linguagens compatíveis *web* e repositórios especializados abertos; normas que proporcionem a descrição de

qualidade e recursos automáticos para citação; e a adoção de mecanismos de atribuição de licenças públicas.

Como exemplificação da adoção de uma ciberinfraestrutura tecnológica para execução de cada estágio de ciclo de vida dos dados, adotou-se o *Data Lifecycle* do *DataONE*.

4.5.1 Primeiro estágio do ciclo: planejar (plan)

O planejamento descreve o mapeamento dos processos e ciberinfraestrutura (metodológica e tecnológica) necessária para a execução de todos os estágios do ciclo de vida. Recomenda-se como uma boa prática adotar uma ferramenta de planejamento para detalhar o projeto de dados. O *Data Management Plan Tool (DMPTool)* é uma ferramenta aberta e livre que pode ser utilizada para o cumprimento dessa fase e fornece guia de boas práticas.

Além da ferramenta de gestão, recomenda-se o uso de um identificador persistente para cada pesquisador envolvido no projeto. Essa recomendação ajuda a garantir que as atribuições de autoria sejam agregadas aos dados de pesquisa.

4.5.2 Segundo estágio do ciclo: coletar (collect)

Os dados de pesquisa são coletados de acordo com as metodologias de cada pesquisa. A natureza dos dados é relacionada com o projeto de pesquisa que a originou. Dessa forma, além da coleta, o pesquisador deve se preocupar com a curadoria em curto prazo dos dados de pesquisa. Para apoiar esta curadoria, é recomendado que se adote uma infraestrutura inicial com ferramentas de comunicação geral, a exemplo do *dropbox*, *google drive*, *software* específico, equipamentos laboratoriais que condicionem e assegurem a proteção dos dados primários.

Para a interface textual, quando necessário, recomenda-se a adoção de linguagens de marcação legíveis por máquina, tais como: HTML e XML. Tais linguagens favorecem a vinculação de dados ou conjuntos de dados e

documentos utilizando *hyperlinks* e *links* como âncoras para a vinculação das entidades.

4.5.3 Terceiro estágio do ciclo: assegurar (*assure*)

Realizados os procedimentos preliminares de qualidade e validação, migram-se os dados de pesquisa para uma ferramenta compatível que favoreça o compartilhamento, curadoria e preservação em longo prazo em uma única interface.

Estimula-se neste estágio a adoção de identificadores para os dados digitais, a exemplo do *Digital Object Identifier (DOI)*. Um sistema de identificação de objetos digitais *on-line* que, além de assegurar a qualidade e identificação, proporciona gerenciar os direitos de propriedade intelectual de dados e objetos digitais.

É relevante adotar procedimentos legais e o levantamento de licenças públicas antes que os dados de pesquisa avancem para o estágio de descrição dos metadados, visto que, ao adotar as recomendações legais, os dados de pesquisa receberão também um *status* (aberto, fechado ou embargado), o qual implicará diretamente a sua disponibilidade para futuro acesso.

Nesse estágio também deve ser considerada a plataforma ou repositório que apoiará o compartilhamento, curadoria e preservação dos dados em longo prazo. Recomenda-se que seja adotada uma plataforma gerenciável de código aberto com possibilidades de agregação, aplicações, configuração de metadados, protocolos de interoperabilidade e ferramentas *web*.

As ferramentas colaborativas podem ser pensadas para a criação de tutoriais, tais como: *wiki*, *prezi* *eslideshare*. A descrição detalhada de cada estágio com a finalidade de prover materiais para recursos educacionais e de boas práticas é recomendada.

4.5.4 Quarto estágio do ciclo: descrever (*describe*)

Esse estágio é de extrema relevância e a escolha de um padrão de metadados adequado para descrição dos dados de pesquisa é essencial. Na

literatura, já estão disponíveis inúmeros padrões de metadados específicos que podem ser adotados para recursos informacionais (impressos e eletrônicos) e para diferentes tipos de dados de pesquisa.

Após a escolha de um padrão compatível com a natureza dos dados de pesquisa é recomendado que, para a descrição de qualidade dos metadados, utilizem-se vocabulários controlados, tesouros e padrões. Vocabulários controlados ajudam na combinação de termos e na descrição de assuntos específicos. Os padrões, a exemplo do W3C-DTF, ajudam a descrever os formatos de data (período) e tempo, e o *Internet Media Type* (IMT) orienta a descrição do tipo, formato e extensão para a normatização de diferentes suportes e mídias.

Recomenda-se ainda a utilização de padrão para estabelecer o idioma, a exemplo da ISO 639-2, que define o padrão de idioma representado por 3 (três) caracteres, e a ISO 3166 para descrição das especificações das coberturas espaciais e temporais. Por fim, este estágio deve garantir a qualidade dos metadados e a geração dos *codes* de metadados que irão ser interoperabilizados.

4.5.5 Quinto estágio do ciclo: preservar (*preserve*)

A adoção de um plano de preservação dos dados primários em curto e longo prazos é recomendada neste estágio. A preservação de dados em longo prazo tem a finalidade de proporcionar *backups* e arquivamento. Proporciona os processos, tais como: conversão, reformatação e salvamento de dados para reuso no futuro, além de pensar no descarte digital em longo prazo. A preservação realizada em um repositório especializado permitirá o armazenamento em longo prazo dos dados de pesquisa.

Recomenda-se reforçar a instalação de licenças públicas abertas, tais como: *Creative Commons* e *Open Database License*. As licenças preveem a revisão e aprovação de licenciamento público autoral para diferentes contextos com atribuições para documentos, códigos, dados e não atribuições. Este estágio deve rever os estágios já realizados a fim de assegurar que a integração seja realizada com qualidade.

4.5.6 Sexto estágio do ciclo: descobrir (*discover*)

Estimula-se a elaboração de parcerias institucionais para garantir que os dados tenham uma maior visibilidade da pesquisa e proporcione o compartilhamento e a colaboração científica. Recomenda-se a adoção de repositórios de dados, *datasets* e registros em diretórios relacionados à curadoria e preservação de dados de pesquisa. Reforça-se a importância dos metadados estarem disponíveis para serem localizados e potencialmente capazes de serem usados. A partir da integração com o repositório especializado, é também possível acessar outros diretórios de dados.

4.5.7 Sétimo estágio do ciclo: integrar (*integrate*)

A adoção dos protocolos de interoperabilidade para a integração dos metadados tem a finalidade de prover visibilidade, novas análises e investigações neste estágio. Podem-se adotar as ferramentas disponíveis pelo *Open Archival Information System (OAIS)* com o intuito de preservar e disponibilizar, por meio de seus protocolos de interoperabilidade, dados, informações e documentos relacionados ao projeto de pesquisa. O *Open Archives Initiative (OAI)* é o protocolo de interoperabilidade adotado para tornar eficiente a disseminação dos conteúdos contidos no projeto. E o *Open Archives Initiative Object Reuse and Exchange (OAI-ORE)* é responsável pela definição de padrões, recursos e especificações que designam e auxiliam na troca de agregações de recursos da *web*.

Estas agregações devem ser adotadas para combinar os objetos distribuídos em vários tipos de mídia e recursos. O *Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)* deve ser adotado como uma opção técnica simples para prover a possibilidade de metadados serem interoperáveis e disponíveis para serviços baseados em padrões abertos HTTP e XML.

4.5.8 Oitavo estágio do ciclo: analisar (*analyze*)

Este estágio está voltado para a análise e avaliações que podem ser realizadas entre os dados de pesquisa por meio de *software*, câmaras de compensação de dados e ferramentas analíticas e métricas especializadas. Todo o processo de pesquisa já está completo, e o pesquisador pode fazer derivações dos seus dados através de publicações em jornais, repositórios, *datasets* e portais de dados. Sugere-se adotar ferramentas de alta *performance* para a análise métrica e visualização de dados.

Este estágio também apoia uma série de iniciativas e abordagens, tentando descrever e associar todo o conteúdo de uma pesquisa em um mecanismo de leitura óptica para que ele possa ser mais facilmente compartilhado e permutado para além do reuso, oportunizando a reprodutibilidade de dados e códigos abertos.

5 CONSIDERAÇÕES PRELIMINARES

Ao admitir o contexto histórico como parte constitutiva da ciência, defende-se a necessidade da reflexão sobre as práticas de pesquisa que vêm sendo concretizadas no cotidiano do Brasil.

Este estudo adota o termo dimensões para revelar as facetas inter-relacionadas e interdependentes que envolvem a ciência aberta. Ao falar sobre a epistemologia da ciência aberta, discorre-se simultaneamente sobre sua morfologia, tipologias, métodos e tecnologias. Ao tratar dos aspectos ético-legais, isso implica considerar o valor simbólico da cultura científica brasileira. E por fim, ao mencionar a faceta política, consideram-se todas as demais. Isto significa que é impossível discutir o fenômeno da ciência aberta sem ter em mente todas as suas facetas latentes e reais.

Cada dimensão apresentada tem sua relação com o objeto 'dado', pois este é o insumo e o ponto central da ciência aberta. Discutir sobre os processos de gestão, compartilhamento, colaboração, uso e reuso dos dados de pesquisa, por meio das dimensões, afastou o olhar reducionista sobre esse

objeto e conduziu o estudo para a visão multidimensional e multirrelacional da ciência aberta.

O estudo reforça a necessidade de desenvolvimento de arcabouço normativo-legal e político-institucional baseado na noção de *soft law*, inicialmente, para atingir dois propósitos: 1) mobilizar o governo e a comunidade científica para entender o que é a ciência aberta e seus benefícios e; 2) formular estratégias e diretrizes para fomentar ciberinfraestrutura necessária para a concretização da ciência aberta. Os primeiros passos podem ser incertos, mas prementes para o futuro da ciência.

As facetas teóricas e técnicas da ciência aberta ainda estão em elaboração devido às múltiplas acepções. Isso se constitui uma oportunidade de estudos na área da Ciência da Informação e da Biblioteconomia, bem como de outras áreas de conhecimento.

A agenda da ciência aberta é urgente e necessita que esforços sejam empreendidos para o desenvolvimento de uma política pública de informação que dê sustentabilidade de práticas científicas orientadas à gestão de dados de pesquisa no cenário brasileiro.

REFERÊNCIAS

ALBAGLI, Sarita. Ciência aberta em questão. In: ALBAGLI, Sarita; MACIEL, Maria Lúcia; ABDO, Alexandre Hannud (Ed.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. p. 9-26.

AULETE, Francisco Julio de Caldas; VALENTE, Antônio L. S. **Dicionário Aulete**. Rio de Janeiro: Lexicon, [2012].

BIRD, Alexander. Thomas Kuhn. In: ZALTA, Edward N. (Ed.). **The Stanford encyclopedia of philosophy**. Fall, 2013. Disponível em: <<http://plato.stanford.edu/archives/fall2013/entries/thomas-kuhn/>>. 2013. Acesso em: 11 set. 2015.

BORGMAN, Christine L. **Big data, little data, no data: scholarship in the networked world**. Cambridge: MIT Press, 2015. Online.

BORGMAN, Christine L. The conundrum of sharing research data. **Journal of the American Society for Information Science and Technology**, New York, v. 63, n. 6, p. 1059-1078, Jun. 2012. Disponível em: <<http://ez45.periodicos.capes.gov.br/doi/10.1002/asi.22634/pdf%3E>>. Acesso em: 17 abr. 2014.

BRAMAN, Sandra. **Change of state**: information, policy, and power. Cambridge, MA: MIT Press, 2009.

BRUYNE, Paul de; HERMAN, Jacques; SCHOUTHEETE, Marc de. **Dinâmica da pesquisa em ciências sociais**: os pólos da prática metodológica. Tradução de Ruth Joffily. Rio de Janeiro: F. Alves, 1977.

BUFREM, Leilah. Configurações da pesquisa em ciência da informação. **Datagrama Zero**, Rio de Janeiro, v. 14, n. 6, dez. 2013. Disponível em: <<http://basessibi.c3sl.ufpr.br/brapci/v/a/14107>>. Acesso em: 17 abr. 2014.

CARVALKO, Joseph R.; MORRIS, Cara. **The science and technology**: guidebook for Lawyers. Chicago: American Bar Association, 2014.

DATA MANAGEMENT FACILITATES SHARING AND REUSE. DataONE Education Module:Data Management. Retrieved Nov12, 2012. Disponível em: <http://www.dataone.org/sites/all/documents/L01_DataManagement.pptx>. Acesso em: 16 fev. 2015.

DATAONE. Data Observation Network For Earth. 2005. Disponível em: <<https://www.dataone.org/>>. Acesso em: 16 fev. 2015.

EUROPEAN COMMISSION. **H2020 programme**: guidelines on open access to scientific publications and research data in horizon 2020. Version 3. 25 Aug. 2016. Disponível em: <http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf>. Acesso em: 2 abr. 2016.

FROSIO, Giancarlo F. **Open access publishing**: a literature review. Nottingham: University of Nottingham, 2014.

GONZÁLEZ DE GÓMEZ, Maria Nélide. Da política de informação ao papel da informação na política contemporânea. **Revista Internacional de Estudos Políticos**: RIEP, Rio de Janeiro, v. 1, n. 1, p. 63-97, abr. 1999.

HEY, Tony. The big idea: the next scientific revolution. **Harvard Business Review**, Brighton, Nov. 2010. Disponível em: <<https://hbr.org/2010/11/the-big-idea-the-next-scientific-revolution>>. Acesso em: 2 abr. 2016.

HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (Ed.). Jim Gray on eScience: a transformed scientific method. In: _____. (Ed.). **The fourth paradigm: data-intensive scientific discovery**. Redmond:Microsoft Research, 2009. p. xvii-xxxi. Disponível em: <<http://digital.library.unt.edu/ark:/67531/metadc31516/>>. Acesso em: 25 abr. 2016.

HOUAISS, Antônio. **Dicionário Houaiss da língua portuguesa**. Rio de Janeiro: Objetiva, 2001.

ICPSR. Inter-University Consortium for Political and Social Research. **Guide to social science data preparation and archiving: best practice throughout the data life cycle**. 5th ed. Ann Arbor:Institute for Social Research University of Michigan, 2012. Disponível em: <<http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf>>. Acesso em: 17 fev. 2015.

JARDIM, José Maria. **Memória e políticas públicas de informação**. Rio de Janeiro: Fundação Casa de Rui Barbosa, 2004. Ciclo Memória e Ciência da Informação.

KROKOSZ, Marcelo. **Outras palavras sobre autoria e plágio**. São Paulo: Atlas, 2015.

KUHN, Thomas S. **The structure of scientific revolutions**. 2nd ed. Chicago: The University of Chicago, 1970. International Encyclopedia of Unified Science, v. 2, n. 2, p. 1-210. Disponível em: <http://projektintegracija.pravo.hr/_download/repository/Kuhn_Structure_of_Scientific_Revolutions.pdf>. Acesso em: 14 set. 2015.

LAFUENTE, Antonio; ESTALELLA, Adolfo. Modos de ciência: pública, abierta y común. In: ALBAGLI, Sarita; MACIEL, Maria Lucia; ABDO, Alexandre Hannud. (Org.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015.

NOW -NETHERLANDS ORGANISATION FOR SCIENTIFIC RESEARCH. **Open science**. Amsterdam, 2016. Disponível em: <<http://www.nwo.nl/en/policies/open+science>>. Acesso em: 20 abr. 2016.

NSF -NATIONAL SCIENCE FOUNDATION. **Long-lived digital data collections: enabling research and education in the 21st century**. Arlington: National Science Board, Sept. 2005. Disponível em: <<http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>>. Acesso em: 20 out. 2015.

OECD - ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **Principles and guidelines for access to research data from public funding**. Paris, 2007. Disponível em: <<http://www.oecd.org/sti/scitech/38500813.pdf>>. Acesso em: 20 set. 2013.

OLIVEIRA, Adriana Carla S. et al. Quarto paradigma científico: adoção do *data life cycle model* para o gerenciamento de dados científicos abertos do experimento *e-quilt prototype*. In: CONFERÊNCIA LUSO-BRASILEIRA-CONFOA, 6., Salvador, 2015. **Anais...** Salvador: UFB, 2015. Disponível em: <<https://drive.google.com/folderview?id=0B2CiJWPT1NhwcmNiZ1BwZWY5TVU&usp=sharing>>. Acesso em: 20 abr. 2016.

OSTROM, Elinor Claire Lin. **Governing the commons**: the evolution of institutions for collective action. Cambridge: Cambridge University Press, 1990. OXFORD ENGLISH DICTIONARY. **Data**. Disponível em: <<http://www.oxforddictionaries.com/definition/english/data?q=data>>. Acesso em: 21 out. 2015.

SAYÃO, Luís Fernando; SALES, Luana Farias. Dados de pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país. **Tendências da Pesquisa Brasileira em Ciência da Informação**, Belo Horizonte, v. 6, n. 1, 2013. Disponível em: <<http://inseer.ibict.br/ancib/index.php/tpbci/article/viewArticle/102>>. Acesso em: 21set. 2015.

SERINGHAUS, Michael R.; GERSTEIN, Mark. B. Publishing perishing? Towards tomorrow's information architecture. **BMC Bioinformatics**, London, v. 8, n. 17, Jan. 2007. Disponível em: <<http://www.biomedcentral.com/1471-2105/8/17>>. Acesso em: 15 abr. 2016.

SILVA, Edilene Maria; GARCIA, Joana Coeli Ribeiro. Política de informação científica e tecnológica no Brasil. **Datagrama Zero**, Rio de Janeiro, v. 15, n. 4, ago. 2014.

TENOPIR, Carolet al. Data sharing by scientists: practices and perceptions. **PLoS ONE**, San Francisco, v. 6, n. 6, jun. 2011. Disponível em: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101>>. Acesso em: 18 jun. 2015.

Title

Open Science: Dimensions to a new scientific practice

Abstract:

Introduction: The practices of e-science and the use and reuse of scientific data have constituted a new scientific work that leads to the reflection on new regulatory, legal, institutional and technological frameworks for open science.

Objective: This study shows the following research question: which dimensions provide sustainability for the formulation of a policy geared to open science and its practices in the Brazilian context? The aim of this study is to discuss the dimensions that support transversely the formulation of a policy for open science and its scientific practices.

Methodology: Theoretically, the study is guided by the fourth scientific paradigm grounded in the e-Science. The methodology is supported by Bufrem's studies (2013), which propose an alternative and multidimensional model for analysis and discussion of scientific research. Technically, the literature review and documentary survey were the methods used on the Data Lifecycle scientific model, laws and international agreements. For this study purpose, five dimensions were proposed, namely: epistemological, political, ethical-legal-cultural, morphological, and technological.

Results: This study understands that these dimensions substantiate an information policy or the development of minimum guidelines for the open science agenda in Brazil.

Conclusions: The dimensions put away the reductionist perspective on survey data and they conducted the study for the multi-dimensional and multi-relational vision of open science.

Keywords: Open science. e-Science. Information policies. Collaboration. Dimensions.

Titulo

Ciencia Abierta: dimensiones a una nueva practica cientifica

Resumen

Introducción: Las prácticas de e-ciencia y uso y reuso de los datos científicos que constituyen un nuevo hacer científico que conduce a la reflexión sobre nuevos marcos normativos, legales, institucionales y tecnológicos dirigidos para una ciencia abierta.

Objective: Este estudio presenta cómo question de investigación: ¿cuáles dimensiones dan sostenibilidad a la formulación de una política orientada a la ciencia abierta y sus prácticas en el contexto brasileño? El objetivo de este estudio es discutir las dimensiones que apoyan transversalmente la formulación de una política direccionada a la ciencia abierta y su prácticas científicas.

Metodología: Teóricamente, el estudio esta orientado por el cuarto paradigma científico basado en e-Ciencia. La metodología del estudio esta apoyada en el estudio Bufrem (2013) que propone un modelo alternativo y multidimensional para el análisis y discusión de investigación científica. Técnicamente se emplearon los métodos de investigación bibliográfica y levantamiento documental sobre el modelo *Data Lifecycle*, las legislaciones y acuerdos internacionales. Para fines de este estudio se propone cinco dimensiones, a saber: epistemológica, política, ético-legal-culturales, morfológicas y la tecnológica.

Resultados: Las dimensiones sustancian una política de información o la elaboración de directrices mínimas para la agenda de la ciencia abierta en Brasil.

Conclusiones: Las dimensiones apartó la mirada reduccionista sobre datos de investigación y condució el estudio para la visión multidimensional y multirrelacional de la ciencia abierta.

Palabra Clave: Ciencia abierta. Política de información. Colaboración. Dimensiones.

Enviado em: 17.07.2016

Aceito em: 20.11.2016.