

COMPARTILHAMENTO DE DADOS NO CONTEXTO DA CIÊNCIA BRASILEIRA: UM ESTUDO INTEGRATIVO

DATA SHARING IN THE BRAZILIAN SCIENCE CONTEXT: AN INTEGRATIVE STUDY

Débora Gomes de Araújo^a

Guilherme Ataíde Dias^b

Marynice de Medeiros Matos Autran^c

RESUMO

Objetivo: O objetivo deste trabalho foi analisar os dados obtidos e integrados a partir da investigação realizada com os professores pesquisadores vinculados aos programas de pós-graduação brasileiros, na área da Ciência da Informação, e os demais abordados na pesquisa intitulada de “Práticas e percepções dos pesquisadores brasileiros”, no que tange as razões pelo não compartilhamento dos seus dados. **Metodologia:** Caracteriza-se como uma pesquisa bibliográfica e exploratória e possui uma abordagem quanti-qualitativa. Os dados foram processados e posteriormente submetidos ao teste qui-quadrado. **Resultados:** Verificou-se que as Ciências Humanas e as Ciências Sociais são as áreas com maiores desafios no contexto do compartilhamento dos dados. A falta de exigência para a publicação de dados e a carência de infraestrutura foram as principais barreiras apresentadas pelos pesquisadores. Ficou constatado que a área da Ciência da Informação necessita de uma infraestrutura que incentive os pesquisadores a compartilharem os seus dados. **Conclusões:** Conclui-se que há a necessidade de implementação de políticas mais efetivas voltadas à disponibilização de dados de pesquisa, de modo a facilitar o uso/reuso dos dados por toda comunidade científica.

Descritores: Compartilhamento de dados. Ciência da Informação. Acesso aberto. Reuso de dados.

^a Doutoranda no Programa de Pós-graduação em Ciência da Informação pela Universidade Federal da Paraíba (UFPB). E-mail: debora.g.de.araujo@gmail.com

^b Docente no Programa de Pós-Graduação em Ciência da Informação da Universidade Federal da Paraíba (UFPB). Doutor em Ciência da Informação pela Universidade de São Paulo (USP). E-mail: guilhermeataide@ccsa.ufpb.br

^c Docente no Programa de Pós-Graduação em Ciência da Informação da Universidade Federal da Paraíba (UFPB). Doutora em Informação e Comunicação em Plataformas Digitais, pela Universidade do Porto (UP). E-mail: marynice.autran@gmail.com

1 INTRODUÇÃO

O registro e a disseminação do conhecimento através de publicações científicas enfrentaram transformações no decorrer do tempo. Segundo Freitas (2006), do período da invenção da imprensa até o século XVII, a ciência era comunicada através de folhetins, em volantes e em jornais diários. Neste período, o conhecimento mais especializado era noticiado por meio de correspondências entre os cientistas ou encaminhadas às suas coletividades. Esta realidade mostra que as descobertas, mesmo de forma incipiente, eram compartilhadas entre os estudiosos da época, embora em um processo restrito.

A troca de correspondências entre os pesquisadores no século XVII contribuiu para o surgimento das publicações científicas. Estas apresentavam como um diferencial o fato de serem direcionadas para um público mais amplo, apesar de especializado. Posteriormente, os periódicos passaram a ser um novo modo de comunicar a ciência. No século XIX, eles floresceram e se especializaram (FREITAS, 2006).

Segundo Mueller (2006), o surgimento das publicações científicas eletrônicas na década de 90, do século XX, veio despertar a esperança nos pesquisadores sobre uma mudança radical no sistema tradicional da comunicação científica. A ideia do acesso aberto começou a ser materializada a partir deste momento. Assim sendo, houve o surgimento dos periódicos científicos eletrônicos de acesso livre, em que o acesso era viabilizado de forma gratuita proporcionando uma maior flexibilidade para o pesquisador. Desta forma, a autora enfatiza a importância do acesso aberto para a comunicação científica, porém adverte sobre ser uma questão desafiante, por implicar mudanças radicais no sistema tradicional e arraigado da comunicação do conhecimento científico.

As transformações ocorridas evidenciadas nesse breve histórico sobre o processo da divulgação da produção do conhecimento científico revelam que o compartilhamento do saber entre os pesquisadores passou de um processo tímido para um movimento que mesmo apresentando obstáculos a serem superados, contribuiu para a eliminação de barreiras no acesso ao conhecimento.

Nos dias atuais, temos o movimento da Ciência Aberta, cujas contribuições possibilitam que os resultados das investigações científicas sejam mais acessíveis,

incluindo o compartilhamento de códigos, dados e artigos científicos. Tal movimento é abrangente e envolve uma gama de aspectos distintos, embora diversas vezes relacionados, que geram impacto em todo o ciclo de vida da pesquisa científica, incluindo a publicação aberta, dados abertos, *software* de código aberto, registros científicos abertos, revisão por pares aberta, disseminação aberta e materiais abertos (BEZJAK *et al.*, 2018).

Cavalcanti e Sales (2017, p.77) esclarecem que “[...] a necessidade de compartilhamento dos dados que subsidiam a pesquisa” é uma questão importante na atualidade, visto que a relevância dos dados de pesquisas vem crescendo no âmbito científico mundial. Através do uso e reúso dos dados, torna-se possível replicar as pesquisas e o desenvolvimento de novos estudos.

A necessidade de compartilhamento dos dados de pesquisas é um assunto pertinente em todas as áreas do saber. Por isso, é imprescindível a conscientização entre os pesquisadores sobre a importância da disponibilização dos seus dados científicos, a criação de políticas que incentivem essa prática, o desenvolvimento tecnológico e a capacitação dos pesquisadores com ferramentas que facilitem o processo. Essas medidas podem colaborar com a credibilidade das investigações científicas, a economia dos recursos públicos e a eliminação do retrabalho na coleta de dados.

Neste contexto, pesquisas vêm sendo realizadas com o objetivo de avançar na compreensão sobre a temática dos dados de pesquisa. No âmbito nacional, Dias, Anjos e Araújo (2019) realizaram um estudo acerca da gestão dos dados de pesquisa na comunidade dos pesquisadores vinculados aos programas de pós-graduação brasileiros na área da Ciência da Informação. Vanz *et al.* (2018) publicaram um relatório intitulado de Acesso aberto a dados de pesquisa no Brasil: práticas e percepções dos pesquisadores. Este relatório apresenta resultados do projeto de pesquisa denominado de Rede de Dados de Pesquisa Brasileira (RDP Brasil), cujo arquivo do conjunto de dados é nomeado de “Práticas e Percepções dos Pesquisadores Brasileiros”, desenvolvido no âmbito da Universidade Federal do Rio Grande do Sul (UFRGS) em parceria com a Universidade Federal do Rio Grande (FURG), sob a coordenação executiva da Rede Nacional de Ensino e Pesquisa (RNP) e do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).

Diante das considerações iniciais, o presente trabalho tem como objetivo

analisar os dados obtidos e integrados a partir da investigação realizada por Dias, Anjos e Araújo (2019) com relação aos pesquisadores vinculados aos programas de pós-graduação brasileiros na área da Ciência da Informação e os pesquisadores brasileiros da pesquisa da RDP Brasil, cujos resultados estão materializados no relatório intitulado de Acesso aberto a dados de pesquisa no Brasil: práticas e percepções dos pesquisadores, no que tange as razões pelo não compartilhamento dos seus dados.

2 O COMPARTILHAMENTO DE DADOS: UMA NOVA PRÁTICA DE COMUNICAR A CIÊNCIA

Há 20 anos, Targino (2000) observou a imersão da sociedade em uma avalanche de dados provenientes de meios tradicionais, como o livro, a revista, o jornal, o rádio, a televisão (TV) ou novas redes eletrônicas de informação. Paralelamente, Hurd (2000) com um olhar visionário revelou que disponibilizar dados de forma aberta poderia contribuir com o desenvolvimento da ciência em lugares marginalizados por meio das tecnologias. No decorrer do tempo, a questão dos dados passou a fazer parte de forma pontual de um novo cenário associado com a comunicação científica, a ciência aberta.

A importância dos dados envolve a compreensão de que eles são matérias-primas essenciais para a ecologia da ciência e são fundamentais para a criação de novos ciclos do conhecimento científico, pois oferecem insumos para um processo iterativo no ciclo de vida da pesquisa, ao propiciar que sejam realizadas de forma contínua descoberta científica e inovação tecnológica (CURTY; AVENTURIER, 2017).

Nos dias atuais, a publicação dos dados ganhou destaque. O número de órgãos responsáveis pelo financiamento das pesquisas que exigem que os dados sejam localizáveis, acessíveis e abertos continua a crescer (BEZJAK *et al.*, 2018). A realidade em questão requer dos pesquisadores uma maior familiaridade com a temática, o que ainda não é uma prática comum para muitos dessa comunidade.

Monteiro e Sant'Ana (2016) revelam que, para ocorrer o compartilhamento eficaz, a gestão adequada dos dados é condição essencial para que as comunidades científicas afins e outras não pertencentes ao grupo alvo se beneficiem dos dados, especificamente se eles forem de origem observacional, em virtude de uma maior

dificuldade para replicação (BORGMAN, 2015).

Para Austin *et al.* (2017), paulatinamente há uma nova compreensão sobre a publicação dos dados assegurar a sua preservação a longo prazo. Este fato possibilita que a pesquisa seja reproduzível, facilita novas descobertas e maximiza o retorno do investimento alocado ao trabalho científico.

Candela *et al.* (2015) enfatizam a publicação de dados como um pré-requisito para compartilhamento e reutilização de dados. No entanto, os esforços com o foco na divulgação dos dados ainda são insuficientes. Diversas barreiras dificultam essa publicação, tais como: a falta de atribuição e recompensas, práticas vagas de citação, questões de qualidade, assim como a carência geral de uma cultura de compartilhamento de dados.

Zimmerman *et al.* (2009) relevam os vários desafios no compartilhamento de dados, a saber: sociais, organizacionais, legais, científicos e técnicos. O reconhecimento e a recompensa por seu trabalho, segundo eles, são os principais obstáculos para o pesquisador. Neste sentido, a prática em questão apresenta numerosos entraves que precisam ser superados, mesmo ao apresentar várias razões para os cientistas não se deterem em formas tradicionais de comunicar a ciência.

Ao observar as perspectivas dos autores supracitados é evidente que a temática de compartilhamento dos dados está entrelaçada com a sua adequada gestão e publicação. Na ausência de um gerenciamento efetivo e de práticas que tornem os dados públicos, a reutilização deles é afetada, conseqüentemente prejudica as novas descobertas científicas. Esse fato pode ser superado a partir de boas práticas que se tornem comuns entre os pesquisadores.

Targino (2000) ao tratar do compartilhamento de dados, revela que os cientistas não são detentores dos resultados da pesquisa, mas sim a humanidade. Os frutos oriundos das investigações científicas são provenientes da colaboração social e, por isso, devem ser compartilhados com todos. Trata-se de adotar um comportamento de promover a ciência com menor restrição possível para posterior reutilização dos dados.

Segundo Candela *et al.* (2015), boas práticas e políticas pertinentes de compartilhamento de dados precisam ser iniciadas com a finalidade de proporcionar a disponibilidade dos dados. É necessário identificar mecanismos para gerar conscientização na comunidade científica dos conjuntos de dados disponíveis,

promover sua compreensão e reutilização de forma eficaz e eficiente.

2.1 BOAS PRÁTICAS PARA A PUBLICAÇÃO E O COMPARTILHAMENTO DE DADOS

Diversos desafios prejudicam o compartilhamento dos dados, para superar esta realidade surgiram diferentes práticas na literatura. Em estudo realizado por Zimmerman *et al.* (2009) é evidenciado que ao agregar e integrar dados dispersos é possível evitar a duplicação de esforços, além de muitas problemáticas de pesquisas terem ao seu dispor um conjunto maior de dados para trabalhar ao possibilitar análises mais sofisticadas e robustas. Outra prática evidenciada pelos autores trata-se de um plano de compartilhamento de dados que começou a ser exigido pelo *National Institutes of Health* (NIH) em 2005 para as propostas que recebiam financiamento. No Brasil, a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) passou a exigir no final de 2017 um Plano de Gestão de Dados (PGD) para algumas modalidades e chamadas.

Para o *Digital Curation Centre-DCC* (2019), os PGDs comumente apresentam os dados que serão criados e como ocorrerá esse processo, descrevem os planos de compartilhamento e preservação levando em consideração o que é apropriado, a natureza dos dados e qualquer restrição que necessite ser incorporada. O DCC fornece uma ferramenta denominada de *DMPonline* que facilita a criação, revisão e compartilhamento de PGD. A ferramenta objetiva suprir os requisitos institucionais e de financiadores. Ela é gratuita e pode ser usada por qualquer pesquisador.

Outras iniciativas de planos de gestão de dados existem, a exemplo do PGD, que é publicado na *Research Ideas and Outcomes* (RIO), da Universidade de *Cambridge*, voltado para área das ciências da vida. A Universidade de *Leeds* também oferece um modelo de PGD. Desta forma segundo o *UK Data Service - Serviço de dados do Reino Unido* (2019), dados devidamente gerenciados são fáceis de compartilhar, têm a possibilidade de reuso em novas pesquisas ou podem validar investigações já existentes.

Zimmerman *et al.* (2009) ainda discutem dois desafios relacionados às diferenças semânticas e metodológicas que prejudicam a integração dos dados – sobre as terminologias e jargões que cada disciplina tem, além do significado diferente que as palavras possuem entre os campos de estudo. Outro tipo de prática

apresentada pelos autores é padronizar os dados antes da sua coleta.

Curty e Aventurier (2017) mostram em seu trabalho três abordagens sobre a publicação de dados: Repositório de dados, publicações ampliadas e artigos de dados. As abordagens serão apresentadas a seguir.

2.1.1 Repositórios de dados científicos

Hurd (2000) ao descrever um novo modelo da comunicação científica para o ano de 2020 previu o funcionamento de repositórios de dados brutos armazenados em servidores para que outras pessoas pudessem acessar as descobertas assim como incluí-las em suas próprias pesquisas. Trata-se de uma previsão que está se concretizando nos dias atuais.

Sanchez, Vechiato e Vidotti (2019) revelam que os repositórios de dados científicos constituem um meio fundamental para demonstrar o que os pesquisadores defendem, uma vez que esses repositórios armazenam desde os dados brutos provenientes da pesquisa até a literatura produzida em vista dos resultados que foram obtidos.

Neste contexto, Monteiro (2017) mostra que os repositórios de dados científicos têm uma parcela fundamental no processo de gestão de dados. Eles tornam viável a representação, o processamento, gerenciamento, disseminação e o acesso aos dados que neles foram depositados, envolvendo métodos de curadoria, fortalecendo a colaboração científica internacional ao propiciar o compartilhamento, acesso, uso e reúso dos dados.

Existem alguns tipos de repositório de dados, conforme encontrados na literatura podemos mencionar, o institucional e o genérico. Citamos como exemplos a Universidade de *Cambridge* e a Universidade de *Leeds* na Inglaterra que possuem repositórios de dados institucionais com foco na descoberta e compartilhamento dos dados. Outros exemplos são repositórios de dados generalistas como o *Dryad*, *Figshare*, *Zenodo* e o *Dataverse*.

Para Curty e Aventurier (2017), os repositórios de dados têm suas funcionalidades voltadas para a citabilidade e encontrabilidade dos dados, porém ainda é considerado um meio menos atrativo quanto ao reconhecimento do trabalho científico no que tange a publicação, o registro e a documentação dos dados. Os

pesquisadores optam por modalidades de publicação que promovam uma maior visibilidade, resultando em mais crédito e citações. Desse modo, mesmo sendo uma ferramenta que possibilita publicar os dados para o seu devido compartilhamento, os repositórios de dados apresentam desafios que precisam ser abordados.

2.1.2 Publicações ampliadas

No final da década de 1990, Meadows (1999) ao discutir sobre o desenvolvimento da Tecnologia da Informação (TI), já previa que com o aumento da capacidade de armazenamento dos computadores, os dados brutos poderiam ser publicados como parte de um relatório de pesquisa associados com a sua análise. Esta futura realidade poderia provocar uma expansão da quantidade de informações científicas em circulação. Logo, o autor com uma visão antecipada do que ocorreria em um futuro próximo mostrou que através dos meios eletrônicos seria possível conectar os dados brutos com as informações tratadas em um processo de geração do conhecimento.

Segundo Santos e Rockembach (2018), no contexto atual de ciência aberta estão emergindo novas práticas de comunicar a ciência como as publicações ampliadas, em virtude do uso intenso dos dados. As suas características vão além do escopo de uma publicação tradicional, interligando-a aos seus próprios dados além de outros documentos e objetos digitais relacionados.

Uma publicação ampliada é uma instância de um objeto digital complexo que se caracteriza pela vinculação de artigos aos dados que subsidiaram a sua criação, bem como à outras informações que auxiliem na contextualização da pesquisa, por exemplo, informações sobre os autores, sobre os equipamentos utilizados, projetos, etc. Essas vinculações podem ter valores semânticos e estarem associados à diversas ontologias. (SAYÃO; SALES, 2016, p. 69).

No que se refere à publicação ampliada, Curty e Aventurier (2017) alertam que ela permite a conexão entre o manuscrito e os dados científicos, porém nesta modalidade os dados não são protagonistas, eles são auxiliares e dependentes do artigo científico. Esta abordagem de publicação não possibilita a citação independente de dados, pela carência de identificadores e metadados pertinentes. Assim sendo, pode apresentar barreiras ao reuso dos dados.

2.1.3 Artigos de dados (*data papers*)

Curty e Aventurier (2017) revelam que, diferentemente do que ocorrem com as publicações ampliadas, os artigos de dados colocam os dados científicos em um lugar de destaque. O seu foco está em delinear, de forma exaustiva, a coleção de dados, com as descrições do contexto, da metodologia e dos procedimentos da investigação e das possíveis aplicabilidades dos dados. A referida modalidade pode ser expressa através da publicação dos artigos de dados em periódicos científicos híbridos, receptivos aos *data papers* ou periódicos voltados à publicação de dados. Porém, a elaboração de um *data paper* necessita de recursos, a exemplo de esforço e tempo dos pesquisadores. Isso pode ser uma barreira para os envolvidos nesse processo, pela preferência na produção de vários artigos científicos, a partir de dados coletados, ao invés de os tornarem públicos.

Segundo Kratz e Strasser (2014), o tamanho e a estrutura de um *data paper* diferem entre os periódicos. No entanto, são unânimes na tendência em apresentar um formato curto e bem estruturado. Para os autores, eles são compreendidos com maior clareza, não pela presença de qualquer informação em particular, mas pela inexistência de análises ou conclusões. Portanto, sua definição está relacionada com a ausência de uma publicação prévia sobre a análise deles.

Neste contexto que envolve o foco na publicação, temos os periódicos dos dados, a exemplo da revista *Scientific Data* (da *Springer Nature*). Essa é de acesso aberto, revisada por pares e voltada para descrições de conjuntos de dados com valor científico. O seu objetivo é fomentar o compartilhamento e a reutilização de dados mais amplos e creditar aqueles que compartilham. Outro exemplo é o *Data Science Journal* que também consiste em um periódico eletrônico, de acesso aberto, revisado por pares, o qual publica artigos sobre gerenciamento, disseminação, uso e reutilização de dados e bancos de dados de pesquisa. Ele envolve diferentes áreas do conhecimento, incluindo ciência, tecnologia, ciências humanas e artes.

De acordo com Monteiro (2017), periódicos que publicam dados científicos surgiram como mais um canal da comunicação científica para que pesquisadores publiquem seus dados. Periódicos de dados são tendências ímpares para tornar público, de forma mais ampliada, os conjuntos de dados. O autor apresenta alguns periódicos com essa natureza: *Biodiversity Data Journal*, *Data in Brief*, da Elsevier,

Earth system Science Data, GeoScience Data Journal, Journal of Open Archeology Data, Journal of Open Psychology, Journal of Physical and chemical Research Data, Journal of Open Research Software e o Open Health Data.

3 METODOLOGIA

A presente pesquisa está alicerçada na investigação empreendida por Dias, Anjos e Araújo (2019), desenvolvida no ano de 2017. O estudo busca aprofundar, nos resultados que não foram explorados no referido trabalho, sobre as razões do não compartilhamento dos dados pelos professores pesquisadores vinculados aos programas de pós-graduação em Ciência da Informação brasileiros. Além disso, o trabalho faz um paralelo com os dados apresentados na RDP Brasil (2019), sobre a mesma temática, acerca das razões do não compartilhamento de dados.

No que se refere ao conjunto dos dados de Dias, Anjos e Araújo (2019), ele foi compartilhado no âmbito do grupo de estudo cadastrado no Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - *Web, Representação do Conhecimento e Ontologias (WRCO)* (esses dados ainda não foram publicados). O conjunto de dados utilizado foi referente a 66 professores pesquisadores pertencentes aos programas brasileiros de pós-graduação em Ciência da Informação que indicaram utilizar um plano de gestão de dados, porém apresentam razões para não compartilhar os seus dados de pesquisa (DIAS; ANJOS; ARAÚJO, 2019).

Para a obtenção do conjunto de dados associados à RDP Brasil (2019), foi acessado o repositório de dados Dataverse da Rede Nacional de Ensino e Pesquisa (RNP). Esse conjunto de dados encontra-se disponível em acesso aberto e utiliza a licença *Creative Commons - Atribuição 4.0 Internacional*. Os dados obtidos foram provenientes de uma pesquisa do tipo *survey*, envolvendo pesquisadores de todo o Brasil. O instrumento de coleta de dados foi um questionário aplicado no ano de 2018, contendo 27 questões, respondido por 4.703 pesquisadores (VANZ *et al.*, 2018). Para fins deste estudo, fizemos um recorte com as respostas dos pesquisadores/professores, de diversas áreas do conhecimento, que responderam não compartilhar dado algum, sendo a amostra constituída de uma população de N=1105 indivíduos (RDP BRASIL, 2019). O conjunto de dados associado à pesquisa foi obtido por meio do *download* do arquivo intitulado: *dados-questionario.tab*,

disponibilizado originariamente no formato *Comma-Separated Values (CSV)*. Após o *download*, os dados RDP Brasil (2019) foram exportados para o *software* de planilha eletrônica *Excel* da empresa *Microsoft*.

Para se estabelecer uma integração entre as duas pesquisas, foram utilizadas apenas as respostas em comum sobre o fato do não compartilhamento de dados, tais como: falta de financiamento; falta de padronização dos dados; as pessoas não precisam de dados; falta tempo para compartilhar os dados; falta infraestrutura para o compartilhamento; não há exigência para a publicação dos dados; o pesquisador não tem permissão para compartilhar os dados; não sabe como compartilhar ou não tem certeza onde compartilhar.

A pesquisa caracteriza-se como exploratória, pela necessidade de aprofundar o conhecimento sobre a temática, e bibliográfica, por buscar materiais com discussões sobre o compartilhamento de dados. Apresenta ainda uma abordagem quanti-qualitativa. Os resultados da integração dos dados obtidos foram dispostos em tabelas. Para a comparação de variáveis categóricas foi utilizado o método associado à estatística inferencial denominado de qui-quadrado.

Portanto, indicamos que esse trabalho fez o reúso dos dados coletados em outras pesquisas já realizadas (VANZ *et al.*, 2018; DIAS; ANJOS; ARAÚJO, 2019), integrando, desta forma, os conjuntos de dados diversos.

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

A análise e discussão dos resultados foi dividida em três partes. Na parte I foi realizada uma análise sobre o motivo para o não compartilhamento dos dados, considerando as grandes áreas do CNPq, tais como: Agrárias, Biológicas, Engenharias, Exatas, Humanas, Letras, Saúde e Sociais. Na parte II, a investigação buscou aprofundar a compreensão acerca da temática na área da Ciência da Informação e na parte III foi realizada uma comparação entre a CI e as áreas apresentadas no relatório, conforme exibido a seguir.

4.1 PARTE I

Nesta etapa da análise, os conjuntos de dados foram integrados (a pesquisa da RDP Brasil (2019), intitulada de *Práticas e percepções dos pesquisadores*

brasileiros e a investigação da gestão de dados dos pesquisadores brasileiros na área da CI) e foram selecionados os motivos em comum para o não compartilhamento dos dados. Desta forma, depois da integração, foram considerados apenas os respondentes professores/pesquisadores de ambos os estudos que alegaram não compartilhar dado algum. Assim, os incluídos nas análises corresponderam a 66 respondentes do questionário aplicados aos pesquisadores brasileiros da CI na área das Ciências Sociais e 1105 indivíduos da RDP-Brasil (2019), compreendendo um total de 1171 indivíduos.

As Tabelas a seguir, apresentam os motivos do não compartilhamento de dados subdividido pelas grandes áreas da Ciência. Na Tabela 1, observa-se que a falta de financiamento é uma das barreiras para o compartilhamento dos dados. As maiores porcentagens obtidas correspondem às Ciências Sociais (4,87%) e Humanas (4,53%), do total de 1171 professores pesquisadores. O teste qui-quadrado não identificou diferença estatisticamente significativa entre as grandes áreas de ter ou não financiamento como impedimento no compartilhamento dos dados.

Tabela 1 – Motivo para o não compartilhamento dos dados: falta de financiamento

Área da Ciência	Falta de financiamento			Teste Qui-quadrado
	Não	Sim	TOTAL	
Agrárias	46 (3,93%)	11 (0,94%)	57 (4,87%)	X ² =9,5009 g.l.=7 p-valor=0,2187
Biológicas	88 (7,51%)	21 (1,79%)	109 (9,30%)	
Engenharias	89 (7,60%)	18 (1,54%)	117 (9,14%)	
Exatas	132 (11,27%)	27 (2,31%)	159 (13,58%)	
Humanas	206 (17,59%)	53 (4,53%)	259 (22,12%)	
Letras	31 (2,65%)	13 (1,11%)	44 (3,76%)	
Saúde	177 (15,12%)	37 (3,16%)	214 (18,28%)	
Sociais	165 (14,09%)	57 (4,87%)	222 (18,96%)	
TOTAL	934 (79,76%)	237 (20,24%)	1171 (100,00%)	

Fonte: Dados da pesquisa (2020)

Quanto à falta de padronização dos dados ser um problema para o compartilhamento dos dados, houve diferença significativa a 5%, com p-valor menor que 0,001, entre as grandes áreas da Ciência, vide Tabela 2. Dentre os respondentes que não compartilharam os dados por este motivo, a maior porcentagem é visualizada na área de Humanas (5,21%), seguida pela área de Ciências Sociais (4,87%). Zimmerman *et al.* (2009) revelam que uma das soluções para enfrentar o desafio do compartilhamento de dados é padronizá-los antes de sua coleta.

Tabela 2 – Motivo para o não compartilhamento dos dados: falta de padronização dos dados

Área da Ciência	Falta de padronização dos dados			Teste Qui-quadrado
	Não	Sim	TOTAL	
Agrárias	49 (4,18%)	8 (0,68%)	57 (4,87%)	$\chi^2=28,398$ g.l.=7 p-valor<0,001*
Biológicas	95 (8,11%)	14 (1,20%)	109 (9,30%)	
Engenharias	81 (6,92%)	26 (2,22%)	117 (9,14%)	
Exatas	108 (9,22%)	51 (4,36%)	159 (13,58%)	
Humanas	198 (16,91%)	61 (5,21%)	259 (22,12%)	
Letras	37 (3,16%)	7 (0,60%)	44 (3,76%)	
Saúde	184 (15,71%)	30 (2,56%)	214 (18,28%)	
Sociais	165 (14,09%)	57 (4,87%)	222 (18,96%)	
TOTAL	917 (78,30%)	254 (21,70%)	1171 (100,00%)	

* Significativo ao nível de 5%.

Fonte: Dados da pesquisa (2020)

Conforme a Tabela 3, apenas 54 professores/pesquisadores, que corresponde a 4,61% do total de 1171 respondentes, alegaram que as pessoas não precisam dos dados da pesquisa e, portanto, esse foi um dos motivos que os levaram a não compartilhar os dados. Quando comparados entre as grandes áreas do conhecimento, houve diferença estatística entre os indivíduos que marcaram “Sim” e “Não” no motivo para o não compartilhamento dos dados (p -valor<0,001). Infere-se que, no geral, existe consciência entre os pesquisadores sobre os dados serem necessários para as pessoas. A esse respeito, Cavalcanti e Sales (2017) enfatizam que o reuso dos dados permite que as pesquisas sejam replicadas, assim como possibilita o desenvolvimento de novos estudos. Isso gera credibilidade nas investigações científicas e evita o retrabalho.

Tabela 3 – Motivo para o não compartilhamento dos dados: as pessoas não precisam de dados

Área da Ciência	As pessoas não precisam dos dados			Teste Qui-quadrado
	Não	Sim	TOTAL	
Agrárias	54 (4,61%)	3 (0,26%)	57 (4,87%)	$\chi^2=33,552$ g.l.=7 p-valor<0,001*
Biológicas	101 (8,63%)	8 (0,68%)	109 (9,30%)	
Engenharias	98 (8,37%)	9 (0,77%)	117 (9,14%)	
Exatas	142 (12,13%)	17 (1,45%)	159 (13,58%)	
Humanas	254 (21,69%)	5 (0,43%)	259 (22,12%)	
Letras	40 (3,42%)	4 (0,34%)	44 (3,76%)	
Saúde	208 (17,76%)	6 (0,51%)	214 (18,28%)	
Sociais	220 (18,79%)	2 (0,17%)	222 (18,96%)	
TOTAL	1117 (95,40%)	54 (4,61%)	1171 (100,00%)	

* Significativo ao nível de 5%.

Fonte: Dados da pesquisa (2020)

A falta de tempo para compartilhar os dados foi um dos motivos citados por

aproximadamente 13% dos pesquisados, dentre os quais 3,50% correspondem à área de Ciências Sociais, seguido de 3,25% da área de Humanas. Também houve diferença estatística a 5% entre as grandes áreas, com o teste qui-quadrado apresentando p-valor igual a 0,0021, conforme apresentado na Tabela 4. É possível observar que, mesmo com diferenças percentuais, a questão do tempo pode ser uma barreira para os pesquisadores partilharem os seus dados. Simultaneamente a esta realidade, segundo Monteiro e Sant’Ana (2018), para ocorrer um compartilhamento efetivo é preciso que haja uma gestão correta dos dados. Este fato requer a utilização do recurso tempo. Assim, para enfrentar tal desafio, Rice e Southall (2016) falam do bibliotecário de dados como um profissional que pode oferecer um suporte para a gestão dos dados, facilitando o trabalho do pesquisador.

Tabela 4 – Motivo para o não compartilhamento dos dados: falta de tempo

Área da Ciência	Falta de tempo para compartilhar os dados			Teste Qui-quadrado
	Não	Sim	TOTAL	
Agrárias	51 (4,36%)	6 (0,51%)	57 (4,87%)	$\chi^2=22,496$ g.l.=7 p-valor=0,0021*
Biológicas	103 (8,80%)	6 (0,51%)	109 (9,30%)	
Engenharias	90 (7,69%)	17 (1,45%)	117 (9,14%)	
Exatas	133 (11,36%)	26 (2,22%)	159 (13,58%)	
Humanas	221 (18,87%)	38 (3,25%)	259 (22,12%)	
Letras	39 (3,33%)	5 (0,43%)	44 (3,76%)	
Saúde	200 (17,08%)	14 (1,20%)	214 (18,28%)	
Sociais	181 (15,46%)	41 (3,50%)	222 (18,96%)	
TOTAL	1018 (86,95%)	153 (13,07%)	1171 (100,00%)	

* Significativo ao nível de 5%.

Fonte: Dados da pesquisa (2020)

A Tabela 5 mostra que, dos motivos relatados como barreira no compartilhamento dos dados, a falta de exigências para publicação dos dados foi o que apresentou maior porcentagem (40,39% do total de 1171 respondentes). Não houve diferença significativa entre as grandes áreas da Ciência. Diante dos resultados apresentados, há evidências de conscientização da comunidade científica em relação às boas práticas e políticas de compartilhamento de dados serem algo que pode partir dos órgãos superiores que fomentam as pesquisas, conforme abordado por Candela *et al.* (2015). Uma vez que, ao promover esta compreensão, pode-se gerar a reutilização dos dados. Assim, é possível que, em um futuro próximo, o plano de gestão de dados venha ser exigido por outros financiadores de pesquisas no Brasil, além da FAPESP. Para Bezjak *et al.* (2018), tem se tornado comum a quantidade de financiadores que passam a exigir que os dados sejam fáceis de ser encontrados,

acessíveis e abertos.

Tabela 5 – Motivo para o não compartilhamento dos dados: falta de exigências para a publicação dos dados

Área da Ciência	Não há exigências para publicação dos dados			Teste Qui-quadrado
	Não	Sim	TOTAL	
Agrárias	33 (2,82%)	24 (2,05%)	57 (4,87%)	$\chi^2=11,392$ g.l.=7 p-valor=0,1224
Biológicas	61 (5,21%)	48 (4,10%)	109 (9,30%)	
Engenharias	59 (5,04%)	48 (4,10%)	117 (9,14%)	
Exatas	87 (7,43%)	72 (6,15%)	159 (13,58%)	
Humanas	159 (13,58%)	100 (8,54%)	259 (22,12%)	
Letras	35 (2,99%)	9 (0,77%)	44 (3,76%)	
Saúde	126 (10,76%)	88 (7,51%)	214 (18,28%)	
Sociais	138 (11,78%)	84 (7,17%)	222 (18,96%)	
TOTAL	698 (59,61%)	473 (40,39%)	1171 (100,00%)	

Fonte: Dados da pesquisa (2020)

O motivo de impedimento para o compartilhamento de dados, com a segunda maior porcentagem (33,05% do total de 1171 respondentes), foi à falta de infraestrutura disponível para os pesquisadores. Pela Tabela 6, é possível observar que houve diferença estatisticamente significativa (p-valor=0,0007) entre as grandes áreas, dentre as quais a maior porcentagem pode ser visualizada na área de Ciências Sociais (8,63%). Depreende-se, diante dos resultados, que sem uma infraestrutura adequada em todas as grandes áreas da ciência, respectivamente no âmbito brasileiro, não é possível fazer o gerenciamento adequado dos dados. Logo isso se torna um obstáculo para que eles sejam reutilizados em novas pesquisas.

Tabela 6 – Motivo para o não compartilhamento dos dados: falta de infraestrutura

Área da Ciência	Falta de infraestrutura para compartilhamento			Teste Qui-quadrado
	Não	Sim	TOTAL	
Agrárias	43 (3,67%)	14 (1,20%)	57 (4,87%)	$\chi^2=25,106$ g.l.=7 p-valor=0,0007*
Biológicas	82 (7,00%)	27 (2,31%)	109 (9,30%)	
Engenharias	77 (6,58%)	30 (2,56%)	117 (9,14%)	
Exatas	111 (9,48%)	48 (4,10%)	159 (13,58%)	
Humanas	168 (14,35%)	91 (7,77%)	259 (22,12%)	
Letras	29 (2,48%)	15 (1,28%)	44 (3,76%)	
Saúde	153 (13,07%)	61 (5,21%)	214 (18,28%)	
Sociais	121 (10,33%)	101 (8,63%)	222 (18,96%)	
TOTAL	784 (66,95%)	387 (33,05%)	1171 (100,00%)	

* Significativo ao nível de 5%.

Fonte: Dados da pesquisa (2020)

Alguns respondentes relataram, conforme expresso na Tabela 7, que a falta de permissão é um dos motivos para que os dados não sejam compartilhados. Entretanto, estes corresponderam apenas a 19,55% do total de 1171 indivíduos.

Houve diferença estatística a 5% entre as grandes áreas, com p-valor do teste qui-quadrado igual a 0,0007, vide Tabela 7. Desta forma, é preciso ultrapassar a barreira citada por Candela *et al.* (2015) sobre a carência de promover uma cultura de compartilhamento dos dados que envolva todos que participam das pesquisas. Segundo Targino (2000), os dados científicos devem ser visualizados não como patrimônio dos pesquisadores, mas sim da humanidade, por serem provenientes das colaborações da sociedade. Assim, é aconselhado que eles devem ser partilhados com todos.

Tabela 7 – Motivo para o não compartilhamento dos dados: falta de permissão para compartilhar os dados

Área da Ciência	O pesquisador não tem permissão para compartilhar os dados			Teste Qui-quadrado
	Não	Sim	TOTAL	
Agrárias	51 (4,36%)	6 (0,51%)	57 (4,87%)	$\chi^2=25,229$ g.l.=7 p-valor=0,0007*
Biológicas	94 (8,03%)	15 (1,28%)	109 (9,30%)	
Engenharias	79 (6,75%)	28 (2,39%)	117 (9,14%)	
Exatas	143 (12,21%)	16 (1,37%)	159 (13,58%)	
Humanas	194 (16,57%)	65 (5,55%)	259 (22,12%)	
Letras	34 (2,90%)	10 (0,85%)	44 (3,76%)	
Saúde	164 (14,01%)	50 (4,27%)	214 (18,28%)	
Sociais	183 (15,63%)	39 (3,33%)	222 (18,96%)	
TOTAL	942 (80,45%)	229 (19,55%)	1171 (100,00%)	

* Significativo ao nível de 5%.

Fonte: Dados da pesquisa (2020)

Não saber como ou não ter certeza onde compartilhar os dados, foi um dos motivos para o não compartilhamento mencionado pelos professores/pesquisadores. A Tabela 8 mostra que houve diferença estatisticamente significativa a 5% (p-valor=0,0119) entre as áreas da Ciência. Note-se que, dentre os 26,39% que alegaram tal problemática como motivo para não compartilhar dados, o grupo da área de Humanas (6,75%) e Ciências Sociais (6,06%), juntos, somam aproximadamente 13% do total de 1171 indivíduos, ou seja, metade dos que optaram pela categoria “Sim”. A publicação de dados é vista como um pré-requisito para o compartilhamento de dados (CANDELA *et al.*, 2015), assim como pode garantir a preservação deles ao longo do tempo (AUSTIN *et al.*, 2017). Neste sentido, existem algumas abordagens que permitem que os dados sejam publicados, Curty e Aventurier (2017) apresentam três opções, tais como: os repositórios de dados científicos, as publicações ampliadas e os artigos de dados, cujo foco desta última é colocar os dados como primordial. Assim, se faz necessário que os pesquisadores brasileiros, das diferentes grandes áreas do

conhecimento, tenham ciência da existência dessas abordagens e adotem essas práticas como forma de compartilhar os seus dados.

Tabela 8 – Motivo para o não compartilhamento dos dados: Não saber como ou não ter certeza onde compartilhar os dados

Área da Ciência	O pesquisador não sabe como compartilhar ou não tem certeza onde compartilhar os dados			Teste Qui-quadrado
	Não	Sim	TOTAL	
Agrárias	42 (3,59%)	15 (1,28%)	57 (4,87%)	$\chi^2=18,01$ g.l.=7 p-valor=0,0119*
Biológicas	80 (6,83%)	29 (2,48%)	109 (9,30%)	
Engenharias	81 (6,92%)	26 (2,22%)	117 (9,14%)	
Exatas	133 (11,36%)	26 (2,22%)	159 (13,58%)	
Humanas	180 (15,37%)	79 (6,75%)	259 (22,12%)	
Letras	38 (3,25%)	6 (0,51%)	44 (3,76%)	
Saúde	157 (13,41%)	57 (4,87%)	214 (18,28%)	
Sociais	151 (12,89%)	71 (6,06%)	222 (18,96%)	
TOTAL	862 (73,62%)	309 (26,39%)	1171 (100,00%)	

* Significativo ao nível de 5%.

Fonte: Dados da pesquisa (2020)

4.2 PARTE II

Nesta etapa, com a finalidade de analisar apenas a área da Ciência da Informação, foram selecionados apenas os professores/pesquisadores de ambas as pesquisas (a pesquisa da RDP -Brasil, Práticas e percepções dos pesquisadores brasileiros, cujo questionário foi enviado para os respondentes no ano de 2018 e a investigação da gestão de dados dos pesquisadores brasileiros na área da CI, com a aplicação do questionário no ano de 2017). Os respondentes alegaram não compartilhar dado algum e serem da área específica da Ciência da Informação. Foram 66 do estudo com os pesquisadores da CI e 19 da pesquisa do relatório, somando 85 respondentes no total.

Note através da Tabela 9, que apenas a falta de financiamento apresentou diferença estatisticamente significativa, a 5%, entre os questionários, com p-valor igual a 0,0373. Assim, existe divergência entre os pesquisadores do ano 2017 com os do ano 2018, quanto à falta de financiamento ser um dos motivos para o não compartilhamento dos dados da pesquisa. A Tabela 9 apresenta que dos 85 respondentes aproximadamente 33% relatam a falta de financiamento como um problema inerente ao compartilhamento dos dados quando considerados ambos os anos, dentre os quais, cerca de 31% se referem ao ano de 2017. Infere-se que no âmbito da CI a falta de financiamento pode ser uma barreira que está sendo

ultrapassada no que se refere ao impedimento no compartilhamento dos dados.

Tabela 9 – Motivos para o não compartilhamento dos dados dispostos por ano de pesquisa

Razão do não compartilhamento dos dados	Ano da pesquisa			Teste Qui-quadrado	
	2017	2018	Total		
Falta de financiamento	Sim	26 (30,59%)	2 (2,35%)	28 (32,94%)	$\chi^2=4,3354$ g.l.=1 p-valor=0,0373*
	Não	40 (47,06%)	17 (20,00%)	57 (67,06%)	
	Total	66 (77,65%)	19 (22,35%)	85 (100,00%)	
Falta de padronização dos dados	Sim	22 (25,88%)	6 (7,06%)	28 (32,94%)	$\chi^2=3,73e-31$ g.l.=1 p-valor=1,0000
	Não	44 (51,76%)	13 (15,29%)	57 (67,06%)	
	Total	66 (77,65%)	19 (22,35%)	85 (100,00%)	
As pessoas não precisam dos dados	Sim	0 (0,00%)	0 (0,00%)	0 (0,00%)	Não se aplica
	Não	66 (77,65%)	19 (22,35%)	85 (100,00%)	
	Total	66 (77,65%)	19 (22,35%)	85 (100,00%)	
Falta de tempo para compartilhar os dados	Sim	18 (21,18%)	0 (0,00%)	18 (21,18%)	Não se aplica
	Não	48 (56,47%)	19 (22,35%)	85 (100,00%)	
	Total	66 (77,65%)	19 (22,35%)	85 (100,00%)	
Falta de infraestrutura para compartilhamento	Sim	41 (48,24%)	13 (15,29%)	54 (63,53%)	$\chi^2=0,0539$ g.l.=1 p-valor=0,8163
	Não	25 (29,41%)	6 (7,06%)	31 (36,47%)	
	Total	66 (77,65%)	19 (22,35%)	85 (100,00%)	
Não há exigências para publicação dos dados	Sim	19 (22,35%)	9 (10,59%)	28 (32,94%)	$\chi^2=1,5413$ g.l.=1 p-valor=0,2144
	Não	47 (55,29%)	10 (11,76%)	57 (67,05%)	
	Total	66 (77,65%)	19 (22,35%)	85 (100,00%)	
O pesquisador não tem permissão para compartilhar os dados	Sim	2 (2,35%)	1 (1,18%)	3 (3,53%)	$\chi^2=2,98e-30$ g.l.=1 p-valor=1,0000
	Não	64 (75,29%)	18 (21,18%)	85 (100,00%)	
	Total	66 (77,65%)	19 (22,35%)	85 (100,00%)	
Não sabe como compartilhar ou não tem certeza onde compartilhar os dados	Sim	16 (18,82%)	5 (5,88%)	21 (24,70%)	$\chi^2=3,88e-31$ g.l.=1 p-valor=1,0000
	Não	50 (58,83%)	14 (16,47%)	64 (75,30%)	
	Total	66 (77,65%)	19 (22,35%)	85 (100,00%)	

* Significativo ao nível de 5%.

Fonte: Dados da pesquisa (2020)

Outro ponto importante a ser destacado é que “As pessoas não precisam dos dados” e “O pesquisador não tem permissão para compartilhar os dados” não foram considerados pelos pesquisadores como motivos relevantes para o não compartilhamento dos dados, apresentando porcentagens que variaram entre 0,00% e 3,53% para a categoria “Sim”. Esta realidade pode indicar uma conscientização dos pesquisadores brasileiros da CI com a importância da temática, uma vez que os profissionais da CI podem, conforme Sant’Ana (2016), contribuir com os usuários no acesso e utilização dos dados ao minimizar a desigualdade informacional que venha ocorrer entre os que são detentores dos dados e os que deles necessitam.

A questão que apresentou as maiores porcentagens na categoria “Sim” foi relacionada à falta de infraestrutura para compartilhamento dos dados, com porcentagem de 63,53% em ambos os questionários, do total de 85 professores/

pesquisadores da área de Ciência da Informação. Embora o percentual tenha reduzido em 2018, os resultados mostram que a infraestrutura ainda não é projetada para atender de forma satisfatória as necessidades dos pesquisadores na partilha dos seus dados. Neste sentido, Borgman (2015) destaca que os dados não possuem significado isoladamente. Eles existem dentro de uma infraestrutura de conhecimento, de uma ecologia de pessoas, práticas, instituições, objetos materiais e relacionamentos. O assunto é relevante não apenas para os estudiosos atuais e futuros, mas também para aqueles que irão usar o conhecimento criado por eles. Desta forma, é notório que a importância dos dados reside dentro de um contexto e precisam estar registrados por meio da sua preservação e gestão. O que requer uma infraestrutura adequada com capacidade humana, tecnológica, organizacional, relacional dentre outras.

4.3 PARTE III

Nesta etapa da análise foram selecionados apenas os respondentes do RDP-Brasil sobre "*Práticas e percepções dos pesquisadores brasileiros*". Os pesquisados alegaram não compartilhar dado algum. Foram 19 respondentes da área de Ciência da Informação e 1086 das demais áreas, somando 1105 indivíduos no total. Não foram usados os dados da investigação sobre a gestão de dados dos pesquisadores brasileiros na área da CI.

A Tabela 10 mostra que apenas a falta de infraestrutura apresentou diferença estatisticamente significativa, a 5%, entre o grupo de Ciência da Informação e os demais, com p-valor igual a 0,0011. Assim, existe divergência entre tais grupos, no que diz respeito à falta de infraestrutura ser um dos motivos para o não compartilhamento dos dados da pesquisa.

A Tabela 10 evidencia que, dos 19 indivíduos da Ciência da Informação, 13 alegaram que a falta de infraestrutura é um problema, o que corresponde a 68,42%. Quando tomamos os respondentes das demais áreas do conhecimento, temos que, os que consideraram a falta de infraestrutura como um problema equivalem a 333 de um total de 1086, o que corresponde a 30,66%, que é menos da metade dos 68,42% do grupo de Ciência da Informação. Isto nos leva a concluir que, em questão de proporcionalidade, a falta de infraestrutura é um problema que interfere no

compartilhamento dos dados, a qual é relatada com mais frequência pelos professores/pesquisadores da área de Ciência da Informação, quando comparados com as demais áreas.

Tabela 10 – Comparação entre a Ciência da Informação e as demais áreas do conhecimento sobre os motivos para o não compartilhamento dos dados

Razão do não compartilhamento dos dados	Área			Teste Qui-quadrado	
	Ciência da Informação	Outros	Total		
Falta de financiamento	Sim	2 (0,18%)	209 (18,91%)	213 (19,09%)	$\chi^2=0,4411$ g.l.=1 p-valor=0,5066
	Não	17 (1,54%)	877 (79,37%)	899 (80,91%)	
	Total	19 (1,72%)	1086 (98,28%)	1105 (100,00%)	
Falta de padronização dos dados	Sim	6 (0,54%)	226 (20,45%)	232 (20,99%)	$\chi^2=0,7369$ g.l.=1 p-valor=0,3906
	Não	13 (1,18%)	860 (77,83%)	873 (79,21%)	
	Total	19 (1,72%)	1086 (98,28%)	1105 (100,00%)	
As pessoas não precisam dos dados	Sim	0 (0,00%)	54 (4,89%)	54 (4,89%)	Não se aplica
	Não	19 (1,72%)	1032 (93,39%)	1051 (95,11%)	
	Total	19 (1,72%)	1086 (98,28%)	1105 (100,00%)	
Falta de tempo para compartilhar os dados	Sim	0 (0,00%)	135 (12,22%)	135 (12,22%)	Não se aplica
	Não	19 (1,72%)	951 (86,06%)	970 (87,78%)	
	Total	19 (1,72%)	1086 (98,28%)	1105 (100,00%)	
Falta de infraestrutura para compartilhamento	Sim	13 (1,18%)	333 (30,14%)	346 (31,32%)	$\chi^2=10,685$ g.l.=1 p-valor=0,0011*
	Não	6 (0,54%)	753 (68,14%)	759 (68,68%)	
	Total	19 (1,72%)	1086 (98,28%)	1105 (100,00%)	
Não há exigências para publicação dos dados	Sim	9 (0,81%)	445 (40,27%)	454 (41,08%)	$\chi^2=0,10646$ g.l.=1 p-valor=0,7442
	Não	10 (0,91%)	641 (58,01%)	651 (58,92%)	
	Total	19 (1,72%)	1086 (98,28%)	1105 (100,00%)	
O pesquisador não tem permissão para compartilhar os dados	Sim	1 (0,09%)	226 (20,45%)	227 (20,54%)	$\chi^2=1,8947$ g.l.=1 p-valor=0,1687
	Não	18 (1,63%)	860 (77,83%)	878 (79,46%)	
	Total	19 (1,72%)	1086 (98,28%)	1105 (100,00%)	
Não sabe como compartilhar ou não tem certeza onde compartilhar os dados	Sim	5 (0,45%)	288 (26,06%)	293 (26,51%)	$\chi^2=1,27e-28$ g.l.=1 p-valor=1,0000
	Não	14 (1,27%)	798 (72,22%)	812 (73,49%)	
	Total	19 (1,72%)	1086 (98,28%)	1105 (100,00%)	

* Significativo ao nível de 5%.

Fonte: Dados da pesquisa (2020)

Conforme identificado nos resultados obtidos, o problema da infraestrutura necessita ser abordado especialmente na área da CI, uma vez que segundo Sant'Ana (2016), é uma ciência que tem a capacidade de cooperar com esse cenário de acesso e uso intenso de dados ao possibilitar caminhos que venham contribuir da melhor forma possível com os atores participantes no processo de gestão de dados e a sua

otimização.

5 CONSIDERAÇÕES FINAIS

O presente estudo fez o reúso dos dados da pesquisa realizada pela RDP Brasil (2019), intitulada de “*Práticas e percepções dos pesquisadores brasileiros*” e da investigação de Dias, Anjos e Araújo (2019). A investigação envolveu os pesquisadores vinculados aos programas de pós-graduação brasileiros na área da Ciência da Informação, no que tange as razões pelo não compartilhamento dos seus dados e, com isso, estabelecer uma integração entre as duas pesquisas. Isso demonstra a importância do acesso aos dados abertos.

A partir da análise dos dados, constatou-se que as Ciências Humanas e as Ciências Sociais se destacaram como áreas de maiores desafios em relação ao compartilhamento dos dados. A falta de exigência para a publicação de dados e a falta de uma infraestrutura para compartilhá-los foram às opções que apresentaram maiores percentuais entre os respondentes das duas pesquisas.

No que se refere à Ciência da Informação, a falta de uma infraestrutura adequada se apresentou como um desafio que precisa ser superado para incentivar os pesquisadores a não “engavetarem” os seus dados, mas torná-los públicos para reutilização.

Percebeu-se também uma forte ligação entre a publicação e o compartilhamento de dados. De fato, gerar uma cultura entre os pesquisadores brasileiros sobre o compartilhamento dos seus dados passa pelo processo de comunicá-los. Para tanto, existem várias práticas, conforme abordadas no referencial teórico do referido trabalho que precisam ser esclarecidas para os pesquisadores de todas as grandes áreas do conhecimento.

REFERÊNCIAS

AUSTIN, C. C.; BLOOM, T.; DALLMEIER-TIESSEN, S.; KHODIYAR, V. K.; MURPHY, F.; NURNBERGER, A.; RAYMOND, L.; STOCKHAUSE, M.; TEDDS, J.; VARDIGAN, M.; WHYTE, A. Key components of data publishing: using current best practices to develop a reference model for data publishing. **International Journal on Digital Libraries**, [s. l.], v. 18, n. 2, p. 77-92, 2017. DOI 10.1007/s00799-016-0178-2. Disponível em: <http://search-ebSCOhost->

com.ez15.periodicos.capes.gov.br/login.aspx?direct=true&db=iih&AN=122919195&lang=pt-br&site=eds-live. Acesso em: 10 jun. 2021.

BEZJAK, S.; CLYBURN-SHERIN, A.; CONZETT, P.; FERNANDES, P.; GÖRÖGH, E.; HELBIG, K.; KRAMER, B.; LABASTIDA, I.; NIEMEYER, K.; PSOMOPOULOS, F.; ROSS-HELLAUER, T.; SCHNEIDER, R.; TENNANT, J.; VERBAKEL, E.; BRINKEN, H.; HELLER, L. **Open Science Training Handbook**. Hannover: German National Library of Science and Technology, 2018. 204 p. Disponível em: <https://www.fosteropenscience.eu/content/open-science-training-handbook/>. Acesso em: 10 dez. 2019.

BORGMAN, C. L. **Big data, little data, no data: scholarship in the networked world**. London: The MIT Press, 2015.

CANDELA, L.; CASTELLI, D.; MANGHI, P.; TANI, A. Data journals: a survey. **Journal of the Association for Information Science and Technology**, v. 66, n. 9, p. 1747-1762, 2015. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.23358>. Acesso em: 10 dez. 2019.

CAVALCANTI, M. T.; SALES, L. F. Gestão de dados de pesquisa: um panorama da atuação da União Europeia. **Biblos**, Rio Grande, v. 31, n. 1, p. 73-98, ago. 2017. DOI: <https://doi.org/10.14295/biblos.v31i1.5789>. Disponível em: <https://periodicos.furg.br/biblos/article/view/5789>. Acesso em: 5 jan. 2020.

CURTY, R. G; AVENTURIER, P. O paradigma da publicação de dados e suas diferentes abordagens. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18., 2017. **Anais [...]**. Marília: UNESP, 2017. Disponível em: <https://www.brapci.inf.br/index.php/res/v/105144>. Acesso em: 8 jan. 2020.

DIAS, G. A.; ANJOS, R. L.; ARAUJO, D. G. A gestão dos dados de pesquisa no âmbito da comunidade dos pesquisadores vinculados aos programas de pós-graduação brasileiros na área da Ciência da Informação: desvendando as práticas e percepções associadas ao uso e reúso de dados. **Liinc em Revista**, Rio de Janeiro, v. 15, n. 2, p. 5-31, nov. 2019. Disponível em: <http://revista.ibict.br/liinc/article/view/4683>. Acesso em: 15 jan.2020.

DIGITAL CURATION CENTRE. **DMPONLINE**. 2020. Disponível em: <https://dmponline.dcc.ac.uk/>. Acesso em: 15 jan. 2020.

FREITAS, M. H. Considerações acerca dos primeiros periódicos científicos brasileiros. **Ciência da Informação**, Brasília, v. 35, n. 3, p. 54-66, set./dez. 2006. Disponível em: http://www.brapci.inf.br/_repositorio/2010/03/pdf_49a367a95e_0009010.pdf. Acesso em: 5 jan. 2020.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO (FAPESP). **Plano de Gestão de Dados**. 2020. Disponível em: <http://www.fapesp.br/gestaodedados/>. Acesso em: 17 jan. 2020.

HURD, J. M. The transformation of scientific communication: a model for 2020. **Journal of the American Society for Information Science**, v. 51, n. 14, p. 1279-1283. 2000. Disponível em: [https://doi-org.ez15.periodicos.capes.gov.br/10.1002/1097-4571\(2000\)9999:9999<::AID-ASI1044>3.0.CO;2-1](https://doi-org.ez15.periodicos.capes.gov.br/10.1002/1097-4571(2000)9999:9999<::AID-ASI1044>3.0.CO;2-1). Acesso em: 10 jun. 2021.

KRATZ, J.; STRASSER, C. Data publication consensus and controversies. **F1000Research**, v. 3, n. 94, 2014. Disponível em: <https://f1000research.com/articles/3-94>. Acesso em: 11 jan. 2020.

MEADOWS, A. J. **A comunicação científica**. Tradução: Antonio Agenor Briquet de Lemos. Brasília: Briquet de Lemos, 1999.

MONTEIRO, E. C. S. A. **Direitos autorais nos repositórios de dados científicos: análise sobre os planos de gerenciamento dos dados**. 2017. 115 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2017. Disponível em: <https://repositorio.unesp.br/handle/11449/149748>. Acesso em: 5 jan. 2020.

MONTEIRO, E. C. S. A.; SANT'ANA, R. C. G. Plano de gerenciamento de dados em repositórios de dados de universidades. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 23, n. 53, p. 160-173, 2018. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v23n53p160/37296/>. ISSN 1518-2924. Acesso em: 12 jan. 2020.

MUELLER, S. P. M. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ciência da Informação**, Brasília, v. 35, n. 2, p. 27-38, maio/ago. 2006. Disponível em: <http://www.scielo.br/pdf/ci/v35n2/a04v35n2.pdf>. Acesso em: 5 jan. 2020.

REDE DE DADOS DE PESQUISA BRASILEIRA (RDP Brasil). **Práticas e Percepções dos Pesquisadores Brasileiros**. v. 2, 2019. Repositórios Piloto da Rede Nacional de Ensino e Pesquisa. Disponível em: <https://hdl.handle.net/20.500.11959/5>. Acesso em: 10 jan. 2020.

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

SANCHEZ, F. A.; VECHIATO, F. L.; VIDOTTI, S. A. B. G. Encontrabilidade da informação em repositórios de dados: uma análise do DataONE. **Informação & Informação**, Londrina, v. 24, n. 1, p. 51-79, mar. 2019. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/30725>. Acesso em: 5 jan. 2020.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da Ciência da Informação. **Informação & Informação**, Londrina, v. 21, n. 2, p. 116-142, maio/ago. 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940/20124>. Acesso em: 20 jun. 2018.

SANTOS, D. B.; ROCKEMBACH, M. Publicações ampliadas: aspectos da integração de dados de pesquisa. **Informação & Sociedade: Estudos**, João Pessoa, v. 28, n. 2, p. 257-269, maio/ago. 2018. Disponível em: <https://periodicos.ufpb.br/index.php/ies/article/view/38558>. Acesso em: 10 jun. 2021.

SAYÃO, L. F.; SALES, L. F. Curadoria digital e dados de pesquisa. **AtoZ: novas práticas em informação e conhecimento**, Curitiba, v. 5, n. 2, p. 67-71, dez. 2016. Disponível em: <https://revistas.ufpr.br/atoz/article/view/49708/30179>. Acesso em: 5 jan. 2020.

TARGINO, M. das G. COMUNICAÇÃO CIENTÍFICA: uma revisão de seus elementos básicos. **Informação & Sociedade: Estudos**, João Pessoa, v. 10, n. 2, 30 jan. 2000. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/326/248>. Acesso em: 6 jan. 2020.

UK DATA SERVICE. **Prepare and manage data**. Disponível em: <https://www.ukdataservice.ac.uk/manage-data>. Acesso em: 5 jan. 2020.

UNIVERSITY OF CAMBRIDGE. **Data repository**. Disponível em: <https://www.data.cam.ac.uk/repository>. Acesso em: 17 jan. 2020.

VANZ, S. A. S.; PASSOS, P. C. S. J.; CAREGNATO, S. E.; PAVÃO, C. G.; BORGES, N. B.; ROCHA, R. P.; GABRIEL JUNIOR, R. F.; AZAMBUJA, L. A. B. **Acesso aberto a dados de pesquisa no Brasil: práticas e percepções dos pesquisadores: relatório 2018**. Porto Alegre: UFRGS, 2018. 91 p. Disponível em: <http://hdl.handle.net/10183/185195>. Acesso em: 7 jan. 2020.

ZIMMERMAN, A.; BOS, N.; OLSON, J. S.; OLSON, G. M. The promise of data in e-research: many challenges, multiple solutions, diverse outcomes. *In*: NICHOLAS, W. J. (ed.). **E-Research: transformation in scholarly practice**. Nova York: Routledge, 2009. p. 222-239.

DATA SHARING IN THE BRAZILIAN SCIENCE CONTEXT: AN INTEGRATIVE STUDY

ABSTRACT

Objective: The objective of this work was to analyse the data obtained and integrated from a research carried out with the Brazilian researchers professors linked to Brazilian Information Science graduate programs and others addressed in the research entitled “Practices and perceptions of Brazilian researchers”, regarding the reasons for not sharing their data. **Methodology:** The survey has a quantitative and qualitative approach and it is characterized as a bibliographic and exploratory research. The data was processed and subsequently submitted to the chi-square test. **Results:** The data obtained indicated that the Humanities and Social Sciences are the areas with the greatest challenges in the context of data sharing. The lack of requirements for data publication and the lack of infrastructure were the main barriers presented by the researchers. It was found that the Information Science area needs an infrastructure that encourages researchers to share their data. **Conclusions:** It is concluded that there is a need to implement more effective policies aimed at making research data

available, in order to facilitate the use/reuse of data by the entire scientific community.

Descriptors: Data Sharing. Information Science. Open Access. Data reuse.

COMPARTIR DATOS EN EL CONTEXTO DE LA CIENCIA BRASILEÑA: UN ESTUDIO INTEGRATIVO

RESUMEN

Objetivo: El objetivo de este trabajo fue analizar los datos obtenidos e integrados a partir de la investigación realizada con los profesores de investigación vinculados a los programas de posgrado brasileños, en el área de Ciencias de la Información, y los demás abordados en la investigación titulada " Prácticas y percepciones de los investigadores brasileños", sobre las razones por las que no comparten sus datos. **Metodología:** Se caracteriza por ser una investigación bibliográfica y exploratoria y tiene un enfoque cuantitativo y cualitativo. Los datos fueron procesados y posteriormente sometidos a la prueba de chi-cuadrado. **Resultados:** Se descubrió que las Humanidades y las Ciencias Sociales son las áreas con mayores desafíos en el contexto del intercambio de datos. La falta de requisitos para la publicación de datos y la falta de infraestructura fueron las principales barreras presentadas por los investigadores. Se descubrió que el área de Ciencias de la Información necesita una infraestructura que aliente a los investigadores a compartir sus datos. **Conclusiones:** Se concluye que existe la necesidad de implementar políticas más efectivas destinadas a hacer que los datos de investigación estén disponibles, a fin de facilitar el uso/reutilización de datos por parte de toda la comunidad científica.

Descriptores: Intercambio de datos. Ciencias de la Información. Acceso abierto. Reutilización de datos.

Recibido em: 27.07.2020

Aceito em: 26.07.2021