

ARTIGO



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 reuso de dados¹

The management of research data within the community of researchers linked to Brazilian post-graduate programs in the area of Information Science: unveiling the practices and perceptions associated with the use and reuse of data

Guilherme Ataíde Dias*

Renata Lemos dos Anjos**

Débora Gomes de Araújo***

RESUMO

A pesquisa investigou as práticas e percepções associadas com a gestão de dados pelos pesquisadores na pós-graduação brasileira na área da Ciência da Informação (CI). O instrumento de pesquisa utilizado foi um questionário semiestruturado, enviado por e-mail para 341 pesquisadores vinculados aos programas de pós-graduação brasileiros em CI. Os dados obtidos foram analisados através de técnicas de estatística descritiva e análise temática. Verificou-se que as práticas de gestão de dados conduzidas pelos pesquisadores precisam ser aprimoradas e que eles possuem postura favorável com relação ao compartilhamento de dados, desde que exista algum controle formal sobre os

ABSTRACT

The research investigated the practices and perceptions associated with data management by researchers in Brazilian postgraduate programs in the Information Science (IC) area. A semi-structured survey was used as the research instrument, it was sent by e-mail to 341 researchers linked to the Brazilian postgraduate programs in CI. The data was analyzed through descriptive statistics techniques and thematic analysis. It was found that the data management practices conducted by the researchers need to be improved and that they have a favorable approach regarding data sharing, provided there is some formal control over them.

¹ Pesquisa realizada com apoio do CNPq (310825/2015-6)

* Doutor em Ciências da Comunicação (Ciência da Informação) pela USP. Universidade Federal da Paraíba – Cidade Universitária. CCSA – Programa de Pós-Graduação em Ciência da Informação. João Pessoa – PB - CEP: 58051-900. E-mail: guilhermeataide@ccsa.ufpb.br

** Mestre em Ciência da Informação pela UFPB. Universidade Federal da Paraíba – Programa de Pós-Graduação em Ciência da Informação – Cidade Universitária. CCSA – Programa de Pós-Graduação em Ciência da Informação. João Pessoa – PB - CEP: 58051-900. E-mail: renatalemosdosanjos@gmail.com

*** Mestre em Gestão de Organizações Aprendentes pela UFPB. Universidade Federal da Paraíba – Cidade Universitária. CCSA – Programa de Pós-Graduação em Ciência da Informação. João Pessoa – PB - CEP: 58051-900. E-mail: debora.g.de.araujo@gmail.com

mesmos.

Palavras-chave: Dados de Pesquisa; Compartilhamento de Dados de Pesquisa; Ciência da Informação; Tecnologia da Informação.

Keywords: Research Data; Research Data Sharing; Information Science; Information Technology.

INTRODUÇÃO

O uso e reuso dos dados de pesquisa pela comunidade de pesquisadores das mais diversas áreas do conhecimento é um fenômeno que vem se consolidando no cenário científico internacional. No Brasil existe pouca literatura científica publicada acerca de como os pesquisadores fazem uso e reuso dos dados de pesquisa e como se configuram as práticas de gestão relacionadas com este insumo por parte dos membros desta comunidade. Citamos como exemplos de trabalhos no contexto luso-brasileiro que focam nesta temática as pesquisas de Veiga (2017) e Veiga et al. (2018).

A pesquisa doravante apresentada teve como objetivo compreender de que forma a comunidade de pesquisadores vinculados aos programas de pós-graduação brasileiros na área da Ciência da Informação fazem a gestão dos dados de pesquisas associados aos seus respectivos processos de investigações científicas. A forma como os pesquisadores lidam com os dados tem impacto direto nas dinâmicas de uso e reuso destes recursos. Assim, entendemos ser importante compreender quais são as práticas e percepções destes pesquisadores com relação aos dados de pesquisa gerados em seus trabalhos (RICE; SOUTHALL, 2016). Ressaltamos também que a Ciência da Informação está classificada na Tabela de Áreas do Conhecimento do CNPq no contexto das Ciências Sociais Aplicadas (CNPQ, 2019), inserindo-se por afinidade, desta forma, em um domínio onde as questões associadas com a gestão dos dados de pesquisa são relevantes e correntemente investigadas (BORGMAN, 2015), reforçando assim a nossa justificativa para estudar a comunidade mencionada.

A motivação para investigar os pesquisadores vinculados aos programas de pós-graduação brasileiros na área da Ciência da Informação partiu também da constatação de que historicamente este grupo está relacionado com a implantação e pesquisas de soluções tecnológicas que possibilitam melhorias nos processos de informação intrínsecos as ações de disseminação de conteúdos científicos. Dentre as soluções tecnológicas que impactaram de forma positiva toda a comunidade científica e que estão associadas de forma direta ou indireta com os pesquisadores e profissionais da área da Ciência da Informação podemos mencionar de forma não exaustiva os sistemas de gerenciamento eletrônico de periódicos científicos, de editoras e de conferências, assim como os sistemas que possibilitam a implantação de repositórios eletrônicos de objetos digitais, dentre outros.

A EMERGÊNCIA DOS DADOS DE PESQUISA NO CONTEXTO CIENTÍFICO

Os dados de pesquisa são insumos essenciais para a geração do conhecimento científico, constituindo-se como a "matéria prima" através da qual os pesquisadores realizam os seus processos de investigação científica. Neste diapasão entendemos ser importante para os nossos fins de investigação definir dados de pesquisa e para tal trazermos a definição da *Organisation for Economic Co-operation and Development* (OECD):

“Dados de pesquisa” são definidos como registros factuais (resultados numéricos, registros textuais, imagens e sons) usados como fontes primárias para pesquisa científica, e que são comumente aceitos na comunidade científica como necessários para validar as descobertas da pesquisa.² (OECD, 2007, p.13).

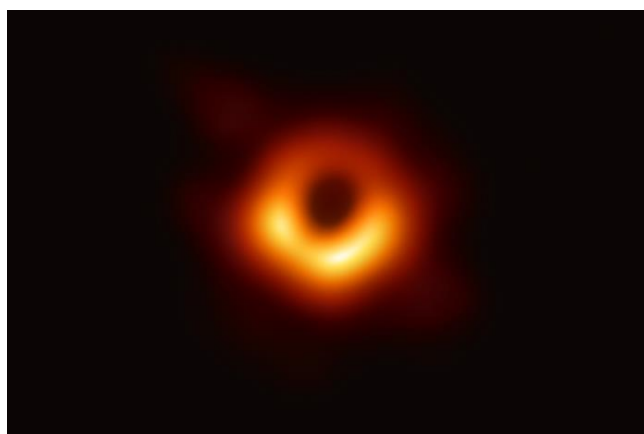
Ainda, conforme apresentado por Sayão e Sales (2015, p. 7-8) e por Borgman (2015, posição 715, *Kindle edition*) com relação a sua origem os dados podem ser classificados como observacionais, computacionais e experimentais. Os itens desta classificação podem estar relacionados com determinadas áreas do conhecimento, os dados observacionais por exemplo são bastante frequentes na área das Ciências Sociais Aplicadas.

Dados sempre foram objetos do domínio da ciência, mas por que então esta importância atribuída aos dados em nossa contemporaneidade? A resposta para esta indagação está inexoravelmente conectada com os avanços ocorridos na segunda metade do Século XX no contexto das Tecnologias da Informação e Comunicação (TIC). Com o acelerado desenvolvimento das TIC, de forma mais notória na área dos semicondutores e das redes de computadores, passou-se a gerar um volume maior de dados e também se tornou possível processá-los de forma mais otimizada. Por outro lado, a enorme quantidade de dados produzida por esta infraestrutura tecnológica criou uma situação ímpar, a tecnologia que gera estes dados é a mesma que torna possível a captura e processamento dos mesmos (BORGMAN, 2015, *Kindle edition*). Assim, como consequência deste contexto, o volume de dados gerados por dispositivos associados com a pesquisa científica chegou ao estrato dos *petabytes* (PB) e continua crescendo.

De forma ilustrativa acerca do dilúvio de dados a que estamos submetidos, podemos mencionar os dados gerados pela rede de radiotelescópios distribuída ao redor do globo associados à iniciativa *Event Horizon Telescope (ETH)* que possibilitou pela primeira vez a renderização através de computadores da imagem (Figura 1) real de um buraco negro. A quantidade de dados provenientes desta iniciativa foi tão grande que a transmissão dos dados por enlaces de rede tornou-se inviável, sendo necessário o transporte dos dados para processamento através de dispositivos físicos de armazenamento (WHITWAM, 2019).

² Texto original em Língua Inglesa: “[...] “research data” are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings.”

Figura 1: Buraco negro



FONTE: Event horizon Telescope (2019)

De forma análoga aos dados de pesquisa usados pela comunidade científica, os dados usados por grandes organizações privadas e pelos governos possuem caráter massivo e são considerados de alto valor estratégico e financeiro. Com relação ao valor financeiro dos dados, o *The Economist*, influente periódico inglês na área de economia é incisivo em comunicar que o recurso mais valioso do mundo não está associado a uma commodity como o petróleo, mas aos dados (THE ECONOMIST, 2017).

Os dados de pesquisa, a exemplo dos capturados pela iniciativa ETH, para terem seu potencial para a geração de novos conhecimentos aproveitados ao máximo, precisam ser compartilhados com as comunidades científicas. Este compartilhamento pode ser feito através de repositórios de dados de pesquisa. De forma ilustrativa e não terminativa mencionamos alguns repositórios de dados correntemente usados pela comunidade científica internacional: O *Dryad Digital Repository*³; *GenBank*⁴; *HARVARD Dataverse*⁵, dentre outros. No Brasil é notável a iniciativa do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) ao implantar uma instância do *Dataverse*.

O IBICT está disponibilizando um repositório de dados de pesquisa que cuida da preservação a longo prazo e de boas práticas de arquivamento, para que os pesquisadores possam compartilhar, manter o controle e obter o reconhecimento de seus dados. O repositório suporta o compartilhamento de dados de pesquisa com citação de dados persistentes, permitindo que eles sejam reproduzidos. O *Dataverse* é um grande repositório aberto de dados de todas as disciplinas, criado pelo Institute for Quantitative Ciências Sociais da Universidade de Harvard. O repositório *Dataverse* do IBICT fornece um meio disponível gratuitamente para depositar e descobrir conjuntos de dados específicos arquivados por colaboradores das instituições participantes da rede Cariniana. (IBICT, 2019, *online*)

³ Disponível a partir da URL <https://datadryad.org/>

⁴ Disponível a partir da URL: <https://www.ncbi.nlm.nih.gov/genbank/>

⁵ Disponível a partir da URL: <https://dataverse.harvard.edu/>

Outros exemplos de iniciativas que podemos considerar como repositórios de dados de pesquisa no Brasil incluem: o *CarpeDien* (<http://carpedien.ien.gov.br/>), associado ao Instituto de Energia Nuclear; o *PPBiodata* (<https://ppbiodata.inpa.gov.br/metacatui>), do Instituto Nacional de Pesquisas Amazônicas; o *GEOINFO* (http://inde.geoinfo.cnpm.embrapa.br/geonetwork_inde/srv/por/catalog.search) da Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) e o *SinBiota* (<http://sinbiota.biota.org.br/>), desenvolvido com apoio da FAPESP (MADEIRO, 2019).

Nos repositórios de dados de pesquisa os pesquisadores podem (devem) disponibilizar além dos seus conjuntos de dados outros recursos que contribuam para o posterior reuso dos mesmos. Estes recursos podem incluir os Planos de Gestão de Dados (PGD), os diversos produtos de *software* usados no contexto das pesquisas, bem como os metadados usados na descrição dos conjuntos de dados.

A publicação de dados em um repositório contribui para que os mesmos sejam reutilizados por outros cientistas. Este procedimento é útil para toda a comunidade por uma série de motivos, dentre os que julgamos mais significativos podemos mencionar: a possibilidade de replicar uma pesquisa já realizada; o reuso dos dados pode contribuir para minimizar os custos e a complexidade no processo de coleta de dados de outras pesquisas que investiguem um mesmo objeto científico; pesquisas que foram custeadas com recursos públicos devem ter seus dados disponibilizados de modo a aumentar o processo de transparência no uso destes recursos, além de possibilitar um incremento na geração de novos conhecimentos científicos. A publicação de dados contribui também para que novos conjuntos de dados sejam gerados por meio de integração com outros conjuntos de dados.

A disponibilização de um determinado conjunto de dados em um repositório de dados de pesquisa é uma ação que, para efetivamente configurar-se como uma iniciativa bem-sucedida, deveria estar associada a um processo de curadoria existente no âmbito da organização ou mesmo no grupo de pesquisa ao qual o cientista está vinculado. Um repositório de dados não é simplesmente um “depósito” onde os dados são armazenados, mas um dispositivo tecnológico que deveria estar integrado com um Ciclo de Vida de Dados (CVD) e seu ecossistema, seja na fase de coleta, recuperação, armazenamento ou descarte (Sant’Ana, 2016).

As opções de CVD disponíveis para os pesquisadores são as mais variadas, os critérios para a escolha de um determinado ciclo vão desde a familiaridade do pesquisador com um CVD específico até a adequabilidade do CVD ao caso concreto de pesquisa. Exemplos relevantes de CVD discutidos no âmbito da comunidade de Ciência da Informação que investiga a temática, incluem: Ciclo de vida dos dados para a Ciência da Informação - CVD-CI (SANT’ANA, 2016,); Ciclo de vida dos dados - *DataONE* (DATAONE, 2018) e o Ciclo de Vida de Dados do *Digital Curation Centre* (DCC) (DIGITAL CURATION CENTRE, 2019).

Sob a perspectiva dos dados de pesquisa no universo aberto, a partir do seu compartilhamento, uso e reuso, evidenciamos a existência da iniciativa FAIR, que apesar não ter sido abordada nesta pesquisa, merece destaque por sua importância no contexto atual. Para um maior aprofundamento da iniciativa FAIR indicamos os trabalhos de Doorn e Dillo (2016), Rodriguez-Iglesias (2016), Dias, Anjos e Rodrigues (2019), GO FAIR (2019), Henning et al. (2019) e Sales et al. (2019).

PERCURSO METODOLÓGICO

O *corpus* da pesquisa constituiu-se pelos pesquisadores vinculados aos programas brasileiros de pós-graduação na área da Ciência da Informação listados na plataforma Sucupira da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) em junho do ano de 2017. Os pesquisadores foram inicialmente contatados através de uma mensagem de correio eletrônico enviada para as suas respectivas caixas postais. Os endereços de correio eletrônico dos pesquisadores foram obtidos através de informações disponibilizadas nas páginas dos programas de pós-graduação aos quais os mesmos estavam vinculados.

O instrumento de pesquisa aplicado foi um questionário semiestruturado⁶, disseminado por correio eletrônico através da ferramenta de *software* livre *LimeSurvey*⁷. Optamos pelo uso de um questionário como instrumento de pesquisa pela possibilidade de obtermos um número razoável de dados em um tempo relativamente curto, além do fato dos respondentes estarem dispersos em uma ampla área geográfica (RICHARDSON, 2017, *online*, *Kindle edition*).

Foram enviados um total de 341 questionários. O período de aplicação e coleta de dados compreendeu o intervalo entre os meses de junho e agosto de 2017. Do total de questionários enviados, obtivemos um retorno de 189 questionários respondidos, fato este que implicou em uma taxa de retorno de 55,4%. Conforme pesquisa de Shi e Fan (2009) verificamos que a taxa de retorno de questionários enviados por correio eletrônico é de 33%. Em pesquisa conduzida por Tai et al. (2018) encontramos que a taxa de retorno de questionários disponibilizados na *Web* e cujo convite para participação foi enviado por correio eletrônico é de 34,8%. A partir destes indicadores obtidos na literatura entendemos que a taxa de retorno dos questionários de nossa pesquisa foi satisfatória. Ao conseguirmos mais da metade (55,4%) de retornos, este índice se encontra acima dos valores encontrados nas pesquisas mencionadas.

Dos 189 questionários retornados, foram considerados efetivamente 187, pois dois pesquisadores optaram por não participar da investigação após a leitura da explicação da pesquisa e do Termo de Consentimento Livre e Esclarecido (TCLE).

Na elaboração do questionário, com a exceção de uma breve explanação sobre o contexto da gestão de dados apresentada após o TCLE, optamos por não incluir para os respondentes, informações adicionais sobre o que viria a ser um Plano de Gestão de Dados (PGD), metadados de descrição, período de embargo e outros conceitos relativos à pesquisa.

Após o encerramento da etapa de preenchimento dos questionários, os dados obtidos foram exportados para um arquivo de texto no formato *comma-separated values* (CSV). Os dados foram posteriormente processados através do *software* de planilha eletrônica *Excel* e *Numbers* das empresas *Microsoft* e *Apple* respectivamente. Para o processamento estatístico dos dados coletados utilizamos uma plataforma voltada para a Ciência dos Dados denominada de *RapidMiner*⁸, o *RStudio* (RSTUDIO, 2019) e o *R* (*software* livre usado para computação estatística e geração de gráficos) (R, 2019, *online*).

⁶ Baseado no questionário intitulado “Research data management assessment: Building an understanding of your tools, data management, and training needs” desenvolvido pelo USGS.

⁷ O *LimeSurvey* pode ser obtido através da URL <https://http://www.limesurvey.org/>

⁸ O *RapidMiner* pode ser obtido através da URL <https://rapidminer.com/>

De modo a estudarmos o nível de concordância acerca do compartilhamento dos dados de pesquisa pelos pesquisadores, conduzimos uma análise de questões usando uma escala do tipo *Likert* de 5 pontos. Foram elaborados dois conjuntos de perguntas. Para cada pergunta havia cinco tipos de resposta diferentes – “Discordo fortemente”, “Discordo”, “Não discordo nem concordo”, “Concordo” e “Concordo fortemente”. Considerando que as questões tratam de itens tipo *Likert*, além da construção de gráficos e da análise descritiva usual por meio de frequências absolutas e relativas foram atribuídos escores às categorias para facilitar a análise.

Observando as relações de ordem intrínsecas às categorias da escala *Likert*, atribuímos uma pontuação para cada opção, aqui definida como Escore Bruto, no qual se atribuiu valor 1 à categoria “Discordo fortemente”, 2 para “Discordo”, 3 para “Não discordo nem concordo”, 4 para “Concordo” e 5 para “Concordo fortemente” (LIKERT, 1932).

O Escore Modificado (EM) foi calculado para viabilizar melhor interpretação dos resultados, modificando apenas a escala original (variando de 1 até 5) para o intervalo [0;100]. Esses escores são dados pela seguinte equação:

$$EM_{ij} = \frac{100(x_{ij} - 1)}{4},$$

onde x_{ij} representa o valor do Escore Bruto referente à resposta do indivíduo i para questão j , para i variando de 1 à 66, que é o tamanho da amostra desta pesquisa, e j variando de 1 à 9 que corresponde à quantidade de questões aplicadas, contemplando o conjunto de questões associados a dois enunciados. O primeiro enunciado se referia ao nível de concordância ao compartilhamento de dados pelo programa de pós-graduação ao qual o pesquisador estava vinculado. O segundo enunciado estava relacionado com relação as percepções do pesquisador com o compartilhamento de dados de uma forma geral (vide Quadro 2 e Quadro 3).

APRESENTAÇÃO E ANÁLISE DOS DADOS

As duas questões iniciais do instrumento de pesquisa tinham como objetivo caracterizar o grupo pesquisado no que diz respeito a faixa etária e o tempo de atuação na pós-graduação. Os dados relacionados com as idades dos pesquisadores e apresentados na Tabela 01 indicam que 82,8% dos respondentes possuem idade maior ou igual a 35 anos e que a faixa etária que representa o maior quantitativo de respondentes situa-se entre os 45 e 54 anos de idade.

Tabela 01: Faixa etária dos pesquisadores

Faixa etária	Quantidade de pesquisadores	Percentual
Menos de 25 anos	0	0,0%
De 25 a 34 anos	11	5,9%
De 35 a 44 anos	44	23,5%
De 45 a 54 anos	67	35,8%

Tabela 01: Faixa etária dos pesquisadores

Faixa etária	Quantidade de pesquisadores	Percentual
De 55 a 64 anos	44	23,5%
65 anos ou mais	21	11,2%
TOTAL	187	100,0%

Fonte: Dados de pesquisa

Os dados apresentados na Tabela 02 indicam a distribuição dos pesquisadores no que diz respeito ao tempo de atuação na pós-graduação em Ciência da Informação. Verificamos que a maior parte dos respondentes (38,0%) possuem pouco menos de 5 anos de atuação na pós-graduação, 27% dos respondentes situam-se na faixa compreendida entre 5 e 9 anos de atuação na pós-graduação. As duas faixas somadas correspondem a um total de 65% dos respondentes. Entendemos que o fato da maior parte dos pesquisadores estarem há no máximo nove 9 anos atuando na pós-graduação pode ser justificado pelo fato da expansão na quantidade de programas de pós-graduação na área da Ciência da Informação nos últimos 10 anos.

Tabela 02 : Tempo de atuação na pós-graduação

Tempo de atuação	Quantidade de pesquisadores	Percentual
Menos de 5 anos	71	38,0%
De 5 a 9 anos	51	27,3%
De 10 a 19 anos	46	24,6%
20 anos ou mais	19	10,2%
TOTAL	187	100,0%

Fonte: Dados de pesquisa

A questão do instrumento de pesquisa referente ao pesquisador realizar ou não a gestão dos dados coletados nas pesquisas revelou que a maior parte (62,6 %) dos respondentes (ver Tabela 03) indicam realizar a gestão dos dados coletados. Se considerarmos que a efetiva prática da gestão de dados oriundos de pesquisas objetivando o compartilhamento e reuso pela comunidade científica é uma prática ainda em processo de consolidação (SCHÖPFEL; PROST, 2016; VILAR; ZABUKOVEC, 2019) e após um olhar inicial nos dados podemos assumir que o número de pesquisadores indicados na pesquisa que realizam esta atividade é significativo, mas

alertamos que a partir das interpretações dos dados obtidos em instâncias seguintes do questionário (ver Tabela 04) este número será reduzido.

Tabela 03 : Realiza gestão dos dados coletados nas pesquisas

Realiza gestão dos dados coletados nas pesquisas	Quantidade de pesquisadores	Percentual
Sim	117	62,6%
Não	70	37,4%
TOTAL	187	100,0%

Fonte: Dados de pesquisa

Os pesquisadores que indicaram não realizar a gestão dos dados coletados nas suas respectivas pesquisas foram solicitados a apresentar o motivo para esta conduta. As respostas apresentadas pelos pesquisadores foram agrupadas em diferentes categorias, sendo possível identificar que o principal motivo indicado para a não realização da gestão dos dados coletados foi o desconhecimento de como fazê-lo (32,9%), seguido por questões culturais (12,9%) e pela falta de tempo (11,4%). O Quadro 01 traz uma amostra das respostas expressas pelos pesquisadores para não realizar a gestão dos dados coletados nas pesquisas.

Quadro 01: Amostra de razões para a não realizar a gestão de dados coletados nas pesquisas

Pesquisador	Razão para não realizar a gestão dos dados coletados nas pesquisas
65_1	"Desconheço métodos adequados de tratamento e armazenamento."
19_1	"Embora perceba a importância da preservação de dados produzidos no âmbito de atividades de pesquisa e da possibilidade de compartilhamento (reutilização) de dados científicos, ainda Não tenho conhecimento de um "lugar" para armazenamento desses dados."
62_1	"Por que nunca fui obrigada a fazer, e de fato, Não sei onde começar! A única coisa que faço é guardar todos meus dados brutos (questionários, transcrições de entrevistas) em pastas - pastas físicas e digitais - estas últimas no meu micro e no Dropbox."

Pesquisador	Razão para não realizar a gestão dos dados coletados nas pesquisas
114_1	"Tenho lido e ouvido falar no tema, porém ainda Não tomei nenhuma iniciativa a respeito. Em minha instituição Não tenho visto discussões ou iniciativas para incentivar os docentes para compartilhamento de dados científicos. Tenho lido e ouvido falar no tema, porém ainda Não tomei nenhuma iniciativa a respeito. Em minha instituição Não tenho visto discussões ou iniciativas para incentivar os docentes para compartilhamento de dados científicos."
156_1	"Porque ainda tenho pouca familiaridade sobre esse procedimento. Preciso conhecer melhor as estratégias necessárias para o preparo e disponibilização dos dados de pesquisa nos repositórios."

Fonte: Dados de pesquisa

A Tabela 04 apresenta a tipologia dos dados usados pelos pesquisadores conforme a classificação dos dados indicada por Sayão e Sales (2015, p. 7-8) e Borgman (2015, posição 715, *Kindle edition*). O campo para resposta no questionário possibilitou que os pesquisadores indicassem qualquer combinação possível entre os tipos de dados usados, ou seja: observacionais, computacionais ou experimentais. Os dados de cunho observacionais foram os mais indicados pelos pesquisadores, compreendendo um percentual de 42,7% das menções. Não causa discrepância o fato dos pesquisadores em Ciência da Informação indicarem majoritariamente o uso de dados observacionais, visto que os mesmos são bastantes associados com a área das Ciências Sociais Aplicadas. Com relação aos dados observacionais, Borgman (2015, posição 715, *Kindle edition*) afirma que os mesmos são considerados os mais importantes para preservar, visto a dificuldade de serem replicados. Em segundo lugar no quantitativo de indicações apresentadas pelos pesquisadores ficaram os dados observacionais e computacionais, com 15,5%. Os dados dos tipos observacionais, computacionais e experimentais ficaram classificados em terceiro lugar (12,0%).

Tabela 04 : Classificação dos dados usados

Classificação dos dados usados	Quantidade dos pesquisadores	Percentual
Observacionais	50	42,7%
Observacionais e computacionais	18	15,4%
Observacionais e experimentais	11	9,4%

Tabela 04 : Classificação dos dados usados

Classificação dos dados usados	Quantidade dos pesquisadores	de Percentual
Observacionais, computacionais e experimentais	14	12,0%
Computacionais	9	7,7%
Computacionais e experimentais	8	6,8%
Experimentais	7	6,0%
TOTAL	117	100,0%

Fonte: Dados de pesquisa

Consideramos que a elaboração de um PGD é uma ação fundamental do processo de gestão de dados de pesquisa, configurando-se como uma ferramenta que não deve ser ignorada, pois provê trilhas que possibilitam o uso futuro dos dados pela comunidade científica. Sayão e Sales (2015, p. 15) complementam indicando que "O PGD descreve o ciclo de vida de gestão para todos os dados que serão coletados, processados ou gerados por um projeto de pesquisa". Cientes da importância do PGD optamos por desconsiderar da pesquisa os respondentes que indicaram a não elaboração desta ferramenta. Esta decisão de pesquisa foi fundada no fato de não considerarmos que um pesquisador está efetivamente fazendo gestão de dados de pesquisa se o mesmo não elabora um PGD. De acordo com o apresentado na Tabela 05, dos 117 pesquisadores que indicaram realizar gestão de dados científicos (Tabela 03), apenas 66 dos mesmos (56,4%) elaboram um PGD no contexto das suas respectivas pesquisas. Desta forma indicamos que o número de pesquisadores que efetivamente fazem gestão dos dados da pesquisa não é significativo no âmbito da população investigada, visto que a elaboração de um PGD é uma ação fundamental no contexto da gestão dos dados.

Tabela 05 : Elabora Plano de Gestão de Dados (PGD)

Elabora Plano de Gestão de Dados (PGD)	Quantidade de pesquisadores	de Percentual
Sim	66	56,4%
Não	51	43,6%

Tabela 05 : Elabora Plano de Gestão de Dados (PGD)

Elabora Plano de Gestão de Dados (PGD)	Quantidade de pesquisadores	de Percentual
TOTAL	117	100,0%

Fonte: Dados de pesquisa

Um PGD pode ser elaborado diretamente em um suporte físico como uma folha de papel ou através de um processador de texto, mas de modo a facilitar e padronizar o processo de elaboração deste documento, configura-se como uma boa prática o uso de uma ferramenta computacional que possibilite a automação das etapas associadas com a criação do PGD. Ferramentas computacionais que automatizam o processo de elaboração do PGD possibilitam a utilização de *templates* desenvolvidos para áreas específicas do conhecimento bem como a geração do PGD em formatos legíveis por máquina, dentre outras funcionalidades. Exemplos de ferramentas computacionais que automatizam o processo de elaboração do PGD incluem o *DMPonline*⁹ do *Digital Curation Centre (DCC)* e o *DMPTool*¹⁰ da Universidade da Califórnia.

O percentual de pesquisadores que indicou fazer uso de uma ferramenta computacional para a criação de um PGD foi de 47%, (vide Tabela 06), percentual um pouco inferior ao dos pesquisadores que indicaram não fazer uso de uma ferramenta computacional. Quando os pesquisadores indicaram fazer uso de uma ferramenta computacional para a elaboração do PGD no preenchimento do questionário eles eram convidados a indicar o nome da ferramenta computacional utilizada para a criação do PGD. Dos 31 pesquisadores que assumiram fazer uso de uma ferramenta computacional para a elaboração de um PGD, apenas 3 indicaram ferramentas que efetivamente se configuram como aplicações computacionais elaboradas especificamente para a criação de um PGD. Exemplos de ferramentas computacionais mencionadas pelos pesquisadores que entendemos não serem próprias para a criação de PGD, incluem: *Mendeley*, *Pajek*, *Dspace*, *SQL*, *Gephi*, dentre outras.

Tabela 06 : Usa ferramenta computacional para criar PGD

Usa ferramenta computacional para criar PGD	Quantidade de pesquisadores	de Percentual
Sim	31	47,0%
Não	35	53,0%

⁹ Acessível a partir da URL <https://dmponline.dcc.ac.uk/>

¹⁰ Acessível a partir da URL <https://dmptool.org/>

Tabela 06 : Usa ferramenta computacional para criar PGD

Usa ferramenta computacional para criar PGD	Quantidade de pesquisadores	Percentual
TOTAL	66	100,0%

Fonte: Dados de pesquisa

Quando indagados acerca do uso de metadados para descrever os seus dados, o grupo de pesquisadores dividiu-se de forma equânime a este respeito, 50% dos respondentes (vide Tabela 07) indicaram fazer uso de metadados para a descrição de dados. Todos os pesquisadores que responderam de forma afirmativa a esta questão foram convidados a indicar quais os padrões de metadados eram utilizados quando da descrição dos dados de pesquisa. Dos 33 pesquisadores que mencionaram fazer uso de metadados, apenas 9 indicaram um padrão de metadados reconhecido pela comunidade científica, no caso o padrão mencionado foi o *Dublin Core*. Os outros 24 pesquisadores escreveram no campo especificado para a resposta elementos que não se configuram como um padrão de metadados ou deixaram o campo em branco. Os resultados indicam que os pesquisadores necessitam de capacitação para que possam ter mais conhecimentos sobre os padrões de metadados existentes e com isso aplicar o que for mais pertinente em suas pesquisas para uma adequada curadoria dos dados.

Tabela 07: Usa metadados para descrever dados

Usa metadados para descrever dados	Quantidade de pesquisadores	Percentual
Sim	33	50,0%
Não	33	50,0%
TOTAL	66	100,0%

Fonte: Dados de pesquisa

O uso de ferramentas que possibilitem explorar, visualizar ou analisar dados contribuem para o seu respectivo processo de gestão. Estas ferramentas podem ser úteis na efetivação das diversas ações presentes nas etapas de um CVD. A Tabela 08 apresenta os resultados da questão que indagou aos pesquisadores se eles faziam uso destas ferramentas. Verificamos que 50% dos respondentes indicaram fazer uso de ferramentas para explorar, visualizar ou analisar dados, os quais, posteriormente, foram solicitados a indicar em um campo de livre escolha quais seriam estas ferramentas. Diversas ferramentas foram indicadas, podemos estratificá-las em três

categorias: *software* de planilhas eletrônicas; pacotes estatísticos e produtos que possibilitam a visualização de redes sociais.

Tabela o8: Usa alguma ferramenta de *software* para explorar, visualizar ou analisar dados

Usa alguma ferramenta de <i>software</i> para explorar, visualizar ou analisar dados	Quantidade de pesquisadores	Percentual
Sim	33	50,0%
Não	33	50,0%
TOTAL	66	100,0%

Fonte: Dados de pesquisa

Com relação as percepções dos pesquisadores sobre o compartilhamento de dados, o Quadro 02 apresenta as descrições das questões aplicadas ao primeiro enunciado e suas respectivas nomenclaturas. As nomenclaturas foram elaboradas com a finalidade de simplificar a discussão no que diz respeito aos resultados encontrados. Para cada pergunta havia cinco tipos de resposta diferentes – “Discordo fortemente”, “Discordo”, “Não discordo nem concordo”, “Concordo” e “Concordo fortemente”.

Quadro 02 - Descrição das questões do primeiro enunciado

Questão	Descrição
E1Q1	Acredito que o meu programa de pós-graduação estaria disposto a colocar pelo menos ALGUNS dos nossos dados Não sensíveis/Não-proprietários em um repositório central de dados sem restrições de acesso.
E1Q2	Acredito que o meu programa de pós-graduação estaria disposto a colocar TODOS os nossos dados Não sensíveis/Não-proprietários em um repositório central de dados sem restrições de acesso.
E1Q3	Acredito que o meu programa de pós-graduação estaria mais inclinado a colocar nossos dados Não sensíveis/Não-proprietários em um repositório central de dados se eu pudesse colocar condições de acesso aos mesmos.

Fonte: Dados de pesquisa

O Quadro 03 apresenta as descrições das questões aplicadas ao segundo enunciado e suas respectivas nomenclaturas, com a finalidade de simplificar a discussão no que diz respeito aos resultados encontrados. Para cada pergunta foram considerados

cinco tipos de resposta diferentes – “Discordo fortemente”, “Discordo”, “Não discordo nem concordo”, “Concordo” e “Concordo fortemente”.

Quadro 03 - Descrição das questões do segundo enunciado

Questão	Descrição
E2Q1	Usaria conjuntos de dados de outros pesquisadores, se esses conjuntos de dados estivessem facilmente acessíveis.
E2Q2	Estou disposto a colocar pelo menos ALGUNS dos meus dados em um repositório central de dados sem restrições.
E2Q3	Estou disposto a colocar TODOS os meus dados em um repositório central de dados sem restrições.
E2Q4	É correto criar novos conjuntos de dados a partir de dados compartilhados
E2Q5	É importante que meus dados sejam citados quando utilizados por outros pesquisadores.
E2Q6	Normalmente não preciso usar conjuntos de dados de outros pesquisadores.

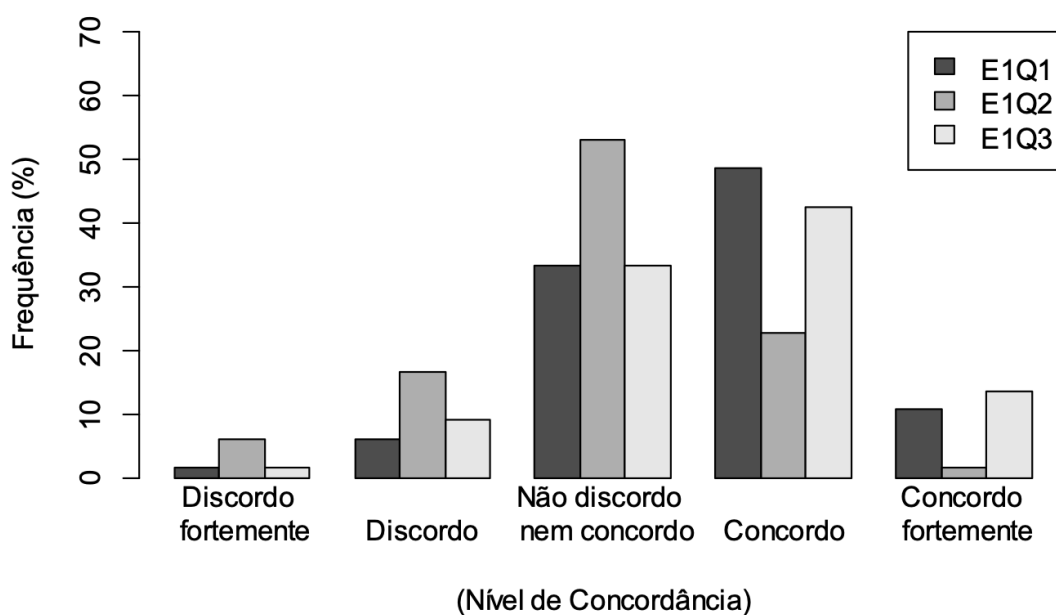
Fonte: Dados de pesquisa

A cultura de compartilhamento de dados já é comum em algumas áreas do conhecimento como a Física, Astronomia, Biologia, entre outras (DAIANE BARRILI; ROCKEMBACH, 2018). No âmbito da Ciência da Informação, Sant’Ana (2016) destaca a importância contributiva deste campo para com o uso e acesso de dados científicos ao buscar levantar fatores e características que promovam um equilíbrio envolvendo os autores participantes do processo e a eficiência na utilização dos dados.

Diante desta realidade, foi possível levantar os resultados das questões voltados para a temática de compartilhamento de dados na área da Ciência da Informação no cenário brasileiro. A princípio, observou-se a distribuição das respostas dispostas em gráficos de barras, como expõem as figuras 02, 03, 04 e 05.

A Figura 02 mostra que as opções “Não discordo nem concordo” e “Concordo” se destacam nas respostas dos pesquisadores em todas as questões relativas ao compartilhamento de dados pelos programas de pós-graduação em Ciência da Informação brasileiros. O que é algo que precisa ser levado em consideração pelos programas em questão, uma vez que possam refletir sobre uma maior disposição em tornar os dados dos pesquisadores acessíveis em repositórios, pois a prática de compartilhamento de dados segundo Sayão e Sales (2014) abre caminhos para que a vasta quantidade de dados provenientes da pesquisa científica da atualidade possa passar pelos processos de coleta, comparação e análise, promovendo novos saberes e novos questionamentos.

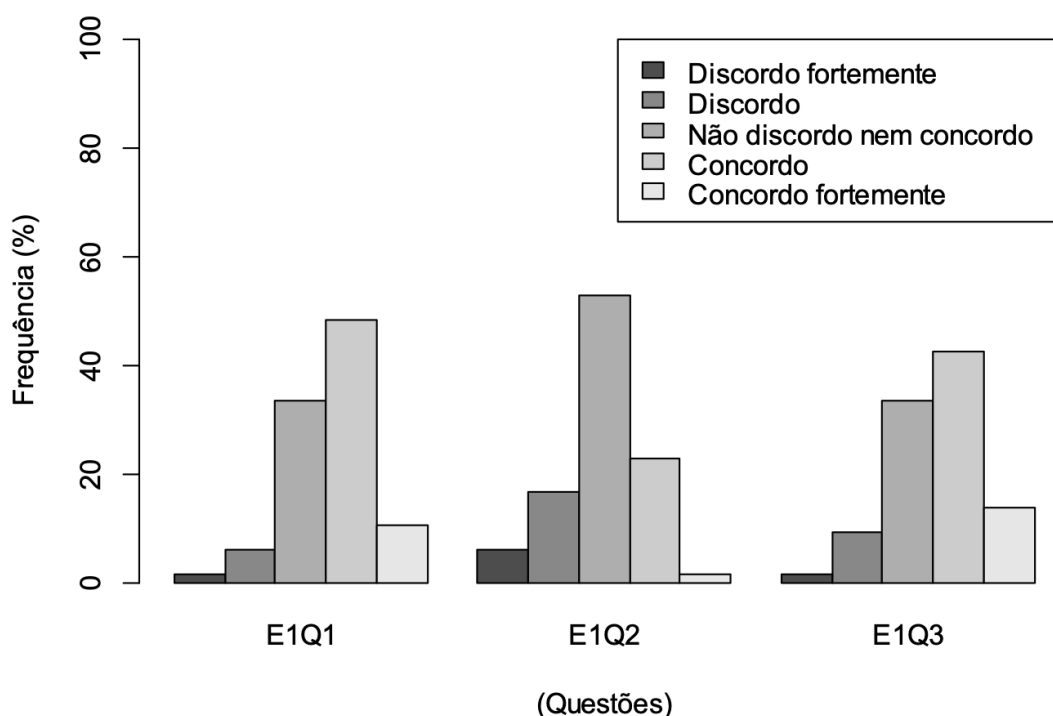
Figura 02 - Frequência relativa das questões do Enunciado 1, subdivididas pelos níveis de concordância.



Fonte: Dados de pesquisa

Na Figura 03, observa-se que a opção central das respostas (“Não discordo nem concordo”) foi a principal escolha dos respondentes na E1Q2 referente à disposição do programa de pós-graduação em colocar todos os dados Não sensíveis/Não-proprietários dos pesquisadores em um repositório central de dados sem restrições de acesso. A este respeito, Patel (2016) aborda o desafio da privacidade, que defende a proteção da confiabilidade dos dados, ou seja, a manutenção da sua confiança é essencial para preservar a privacidade das pessoas. Desta forma, essa é uma questão desafiadora e que deve ser observada pelos programas ao gerenciar os seus dados.

Figura 03 - Frequência relativa dos níveis de concordância, subdivididos pelas questões (Enunciado 1).

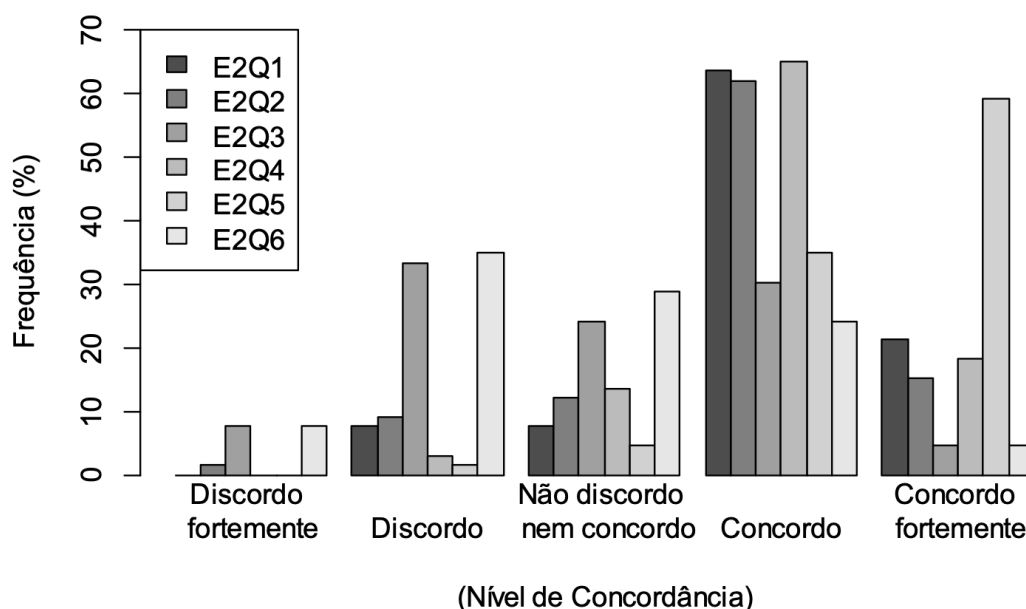


Fonte: Dados de pesquisa

Já com relação à E1Q1 e E1Q3, “Concordo” foi a opção mais evidenciada em ambas. Este resultado indica que os pesquisadores concordam que o programa de pós-graduação ao qual pertencem estaria mais inclinado a colocar pelo menos alguns de seus dados Não sensíveis/Não-proprietários em um repositório central de dados com ou sem restrições de acesso. Os resultados revelam que as condições de acesso poderiam ser elencadas pelos próprios pesquisadores. Desta forma, estes devem estar bem informados das restrições que precisam ser estabelecidas para uma adequada disponibilização dos dados, ou seja, revela uma mentalidade voltada para a busca de conhecimento sobre a temática.

As Figuras 04 e 05 são relativas aos resultados encontrados para as questões de percepção pessoal dos pesquisadores acerca do compartilhamento de dados.

Figura 04 -Frequência relativa das questões do Enunciado 2, subdivididas pelos níveis de concordância.



Fonte: Dados de pesquisa

Os níveis de concordância (Figura 04) demonstraram comportamento semelhante nas questões E2Q1, E2Q2 e E2Q4, isso pode ser visualizado também na Figura 05. Maiores frequências na categoria “Concordo”, nestas questões, evidenciam que os pesquisadores usariam conjuntos de dados de outros pesquisadores, caso os dados fossem de fácil acesso; que estariam dispostos a colocar pelo menos alguns dos seus dados em um repositório central de dados sem restrições; e que julgam correto criar novos conjuntos de dados a partir de dados compartilhados. Infere-se diante dos resultados, que se os dados forem corretamente curados, esses pesquisadores mostram uma pré-disposição voltada à prática da utilização dos dados de outros pesquisadores, assim como também da disponibilização dos seus.

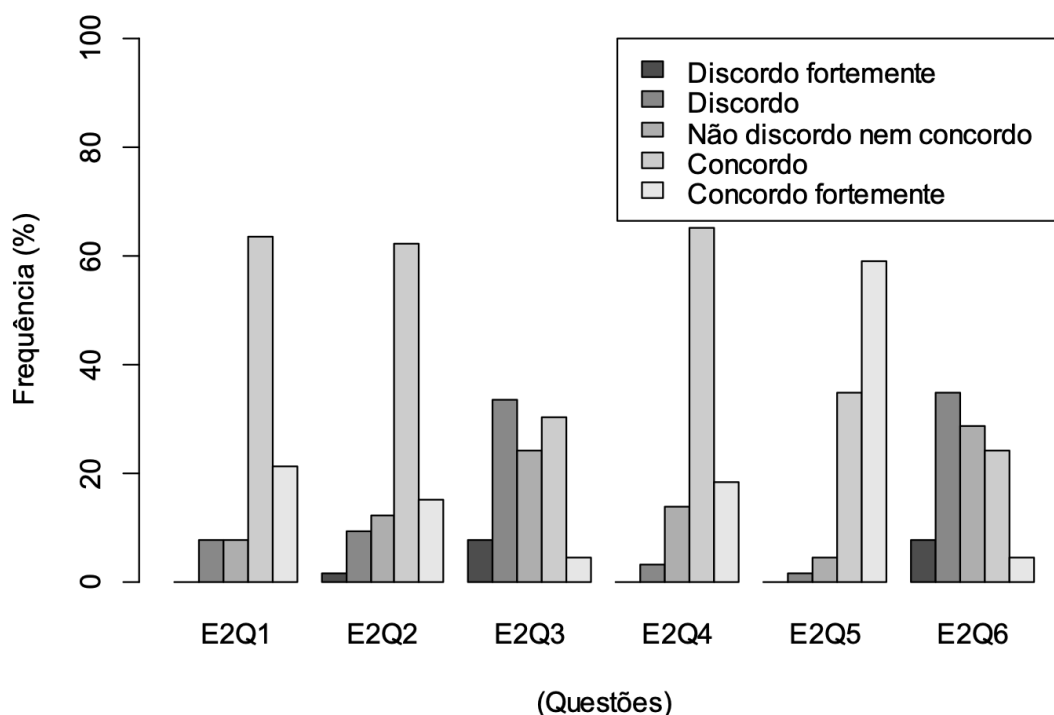
É preciso levar em consideração o que é destacado por Sayão e Sales (2014) ao mostrarem que não é necessário apenas disponibilizar as coletas de dados de

pesquisas na Web, é preciso que esses recursos venham ser acessados e reutilizados na atualidade e no futuro, para tanto requer informações contextuais – semânticas e estruturais atreladas aos dados digitais de forma que eles estejam auto descritos. Além do mais, a geração e disponibilização desses dados não são suficientes para obter valor científico, eles precisam ser comunicados formalmente (DAIANE BARRILI; ROCKEMBACH, 2018).

Um resultado que se destaca é referente a E2Q5 em que os pesquisadores concordam fortemente que é importante existir a citação de dados para quem disponibilizou os dados compartilhados, quando utilizados por outros pesquisadores. Ou seja, receber os créditos pelo seu trabalho. Desta forma, políticas nesse sentido devem ser desenvolvidas para estimular os pesquisadores a disponibilizarem os seus dados. Diante deste contexto, dentre os benefícios elencados por Bloom (2013 apud PATEL, 2016) sobre o compartilhamento de dados, tais quais: reutilização, confiança, transparência e economia de tempo, está o de citação, que se relaciona aos créditos e o reconhecimento por meio da citação de dados voltados para aqueles que os coletam e fazem as análises dos dados.

Duas questões que apresentaram comportamento diferenciado das demais foram E2Q3 e E2Q6, nas quais as maiores porcentagens predominaram na opção “Discordo”. Evidenciando o receio que os pesquisadores têm em disponibilizar todos os seus dados em um repositório central de dados sem restrições, embora concordem que necessitam usar conjuntos de dados de outros pesquisadores em suas análises. Desta forma, julgam importante fazer uso de dados de outros pesquisadores, pois contribui para reduzir o esforço de pesquisa, reduzindo o desperdício de recursos empregados em estudos já existentes. Assim, os pesquisadores podem ampliar os estudos ou abrir o horizonte para novas pesquisas. O que Sayão e Sales (2014) destacam de uma forma objetiva que evita o custo da duplicação de esforços, possibilita novas visões em contextos científicos distintos para esses dados, sendo possível ainda uma interpretação com um maior valor de criatividade agregado, ao possibilitar novos caminhos a serem desbravados.

Figura 05 - Frequência relativa dos níveis de concordância, subdivididos pelas questões do Enunciado 2.



Fonte: Dados de pesquisa

Observa-se que as maiores porcentagens de respondentes nos níveis de concordância, Enunciado 1 (Tabela 08), estão nas categorias “Concordo” para E1Q1 e E1Q3, com 48,48% e 42,42%, respectivamente, e “Não discordo nem concordo” em E1Q2, com 53,03%. No Enunciado 2, as maiores porcentagens estão nas categorias “Concordo” para E2Q1 (63,64%), E2Q2 (62,12%) e E2Q4 (65,15%), “Concordo fortemente” em E2Q5 (59,09%) e “Discordo” em E2Q3 (33,33%) e E2Q6 (34,85%). Essas porcentagens maiores estão destacadas de azul nas respectivas Tabelas 08 e 09.

Tabela 08 - Frequências absolutas e relativas (%) dos níveis de concordância nas questões do Enunciado 1.

	Discordo Fortemente		Discordo		Não Discordo nem Concordo		Concordo		Concordo Fortemente	
	n	%	n	%	N	%	N	%	n	%
E1Q1	1	1,52	4	6,06	22	33,33	32	48,48	7	10,61
E1Q2	4	6,06	11	16,67	35	53,03	15	22,73	1	1,52
E1Q3	1	1,52	6	9,09	22	33,33	28	42,42	9	13,64

Fonte: Dados da pesquisa

No que se refere a Tabela 09, estão destacadas de vermelho as menores porcentagens que se encontram nas extremidades das categorias, em “Discordo fortemente” e “Concordo fortemente”, variando de 0,00% até 4,55%.

Tabela 09 - Frequências absolutas e relativas (%) dos níveis de concordância nas questões do Enunciado 2.

	Discordo Fortemente		Discordo		Não concordo		Concordo		Concordo Fortemente	
	n	%	n	%	N	%	N	%	n	%
E2Q1	0	0,00	5	7,58	5	7,58	42	63,64	14	21,21
E2Q2	1	1,52	6	9,09	8	12,12	41	62,12	10	15,15
E2Q3	5	7,58	22	33,33	16	24,24	20	30,03	3	4,55
E2Q4	0	0,00	2	3,03	9	13,64	43	65,15	12	18,18
E2Q5	0	0,00	1	1,52	3	4,55	23	34,85	39	59,09
E2Q6	5	7,58	23	34,85	19	28,79	16	24,24	3	4,55

Fonte: Dados da pesquisa

Quanto aos Escores modificados, as médias das questões do Enunciado 1 variaram de 49,24 até 65,15, e os desvios padrões de 20,51 à 22,39 (Tabela 10). Destaca-se também as medianas de E1Q1 e E1Q3, com valor de 75,00 cada, isto indica que no mínimo 50% da amostra, nestas questões, apresentam alto nível de concordância, como já foi mencionado nos resultados gráficos

Tabela 10 - Estatísticas descritivas dos Escores Modificados das questões do Enunciado 1.

	Mínimo	Mediana	Média	Máximo	Desvio Padrão
E1Q1	0,00	75,00	65,15	100,00	20,51
E1Q2	0,00	50,00	49,24	100,00	21,02
E1Q3	0,00	75,00	64,39	100,00	22,39

Fonte: Dados da pesquisa

A Tabela 11 traz os valores dos Escores modificados para o Enunciado 2, na qual observa-se variação de médias entre 45,83 e 87,88, e desvios padrões entre 16,55 e 26,58. Destaca-se a mediana de E2Q5, com o mais alto valor dos escores (100,00), indicando que pelo menos 50% da amostra, nessa questão, concorda fortemente com o respectivo tema abordado (“É importante que meus dados sejam citados quando utilizados por outros pesquisadores”).

Tabela 11 - Estatísticas descritivas dos Escores Modificados das questões do Enunciado 2.

	Mínimo	Mediana	Média	Máximo	Desvio Padrão
E2Q1	25,00	75,00	74,62	100,00	19,36
E2Q2	0,00	75,00	70,08	100,00	21,58
E2Q3	0,00	50,00	47,73	100,00	26,58
E2Q4	25,00	75,00	74,62	100,00	16,69
E2Q5	25,00	100,00	87,88	100,00	16,55
E2Q6	0,00	50,00	45,83	100,00	25,79

Fonte: Dados da pesquisa

De modo geral, a análise evidencia que os pesquisadores concordam com a importância da disponibilização dos dados, mas há uma preocupação com a questão de confidencialidade e devidos créditos aos responsáveis pelas informações fornecidas, por isso essas devem ser questões não negligenciadas pelos programas de Pós-Graduação em CI no cenário brasileiro. De modo a contribuir com a qualidade e efetividade da pesquisa, que a cada dia possa gerar impactos positivos na sociedade diante de uma preservação, gerenciamento e disseminação dos dados de pesquisa que promova o conhecimento científico, sendo este segundo Sayão e Sales (2014) patrimônio da humanidade.

Com o objetivo de identificar os motivos para que os dados de suas respectivas pesquisas não fossem compartilhados, o questionário submetido aos pesquisadores apresentou uma questão indicando diversas possibilidades de razões para o não compartilhamento dos dados. Nesta questão os pesquisadores tinham a possibilidade de marcar todas as alternativas que julgassem pertinentes para justificar o não compartilhamento de dados.

A razão mais indicada pelos 66 pesquisadores que continuaram na resolução do questionário para o não compartilhamento dos dados foi a “Falta de infraestrutura técnica” (62,1%), em segundo lugar foi apontado a “Ausência de pessoal de apoio” (51,5%), em terceiro lugar foi indicado a “Inexistência de incentivos para a disponibilização dos dados” (48,4%) e em quarto lugar foi mencionado (39,3%) a “Falta de financiamento”.

Um dilema importante que o pesquisador enfrenta após a coleta dos dados de pesquisa é a decisão de disponibilizá-los ou não de imediato para o uso por outros pesquisadores. Liberar os dados durante o período em que os mesmos são analisados pode comprometer o investimento feito pelo pesquisador e sua equipe. Outros de posse destes dados podem se adiantar e publicar resultados de pesquisa a partir de dados coletados com o esforço de outros pesquisadores. A partir destes fatos entendemos que é necessária a possibilidade dos pesquisadores estabelecerem um período de embargo para os dados por eles gerados. Um período de embargo também se justifica no caso em que os dados coletados possuem restrições legais ou éticas que justifiquem a não imediata liberação para toda a comunidade, como por exemplos dados associados à saúde humana ou que envolvam menores. Com relação ao período de embargo, Borgman (2015, posição 503, *Kindle edition*) explica que:

Em algumas áreas, os pesquisadores recebem períodos de embargo - também chamados de períodos proprietários - para controlar seus dados antes de liberá-los. O tempo que os pesquisadores podem controlar seus dados geralmente varia de alguns meses a alguns anos. O período deve ser longo o suficiente para analisar os dados e publicar suas descobertas, mas curto o suficiente para encorajar a divulgação de dados para a comunidade.¹¹

Conforme ilustrado na Tabela 12, a maioria dos pesquisadores entrevistados (92,4%) entendem que eles devem ter a autonomia para definir um período de embargo sobre os seus dados. Este comportamento é compreensível, os pesquisadores não querem ou não podem perder o controle sobre os dados por eles coletados, visto a necessidade de primazia pela escrita das primeiras publicações e o consequente reconhecimento pela comunidade das descobertas associadas aos conjuntos de dados.

Tabela 12: Pesquisadores devem ter autonomia para definir período de embargo

Pesquisadores devem ter autonomia para definir período de embargo	Quantidade de pesquisadores	Percentual
Sim	61	92,4%
Não	5	7,6%
TOTAL	66	100,0%

Fonte: Dados de pesquisa

Os pesquisadores que indicaram “sim” acerca da possibilidade da definição de um período de embargo foram convidados a apontar qual seria um período razoável para o embargo de dados conforme apresentado na Tabela 13 e ilustrado na Figura 06.

Tabela 13 : Período razoável para o embargo de dados

Tempo de atuação	Quantidade de pesquisadores	Percentual
------------------	-----------------------------	------------

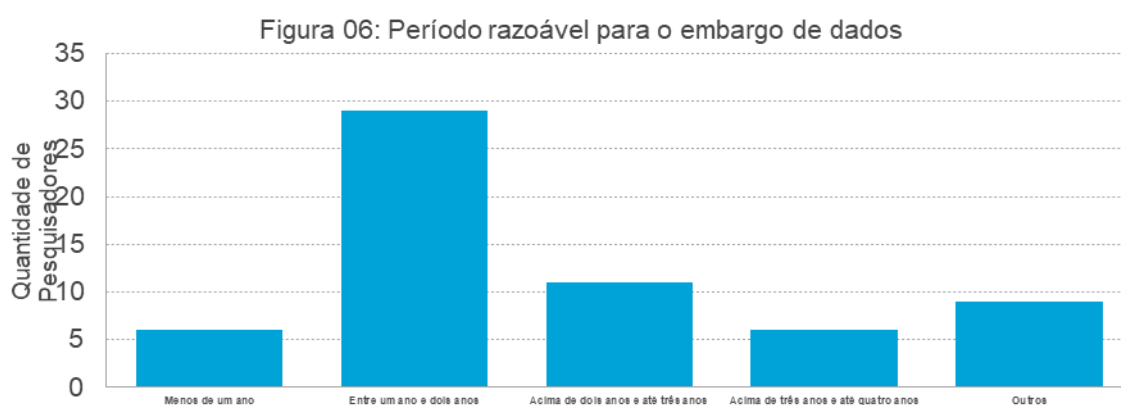
¹¹ Texto original em Língua Inglesa: "In some fields, investigators are given embargo periods—also called proprietary periods—to control their data before releasing them. The length of time investigators can control their data typically ranges from a few months to a few years. The period is intended to be long enough to analyze the data and to publish their findings, but short enough to encourage the release of data to the community."

Tabela 13 : Período razoável para o embargo de dados

Tempo de atuação	Quantidade de pesquisadores	de Percentual
Menos de um ano	6	9,8%
Entre um ano e dois anos	29	47,5%
Acima de dois anos e até três anos	11	18,0%
Acima de três anos e até quatro anos	6	9,8%
Outros	9	14,8%
TOTAL	61	100,0%

Fonte: Dados de pesquisa

A maioria dos pesquisadores (47,5%), quase a metade dos respondentes indicaram que o período razoável para o embargo de dados seria entre um e dois anos. A partir das considerações de Borgman (2015), entendemos que o lapso temporal indicado de forma majoritária (Entre um ano e dois) pelos pesquisadores na Tabela 13 é uma estimativa razoável para o tempo necessário para a análise dos dados e a publicação das descobertas em periódicos pelos pesquisadores.



Fonte: Dados de pesquisa

CONCLUSÕES

Os resultados desta pesquisa contribuem para desvendar de que forma os dados obtidos a partir de pesquisas são geridos pelos pesquisadores da área da Ciência da

Informação e quais são as suas percepções acerca do assunto. Ressaltamos o caráter inovador da investigação conduzida, visto que até o momento são escassas as pesquisas que abordem o tema da gestão de dados de pesquisa no âmbito dos pesquisadores associados com a área da Ciência da Informação no Brasil.

A partir do processamento dos dados coletados, foi possível constatar que dos 187 pesquisadores que efetivamente foram considerados para a pesquisa, apenas 66 (35,2%) deles, menos da metade, foram considerados como executores de algum tipo de gestão de dados de pesquisa. Efetivamente, este número ainda pode ser menor, visto que a pesquisa conduzida, pelas suas características, na maior parte constituída por questões de cunho quantitativo, não possibilitou abordar mais detalhadamente de que forma a gestão dos dados se configura efetivamente nos ambientes de trabalho dos respondentes.

Os dados obtidos através do instrumento de pesquisa indicaram o desconhecimento dos respondentes sobre uma efetiva gestão dos dados de pesquisa. Este fato é um motivo para que sejam realizadas novas pesquisas na área da Ciência da Informação sobre a questão da gestão dos dados de pesquisa, bem como uma alerta, considerando que os respondentes são pesquisadores desta mesma área.

No que diz respeito ao desconhecimento e aos desafios associados com a gestão dos dados de pesquisa, indicamos que esta situação não é exclusiva da população aqui investigada. As pesquisas de Curty (2015), Aydinoglu, Dogan e Taskin (2017) e Elsayed e Saleh (2018) contribuem para referendar esta afirmação.

A partir dos achados desta investigação, sugerimos que ações de estímulo à pesquisa e à educação no contexto da gestão de dados poderiam ser desenvolvidas pelos órgãos federais e estaduais brasileiros de fomento à pesquisa. Entendemos também que para uma melhor disseminação dos conceitos intrínsecos a gestão dos dados de pesquisa é importante que sejam realizadas atualizações nas grades curriculares dos cursos de pós-graduação brasileiros em Ciência da Informação.

Artigo recebido em 02/05/2019 e aprovado em 29/10/2019.

REFERÊNCIAS

AYDINOGLU, A. U.; DOGAN, G.; TASKIN, Z. Research data management in Turkey: perceptions and practices. *Library Hi Tech*, v. 35, n. 2, p. 271-289, 2017. Disponível em: <https://doi.org/10.1108/LHT-11-2016-0134>. Acesso em: 29 out. 2019.

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

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO - CNPQ. *Tabela de Áreas do Conhecimento*. 2019. Disponível em: <http://www.cnpq.br/documents/10157/186158/TabeladeAreasdoConhecimento.pdf>. Acesso em: 28 out. 2019.

CURTY, R. G. *Beyond “data thrifting”*: an investigation of factors influencing research data reuse in the social sciences. 2015. 308f. Tese (Doutorado em Filosofia – PhD) – School of Information Studies / Syracuse University, New York, 2015. Disponível em: <https://surface.syr.edu/etd/266/>. Acesso em: 29 outubro 2019.

DAIANE BARRILI, D.S.; ROCKEMBACH, M. Publicações ampliadas: aspectos da integração de dados de pesquisa. *Informação & Sociedade*, João Pessoa, v. 28, n. 2 2018.

DATAONE. *Data life cycle*. 2019. Disponível em: <https://www.dataone.org/data-life-cycle>. Acesso em: 23 abr. 2019.

DIAS, G. A.; ANJOS, R. L.; RODRIGUES, A. A. Os princípios FAIR: viabilizando o reuso de dados científicos. In: DIAS, A. D.; OLIVEIRA, B. M. J. F (org.). *Dados científicos: perspectivas e desafios*. João Pessoa: Editora UFPB, 2019. p. 177-187.

DIGITAL CURATION CENTRE. 2019. *curation lifecycle model*. Disponível em: <http://www.dcc.ac.uk/resources/curation-lifecycle-model>. Acesso em: 23 abr. 2019.

DOORN, P.; DILLO, P. *Fair data in trustworthy data repositories Webinar*. EUDAT, 2016. Disponível em: <https://eudat.eu/events/webinar/fair-data-in-trustworthy-data-repositories-webinar>. Acesso em: 29 out. 2019.

ELSAYED, A. M.; SALEH, E. I. Research data management and sharing among researchers in Arab universities: An exploratory study. *IFLA Journal*, v. 44, n. 4, p. 281-299, 2018. Disponível em: <https://doi.org/10.1177/0340035218785196>. Acesso em: 29 out. 2019.

Event horizon Telescope. *Astronomers capture first image of a black hole*. 2019. Disponível em: <http://www.oecd.org/sti/inno/38500813.pdf>. Acesso em: 22 abr. 2019.

GO FAIR. *FAIR principles*. 2019. Disponível em: <https://www.go-fair.org/fair-principles/>. Acesso em: 29 out. 2019.

HENNING, P. C. et al. GO FAIR e os princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta. *Em Questão*, v. 25, n. 2, p. 389-412, 2019. Disponível em: <https://www.redalyc.org/service/redalyc/downloadPdf/4656/465658944018/7>. Acesso em: 29 out. 2019.

LIKERT, R. A technique for the measurement of attitudes. *Archives of Psychology*, n. 140, p. 1-55, 1932.

MADEIRO, A. L. de F. Repositórios de dados de pesquisa: investigando sua adoção nos institutos brasileiros de pesquisa. In: *INFORMAÇÃO, dados e tecnologia*. 1 ed. Tupã: Editora da Faculdade de Ciências e Engenharia - UNESP, 2019. V.1, p. 282 - 290. Disponível em: http://dadosabertos.info/enhanced_publications/idt/enhanced_paper.php?id=3&lang=pt_BR#accesspoint. Acesso em: 21 abr. 2019.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO - OCDE. *OECD principles and guidelines for access to research data from public funding*. 2015. Disponível em: <http://www.oecd.org/sti/inno/38500813.pdf>. Acesso em: 19 abr. 2019.

PATEL, D. Research data management: a conceptual framework. *Library Review*, v. 65, n. 4/5, p. 226-241, maio 2016. Disponível em: <https://www-emeraldinsight-com.ez15.periodicos.capes.gov.br/doi/full/10.1108/LR-01-2016-0001>. Acesso em: 26 abr. 2019.

R. *The R project for statistical computing*. 2019. Disponível em: <https://www.r-project.org/>. Acesso em: 25 abr. 2019.

R STUDIO. *Products*. Disponível em: <https://www.rstudio.com/products/rstudio/download/>. Acesso em: 28 abr. 2019.

- RICE, R.; SOUTHALL, J. *The data librarian's handbook*. London: Facet Publishing, 2016.
- RICHARDSON, R. J. *Pesquisa social: métodos e técnicas*. 4. ed. São Paulo: Atlas, 2017.
- RODRIGUEZ-IGLESIAS, A. et al. Publishing FAIR data: an exemplar methodology utilizing PHI-base. *Frontiers in plant Science*, v. 7, n. 641. 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/27433158>. Acesso em: 29 out. 2019.
- SALES, L. et al. GO FAIR Brazil: a challenge for brazilian data science. *Data Intelligence*, v. 2, n. 1, p. 234-245, 2019. Disponível em: <http://www.data-intelligence-journal.org/p/52/#reference>. Acesso em: 29 outubro 2019.
- SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. *Informação & Informação*, v. 21, n. 2, p. 116–142, dez. 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940/20124>. Acesso em: 21 abr. 2019.
- SAYÃO, L. F.; SALES, L. Farias. Dados abertos de pesquisa: ampliando o conceito de acesso livre. *Revista Eletrônica de Comunicação, Informação e Inovação em Saúde*, v. 8, n. 2, junho 2014. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/611>. Acesso em: 26 abr. 2019.
- SAYÃO, L. F.; SALES, L. F. *Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores*. Rio de Janeiro: CNEN/IEN, 2015. Disponível em: http://www.cnen.gov.br/images/CIN/PDFs/GUIA_DE_DADOS_DE_PESQUISA.pdf. Acesso em: 9 abr. 2019.
- SCHÖPFEL, J.; PROST, H. Research data management in social sciences and humanities: a survey at the University of Lille (France). *LIBREAS*, n. 29, 2016. Disponível em: <https://libreas.eu/ausgabe29/09schoepfel/>. Acesso em: 8 abr. 2019.
- SHI, T.; FAN, X. Comparing response rates in e-mail and paper surveys: A meta-analysis. *Educational Research Review*, v.4, n.1, p. 26-40, 2009. Disponível em: <https://doi.org/10.1016/j.edurev.2008.01.003>. Acesso em: 3 abr. 2019.
- TAI, X. et al. Comparison of response rates on invitation mode of a web-based survey on influenza vaccine adverse events among health care workers: a pilot study. *BMC Medical Research Methodology*, v.18, 2018. Disponível em: <https://doi.org/10.1016/j.edurev.2008.01.003>. Acesso em: 3 abr. 2019.
- THE ECONOMIST. *The world's most valuable resource is no longer oil, but data*. 2017. Disponível em: <http://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource>. Acesso em: 21 abr. 2019.
- VEIGA, V. S. de O. *Percepção dos pesquisadores portugueses e brasileiros da área de Neurociências quanto ao compartilhamento de artigos científicos e dados de pesquisa no acesso aberto verde: custos, benefícios e fatores contextuais*. 2017. 294 f. Tese (Doutorado em Ciências) - Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, 2017. Disponível em: <https://www.arca.fiocruz.br/handle/icict/26842>. Acesso em: 01 abr. 2019.
- Veiga, V. S. de O.; Da Silva, C. H.; BORGES, M. M.; Borges, P. Compartilhamento e acesso a dados de pesquisa em Humanidades Digitais. *Cadernos BAD*, n.1, p. 3-14, 2018. Disponível em: <https://www.bad.pt/publicacoes/index.php/cadernos/article/view/1932>. Acesso em: 01 abr. 2019.

VILAR, P., ZABUKOVEC, V. Research data management and research data literacy in Slovenian science. *Journal of Documentation*, v. 75, n. 1, p. 24-43, 2019. Disponível em: <https://doi.org/10.1108/JD-03-2018-0042>. Acesso em: 08 abr. 2019.

WHITWAM, R. *It took half a ton of hard drives to store the black hole image data*. 2019. Disponível em: <https://www.extremetech.com/extreme/289423-it-took-half-a-ton-of-hard-drives-to-store-ehh-black-hole-image-data>. Acesso em: 28 abr. 2019.