



Práticas e percepções dos pesquisadores brasileiros sobre serviços de acesso aberto a dados de pesquisa

Practices and perceptions of Brazilian researchers on open access to research data

Sônia Elisa Caregnato*

Samile Andrea de Souza Vanz**

Caterina Groposo Pavão***

Paula Caroline Schifino Jardim Passos****

Eduardo Nunes Borges*****

Renê Faustino Gabriel Júnior*****

Luís Alberto Azambuja*****

Rafael Port da Rocha*****

* Doutora pela University of Sheffield, Inglaterra. Professora do Programa de Pós-Graduação em Comunicação e do Programa de Pós-Graduação em Ciência da Informação, ambos da Universidade Federal do Rio Grande do Sul (UFRGS). Endereço: Rua Ramiro Barcelos, 2705, sala 212, Campus Saúde, CEP 90.035-007, Porto Alegre, RS. Telefone: (51) 3308.5737. E-mail: sonia.caregnato@gmail.com

** Doutora em Comunicação e Informação pela Universidade Federal do Rio Grande do Sul (UFRGS). Professora do Programa de Pós-Graduação em Comunicação e do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Rio Grande do Sul (UFRGS). Endereço: Rua Ramiro Barcelos 2705 sala 214, CEP 90035-007, Porto Alegre, RS. Telefone: (51) 3308-5434. E-mail: samile.vanz@ufrgs.br

*** Doutora em Comunicação e Informação pela Universidade Federal do Rio Grande do Sul (UFRGS). Professora do Curso de Biblioteconomia e do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Rio Grande do Sul (UFRGS). Endereço: Rua Ramiro Barcelos, 2705 - Campus Saúde - Porto Alegre - RS - CEP - 90035-007. Telefone: (51) 3308-5336. E-mail: caterina@cpd.ufrgs.br

**** Doutora, com pós-doutorado, em Comunicação e Informação pela Universidade Federal do Rio Grande do Sul. Endereço: Rua Ramiro Barcelos, 2705, Campus Saúde, CEP 90.035-007, Porto Alegre, RS. Telefone: (51) 3308.5942 Email: paulacarolinejardim@gmail.com

***** Doutor em Computação pela Universidade Federal do Rio Grande do Sul (UFRGS). Professor do Programa de Pós-Graduação em Computação da Universidade Federal do Rio Grande (FURG). Endereço: Av. Itália, km 8, Campus Carreiros, CEP 96203-900, Rio Grande, RS. Telefone: (53) 3237-3069. E-mail: eduardoborges@furg.br

***** Doutor em Ciência da Informação pela Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp). Professor do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Rio Grande do Sul (UFRGS) Endereço: Rua Ramiro Barcelos, 2705, Campus Saúde, Santo Antônio, CEP 90.035-007, Porto Alegre, RS. Telefone: (51) 3308.5942. E-mail: rene.gabriel@ufrgs.br

***** Mestre em Engenharia de Computação pela Universidade Federal do Rio Grande (FURG). Técnico em Tecnologia da Informação da Universidade Federal do Rio Grande (FURG). Endereço: Av. Itália, km 8, bairro Carreiros, CEP 96203-900, Rio Grande, RS, Brasil. Telefone: (53) 3233-6580. E-mail: cpdlaa@furg.br

***** Doutor em Computação pela Universidade Federal do Rio Grande do Sul (UFRGS). Professor do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Rio Grande do Sul (UFRGS) Endereço: Rua Ramiro Barcelos, 2705, Campus Saúde, CEP 90.035-007, Porto Alegre, RS. Telefone: (51) 3308.5942. E-mail: rafael.rocha@ufrgs.br

RESUMO

O artigo apresenta análise exploratória das práticas e das percepções a respeito do acesso aberto a dados de pesquisa embasada em dados coletados por meio de survey, realizada com pesquisadores brasileiros. As 4.676 respostas obtidas demonstram que, apesar do grande interesse pelo tema, evidenciado pela prevalência de variáveis relacionadas ao compartilhamento e ao uso de dados e aos repositórios institucionais, não há clareza por parte dos sujeitos sobre os principais tópicos relacionados. Conclui-se que, apesar da maioria dos pesquisadores afirmar que compartilha dados de pesquisa, a disponibilização desses dados de forma aberta e irrestrita ainda não é amplamente aceita.

Palavras-chave: Dados Abertos de Pesquisa; Compartilhamento de Dados; Reuso de Dados.

ABSTRACT

This article presents an exploratory analysis of the practices and perceptions regarding open access to research data based on information collected by a survey with Brazilian researchers. The 4,676 responses show that, despite the great interest in the topic, evidenced by the prevalence of variables related to data sharing and use and to institutional repositories, there is no clarity on the part of the subjects on the main related topics. We conclude that, although the majority of the researchers share research data, the availability of this data in an open and unrestricted way is not yet widely accepted.

Keywords: Open Research Data; Data Sharing; Data Reuse.

INTRODUÇÃO

A visibilidade e a reputação da pesquisa científica estão, cada vez mais, relacionadas à abertura dos dados coletados, gerados e utilizados pelos pesquisadores. O compartilhamento de dados de pesquisa com a comunidade científica, ou até mesmo com o público em geral, permite que tais informações sejam escrutinadas, conferidas e reutilizadas em diferentes contextos. Esse movimento em direção a uma ciência aberta, para além das publicações em formato livre, já se encontra em estado avançado no cenário internacional.

Buscando articular ações relacionados ao Acesso Aberto a Dados de Pesquisa (AADP) no âmbito da pesquisa brasileira, a Rede Nacional de Ensino e Pesquisa (RNP) e o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) iniciaram, em 2017, uma parceria com o Centro de Documentação e Acervo Digital de Pesquisa (CEDAP), da Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul (UFRGS), e com o Grupo de Pesquisa em Gerenciamento de Informações, do Centro de Ciências Computacionais (C3) da Universidade Federal do Rio Grande (FURG), para investigar as práticas dos pesquisadores, as tecnologias envolvidas e a oferta de serviços para dados de pesquisa, especialmente na forma de repositórios para esse tipo de conteúdo.

Entre os objetivos do projeto conjunto, que ainda se encontra em andamento sob o título Rede de Dados de Pesquisa Brasileira (RDP Brasil), estava a identificação de potenciais usuários nacionais de serviços de AADP no Brasil. Para conhecer o cenário brasileiro, foi realizada uma *survey* junto aos pesquisadores de todas as áreas do conhecimento e de todas as instituições brasileiras.

Partiu-se do pressuposto que os potenciais usuários desse tipo de serviço no país são tanto aqueles que já atentaram para a necessidade de se estabelecerem processos de gestão de dados de pesquisa para atingir a finalidade de disponibilizar, acessar e

reutilizar os resultados da ciência nacional, quanto os que ainda não conhecem as dimensões do fenômeno e suas possíveis consequências. Assim, o levantamento, além de mapear a realidade brasileira, também serviu para chamar a atenção para o tema em âmbito nacional.

Inserido nesse contexto, este artigo tem por objetivo apresentar os resultados do estudo que buscou identificar as percepções dos pesquisadores brasileiros sobre o acesso aberto a dados de pesquisa e as práticas já em curso nas instituições nacionais. O relatório completo da pesquisa pode ser encontrado no endereço <http://hdl.handle.net/10183/185195>. Os dados da pesquisa estão disponíveis em formato aberto em <https://hdl.handle.net/20.500.11959/5>.

Adicionalmente, e de forma a aplicar os princípios defendidos, os dados resultantes do estudo foram disponibilizados em formato aberto, atendendo à alguns princípios FAIR (GOFAIR, 2019), que preconizam que os dados devem ser encontráveis (*findable*), acessíveis (*accessible*), interoperáveis (*interoperable*) e reutilizáveis (*reusable*), como o uso de identificador persistente, de metadados claramente descritos e ricos, de identificação de proveniência, entre outros.

PROCEDIMENTOS METODOLÓGICOS

O estudo “Práticas e percepções sobre acesso aberto a dados de pesquisa” foi elaborado com base em levantamentos similares realizados fora do Brasil sobre compartilhamento e reuso de dados de pesquisa, especialmente os trabalhos de Beagrie et al. (2009), Tenopir et al. (2011), Kim e Adler (2015), Kim e Zhang (2015), Parham, Bodnar e Fuchs (2012), Rowley et al. (2017), Wallis et al. (2013) e Digital Curation Center (2009). O questionário foi organizado no software Survey Monkey em uma estrutura composta por oito questões demográficas e mais dois eixos principais, contendo 13 questões acerca das práticas dos pesquisadores sobre dados de pesquisa e seis questões acerca de percepções em relação ao compartilhamento e ao uso de dados. No total, foram 27 questões, as quais os participantes levaram em média 14 minutos para responder.

O questionário online foi enviado, em março de 2018, a três grupos de endereços:

- a) 71.712 e-mails válidos de 48.335 líderes e 27.869 vice-líderes dos grupos de pesquisa cadastrados no Diretório de Grupos de Pesquisa do CNPq;
- b) 4.677 e-mails institucionais de coordenadores dos Programas de Pós-Graduação cadastrados na CAPES, dentre um universo de 4.177 programas (GEOCAPES, 2016);
- c) 249 e-mails de coordenadores de Institutos Nacionais de Ciência e Tecnologia.

O questionário foi fechado em abril de 2018, tendo, portanto, ficado disponível por um mês. O número total de respondentes foi 4.676, dos quais 1.054 (22,54%) demonstraram interesse em participar de entrevista posterior e 1.655 (35,39%) demonstraram interesse em participar da lista discussão do RDP Brasil. As respostas não eram obrigatórias e o respondente tinha liberdade para abandonar o questionário a qualquer momento, o que resultou em um volume de respostas diferente para cada questão. Algumas perguntas permitiram respostas e comentários em texto livre.

Os dados foram analisados utilizando-se da estatística descritiva e com apoio do software Microsoft Excel. Cada uma das perguntas foi analisada individualmente e apresentada em gráfico ou tabela. As frequências dos valores relativos, ou seja, o

número de respondentes para cada alternativa, podem ser lidas a partir de gráficos de barras simples. Nos gráficos de barras empilhadas, a distribuição relativa das respostas é representada, em sua grande maioria, pelas grandes áreas do conhecimento. A soma das frequências relativas dos valores de característica sempre produz 100%. Para algumas questões selecionadas, foram criadas tabelas de relações cruzadas. Essas foram elaboradas com a finalidade de localizar dependências de acordo com grupos de pessoas em suas respostas às perguntas.

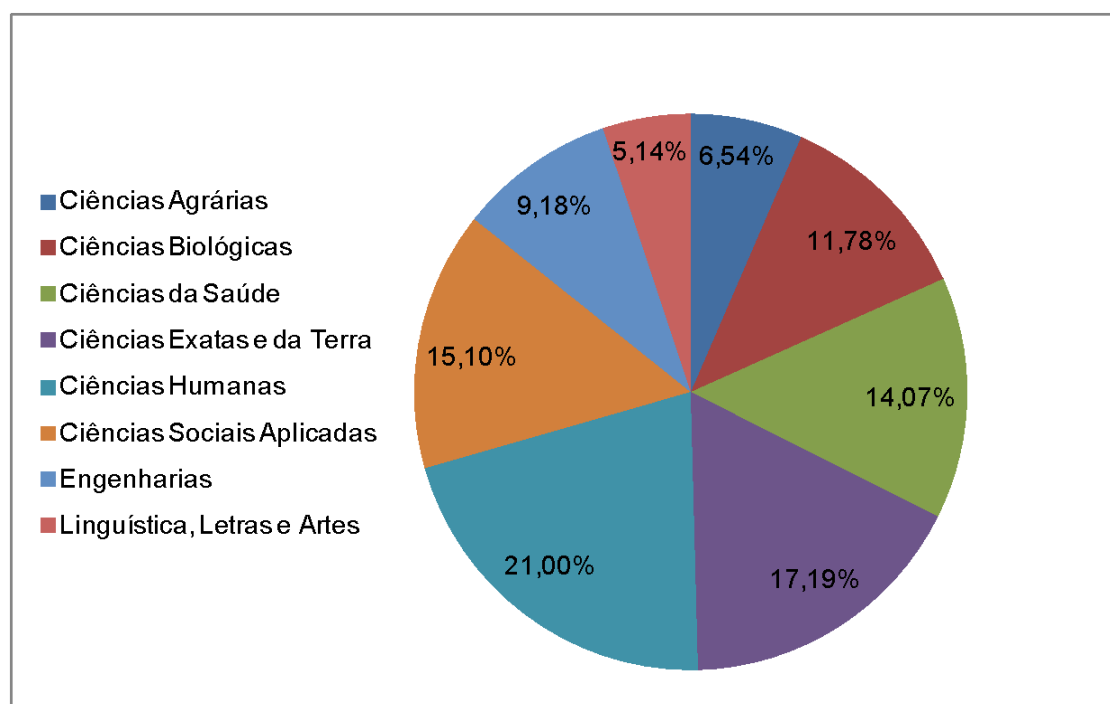
RESULTADOS E DISCUSSÃO

O grande número de respondentes (4.676), no curto período em que a *survey* ficou disponível (30 dias), revela o interesse que o tema desperta entre os pesquisadores brasileiros. Além do volume de respondentes, chama a atenção que 35,39% deles tenham interesse em receber informações sobre o projeto por e-mail e que 22,54% aceitariam participar de uma entrevista para aprofundar a discussão.

Os respondentes têm em média 48 anos de idade (mediana 47). O respondente mais velho tem 86 anos e o mais jovem, 20 anos. A grande maioria (79,50%) se declarou professor pesquisador, 10,78% se declarou pesquisador e 9,72%, aluno de pós-graduação. Os respondentes estão vinculados a mais de 500 instituições, sendo que as maiores representatividades advêm da Universidade de São Paulo (USP), da Universidade Federal do Rio Grande do Sul (UFRGS), da Universidade Federal de Minas Gerais (UFMG) e da Universidade Estadual Paulista (UNESP). Embrapa e Fiocruz foram as instituições de pesquisa com participação mais expressiva na *survey*. No entanto, há respondentes da Petrobrás, do Centro Brasileiro de Pesquisas Físicas (CBPF), dos Centros Tecnológicos (CEFET) e de institutos federais de educação (IF) de vários estados. Alguns respondentes informaram situação de desemprego e outros de aposentadoria. Curiosamente, alguns respondentes do exterior também participaram, informando universidades de filiação americanas, espanholas e portuguesas. Infere-se que se trata de pesquisadores brasileiros em estágio, missão ou pós-doutorado em instituição do exterior.

Observa-se, na Figura 1, a representatividade das oito áreas do conhecimento conforme classificação do CNPq. Entre os 4.357 respondentes desta questão, há maior volume de respostas das Ciências Humanas (21,00%), seguida por Ciências Exatas e da Terra (17,19%) e Ciências Sociais Aplicadas (15,10%).

Figura 1: Respondentes de acordo com grandes áreas do CNPq



Fonte: Elaborada pelos autores com base nos dados da pesquisa.

A representatividade das áreas no questionário corresponde, aproximadamente, a representatividade das áreas no Brasil, de acordo com dados do Censo CNPq 2016, com exceção das Ciências Exatas e da Terra, que tem participação em 10,00% dos grupos de pesquisa brasileiros e, no entanto, perfaz 17,19% dos questionários (CONSELHO NACIONAL ..., 2016, online).

A Tabela 1 apresenta dados acerca de sexo e de escolaridade. A grande maioria dos respondentes possui formação de doutorado (45,20%) e pós-doutorado (36,95%). Os mestres perfazem 14,91% do conjunto e os graduados compõem 2,94%.

Tabela 1: Nível de escolaridade e sexo dos pesquisadores

Nível de Escolaridade	Masculino		Feminino		Total	
	Quantidade	Porcentagem	Quantidade	Porcentagem	Quantidade	Porcentagem
Graduação	77	60,16%	51	39,84%	128	2,94%
Mestrado	379	58,40%	270	41,60%	649	14,91%
Doutorado	1092	55,49%	876	44,51%	1968	45,20%
Pós-Doutorado	924	57,43%	685	42,57%	1609	36,95%
Total	2472	56,78%	1882	43,22%	4354	100,00%

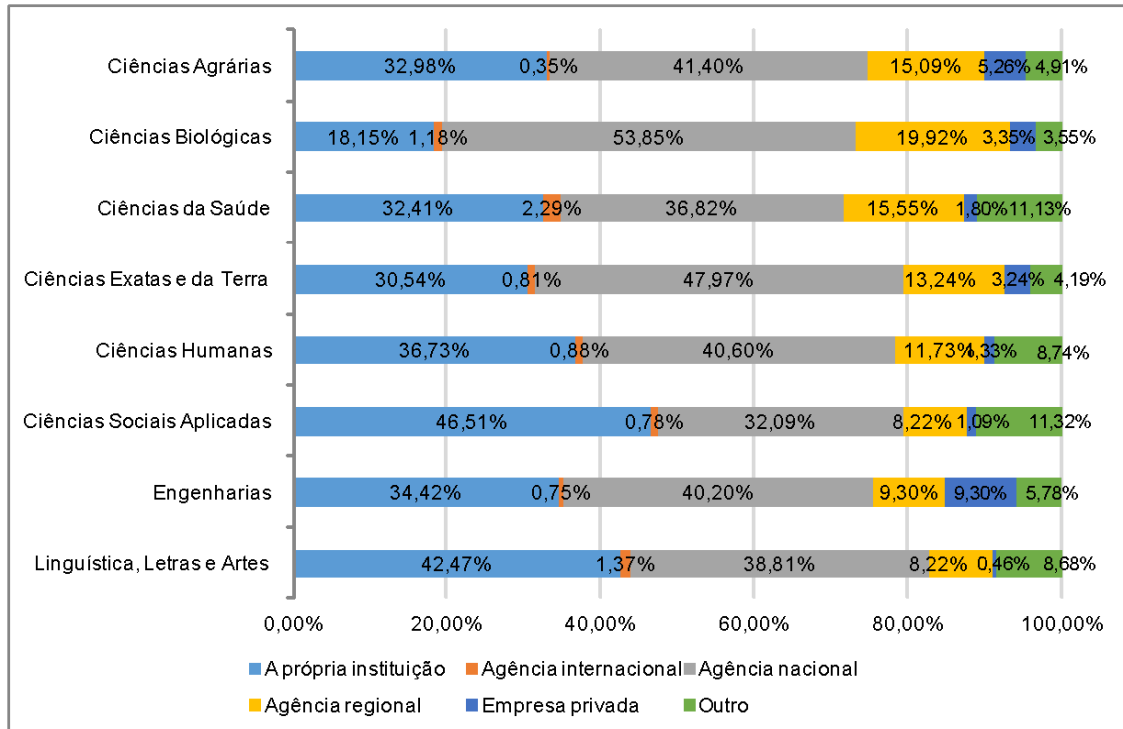
Fonte: Elaborada pelos autores com base nos dados da pesquisa.

A porcentagem de mulheres e de homens respondentes do questionário foi 43,22% e 56,78%, respectivamente. Esse resultado difere um pouco do quadro brasileiro. Conforme o relatório *Gender in the Global Research Landscape*, da editora Elsevier, a

participação na ciência é considerada equitativa entre ambos os sexos, sendo que os homens correspondem a 51% dos autores de pesquisas e as mulheres, a 49% (ACADEMIA BRASILEIRA DE CIÊNCIAS, 2017).

Uma das questões buscou identificar a fonte de apoio financeiro recebido pelo pesquisador, pois estudos como o de Kim e Stanton (2015) salientam que uma das formas de incentivar a abertura dos dados de pesquisa é a exigência por parte das agências financiadoras de um plano de gestão dos dados antes da concessão. A Figura 2 apresenta a relação entre grande área e financiadores da pesquisa.

Figura 2: Relação entre grandes áreas CNPq e principais financiadores de pesquisa



Fonte: Elaborada pelos autores com base nos dados da pesquisa.

As agências nacionais de financiamento são responsáveis por percentuais que variam de 32,09% (Ciências Sociais Aplicadas) a 53,85% (Ciências Biológicas). O financiamento por parte de empresas privadas é bastante reduzido para todas as áreas, com leve diferença para as Engenharias. O mesmo acontece com o financiamento por agências internacionais, que apresenta leve diferença para as Ciências da Saúde.

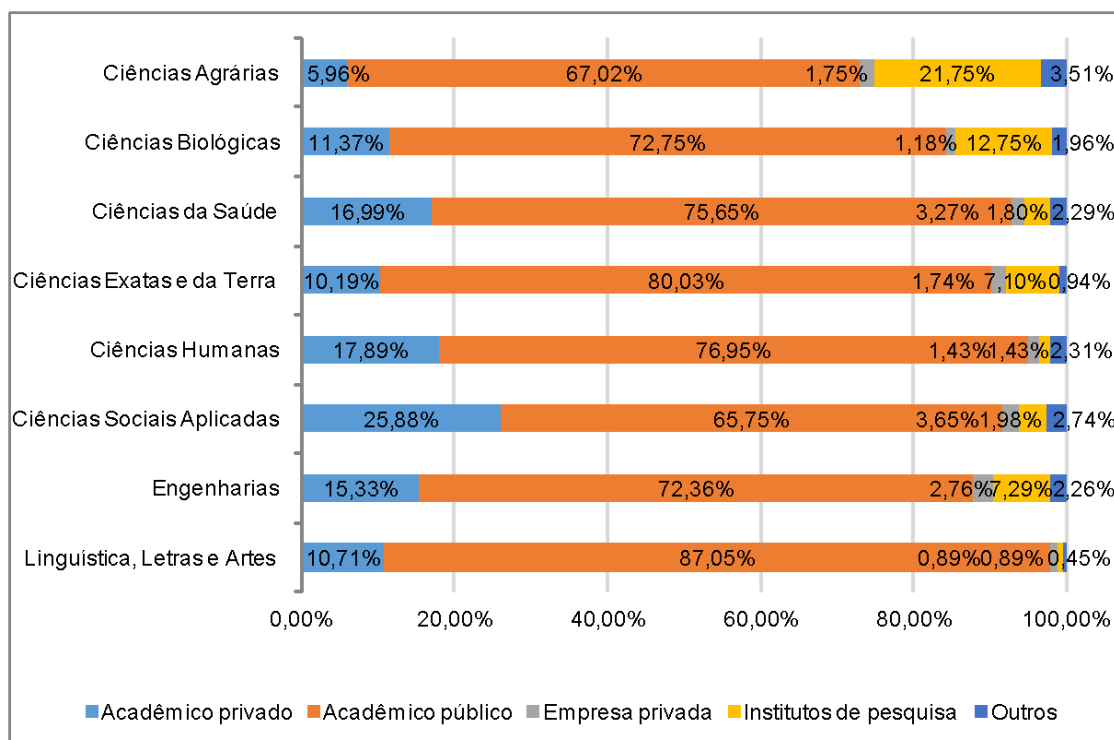
As formas de financiamento que não se enquadram nas alternativas apresentadas, classificadas na opção "outros", perfizeram um percentual de 7,31% do total de respostas. A análise dessa opção revela que 87% dos respondentes indicaram realizar suas pesquisas com recursos próprios.

Os resultados da pesquisa conduzida por Tenopir et al. (2011) demonstram que os próprios pesquisadores reconhecem que pesquisas financiadas com recursos públicos devem gerar resultados que são propriedade pública e que, dessa forma, devem ser preservados como tal.

Em relação aos setores aos quais os pesquisadores estão vinculados, o setor acadêmico público aparece em 74,52% das respostas; o acadêmico privado, em 15,53%;

institutos de pesquisa, em 6,19%; empresa privada, em 2,06%. Os institutos de pesquisa despontam na área de Ciências Agrárias, o que pode corresponder a grande participação da Embrapa na pesquisa brasileira. A Figura 3 apresenta os setores por área do conhecimento.

Figura 3: Relação grandes áreas CNPq com os setores aos quais os pesquisadores estão vinculados

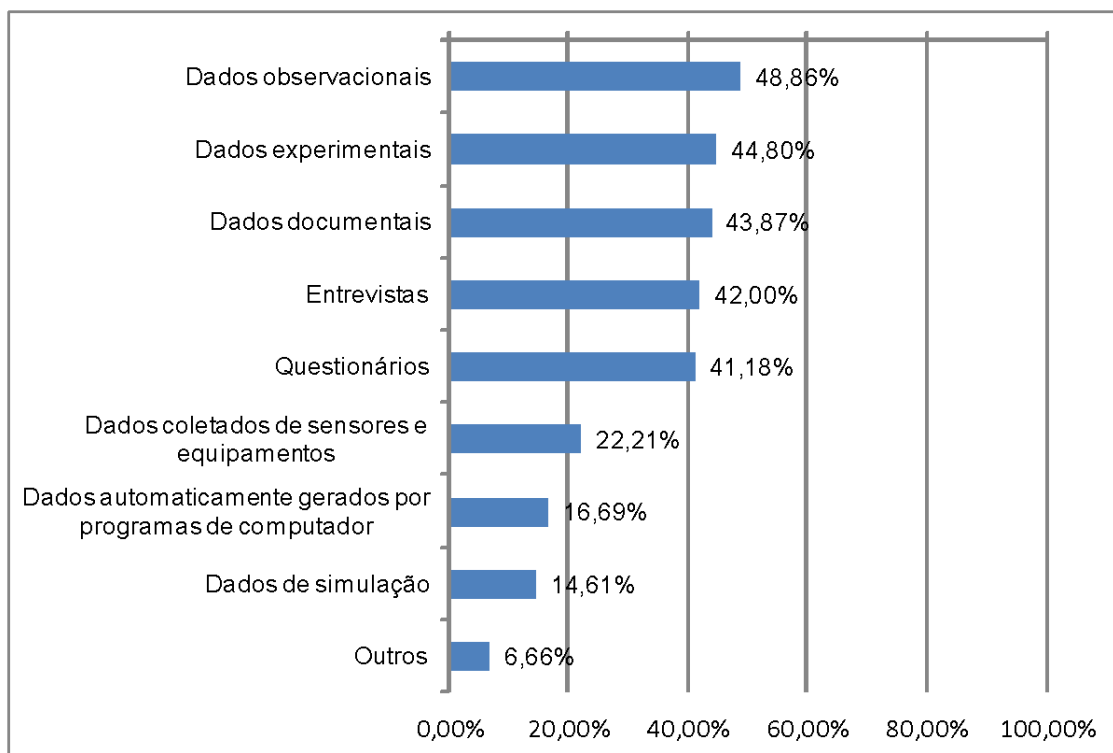


Fonte: Elaborada pelos autores com base nos dados da pesquisa.

A partir das questões subsequentes, buscou-se identificar as práticas de uso, de reuso e de compartilhamento de dados de pesquisa. Inicialmente, os pesquisadores selecionaram os termos que melhor descreviam os seus dados, bem como os principais tipos de dados produzidos. Salienta-se que a tipologia de dados utilizada como opção de resposta não era baseada em conceitos e termos únicos, o que talvez possa ter gerado problemas na seleção. A lacuna terminológica sobre o tema foi observada por Sales e Sayao (2019), que propuseram uma taxonomia baseada nos preceitos da área da organização do conhecimento para a classificação dos tipos de dados de pesquisa, o que pode auxiliar em futuros trabalhos.

A Figura 4 mostra que mais de 48% dos respondentes alegam que o melhor termo para nomear os dados por eles produzidos são “Dados observacionais”, seguido de “Dados experimentais”, “Dados documentais”, “Entrevistas” e “Questionários”.

Figura 4: Termos que melhor descrevem os dados produzidos pelos respondentes em todas as áreas do conhecimento



Fonte: Elaborada pelos autores com base nos dados da pesquisa.

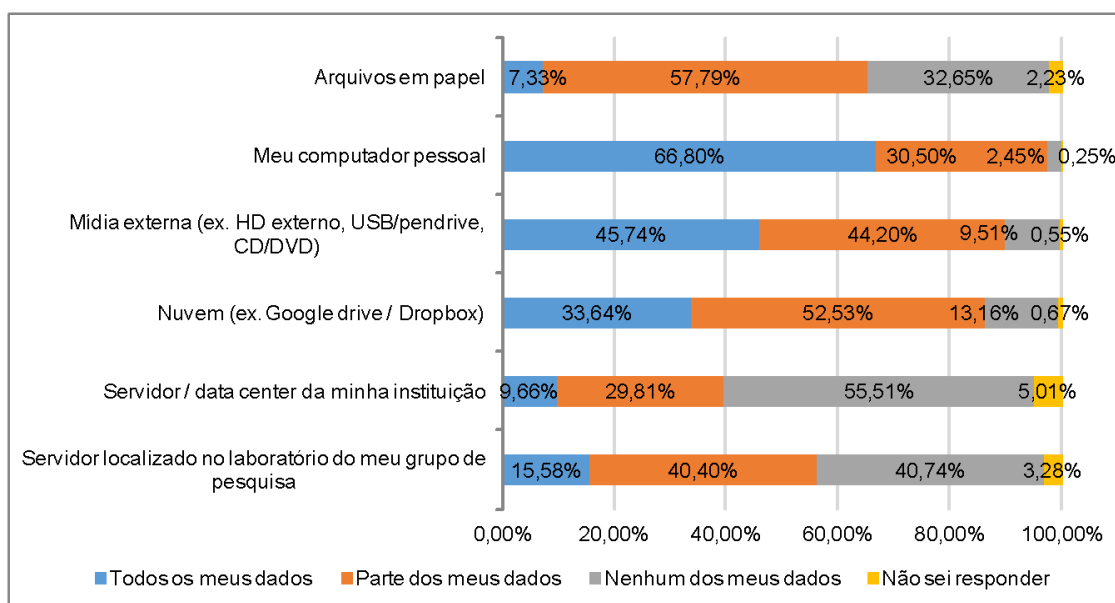
O resultado por área mostra que os “Dados observacionais” são utilizados em maior escala pelas Ciências Humanas (29,76%) e pelas Ciências Sociais Aplicadas (21,36%). Referente ao termo “Dados de Simulação”, as duas áreas que obtiveram o maior percentual foram Ciências Exatas e da Terra (41,77%) e Engenharias (32,19%).

Acerca do volume de dados produzidos por ano, em média 32,90% dos entrevistados não souberam responder. Ao comparar estes resultados com resultados de uma *survey* realizada na Áustria, na qual apenas 7% dos pesquisadores não souberam responder (BAUER et al., 2015), percebe-se a despreocupação ou o desconhecimento do pesquisador brasileiro em relação ao armazenamento de seus dados. Observa-se que o percentual de respondentes que não soube responder é grande nas Ciências da Saúde (46,56%), Ciências Humanas (43,62%) e Linguística, Letras e Artes (52,15%).

Na perspectiva das pesquisas que geram dados, pode-se observar que 81,19% dos pesquisadores trabalham em colaboração. A maioria dos grupos (55,97%) é formada por até cinco pessoas. Os grupos com maior número de colaboradores, formados de seis a 12 pessoas, foram apontados por 17,75% dos respondentes, e apenas 7,47% dos respondentes apontaram grupos formados por mais de 12 pessoas. A realização de pesquisa individualmente foi mencionada por 18,81% dos respondentes.

A forma como os dados são armazenados pelos pesquisadores também foi investigada (Figura 5), pois como afirma Tenopir et al. (2011), o encorajamento do compartilhamento e do reuso de dados de pesquisa começa com boas práticas em todas as fases do ciclo de vida do dado.

Figura 5: Locais utilizados para armazenamento de dados



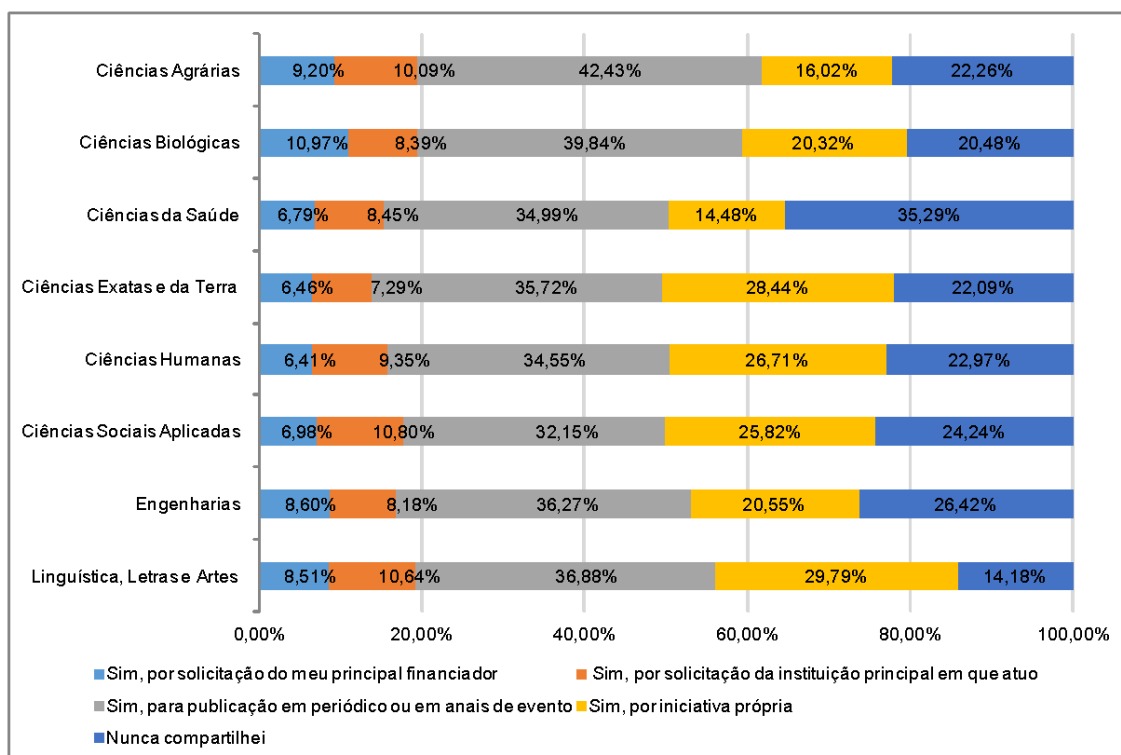
Fonte: Elaborada pelos autores com base nos dados da pesquisa.

As formas de armazenamento de dados de pesquisa exibem um quadro heterogêneo. Dentre as respostas, destaca-se que 57,79% dos respondentes guardam parte dos seus dados em arquivos em papel, e que o computador pessoal armazena todos os dados de 66,8% dos pesquisadores. Outro dado relevante levantado no questionário diz respeito ao armazenamento de dados no Servidor/data center da instituição, tendo em vista que 55,51% dos respondentes alega não armazenar nenhuma parte de seus dados neles. A nuvem também é muito utilizada pelos respondentes, onde guardam todos os seus dados (33,64%) ou parte deles (52,53%).

Os resultados são similares aos apontados pela survey da Áustria, na qual 71% dos pesquisadores afirmou armazenar seus dados de pesquisa em seu computador, 64% armazena em mídias externas como disco ou pen drive, e apenas 39% afirmou utilizar servidores da instituição (BAUER et al., 2015). Assim como na Áustria, percebe-se, no Brasil, a falta de políticas e de diretrizes uniformes para a preservação e a segurança de dados de pesquisa - a matéria prima do empreendimento científico.

Em seguida, os respondentes foram questionados se já haviam compartilhado seus dados de pesquisa em acesso aberto. Foram vinculadas quatro razões à resposta “Sim”, além da opção de resposta “Nunca compartilhei”. A Figura 6 apresenta a relação entre as grandes áreas, a experiência do compartilhamento e o motivo que leva os pesquisadores a compartilhar seus dados em acesso aberto.

Figura 6: Motivos para o compartilhamento de dados de pesquisa em acesso aberto por grandes áreas CNPq



Fonte: Elaborada pelos autores com base nos dados da pesquisa.

Em média, quase um quarto dos respondentes (23,49%) nunca compartilhou dados de pesquisa, com percentuais que variam de 14,18% na Linguística, Letras e Artes até 35,29% nas Ciências da Saúde.

Dentre os principais motivos apontados pelos respondentes para compartilhar dados de pesquisa em acesso aberto, destaca-se a necessidade para publicação em periódico ou em anais de evento, indicada por 36,60% dos respondentes – com variações relevantes por áreas específicas, como as Ciências Agrárias, onde o percentual é de 42,43%. O compartilhamento por iniciativa própria também acontece e é destaque para algumas áreas. Nas Ciências da Saúde, esta motivação é baixa (14,48%).

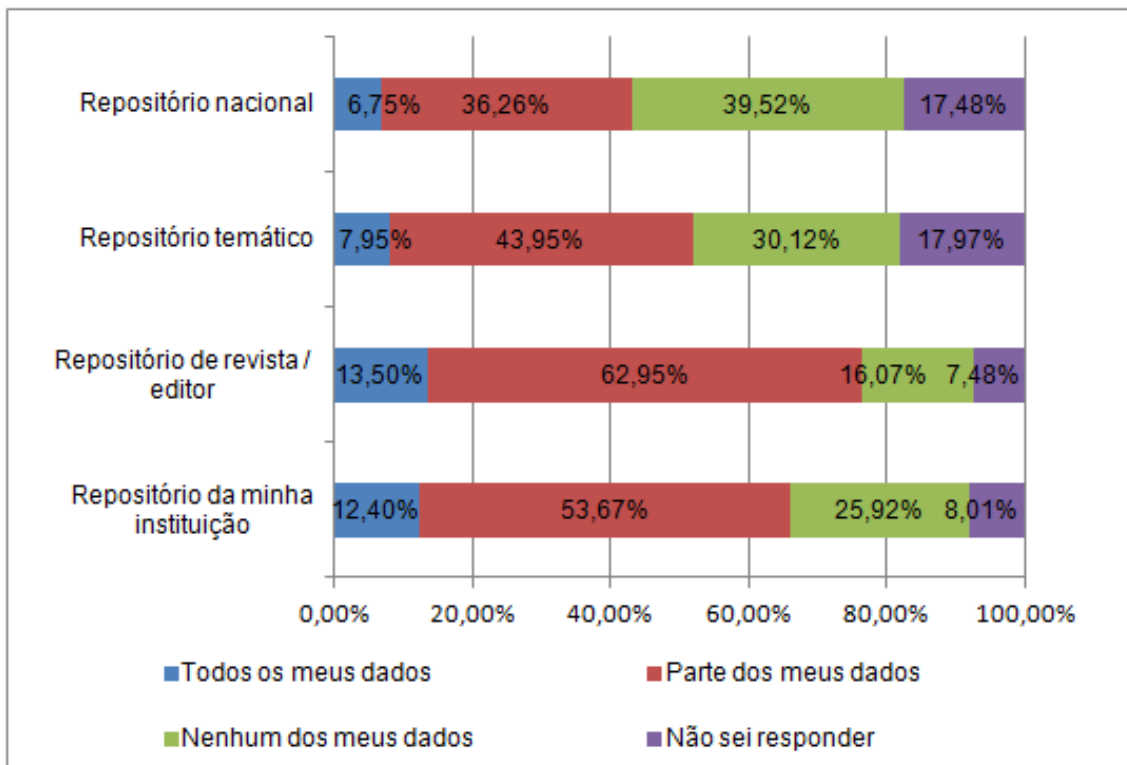
É importante destacar que a literatura internacional relata que o compartilhamento de dados de pesquisa não é uma prática comum entre cientistas sociais, e que, quando acontece, alegam-se motivações pessoais (KIM; ADLER, 2015). A comunidade brasileira de Ciências Sociais Aplicadas respondeu que a solicitação do financiador da pesquisa, da instituição de atuação ou do periódico/evento de publicação dos resultados correspondem a metade de sua motivação para o compartilhamento de dados de pesquisa.

A prática de compartilhamento de dados, de fato, não é algo tão comum na pesquisa brasileira. Tal efeito pode ser evidenciado no baixo percentual de respondentes que indicaram compartilhar todos os seus dados de pesquisa (9,19%), e o alto percentual de respondentes que não compartilham dado algum (37,02%). O compartilhamento de parte dos dados é uma prática apontada por 53,79% dos

respondentes. Entre a comunidade austríaca, 22% dos respondentes não compartilha (BAUER et al., 2015).

Uma das questões teve por finalidade identificar o tipo de repositório em que os pesquisadores depositam os dados que estão dispostos a compartilhar abertamente. Os resultados são apresentados na Figura 7.

Figura 7: Tipo de repositório de dados de pesquisa em acesso aberto utilizados pelos respondentes



Fontes: Elaborada pelos autores com base nos dados da pesquisa.

Observa-se que 62,95% dos respondentes indicaram armazenar parte de seus dados junto a “Repositório de revista/editor”, seguido pelo repositório da própria instituição (53,67%). Ao somar os pesquisadores que responderam “Nenhum dos meus dados” e “Não sei responder” temos um total médio de 40,64%, o que reitera resultados anteriores. Um percentual alto (39,52%) indicou que nenhum de seus dados está armazenado em repositório nacional.

A necessidade de publicar os resultados da pesquisa antes do compartilhamento foi apontada por 52,73% dos respondentes como motivo para não compartilhar os dados ou compartilhar parcialmente. Na Figura 8, também aparecem como razões para tal a falta de exigência (31,94%), de infraestrutura (31,14%) e de financiamento (24,86%), seguidas de outros motivos.

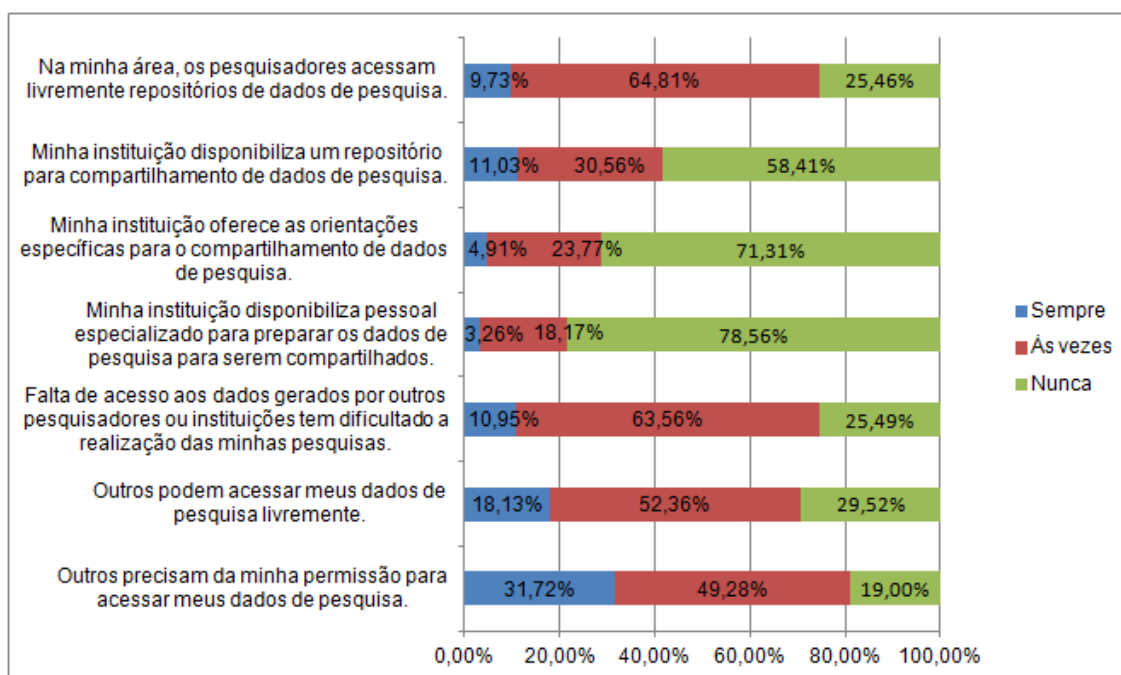
Figura 8: Razões para não compartilhar ou compartilhar parcialmente os dados de pesquisa



Fonte: Elaborada pelos autores com base nos dados da pesquisa.

Os respondentes foram instigados a apontar a frequência (sempre, às vezes e nunca) de determinadas situações e atitudes relacionadas ao compartilhamento de dados de pesquisa. Os resultados são apresentados na Figura 9.

Figura 9: Frequência de situações e de atitudes relacionadas ao compartilhamento de dados de pesquisa



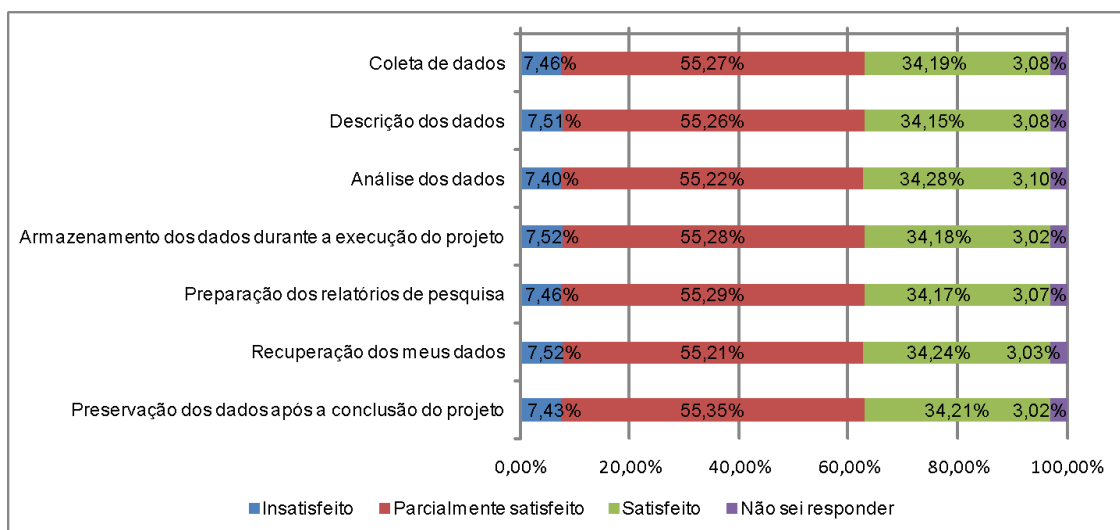
Fonte: Elaborada pelos autores com base nos dados da pesquisa.

Os pesquisadores brasileiros parecem ter um perfil mais conservador no que diz respeito ao compartilhamento e ao acesso a dados de pesquisa. Observa-se que um percentual pequeno (9,73%) dos respondentes afirma que, em sua área, os pesquisadores sempre acessam livremente os repositórios de dados de pesquisa. Dos pesquisadores, 31,72% afirmam que sempre solicitam a permissão para o acesso a seus próprios dados de pesquisa, e que o acesso é livre em apenas 18,13% dos casos.

Resultados da survey da Áustria demonstram que, naquele país, mais da metade dos pesquisadores permitem acesso aos dados de pesquisa mediante autorização, com diferenças importantes entre as áreas do conhecimento: pesquisadores das Humanas e das Ciências Sociais são mais restritivos, enquanto os pesquisadores da Matemática têm um comportamento mais aberto. De maneira geral, o acesso a dados de pesquisa austríacos era bastante restrito no momento da aplicação da survey (BAUER et al., 2015).

Os pesquisadores foram questionados quanto à sua satisfação na organização de seus dados (Figura 10).

Figura 10: Nível de satisfação do pesquisador com a organização dos dados de pesquisa



Fonte: Elaborada pelos autores com base nos dados da pesquisa.

Outro dado que chama a atenção é que a disponibilização de repositório para compartilhamento de dados de pesquisa nunca acontece para 58,41% dos respondentes, indicando a falta de oferta por parte das instituições brasileiras. No que diz respeito a oferta de orientação e de pessoal especializado no tema, a situação é ainda mais crítica: 71,31% dos respondentes apontaram que a instituição nunca oferece orientações e 78,56% aponta que a instituição não oferece pessoal especializado para preparar os dados a serem compartilhados.

Outro aspecto importante explorado pelo questionário relaciona-se com a familiaridade dos respondentes com a gestão de dados de pesquisa. A análise dos dados por grandes áreas do CNPq mostrou-se bastante homogênea quanto às possibilidades de conhecimento. Observa-se que, em média, 12,20% dos respondentes afirmaram conhecer bastante sobre gestão de dados de pesquisa, 62,28% conhecem pouco e 25,52% afirmaram desconhecer o tema.

Quando perguntados se já elaboraram um plano de gestão de dados, os respondentes apresentaram perfil bastante variado, conforme Tabela 2.

Tabela 2: Elaboração de plano de gestão de dados por grandes áreas CNPq

Áreas	Sim, por solicitação do meu principal financiador %	Sim, por solicitação da instituição principal em que eu atuo %	Sim, para publicação em periódico %	Sim, por outras razões %	Nunca elaborei %
Ciências Agrárias	7,74	7,23	9,94	6,15	6,39
Ciências Biológicas	20,00	13,86	13,08	10,40	11,21
Ciências da Saúde	16,77	13,86	18,05	14,66	13,04
Ciências Exatas e da Terra	16,13	11,45	13,63	19,39	17,39
Ciências Humanas	12,90	24,10	18,23	20,57	21,71
Ciências Sociais Aplicadas	9,03	12,65	11,60	10,64	15,61
Engenharias	14,19	12,05	10,13	10,40	9,39
Linguística, Letras e Artes	2,58	4,22	4,97	7,09	4,86
Média	12,42	12,42	12,45	12,41	12,45

Fonte: Elaborada pelos autores com base nos dados da pesquisa.

Observa-se que o percentual de respondentes que nunca elaborou um plano de gestão de dados é baixo em áreas como Ciências Agrárias (6,39%), e Linguística, Letras e Artes (4,86%). O percentual é maior nas áreas de Ciências Exatas e da Terra (17,39%) e de Ciências Humanas (21,71%).

Dentre os 4.679 respondentes, 573 (13,72%) consideram-se “Respondentes Experientes”, ou seja, aqueles que já depositaram parte ou todos seus dados em repositório temático, institucional, nacional ou de revista e fizeram Plano de Gestão de Dados por solicitação do financiador, por solicitação da instituição, para publicação em periódico ou por outras razões. Os dados são apresentados na Tabela 3.

Tabela 3: Perfil de experiência dos pesquisadores

Área do conhecimento	Já depositou em repositório e fez plano %	Já fez plano %	Já depositou em repositório %	Já compartilhou %
Ciências Agrárias	18,21	22,68	57,73	60,48
Ciências Biológicas	15,69	19,37	56,78	58,33
Ciências da Saúde	15,09	19,64	43,01	42,53
Ciências Exatas e da Terra	10,78	12,38	55,92	58,98
Ciências Humanas	11,53	13,92	53,21	54,73
Ciências Sociais Aplicadas	10,25	12,06	48,41	50,67
Engenharias	15,00	17,00	52,25	55,00
Linguística, Letras e Artes	13,21	14,53	59,03	62,11

Média	13,72	16,44	53,29	55,35
--------------	-------	-------	-------	-------

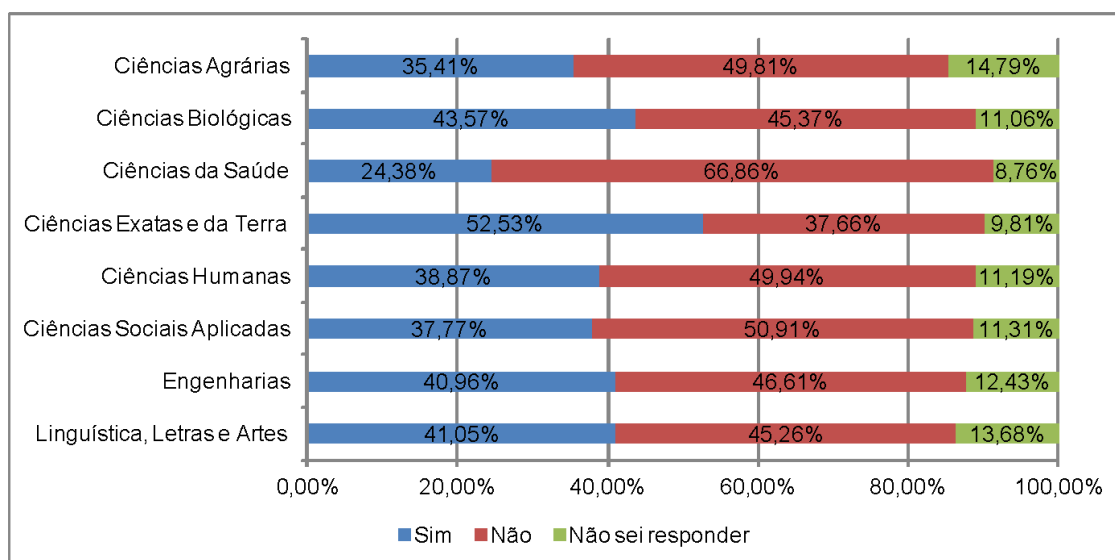
Fonte: Elaborada pelos autores com base nos dados da pesquisa.

Quando perguntados se utilizam algum repositório para acessar dados de pesquisa de outros pesquisadores, 31,71% dos respondentes afirmaram que sim, enquanto 68,29% responderam que não (Figura 11). Ao serem solicitados para especificar o nome do repositório utilizado, observa-se que a grande maioria dos respondentes informou bases de dados como Scielo e Web of Science; portais, como Portal de Periódicos da Capes; redes sociais acadêmicas como Mendeley, ResearchGate e Academia.edu; repositórios institucionais; e revistas científicas. Isso denota o desconhecimento por parte dos pesquisadores brasileiros do que seja um repositório de dados de pesquisa. Tal fato gera uma interpretação diferente dos resultados sumarizados na Figura 11, porque o conceito de repositório de dados é equivocado para muitos pesquisadores. Alguns pesquisadores informaram exemplos de repositórios utilizados, como o Dryad, GenBank, ITIS.

A título de contextualização, o relatório da *survey* Austríaca também relata inúmeros exemplos equivocados mencionados pelos pesquisadores (BAUER et al., 2015), denotando que este ainda é um tema que precisa ser mais explorado internacionalmente.

A última questão do eixo de práticas perguntou se os respondentes já utilizaram em suas pesquisas dados abertos compartilhados por outros grupos de pesquisa. Em média, 49,36% dos respondentes nunca utilizou dados compartilhados por outros grupos; 39,55% já fez uso deste tipo de dado e 11,09% não soube responder. Os resultados detalhados por grande área do CNPq são apresentados na Figura 11.

Figura 11: Utilização de dados elaborados por outros grupos de pesquisa por grandes áreas CNPq



Fonte: Elaborada pelos autores com base nos dados da pesquisa.

É surpreendente que as Ciências da Saúde, área que se caracteriza pela grande produção de dados e de séries longitudinais extensas, apresente o maior percentual de respondentes que não utilizam dados compartilhados por outros grupos (66,86%),

talvez por questões éticas relacionadas à pesquisa envolvendo seres humanos e por uma cultura de confidencialidade desses dados. A área que se destaca por ter respondido que já utilizou dados elaborados por outros grupos de pesquisa é a Ciências Exatas e da Terra (52,53%).

No que diz respeito às percepções dos pesquisadores, observa-se que grande parte dos respondentes utilizariam dados de outros pesquisadores caso estivessem acessíveis (63,38% concordaram e 29,92% concordaram parcialmente com a afirmativa). Da mesma maneira, a maioria dos respondentes compartilharia pelo menos parte de seus dados de pesquisa em um repositório de acesso aberto sem restrições (54,82% concordaram e 35,92% concordaram parcialmente).

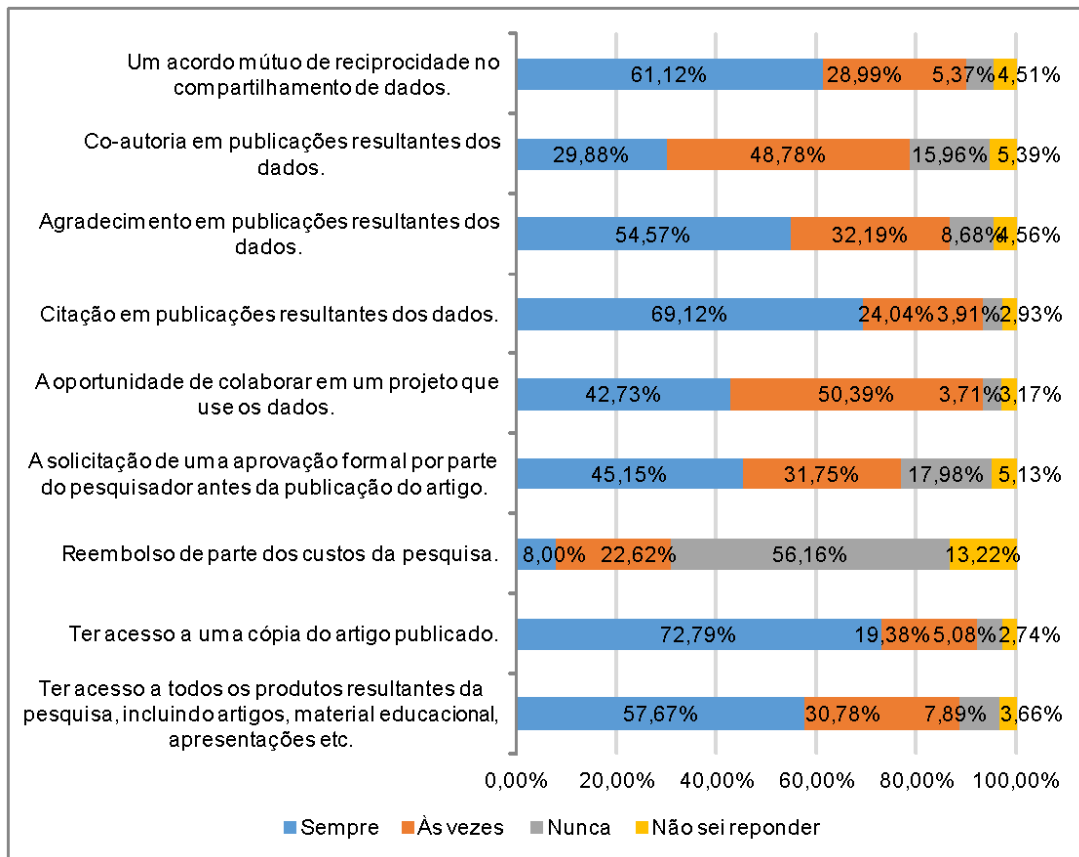
Já o compartilhamento de todos os seus dados de pesquisa em um repositório de acesso aberto sem restrições recebeu respostas um pouco distintas: 21,86% concordaram, 42,68% dos respondentes concordaram parcialmente, e 31,71% discordaram da afirmativa. Esse resultado demonstra que a ideia de compartilhamento total e irrestrito ainda não é aceita pelos respondentes.

Grande parte dos respondentes considera que o compartilhamento de dados pode ampliar o reconhecimento acadêmico, por exemplo, por meio da citação (65,90% concordam e 26,64% concordam parcialmente com a afirmativa). Esses resultados sem dúvida demonstram que há receptividade à ideia do compartilhamento e do reuso de dados por parte dos pesquisadores brasileiros.

Um aspecto importante a considerar é expectativa dos pesquisadores em relação ao compartilhamento de seus dados. Em estudo longitudinal com pesquisadores americanos das áreas da computação e da engenharia, Wallis et al. (2013) comentaram que os participantes desejavam compartilhar seus dados, desde que observadas determinadas condições, por exemplo, direito de publicar primeiro, atribuição adequada da fonte (citação e agradecimentos), familiaridade entre quem compartilha e quem usa, entre outros.

Neste sentido, quando perguntados sobre o retorno esperado a partir do compartilhamento de dados, os pesquisadores brasileiros responderam o que consta na Figura 12.

Figura 12: Retorno esperado a partir do compartilhamento de dados de pesquisa



Fonte: Elaborada pelos autores com base nos dados da pesquisa.

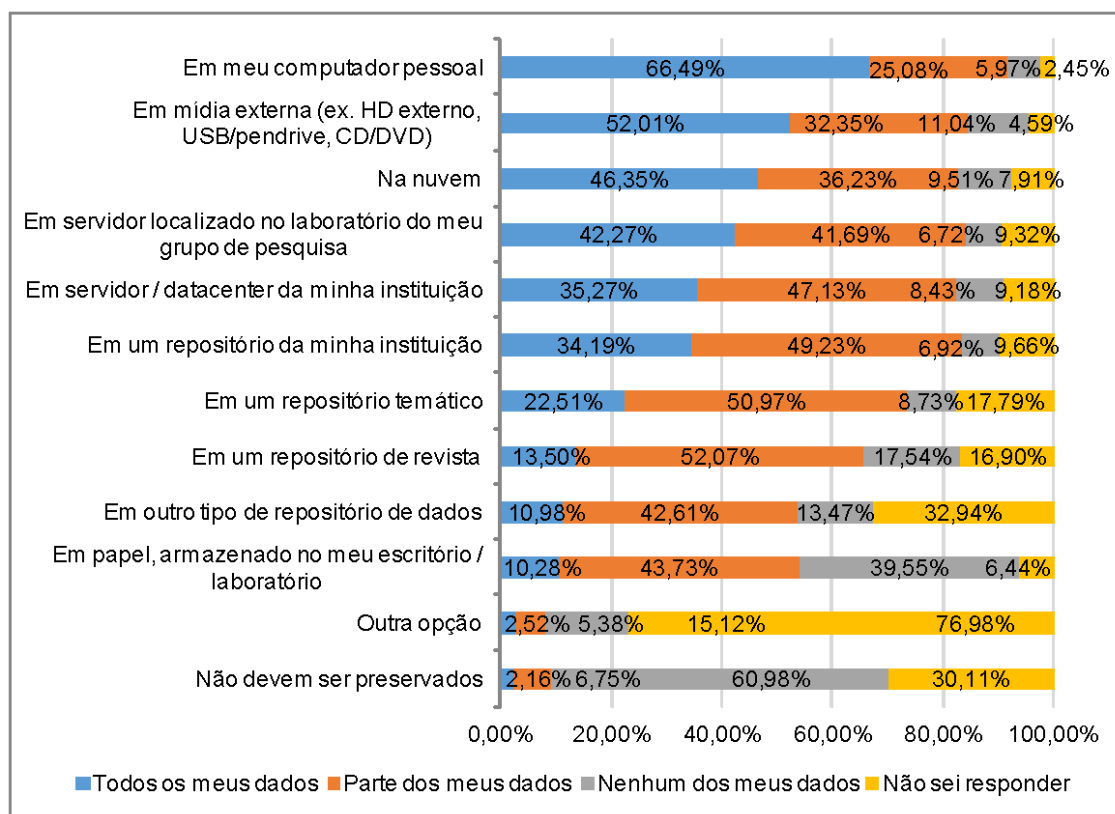
Observa-se que grande parte dos pesquisadores espera que sempre ocorra um acordo de reciprocidade no compartilhamento de dados (61,12%), a citação em publicações resultantes dos dados (69,12%) e o acesso a uma cópia do artigo publicado (72,79%).

A coautoria em publicações resultantes dos dados é uma expectativa alta: 29,88% dos pesquisadores responderam que a esperam sempre e 48,78% responderam que a esperam às vezes. Da mesma maneira, os agradecimentos em publicações resultantes dos dados também são esperados: 54,57% respondeu que os esperam sempre e 32,19% respondeu os esperam às vezes.

Inferese que exista um desejo de parceria e de colaboração entre os respondentes, tendo em vista que pouco mais de 90% deles esperam, sempre ou às vezes, uma oportunidade de colaborar em um projeto que use os dados. Mais da metade dos respondentes afirma que nunca espera o reembolso de parte dos custos da pesquisa. Os retornos esperados pela comunidade científica brasileira estão alinhados às motivações apontadas pela comunidade austríaca para compartilhar dados, quais sejam: visibilidade, impacto, colaboração e reconhecimento (BAUER et al., 2015).

A Figura 13 apresenta resultados acerca do local onde os respondentes consideram adequado preservar seus dados.

Figura 13: Local onde os dados de pesquisa deveriam ser preservados



Fonte: Elaborada pelos autores com base nos dados da pesquisa.

Observa-se que a maioria dos pesquisadores considera o computador pessoal como local adequado para preservação de seus dados de pesquisa – 66,49% dos respondentes apontaram este suporte para preservação de todos os seus dados e 25,08%, para parte de seus dados. Em seguida, aparecem as mídias externas, a nuvem, o servidor do laboratório do grupo de pesquisa, o servidor da instituição e somente depois aparecem os repositórios da instituição (34,19%), repositório temático (22,51%), repositório de revista (13,50%) e outros tipos de repositórios (10,98%). Surpreendentemente, o papel segue sendo o suporte para preservação de todos ou parte dos dados de 64,01% dos respondentes.

A Tabela 4 apresenta resultados acerca de quem é considerado responsável pela oferta de serviços de apoio a gestão dos dados de pesquisa.

Tabela 4: Responsabilidade pela oferta de serviços de apoio a gestão dos dados de pesquisa

Áreas	Agência de financiamento %	Agência não governamental %	Associação científica %	Ente governamental (ex, instituição subordinada ao MCTIC) %	Instituição de atuação do pesq. %	Não acho necessário apoio %	Não sei responder %
Ciências Agrárias	17,80	0,85	5,08	15,25	50,42	1,27	7,63
Ciências Biológicas	19,71	1,92	5,29	15,38	49,04	0,72	6,49

Ciências da Saúde	14,37	1,01	8,50	14,98	51,42	1,21	7,29
Ciências Exatas e da Terra	19,32	2,05	8,55	13,33	40,34	2,91	11,97
Ciências Humanas	17,47	0,67	8,20	11,69	50,67	1,21	8,87
Ciências Sociais Aplicadas	16,44	1,16	10,44	10,64	48,16	1,35	9,48
Engenharias	20,18	0,61	9,17	12,54	44,04	0,92	11,31
Linguística, Letras e Artes	24,71	0,00	6,32	8,05	51,72	0,00	5,75
Total	18,15	1,14	8,04	12,88	47,82	1,37	9,01

Fonte: Elaborada pelos autores com base nos dados da pesquisa.

Quando perguntados acerca de quem deve oferecer serviços de apoio na gestão dos dados de pesquisa, os respondentes apontaram prioritariamente a instituição de atuação do pesquisador, com média de 47,82% das respostas para todas as áreas. Logo após, a agência de financiamento aparece como segunda responsável pelo apoio, com média de 18,15% das respostas.

A conclusão deste artigo apresenta a interpretação dos resultados obtidos a partir dos eixos 1 e 2 do questionário.

CONCLUSÕES

O estabelecimento de repositórios para dados de pesquisas é uma estratégia vantajosa para o Brasil, tendo em vista que amplia a visibilidade e a reputação dos pesquisadores e, por consequência, das instituições de pesquisa às quais estão vinculados. A iniciativa resulta em benefícios para a sociedade. Para criação da infraestrutura adequada para a curadoria de dados, é preciso inicialmente obter-se informações a respeito dos pesquisadores, aqueles que na prática coletam, criam e utilizam estes dados. Partindo deste pressuposto, este projeto de pesquisa foi desenvolvido.

A pesquisa teve como objetivo elucidar a situação do acesso a dados de pesquisa no Brasil: se os pesquisadores disponibilizam seus dados e se reutilizam dados abertos, e o que pensam a respeito dessa prática. Os resultados mostraram que em média, 49,36% dos respondentes nunca utilizou dados compartilhados por outros grupos; e quase um quarto dos respondentes (23,49%) nunca compartilhou dados de pesquisa, com percentuais que variam de 14,18% na Linguística, Letras e Artes até 35,29% nas Ciências da Saúde.

Entende-se que a gestão e a disponibilização de dados de pesquisa é a chave para que esses possam ser reutilizados por pesquisadores brasileiros ou estrangeiros em futuras pesquisas. Apesar de 57,79% dos respondentes armazenarem parte dos seus dados em arquivos em papel e 66,80% guardarem todos os dados em seu computador pessoal, há um entendimento geral de que os dados devam ser compartilhados. Por exemplo, 92,54% concordam total ou parcialmente que o compartilhamento pode ampliar o reconhecimento acadêmico, por exemplo, por meio da citação para os conjuntos de dados. Entretanto, entre os motivos para não compartilharem, destacam-se a necessidade de publicar antecipadamente os resultados da pesquisa (52,73%), a falta de exigência (31,94%), a falta de infraestrutura (31,14%) e a falta de financiamento (24,86%).

Fica evidente o problema de infraestrutura para o armazenamento e o compartilhamento dos dados de pesquisa nas instituições brasileiras ao analisar que 71,31% dos respondentes apontaram que a instituição de trabalho nunca oferece orientações sobre o tema e que 78,56% responderam também que a instituição não oferece pessoal especializado para preparar os dados a serem compartilhados.

Outra inferência possível a partir da análise das respostas diz respeito à responsabilidade pelo armazenamento e pela custódia dos dados. Parece ser fundamental o apoio de especialistas treinados para a gestão de dados de pesquisa, de modo a garantir a disponibilização destes dados para projetos futuros. Reflexões acerca de propriedade e de direito sobre os dados, se pertencem ao pesquisador responsável pelo projeto ou à instituição de filiação, precisam entrar na agenda de discussão da comunidade científica brasileira.

Por fim, o fato de mais de 32,90% dos respondentes desconhecer o volume total de dados que produz anualmente indica a necessidade de uma pesquisa em profundidade para melhor compreender as necessidades de espaço de armazenamento para cada área do conhecimento.

De uma forma geral, as respostas indicam desorientação por parte dos pesquisadores acerca dos conceitos de dado de pesquisa, das definições de repositórios e das práticas de disponibilização e de reuso. Dos 31,71% dos respondentes que afirmaram utilizar algum repositório para acessar dados de pesquisa de outros pesquisadores, grande parte informou bases de dados, portais, redes sociais acadêmicas, repositórios institucionais e revistas científicas. Este fato demonstra o desconhecimento por parte dos pesquisadores brasileiros do que seja um repositório de dados de pesquisa, assim como do próprio conceito de dados de pesquisa

Em relação às percepções dos pesquisadores, 21,86% concordaram em compartilhar todos os seus dados de pesquisa em um repositório de acesso aberto sem restrições, 42,68% dos respondentes concordaram parcialmente, e 31,71% discordaram da afirmativa. Este resultado demonstra que a ideia de compartilhamento total e irrestrito ainda não é aceita pelos respondentes. Nesse sentido, a pesquisa promove a discussão do tema entre a comunidade brasileira e pode contribuir para despertar o interesse e para minimizar dúvidas sobre compartilhamento e reuso de dados abertos de pesquisa.

A partir do diagnóstico apresentado, constata-se que o acesso aberto a dados de pesquisas no Brasil ainda é incipiente, em parte pela falta de familiaridade da comunidade científica com o tema, mas também pela ausência de infraestrutura que dê suporte a empreendimentos nessa direção. No entanto, esses problemas podem ser enfrentados por meio de políticas, projetos institucionais e pesquisas científicas.

O elemento motivador para a realização de tais ações é a receptividade dos pesquisadores brasileiros e o desejo de parcerias e colaboração, conforme demonstrados nas respostas dadas a este estudo.

Artigo recebido em 09/07/2019 e aprovado em 19/11/2019.

REFERÊNCIAS

ACADEMIA BRASILEIRA DE CIÊNCIAS. *Elsevier lança, na ABC, relatório sobre a presença de mulheres na ciência*. [S.l.], 26 jun. 2017. Disponível em:

<http://www.abc.org.br/2017/06/26/elsevier-lanca-na-abc-relatorio-sobre-a-presenca-de-mulheres-na-ciencia/>. Acesso em: 08 jun. 2019.

BAUER, B. *et al.* *Researchers and their data: results of an Austria survey: report 2015*. 2015. Disponível em: <https://phaidra.univie.ac.at/o:409318>. Acesso em: 08 jun. 2019.

BEAGRIE, N.; BEAGRIE, R.; ROWLANDS, I. Research data preservation and access: the views of researchers. *Ariadne*, v.60, July 2009. Disponível em: <http://www.ariadne.ac.uk/issue/60/beagrie-et-al/>. Acesso em: 08 jun. 2019.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO - CNPQ. Diretório dos Grupos de Pesquisa no Brasil Lattes. *Censo atual*. 2016. Disponível em: <http://lattes.cnpq.br/web/dgp/censo-atual/>. Acesso em: 08 jun. 2019.

DIGITAL CURATION CENTRE. *Data asset framework: implementation guide* [online]. 2009. Disponível em: https://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf. Acesso em: 08 jun. 2019.

KIM, Y.; ADLER, M. Social scientists' data sharing behaviors: investigating the roles of individual motivations, institutional pressures, and data repositories. *International Journal of Information Management*, v. 35, n. 4, p. 408-418, 2015.

KIM, Y.; STANTON, J. M. Institutional and individual factors affecting scientists' data sharing behaviors: a multilevel analysis. *Journal of the Association for Information Science and Technology*, v. 67, n. 4, p. 776-799, 2016.

KIM, Y.; ZHANG, P. Understanding data sharing behaviors of STEM researchers: The roles of attitudes, norms, and data repositories. *Library & Information Science Research*, v. 37, n. 3, p. 189-200, 2015.

PARHAM, S. W.; BODNAR, J.; FUCHS, S. *Supporting tomorrow's research: assessing faculty data curation needs at Georgia Tech*. C&RL News. Jan. 2012. p. 10-13. Disponível em: https://smartech.gatech.edu/bitstream/handle/1853/48706/Parham_CRLNews.pdf?sequence=1&isAllowed=y. Acesso em: 08 jun. 2019.

ROWLEY, J. *et al.* Academic's behaviors and attitudes towards open access publishing in scholarly journals. *Journal of the Association for Information Science and Technology*, v. 68, n. 5, p. 1201-1211, May 2017.

SALES, L. F.; SAYÃO, L. F. Uma proposta de taxonomia para dados de pesquisa. 2019. In: ISKO IBÉRICO 2019, 2019, Barcelona, Espanha. *Anales electrónicos* [...]. Disponível em: <http://eprints.rclis.org/38366/>. Acesso em: 08 jul. 2019.

TENOPIR, C. *et al.* Data sharing by scientists: practices and perceptions. *PLoS ONE*, v. 6, n. 6, 2011. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101>. Acesso em: 08 jul. 2019.

WALLIS, J. C.; ROLANDO, E.; BORGMAN, C. L. If we share data, will anyone use them?: data sharing and reuse in the long tail of science and technology. *PLoS ONE*, v. 8, n. 7, 2013. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0067332>. Acesso em: 08 jul. 2019.