



UnB

Universidade de Brasília
Faculdade de Ciência da Informação
Programa de Pós-Graduação em Ciência da Informação

**PRÁTICAS DE COMUNICAÇÃO DE DADOS DE PESQUISA: A PERCEPÇÃO
DE BOLSISTAS DE PRODUTIVIDADE BRASILEIROS DO CAMPO DA
GENÉTICA**

Brasília
Janeiro de 2022

PRISCILA RODRIGUES DOS SANTOS

**PRÁTICAS DE COMUNICAÇÃO DE DADOS DE PESQUISA: A PERCEPÇÃO
DE BOLSISTAS DE PRODUTIVIDADE BRASILEIROS DO CAMPO DA
GENÉTICA**

Dissertação apresentado ao Programa de Pós-Graduação em Ciência da Informação (PPGCIInf) da Universidade de Brasília (UnB) como requisito parcial para obtenção do título de Mestra em Ciência da Informação.

Orientador: Prof. Dr. Fernando Leite
Coorientadora: Prof^a. Dr^a. Michelli Costa

Brasília
Janeiro de 2022



FOLHA DE APROVAÇÃO

Título: “Práticas de comunicação de dados de pesquisa: a percepção de bolsistas de produtividade brasileiros do campo da genética”

Autora: Priscila Rodrigues dos Santos

Área de concentração: Gestão da Informação

Linha de pesquisa: Comunicação e Mediação da Informação

Dissertação submetida à Comissão Examinadora designada pelo Colegiado do Programa de Pós-graduação em Ciência da Informação da Faculdade em Ciência da Informação da Universidade de Brasília como requisito parcial para obtenção do título de **MESTRE** em Ciência da Informação.

Dissertação aprovada em: 17 de dezembro 2021.

Presidente (UnB/PPGCINF): Dr. Fernando César Lima Leite

Membro Externo (IBICT-UFRJ): Dra. Luana Farias Sales

Membro Interno (UnB/PPGCINF): Dr. João de Melo Maricato

Suplente (UnB/PPGCINF): Dra. Fernanda Passini Moreno

Em 10/11/2021.



Documento assinado eletronicamente por **Fernando César Lima Leite, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 21/12/2021, às 16:21, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Joao de Melo Maricato, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 22/12/2021, às 11:37, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **LUANA FARIAS SALES MARQUES, Usuário Externo**, em 07/01/2022, às 21:29, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



A autenticidade deste documento pode ser conferida no site http://sei.unb.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **7372167** e o código CRC **E80186D9**.

Ficha catalográfica elaborada por Priscila Rodrigues dos Santos CRB1/3381

S237 Santos, Priscila Rodrigues dos

Práticas de comunicação de dados de pesquisa: a percepção de bolsistas de produtividade brasileiros do campo da genética / Priscila Rodrigues dos Santos; orientador Fernando César Lima Leite; coorientadora Michelli Pereira da Costa. – Brasília, 2022.

158p.

Dissertação (Mestrado - Mestrado em Ciência da Informação) --
Universidade de Brasília, 2022.

1. Comunicação científica. 2. Ciência aberta. Ciclo de vida dos dados.
3. Dados científicos. 4. Gestão de dados de pesquisa. I. Leite, Fernando César
Lima, orient. II. Costa, Michelli Pereira da, coorient. III. Título.

AGRADECIMENTOS

Primeiramente agradeço a Deus e aos meus mentores por terem me dado discernimento e acalmado meu coração em momentos de dificuldades, e por me manter firme nessa trajetória rumo a algo que tanto desejei.

Ao meu namorado, João Marcos, por me acompanhar nessa caminhada, por ser um ótimo ouvinte e me acalantar sempre que eu precisei. Sem sombra de dúvidas essa conquista é dele também.

Aos meus pais que dentro da sua simplicidade, mesmo não entendendo muito bem as loucuras da Pós-Graduação, se mantiveram firmes em me ajudar dentro do possível. A minha não tão pequena Pietra, que é o meu motivo há 9 anos para sempre evoluir e ser o melhor exemplo possível a ela.

Ao meu orientador, Prof^o Fernando Leite, pela oportunidade de adentrar esse mundo da pesquisa.

A minha coorientadora, Prof^a Michelli Costa, por cada palavra amiga e compreensão, por sempre devolver com afeto cada aflição que eu levava para as nossas reuniões.

Expresso minha gratidão a Prof^a Luana Sales pelas excelentes contribuições para o trabalho e a abertura para o diálogo.

Ao Prof^o João Maricato, pelas valorosas contribuições e reflexões oferecidas ao enriquecimento deste trabalho.

A minha xuzinha, Jan Barcelos, por ao longo desses vários anos ter sido uma ótima amiga e ainda mais nesses anos de mestrado, por sempre me ouvir e dividir comigo sua trajetória na academia.

Ao Ton e Jaque, por me acompanharem desde a graduação e me mostrarem por muitas vezes a importância dessa etapa em minha vida.

Ao Fernando Pereira, por quase que diariamente conversar comigo sobre a minha dissertação e sempre acreditar na minha capacidade.

Ao melhor coaching desmotivacional do mundo, Fred, por me fazer rir e reclamar junto comigo das mazelas da vida.

Aos meus colegas do Ibict que de alguma forma contribuíram para essa pesquisa, seja com conversas, incentivos ou reflexões.

RESUMO

O presente estudo investiga as atividades de comunicação dos dados de pesquisa sob a percepção dos pesquisadores bolsistas de produtividade brasileiros do campo da Genética. O estudo teve como objetivo geral analisar as práticas de comunicação dos dados de pesquisa entre os pesquisadores brasileiros da Genética. A fundamentação teórica foi traçada por debates à luz da comunicação científica, com foco na comunicação dos dados de pesquisa e seus macro processos, produção, comunicação e reuso. Metodologicamente, o estudo utiliza como concepção filosófica o construtivismo social, em que se considera prioritariamente a perspectiva dos participantes sobre o assunto e os diferentes contextos vivenciados por eles. A abordagem adotada na pesquisa foi fundamentalmente qualitativa, com característica descritiva, levantada através de entrevista e amostragem intencional não probabilísticas. A amostra selecionada foi levantada a partir da aplicação da técnica de amostragem intencional não probabilística e estratificada. Assim, a amostra é composta por 12 pesquisadores bolsistas de produtividade cadastrados na Plataforma Carlos Chagas. Os dados foram coletados através de entrevistas semiestruturadas que ocorreram por meio da plataforma Microsoft Teams que permite realizar gravações. Para a codificação, categorização e análise dos dados coletados foi utilizado o software NVivo (edição Starter) para auxiliar nestas atividades. Entre os resultados, foi possível perceber que as variadas disciplinas que compõem o campo da Genética, as práticas de produção/obtenção, compartilhamento e reuso de dados são feitas de diferentes maneiras, não sendo práticas uniformes. Também foi possível levantar as problemáticas enfrentadas pelos pesquisadores, como, a falta de apoio técnico e institucional, dados sem qualidade e a alta competitividade relacionada a tais práticas.

Palavras-chave: Comunicação científica. Ciência aberta. Ciclo de vida dos dados. Dados científicos. Gestão de dados de pesquisa.

ABSTRACT

This study investigates the communication activities of research data under the perception of Brazilian researchers with productivity grants of Genetics. The objective was to analyze the communication practices of research data among Brazilian researchers in Genetics. The literature review was carried out in the light of scientific communication, focusing on the communication of research data and its macro processes (production, communication and reuse). As a methodology, the study uses social constructivism as a philosophical concept, considering primarily the participants' perspective on the subject and the different contexts they experience. The approach adopted in the research was fundamentally qualitative, descriptive, raised through interviews and non-probabilistic intentional sampling. The sample was selected from the application of the non-probabilistic and stratified intentional sampling technique. Thus, the sample was composed of 12 productivity scholarship researchers registered in the Carlos Chagas Platform. Data were collected through semi-structured interviews using Microsoft Teams. NVivo (Starter Edition) was used for the coding, categorization and analysis of the data collected. Among the results, it can be seen that in the various disciplines belonging to the Genetics, the practices of production, sharing and reuse of data are done in different ways, that is, not uniformly. In addition, there were observed problems faced by the researchers, such as the lack of technical and institutional support, poor quality data and the high competitiveness.

Keywords: Scientific communication. Open science. Data life cycle. Scientific data. Research data management.

LISTA DE FIGURAS

Figura 1 - Produção brasileira em nível mundial.	21
Figura 2 - Comparativo da produção científica brasileira com a América Latina e o mundo	22
Figura 3 - Gastos com pesquisa científica no Brasil	24
Figura 4 - Ações em pesquisa e inovação para o enfrentamento a Covid-19.....	27
Figura 5 - Áreas e linhas beneficiadas com os editais para o enfrentamento da Covid-19.	27
Figura 6 - Modelo Unisist.	36
Figura 7 - Taxonomia da Ciência Aberta.	40
Figura 8 - Ciclo de vida dos dados de pesquisa.	47
Figura 9 - Ciclo de vida dos dados.	49
Figura 10 - Pirâmide da gestão de dados.....	51
Figura 11 - Princípios FAIR.	59
Figura 12 - Desdobramentos do reuso dos dados de pesquisa.....	61
Figura 13 - Estrutura fechada para amostra.	77
Figura 14 - Categorização e codificação dos dados.	83
Figura 15 - Análise de dados na pesquisa qualitativa.....	84

LISTA DE TABELAS

Tabela 1 - Distribuição de bolsas de pesquisa.....	24
Tabela 2 - Pesquisadores da Genética contemplados em editais universais.....	25
Tabela 3 - Cinco escolas de pensamento de ciências abertas.....	42
Tabela 4 - Etapas da curadoria de dados de pesquisa.....	47
Tabela 5 - Natureza dos dados de pesquisa.....	54
Tabela 6 - Tipos de dados de pesquisa.....	56
Tabela 7 - Licenças para o reuso de dados de pesquisa.....	63
Tabela 8 - Princípios para citação de dados.....	64
Tabela 9 - Quantidade de pesquisadores por estado e local de atuação.....	76
Tabela 10 - Descrição dos pesquisadores.....	79
Tabela 11 - Procedimentos de coleta de dados.....	83
Tabela 12 - Relação entre os objetivos e procedimentos metodológicos adotados.....	85
Tabela 13 - Características dos dados.....	91
Tabela 14 - Realização da Pesquisa Científica na Genética.....	128
Tabela 15 - Produção de dados de pesquisa na Genética.....	129
Tabela 16 - Aspectos que influenciam a produção de dados de pesquisa.....	130
Tabela 17 - Compartilhamento de dados de pesquisa.....	131
Tabela 18 - Aspectos que estimulam e inibem o compartilhamento de dados de pesquisa.....	132
Tabela 19 - Exigência do compartilhamento dos dados de pesquisa.....	133
Tabela 20 - Experiência e percepção sobre o reuso dos dados de pesquisa.....	134
Tabela 21 - Fatores que favorecem e inibem o reuso dos dados de pesquisa.....	135

LISTA DE GRÁFICOS

Gráfico 1 - Recursos destinados ao MCTI	23
------------------------------------------------------	----

LISTA DE SIGLAS

AHDS	Arts and Humanities Data Service
ANDS	Australian National Data Service
BOIA	Budapest Open Access Initiative
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CGEE	Centro de Gestão e Estudos Estratégicos
CNPQ	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CODATA	Committee on Data for Science and Technology
COICT	Coordenação de Indicadores de Ciência, Tecnologia e Inovação
DCI	<i>Data Citation Index</i>
DFG	German Research Foundation
ESDS	Economic and Social Data Service
EC	European Commission
FAIR	Findable, Accessible, Interoperable e Reusable
FAP	Fundação de Amparo à Pesquisa
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FIOCRUZ	Fundação Oswaldo Cruz
FNDCT	Fundo Nacional de Desenvolvimento Científico e Tecnológico
FOSTER	Facilitate Open Science Training For European Research
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
ICB	Instituto de Ciências Biológicas
ICSU	International Council of Science
IPEA	Instituto de Pesquisa Econômica Aplicada
MCTI	Ministério da Ciência, Tecnologia e Inovações
ME	Ministério da Educação
MOOC	Massive Open on-line Course
NAS	National Academy of Science
NHMRC	National Health and Medical Research Council
NIH	National Institutes of Health

NSF	National Science Foundation
OAD	Open Access Directory
OECD	Organisation for Economic Co-Operation and Development
P&D	Pesquisa e Desenvolvimento
PERX	Pilot Engineering Repository Xsearch
PGH	Projeto Genoma Humano
PIB	Produto Interno Bruto
TI	Tecnologia da Informação
TICS	Tecnologias de Informação e Comunicação
UFMG	Universidade Federal de Minas Gerais
UNB	Universidade de Brasília
UNESCO	Organização das Nações Unidas para a Educação, a Ciência e a Cultura
WDC	Word Data Center

SUMÁRIO

1 INTRODUÇÃO	16
1.1 Contexto da Pesquisa	20
1.2 Objetivo geral	28
1.3 Objetivos específicos	28
2 JUSTIFICATIVA	29
3 REVISÃO DE LITERATURA	32
3.1 Comunicação científica	32
3.2 Ciência aberta	38
3.3 Dados de pesquisa	45
3.3.1 <i>Ciclo de vida dos dados de pesquisa</i>	46
3.3.2 <i>Gestão dos dados de pesquisa</i>	50
3.3.3 <i>Comunicação de dados</i>	52
3.4 Produção de dados de pesquisa	54
3.5 Compartilhamento de dados de pesquisa	57
3.6 Reuso de dados de pesquisa	60
3.6.1 <i>Características do uso/reuso dos dados de pesquisa</i>	60
3.6.2 <i>Licenças abertas</i>	62
3.6.3 <i>Crítérios para citação de dados</i>	64
3.6.4 <i>Estudos relacionados à abertura dos dados de pesquisa</i>	65
4 METODOLOGIA	73
4.1 Caracterização da pesquisa	73
4.2 Procedimentos metodológicos	75
4.2.1 <i>Amostragem da pesquisa</i>	76
4.2.2 <i>Métodos e técnicas para coleta e análise dos dados</i>	81
5 ANÁLISE E DISCUSSÃO DOS RESULTADOS	86
5.1 Levantamento	86
5.1.1 <i>Produção/obtenção de dados de pesquisa</i>	86
5.1.2 <i>Compartilhamento de dados de pesquisa</i>	97
5.1.3 <i>Reuso dos dados de pesquisa</i>	117
6 SÍNTESE DOS RESULTADOS DA PESQUISA	128
6.1 Produção/obtenção de dados de pesquisa	128
6.2 Compartilhamento de dados de pesquisa	130
6.3 Reuso dos dados de pesquisa	133
7 CONCLUSÃO	135

7.1 Identificar as práticas de produção de dados de pesquisa no campo da Genética.....	135
7.2 Identificar as práticas de compartilhamento de dados de pesquisa no campo da Genética	136
7.3 Identificar as práticas de reuso de dados de pesquisa no campo da Genética.....	137
7.4 Limitações do estudo.....	139
7.5 Sugestões para futuras pesquisa	139
REFERÊNCIA	140
ANEXO 1.....	157

1 INTRODUÇÃO

Nos últimos anos, um movimento vem ganhando cada vez mais visibilidade entre a comunidade científica nacional e internacional, a esse movimento é dado o nome de Ciência Aberta. A Ciência Aberta abrange diferentes dimensões, tais como acesso aberto, dados abertos, ciência cidadã, revisão por pares abertos, código aberto, citações abertas, caderno aberto de laboratório, recursos educacionais abertos e Redes sociais científicas. Entre tais dimensões, encontram-se a de dados de pesquisa, que constituem insumos para a produção do conhecimento por meio da pesquisa científica em diversos campos.

Com a expansão cada vez maior da atividade científica, os dados produzidos por pesquisadores vêm aumentando consideravelmente nos últimos tempos graças ao crescimento da ciência, do acesso aberto e da criação de ferramentas para a visualização desses documentos (NAVARRO; CONEGLIAN; SANTARÉM, 2018).

Em razão do fato de serem produzidos e utilizados em diferentes áreas, o entendimento e o conceito de dados de pesquisa variam consideravelmente em função da disciplina ou área em que são empregados. Nesse sentido, Borgman (2012) aponta as formas e técnicas de obtenção dos dados de pesquisa conforme as áreas do conhecimento e os formatos de apresentação que os dados podem adquirir.

As fontes de dados também variam amplamente. Nas ciências físicas e da vida, a maioria dos dados são coletados ou produzidos por pesquisadores, como observações, experimentos ou modelos. Nas ciências sociais, pesquisadores podem coletar ou produzir seus próprios dados ou pode obter dados de outras fontes, como registros públicos de atividade econômica [...] os dados de humanidades geralmente são extraídos de registros de cultura humana, sejam materiais de arquivo, documentos publicados ou artefatos (BORGMAN, 2012, p. 1061, tradução nossa).

Nesse mesmo sentido, Martínez-Uribe e MacDonald (2008, p. 274, tradução nossa) apresentam outras formas de dados em variados campos de pesquisa:

[...] dados nas ciências naturais são medições de precipitação pluviométrica, observações bases de dados de modelos genéticos astronômicos ou estruturas cristalográficas. Nas ciências sociais, os dados são gerados através de pesquisas de opinião ou Mapas com informações de censo georreferenciadas. Nas humanidades, eles podem incluir fotografias de antigas escrevendo em pedra e em medicina neuroimagem que eles capturam a atividade cerebral.

No campo da Tecnologia da Informação (TI), por exemplo, que tem uma grande produção de dados em várias de suas vertentes, alguns desses usos de dados ocorrem em “[...] banco de dados, arquivo de dados, mineração de dados, dados descritivos, metadados, dados vinculados e agora big data” (HJORLAND, 2018, online). Uma outra forma de apresentação dos dados de pesquisa é por meio de formatos numéricos, textos, *softwares*, áudios, equações, simulações, vídeos, entre outros (SAYÃO; SALES, 2015). Sayão e Sales (2015) consideram que esses dados podem adquirir valor duradouro e imediato como também ao longo do tempo.

Ainda de acordo com Sayão e Sales (2014), o entendimento sobre o que são dados de pesquisa varia conforme a percepção do pesquisador e especialmente entre as áreas de conhecimento. Corroborando com os autores, Borgman (2012) declara que os dados de pesquisa possuem características inerentes que os ligam às disciplinas que os produzem.

Outro aspecto interessante sobre os dados é o seu compartilhamento. Milham *et al.* (2017) apontam que a abertura de dados é cada vez mais aconselhada a ser feita pelos pesquisadores, pois essa ação acelera a ciência, facilita a colaboração, a transparência e a reprodutibilidade. Além desses benefícios que o compartilhamento de dados traz para a comunidade acadêmica, outra vantagem é apontada por Veiga (2017, p. 47):

O compartilhamento de dados de pesquisa possibilita ainda o conhecimento de dados negativos. Na maioria das vezes, esses dados não são publicados, não geram um artigo tradicional, mas são fundamentais para o avanço da ciência, pois constituem aprendizado, explicitam equívocos que não precisam ser repetidos e, desta forma, proporcionam economia de recursos.

Questões relacionadas ao compartilhamento de dados de pesquisa são encontradas em estudos e modelos de comunicação científica como de Garvey e Griffith (1979), desenvolvidos dentro da Ciência da Informação, onde são investigadas as práticas de produção, compartilhamento e reuso dessas informações produzidas pelos pesquisadores de diferentes disciplinas. Na última década tem se criado modelos de comunicação de dados de pesquisa e práticas que permitem um compartilhamento mais rápido, seguro e transparente da informação, além dos benefícios dessa atividade.

Pesquisadores também são diretamente beneficiados com o compartilhamento de seus dados. Isso potencializa a visibilidade e seu reconhecimento, pois uma vez que os dados são depositados em um repositório, ele recebe um identificador persistente e uma sugestão de citação, dessa forma o pesquisador que os coletou continua tendo reconhecimento pelo seu trabalho em publicações futuras (VEIGA, 2017).

Todavia, apenas o compartilhamento dos dados não é suficiente para garantir que ele seja usado em outras pesquisas. Por esta razão, é importante considerar aspectos relacionados com o uso e reuso dos dados de pesquisa, juntamente com a gestão desses dados, conforme menciona Borgman (2015). A autora afirma que os benefícios que o compartilhamento dos dados pode oferecer, só serão efetivos se esses dados estiverem bem descritos e recuperáveis em um banco de dados e/ou repositório, possibilitando o reuso por outros pesquisadores em novos estudos.

Segundo Medeiros e Caregnato (2012, p. 319) “o reuso de dados científicos pode ser visto, a grosso modo, como a utilização de dados já utilizados, ou seja, são dados utilizados mais de uma vez, mas com intenções originais”. Por outro lado, o reuso de dados traz dificuldades, como por exemplo, identificar em quais condições foram produzidos (FANIEL; JACOBSEN, 2010). Awre *et al.* (2015) trazem outras dificuldades enfrentadas para o reuso dos dados, como a falta de padronização na linguagem e métodos para descrição, a própria diversidade e complexibilidade da pesquisa, a ausência de investimentos, a representação da pesquisa nas instituições, entre outros.

Para que os dados sejam compartilhados e reutilizados, é importante que eles passem por uma curadoria. HEY, TANSLEY e TOLLE (2009, p. xvii) afirma que após os “dados serem capturados, necessita-se que sejam curados antes que se inicie qualquer tipo de análise”. Dessa forma, como ressaltam Medeiros e Caregnato (2012), a curadoria dos dados é um procedimento de gestão de dados científicos que permite agregação de valor aos dados de pesquisa, contribuindo para o seu reuso.

Reforçando o entendimento sobre curadoria, Sayão e Sales (2016, p. 70) afirmam que a gestão de dados em grande escala implica:

[...] adicionar valor aos dados. Nessa direção, os dados precisam ser avaliados, analisados, enriquecidos com anotações, ligados por

hiperlinks com outros recursos, comentados, agregados, “regerados” e descartados. Este papel, que exige conhecimento profundo da área, é atribuído ao pesquisador e a outros especialistas da área.

Entretanto, apesar da importância dessas atividades para o avanço da ciência, procedimentos como o compartilhamento e o reuso dos dados de pesquisa não são muito comuns entre os pesquisadores. Como exposto por Curty (2017), a prática de reuso de dados está longe de ser consolidada no ambiente científico. A atividade de aproveitamento de dados é mais comum nas ciências biológicas, exatas, astronomia e da saúde (TENOPIR *et al.*, 2015; BORGMAN *et al.*, 2016; SIELEMANN; HAFNER; PUCKER, 2020; SAYÃO; SALES, 2014), em especial no campo da Genética por ser uma área ativa na atividade de compartilhamento de dados e que conta com um dos maiores bancos de dados referência na área, o GenBank.

Ainda assim, nem mesmo dentro das áreas que têm certo costume nas práticas de comunicação de dados, os pesquisadores se sentem confortáveis ou têm o conhecimento necessário para realizarem o compartilhamento e o reuso dos dados de pesquisa de forma mais efetiva. É o caso da Genética, mesmo com iniciativas tais como o Programa Genoma Humano (PGH) e a política de comunicação dos dados de pesquisa do National Institutes of Health (NIH), é observado na literatura, receios por parte dos pesquisadores sobre o compartilhamento e reuso de dados gerados por outros pesquisadores, principalmente por não possuírem habilidades e experiências na gestão de dados e em realizar um compartilhamento eficaz desse material (CAMPBELL *et al.*, 2002; CAMPBELL *et al.*, 2003; TENOPIR *et al.*, 2011; BARONE *et al.* 2017; FEDERER *et al.* 2015; TENOPIR *et al.* 2016).

No estudo realizado por Campbell (2003) com geneticistas e outros pesquisadores das Ciências da Vida, o autor observou que a retenção de dados pelos pesquisadores de Genética é mais comum do que se pensava. Além dessa descoberta, o autor supracitado identificou alguns motivos para a retenção desses dados, como a quantidade de esforço delicada a produção dos dados, proteção dos seus pares e/ou até mesmo relacionada a sua capacidade de publicar. Há casos, que pesquisadores da área da Genética não fornecem os dados de pesquisa de forma nenhuma, mesmo que o solicitante seja colaborador da pesquisa em questão (KURATA; MATSUBAYASHI; MINE, 2017).

A idade do pesquisador também influencia diretamente a prática do compartilhamento dos dados, autores como Zuiderwijk e Spiers (2019), Jarolímková, Drobíková e Souček (2018) e Costa (2017) tratam sobre o tema, onde os pesquisadores mais jovens se mostram mais favoráveis ao compartilhamento dos seus dados do que pesquisadores com a idade mais avançada. Drobíková e Souček (2018) discorrem que os pesquisadores entre os 36 anos a 45 anos, cerca de 39% dos respondentes do seu estudo são os que mais compartilham. Como explicação para o compartilhamento ser realizado mais facilmente nesta faixa etária, é o fato de pesquisadores terem mais familiaridade e facilidade em utilizarem os sistemas com essa finalidade. Outro ponto a ser considerado é que mais de 50% das descobertas não são compartilhadas (FERGUSON *et al.*, 2014).

Como discorre Sayão e Sales (2019), o compartilhamento de dados costuma acontecer apenas entre os pares mais próximos ou nos sites da própria instituição. Conforme ainda os autores supracitados, os dados da cauda longa, onde é produzido diversos conjuntos de dados, os dados de pesquisa frequentemente estão espalhados em projetos, laboratórios, grupos de pesquisa, instituições e pesquisadores individuais e são bastante heterogêneos.

Dada a relevância do tema e a importância das práticas de produção, compartilhamento e reuso dos dados de pesquisa, o presente trabalho se propõe a responder à seguinte pergunta: **Quais são as práticas de comunicação dos dados de pesquisa entre os pesquisadores da Genética?**

1.1 Contexto da Pesquisa

A produção científica tem um papel importante no avanço da ciência, e é por meio dela também que se avalia a qualidade e a quantidade de conhecimento produzido em uma determinada disciplina. Além de “[...] permitir o entendimento dos processos de produção, difusão e uso do conhecimento, também pode orientar o desenvolvimento e a adaptação de políticas científicas, tecnológicas e de inovação” (SANTIN; VANZ; STUMPF, 2015, online).

De acordo com o *Ranking* Iberoamericano de instituições de Educação Superior publicado em 2021 pelo SCImago Institutions Ranking, até o ano de 2019 o Brasil era o único país latino-americano a investir mais de 1% do seu Produto Interno

Bruto (PIB) em Pesquisa e Desenvolvimento. A produção científica brasileira é gerada principalmente pelas faculdades e universidades (91% da produção), desse total, cerca de 141 instituições disponibilizaram pelo menos a metade das suas publicações em acesso aberto (SCIMAGO INSTITUTIONS RANKING, 2021)

A produção científica brasileira tem crescido no cenário global nos últimos seis anos, como aponta o estudo feito pelo Centro de Gestão e Estudos Estratégicos (CGEE) em junho de 2021. Segundo os dados apresentados pelo relatório, o Brasil entre os anos de 2015 a 2020 apresentou um crescimento de 32,2% da sua produção científica enquanto o resto do mundo foi de 27,1% (Figura 1). O Brasil responde por 3,2% da produção científica mundial, cerca de 372 mil publicações (CGEE, 2021).

Figura 1 - Produção brasileira em nível mundial.



Fonte: CGEE (2021, p. 22)

Em levantamento feito pela Coordenação de Indicadores de Ciência, Tecnologia e Inovação (COICT) do Ministério da Ciência, Tecnologia e Inovações (MCTI) com periódicos científicos indexados na *Scopus*, o número de artigos brasileiros cresceu em todas as áreas do conhecimento nos últimos vinte anos. As áreas de Bioquímica, Genética e Biologia molecular em particular, apresentaram um aumento considerável em sua produção, representando mais de 50% da produção na América Latina nos últimos 14 anos (Figura 2).

Figura 2 - Comparativo da produção científica brasileira com a América Latina e o mundo.

Área do Conhecimento	Ano	Brasil	América Latina	Mundo	% do Brasil em relação à América Latina	% do Brasil em relação ao Mundo
Bioquímica, genética e biologia molecular	2000	2.079	4.987	176.741	41,7%	1,2%
	2001	2.151	4.981	177.758	43,2%	1,2%
	2002	2.564	5.734	181.742	44,7%	1,4%
	2003	2.856	6.172	189.487	46,3%	1,5%
	2004	3.074	6.606	197.326	46,5%	1,6%
	2005	3.247	6.823	207.379	47,6%	1,6%
	2006	3.881	7.778	216.968	49,9%	1,8%
	2007	4.303	8.373	230.230	51,4%	1,9%
	2008	4.665	8.976	237.005	52,0%	2,0%
	2009	4.719	9.200	249.681	51,3%	1,9%
	2010	5.083	9.631	250.653	52,8%	2,0%
	2011	5.365	10.232	263.317	52,4%	2,0%
	2012	5.901	11.254	276.099	52,4%	2,1%
	2013	6.326	11.830	286.956	53,5%	2,2%
	2014	6.549	12.323	299.351	53,1%	2,2%
	2015	7.026	12.909	304.298	54,4%	2,3%
	2016	7.361	13.913	306.723	52,9%	2,4%
	2017	7.840	14.557	304.543	53,9%	2,6%
	2018	8.294	15.476	327.972	53,6%	2,5%
2019	8.717	16.480	339.005	52,9%	2,6%	
2020	10.087	18.927	384.022	53,3%	2,6%	

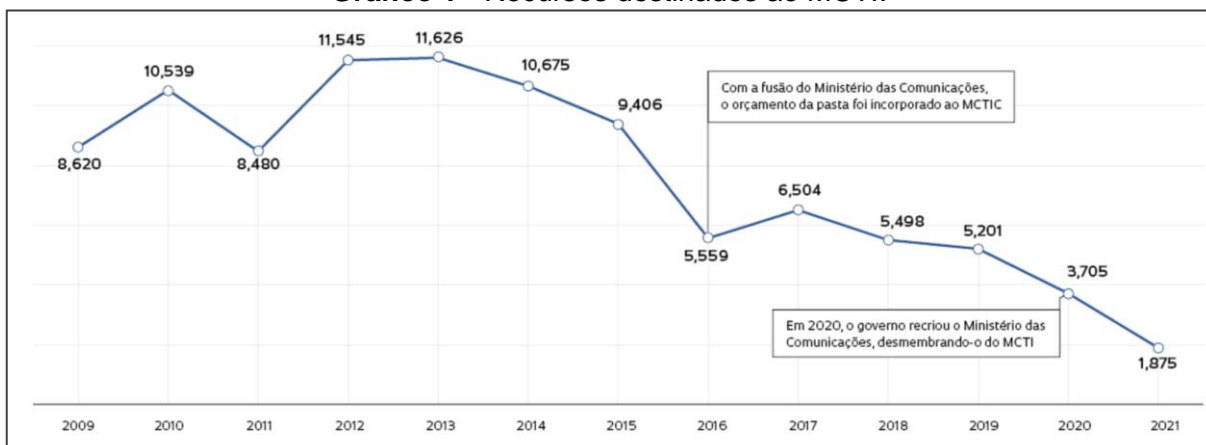
Fonte: COICT adaptado do SJR SCImago Journal & Country Rank (2021)

Essa representatividade crescente da pesquisa científica brasileira em nível mundial ocorre principalmente pelos investimentos feitos pelos governos com concessão de

[...] bolsas, concessão de auxílios à pesquisa e à infraestrutura, subvenção econômica, empréstimos, renda variável, compra do Estado com margem de preferência local, encomenda tecnológica, incentivos fiscais, bônus tecnológico, títulos financeiros, cláusula de PD&I de Agências Reguladoras. (FRANCISCO; ZUCATTO, 2018, p. 44)

A disponibilização de bolsas de pesquisa pelas agências de fomento como o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) e as Fundações de Amparo à Pesquisa (FAPs) são a principal forma de apoio aos pesquisadores.

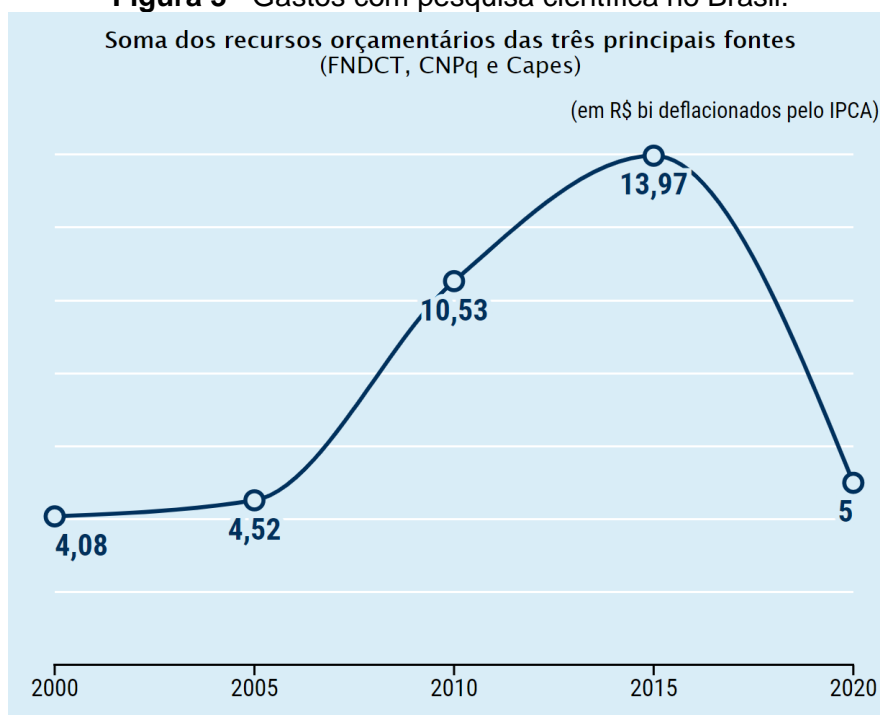
Entretanto, apesar desse crescimento da produção científica brasileira, ao decorrer dos anos a verba destinada a esse tipo de auxílio vem diminuindo gradativamente. No período de 2012 a 2021 o MCTI teve uma redução de 84% de seu orçamento para a pesquisa, passando de R\$ 11,5 bilhões para somente 1,8 bilhão de reais (JORNAL USP, 2021, online). No Gráfico 1 é possível observar a redução de recursos que o MCTI vem sofrendo nos últimos nove anos.

Gráfico 1 - Recursos destinados ao MCTI.

Fonte: Fapesp (2021)

Esses cortes financeiros na ciência brasileira ficaram mais evidentes a partir de 2016, como aponta Mazza (2019, online), os eixos mais afetados foram o de Pesquisa e Desenvolvimento (P&D) “[...] com a redução de recursos para o pagamento de bolsas da Capes e do CNPq e o enorme bloqueio de recursos do Fundo Nacional de Desenvolvimento Científico e Tecnológico (FNDCT), que chega a mais de 80% em 2019”. Mesmo em 2021 os valores alocados para o P&D serem de R\$ 50,666 milhões, quase 75 % a mais do que 2020, esse valor depende de créditos suplementares (SOCIEDADE BRASILEIRA PARA O PROGRESSO DA CIÊNCIA, 2021, online).

No Brasil, as duas agências de fomento que são as principais fontes de recursos dos pesquisadores brasileiros são ligadas ao MCTI, sendo elas o CNPq e o Fundo Nacional de Desenvolvimento Científico e Tecnológico (FNDCT). Outra importante fonte de financiamento de pesquisa é a Capes, ligada ao Ministério da Educação (ME). Somando os recursos destinados à pesquisa dessas três agências, em 2015 se teve o maior somatório de recursos, sendo de R\$ 13,97 bilhões, no último levantamento feito referente ao ano de 2020, essa soma cai para R\$ 5 bilhões (Figura 3).

Figura 3 - Gastos com pesquisa científica no Brasil.

Fonte: Agência Senado adaptado de SBPC (2020), Unesco (2020)

Em consequência desses cortes, a quantidade de pesquisadores que são contemplados com as bolsas que são disponibilizadas por nível (mestrado, doutorado, pós-doutorado, iniciação científica, entre outros) vem diminuindo gradativamente. A Tabela 1 traz a distribuição de bolsas por nível da Capes de 2010 a 2020.

Tabela 1 - Distribuição de bolsas de pesquisa.

Ano	Mestrado	Mestrado profissional	Doutorado	Pós Doutorado	Iniciação científica	Total
2010	33.357	-	21.941	2.734	-	58.032
2011	41.054	1.215	26.108	3.580	-	71.957
2012	43.591	2.914	27.589	3.663	-	77.757
2013	45.754	2.674	32.111	6.217	-	86.756
2014	48.113	-	39.954	6.879	9.968	104.914
2015	49.353	-	42.779	7.486	4.891	104.509
2016	47.830	-	43.188	6.999	1.654	99.671
2017	47.510	-	44.316	7.168	1.748	100.742
2018	47.478	-	44.530	7.075	1.498	100.581
2019	44.232	-	43.327	6.237	910	94.706

2020	46.105	-	43.497	5.281	-	94.883
-------------	--------	---	--------	-------	---	--------

Fonte: Adaptado do Geocapes (2021)

Em consequência da diminuição de recursos, Capes e CNPq vem tornando público ao passar do tempo menos editais para financiar projetos de pesquisa. Polo e Tavares (2019) analisaram as chamadas públicas do Edital Universal do CNPq referente aos anos de 2010, 2011, 2012, 2013, 2014 e 2016, o ano de 2015 não foi contemplado por falta de recursos da Instituição. É possível inferir no estudo dos pesquisadores que a quantidade de trabalhos contemplados sofreu uma redução.

Ao analisar os Editais Universais dos últimos 10 anos (2011-2020), o CNPq publicou apenas 6 editais, contemplando os anos de 2011, 2012, 2013, 2014, 2016, e 2018. Os editais beneficiam projetos em três faixas de valores, de até R\$ 30 mil, de R\$ 30 a R\$ 60 mil e R\$ 60 mil a R\$ 120 mil. Ao analisar apenas a área da Genética, é possível observar que ocorreu uma redução na quantidade de pesquisadores contemplados com os recursos (Tabela 2).

Tabela 2 - Pesquisadores da Genética contemplados em editais universais.

Ano	Até 30 mil reais	30-60 mil reais	60-120 mil reais	Total de beneficiados
2011	52	21	8	81
2012	50	25	18	93
2013	51	25	21	97
2014	21	21	18	60
2016	33	27	22	81
2018	40	20	15	75

Fonte: Elaboração própria (2021)

Os anos que tiveram a maior quantidade de pesquisadores favorecidos na faixa de R\$ 30 mil foram nos anos de 2011, 2012 e 2013. Já na faixa de R\$ 30 a R\$ 60 mil foram os anos de 2012, 2012 e 2017 com respectivamente 25, 25 e 27 pesquisadores contemplados. Os anos de 2012, 2013, 2014 e 2016 tem a maior quantidade de pesquisadores beneficiados com valores na faixa de R\$ 60 mil a R\$ 120 mil. É possível inferir que mesmo em algumas faixas (R\$ 30 mil, de R\$ 30 a R\$ 60 mil e R\$ 60 mil a R\$ 120 mil) dos anos de 2016 e 2018 mostrando uma maior

quantidade de pesquisadores contemplados pelos recursos do CNPq, no somatório geral, a quantidade de pesquisadores é inferior à anos como 2011, 2012 e 2013.

Explorando ainda a questão do fomento à pesquisa, os autores Sidone, Haddad, Mena-Chalco (2016), Melo, Santana, Silva (2018), Souza *et al.* (2020) e MCTI (2011) abordam sobre o acúmulo dos recursos em algumas regiões brasileiras, como é o caso das regiões Sul e Sudeste. Conforme os autores supracitados, as regiões Sul e Sudeste são responsáveis por uma produção de conhecimento científico-tecnológico superior, o que acarreta em uma maior capacidade de captação de recursos das agências de fomento.

Além da escassez de recursos, existem outras problemáticas enfrentadas pelos pesquisadores para desenvolver suas pesquisas no Brasil, como aponta Souza *et al.* (2020) em seu estudo, os pesquisadores veem como desafios a ser transposto:

- Burocracia nos processos de planejamento, execução e avaliação
- Falta de uma equipe de apoio para auxiliar o pesquisador
- Acúmulo de funções do pesquisador (ensino, pesquisa, extensão etc.)
- Infraestrutura de pesquisa deficitária
- Valorização excessiva de artigos em detrimento a outros indicadores
- Baixa interação entre pesquisadores de áreas diferentes
- Baixa interação com empresas e/ou demandas sociais
- Baixa divulgação e apropriação do conhecimento produzido
- Baixa relevância das pesquisas em termos científicos e/ou tecnológicos
- Baixo comprometimento de pesquisadores e/ou alunos de pós-graduação

Um outro agravante é que desde o início de 2020 o mundo está passando por uma pandemia de coronavírus e conseqüentemente os editais para pesquisa estão ficando cada vez mais restritos. Como demonstra o Relatório de atividades da Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp) (2020), os valores destinados ao fomento foram menores que os recebidos em 2019, mesmo em meio a pandemia. Dos recursos recebidos a Fapesp destinou “R\$ 978,3 milhões ao fomento de 21.233 projetos de pesquisa – 45% na área de Ciências da Vida, 37% na de Ciências Exatas e da Terra e Engenharias, 10% interdisciplinares e 7,5% na área de Ciências Humanas” (AGÊNCIA FAPESP, 2021, online).

Em Nota Técnica elaborada pelo Instituto de Pesquisa Econômica Aplicada (Ipea) em 2020 mostra que os recursos destinados à pesquisa e inovação frente à crise da Covid-19 pelo Brasil (US\$ 104,0) é bem inferiores a outros países como Estado Unidos (US\$ 6.172,15), Alemanha (US\$ 2.385,00), Canada (US\$ 995,9), Reino Unido (US\$ 1.661-1.725). A nota traz ainda dados sobre as principais ações em pesquisa e inovação realizadas pelos países para enfrentar a crise causada pela Covid-19 (Figura 4).

Figura 4 - Ações em pesquisa e inovação para o enfrentamento a Covid-19.

AÇÃO ADOTADA		ESTADOS UNIDOS	REINO UNIDO	CANADÁ	ALEMANHA	BRASIL
Novos recursos para P&D e inovação	Em moeda local (bilhões)	6,1	1,31-1,35	1,3	2,1	0,47
	Em US\$ (bilhões)	6,1	1,66-1,72	0,97	2,34	0,10
	Orçamento federal em P&D (%) ¹	4,1	10,8	11,8	6,3	1,8
Grupo científico de assessoramento ao governo		Não	Sim	Sim	Sim	Não
Mecanismo de aprovação rápida de projetos de pesquisa e inovação relacionados à Covid-19		Sim	Sim	Não	Não	Não

Fonte: Ipea (2020, p. 18)

Quanto à quantidade de editais publicados para o enfrentamento da Covid-19, até abril de 2020, foram publicados 23 editais de P&D, as principais áreas temáticas beneficiadas pelos editais foram 8, com foco em 14 linhas (Figura 5).

Figura 5 - Áreas e linhas beneficiadas com os editais para o enfrentamento da Covid-19.

Grandes áreas temáticas	Linhas temáticas	N	%*
Insumos para enfrentamento da COVID-19	Testes diagnósticos	14	60,9
	Equipamentos e dispositivos médicos	9	39,1
	Medicamentos, vacinas e produtos imunobiológicos	11	47,8
	Outros insumos e novos materiais	3	13,0
Prevenção e segurança sanitária	Prevenção e redução do contágio	12	52,2
Tecnologia da informação e comunicação (TIC)	Aplicação das TIC	16	69,6
	Telemedicina	12	52,2
Sistemas de saúde	Infraestrutura	3	13,0
	Políticas, sistemas de redes e serviços de saúde	10	43,5
Conhecimento epidemiológico da doença	Desenvolvimento de cenários e modelagens	6	26,1
	Epidemiologia	11	47,8
Caracterização da doença e sua história natural	Virologia, imunidade e inflamação, modelos experimentais, aspectos clínicos	8	34,8
Impactos socioeconômicos	Ciências sociais, pesquisas econômicas e de comportamento	7	30,4
Outras áreas de P&D/S com aplicação na área da saúde	Fomentos que não se encaixem nas categorias acima	12	52,2

Fonte: Silva *et al.* (2020)

A pandemia também revela outra faceta da ciência, com os esforços direcionados ao combate a Covid-19, outras áreas de pesquisa ficam desassistidas, é o caso, por exemplo, do Instituto de Ciências Biológicas (ICB) da Universidade

Federal de Minas Gerais (UFMG), que perdeu recursos e ficou com 70% da estrutura parada (OLIVEIRA, 2020).

Isso pode ser sentido também em nível mundial, como aborda os autores Subramanya, Lama e Acharya (2020) e Radecki e Schonfeld (2020), muitas atividades de pesquisa tradicionais foram suspensas e houve também o redirecionamento do financiamento atual para o COVID-19. Como discorre Subramanya, Lama e Acharya (2020), muitos financiadores em todo o mundo foram solicitados a tomar providências para lidar com uma redução nos orçamentos devido ao redirecionamento de financiamento.

É perceptível que pesquisas em áreas que não tem ligação com a Covid-19 foram prejudicadas no mundo inteiro. No Brasil, além da escassez frequente de editais para o desenvolvimento de pesquisa, os pesquisadores enfrentam dificuldades também na hora de concorrer aos editais disponibilizados devido a série de critérios a serem atendidos, pois os principais editais publicados entre 2019 e 2020 foram destinados à pesquisa do Covid-19 e pesquisadores que não tinham essa expertise ficaram impossibilitados de concorrer aos recursos.

Diante das problemáticas aqui discutidas, pretende-se a partir da análise dos dados com base nas experiências dos pesquisadores sobre comunicação de dados de pesquisa na Genética, oferecer contribuições nas práticas de produção, compartilhamento e reuso dos dados de pesquisa para o desenvolvimento de uma ciência efetiva independente dos obstáculos gerados pela ausência de recursos ou devida a situação peculiar vivida atualmente, causada pela Covid-19. Para que essas questões sejam respondidas, o objetivo geral e os objetivos específicos desta pesquisa são:

1.2 Objetivo geral

Este trabalho tem como objetivo geral analisar as práticas de comunicação dos dados de pesquisa entre os pesquisadores brasileiros da Genética.

1.3 Objetivos específicos

1. Identificar as práticas de produção de dados de pesquisa no campo da Genética;

2. Identificar as práticas de compartilhamento de dados de pesquisa no campo da Genética;
3. Identificar as práticas de reuso de dados de pesquisa no campo da Genética.

2 JUSTIFICATIVA

A ciência é construída de forma colaborativa, visto que “[...] nenhuma ciência brota do cérebro de fundadores. Ela resulta da investigação realizada por muitos estudiosos, sobre problemas que aparecem em determinadas condições históricas” (VIEIRA, 1996, p. 15). A Ciência Aberta é a visão moderna desse pensamento, em que se prioriza o colaborativismo e o compartilhamento de informações para uma evolução mais ágil do conhecimento científico. Trata-se de uma abordagem constituída de inúmeras dimensões entre as quais os dados de pesquisa, que são tidos como parte essencial dos resultados de investigação.

A ideia de se compartilhar dados de pesquisa não é nova. Como levantado por Ruttenberg e Rishbeth (1994), mais recentemente, na década de 1950 já existiam registros de ações direcionadas à comunicação dos dados. Em 1980, o livro intitulado *Sharing Research Data* foi publicado e nele os assuntos de curadoria dos dados de pesquisa e seu compartilhamento foram discutidos (FIENBERG; MARTIN; STRAF, 1985). O compartilhamento dos dados frutos de pesquisas dispôs de maior visibilidade nos anos de 1990 devido ao Projeto Genoma Humano (PGH), cujo propósito era o sequenciamento do genoma humano, e para que isso ocorresse, pesquisadores do mundo inteiro foram convidados a compartilharem os seus dados. Ao longo dos anos, com os avanços da *web*, pesquisadores compartilham seus dados por diferentes motivos, seja por exigências da instituição às quais estão ligados, seja por exigência das agências de fomento que financiam suas pesquisas ou simplesmente por motivações pessoais.

Além dos pontos abordados acima, a internet se mostra um ambiente fértil para a comunicação dos dados de pesquisa. No que tange ao compartilhamento dos dados, para que esse insumo tenha um maior alcance e conseqüentemente sejam reutilizados por outros pesquisadores, existem fatores que podem interferir neste processo e devem ser levados em conta, como salientam Torino, Trevisan e Vidotti (2019, p. 41), a utilização dos dados:

[...] está diretamente atrelada a variáveis, como a descoberta, a compreensão, a veracidade e a autenticidade, entre outros. Assim, é necessário estabelecer um conjunto de requisitos padronizados para que fornecedores e consumidores de dados, que podem alternar seus papéis em determinados momentos, possam localizar conjuntos de dados, acessá-los, compreendê-los, considerá-los confiáveis, processá-los, torná-los interoperáveis e ainda conectá-los com outros dados ampliando o contexto, exercendo todas as possibilidades de reuso.

Ao analisar a literatura científica que trata sobre dados de pesquisa e o efetivo acesso a esse tipo de recurso, verifica-se que a documentação dos dados “[...] tornou-se importante condição para reforçar a investigação científica aberta” (SOUSA; DIAS; SOUSA, 2020, p. 122). Os dados de pesquisa abrangem todas as áreas do conhecimento que envolvem pesquisa, tornando-se insumos importantes, pois são “[...] registros factuais [...] usados como fonte primária para a pesquisa científica e que são comumente aceitos pelos pesquisadores como necessários para validar os resultados do trabalho científico” (ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD), 2007, p. 13, tradução nossa).

A produção de dados é feita a partir dos registros coletados, que podem vir de telescópios que produzem dados a partir de observações do céu, instrumentos meteorológicos que realizam leituras do clima de vários locais ou até mesmo do rastreamento das variações Genéticas entre animais e plantas. Esses dados são gerados em grandes quantidades, podendo ser medido em trilhões de bytes (MARCUM; GEORGE, 2010).

Com esse elevado volume de dados sendo produzidos constantemente, os dados de pesquisa, para serem compartilhados e posteriormente utilizados em novas investigações, precisam estar envolvidos nas boas práticas de gestão de dados, as quais são processos que compreendem a geração, organização, armazenamento e descrição dos dados de pesquisa, possibilitando assim a sua preservação digital e a sua confiabilidade.

Marcum e George (2010) salientam que a forma de se fazer ciência vem se modificando, e diferentes produtos de pesquisa vêm ganhando importância. Ainda conforme os autores, cada vez mais o valor científico não é colocado nas publicações resultantes de um projeto de pesquisa, mas sim na geração de dados, processos esses que ocorrem mais cedo no ciclo de vida da pesquisa. A disponibilidade dos

dados de pesquisa na internet, seja por meio de bancos de dados, repositórios ou periódicos científicos tradicionais, demonstra que as áreas da ciência vêm se transformando em meio a *web* e suas ferramentas, e o acesso a esse tipo de matéria fornece *insights* e avanços em pesquisas de forma mais ágil, favorecendo disciplinas que até então não tinham fácil acesso.

Para os campos informacionais, como a Ciência da Informação, Sant'Ana (2019) aponta que a gestão dos dados apresenta uma oportunidade, visto que devem

[...] atender agora às especificidades que o acesso a dados acrescenta e pode responder a questões relacionadas às dimensões *input* e *output*, por exemplo, e, ainda, refletir sobre possíveis requisitos, barreiras, atores, relações e assimetrias resultantes do processo. O cenário dos campos informacionais também pode contribuir para a percepção da relação entre os diversos atores envolvidos no universo das pesquisas científicas (SANT'ANA, 2019, p. 28).

Embora o compartilhamento dos dados traga benefícios para a comunidade científica, a sua prática não é tão comum entre os pesquisadores, como aponta Pontika *et al.* (2015), mesmo que exista a aceitação do compartilhamento de dados em alguns campos do conhecimento, como é o caso da Genética, fatores dificultam a comunicação desses dados, tais como os diferentes métodos de produção e compartilhamento que podem mudar de um pesquisador para outro, mesmo que ambos pertençam ao mesmo campo de estudo. Concordando com o entendimento de Pontika *et al.* (2015) sobre as dificuldades enfrentadas no compartilhamento de dados pelos investigadores, Borgman (2012), ela expressa ainda que os pesquisadores adotem ações para o compartilhamento dos seus dados, porém, existe a falta de entendimento em relação a como esse compartilhamento deve ser feito e sob quais condições os dados devem ser analisados e interpretados.

Outra questão é a falta de literatura sobre a prática de uso e reuso dos dados entre os pesquisadores. Segundo Dias, Anjos e Araújo (2019, p. 6), “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”. Posto isso, o uso e reuso dos dados se mostra um campo produtivo para discussões, apresentando assim possíveis lacunas no processo de comunicação desses dados de pesquisa.

Com tudo isso exposto, o presente estudo poderá contribuir na compreensão de como esse fenômeno de produção, compartilhamento e reuso dos dados de pesquisa provocados pelo movimento da *web* e suas ferramentas e incentivada pela Ciência Aberta, se dá por meio das perspectivas dos pesquisadores, sendo esses os atores essenciais para o desenvolvimento da ciência. Além da contribuição teórica desta pesquisa, também poderá colaborar nos debates que envolvem os temas em questão, fazendo-se um campo fértil para estudos futuros.

3 REVISÃO DE LITERATURA

A revisão de literatura é uma parte importante no desenvolvimento de um trabalho científico (MATTOS, 2018; MARTINS, 2018). Na revisão de literatura, busca-se apresentar aos leitores estudos realizados anteriormente por outros pesquisadores com temas semelhantes e como eles dialogam com a pesquisa que está em andamento (CRESWELL, 2010).

Buscando alcançar os objetivos descritos neste trabalho, esta seção irá abordar como temas principais a Comunicação científica, Ciência aberta, Dados de pesquisa, Produção, Compartilhamento e Uso/reuso dos dados.

3.1 Comunicação científica

No decorrer dos séculos, a comunicação científica passou por diferentes momentos, o primeiro deles teve início na Academia Grega, posteriormente passando pelas cartas e relatórios científicos utilizados por pesquisadores como meio popular de comunicação entre seus pares, entretanto o acesso a essas informações era demorado, devido aos meios utilizados para sua disseminação na época. (ANDRADE, 2015).

O termo comunicação científica propriamente dito teve origem na década de 1940 e é atribuído ao físico e historiador John Bernal. A expressão comunicação científica apareceu pela primeira vez no livro *The social function of Science* em que Bernal trazia as atividades de produção da informação, disseminação e seu uso, indo desde o desenvolvimento da ideia até a apresentação dos resultados aos pares (CARIBÉ, 2015; ROSA; GOMES, 2010).

Garvey e Griffith (1979) definem a comunicação científica como um conjunto de tarefas ligadas à produção, disseminação e uso da informação. A geração da informação ocorre na fase da pesquisa e a sua disseminação por meio de transferência de informação por canais de comunicação, que podem ser formais, ou informais.

Para Bueno (2010) a comunicação científica tem como princípio básico a disseminação especializada da informação entre os pares, com a intenção de tornar o conhecimento familiar na comunidade científica, tais como os avanços alcançados, elaboração de novas hipóteses ou aperfeiçoamento das já existentes, reforçando o que é a comunicação científica. Logo, a comunicação científica cumpre função crucial para a ciência, mobilizando pesquisadores para o debate como parte do processo de criação e legitimação do conhecimento científico.

Em uma definição contemporânea, Monteiro e Lucas (2019, p. 14) definem a comunicação científica “[...] pelo intercâmbio contínuo de informações entre os membros de uma comunidade científica, integrando a estrutura social que esta comunidade se constitui”, dessa forma, é uma parte primordial para a ciência, visto que possibilita que os pesquisadores individuais unam forças para avançar em pesquisas, favorecendo a visibilidade do produto (produção científica) e os produtores (pesquisadores) e assim contribuir com a sua credibilidade nas comunidades científicas à qual pertencem (TARGINO, 1998).

A formalização da comunicação científica conforme exposto por Rosa e Barros (2018) foi estabelecida em 1640 sob o olhar da *Royal Society de Londres*, com o objetivo de apresentar artigos, anunciar resultados de pesquisas e divulgar experiências. Essas trocas de correspondências permitiam que aqueles que não pudessem participar presencialmente das reuniões conseguissem ficar a par do que foi discutido e se envolver nos debates (SCHAUDER, 1994; MEADOWS, 1999).

De acordo com Rosa e Gomes (2010), uma vez que as trocas de correspondências não supriam mais as necessidades dos pesquisadores, surgiram os periódicos científicos como caminho mais eficaz de disseminar a informação em maior escala. Os primeiros periódicos científicos surgiram na metade do século XVII, sendo eles o *Journal des Sçavans em Paris* e a *Philosophical Transactions of the Royal Society of London* em Londres.

Meadows (1999) destaca que a comunicação científica por meio de periódicos surgiu por várias razões, algumas

[...] eram específicas (como a expectativa de seus editores de que teriam lucro); algumas, gerais (como a crença de que para fazer novos descobrimentos era preciso que houvesse um debate coletivo). O motivo principal, contudo, encontra-se nessa necessidade de comunicação. Do modo mais eficiente possível, com uma clientela crescente interessada em novas realizações. (MEADOWS, 1999, p. 7)

Menzel, Somers e Glaser (1958) fez uma síntese das atribuições da comunicação científica, dando ênfase ao seu papel para o desenvolvimento da ciência, são elas:

1. Fornecer respostas a perguntas específicas;
2. Contribuir para a atualização profissional do cientista no campo específico de sua atuação;
3. Estimular a descoberta e a compreensão de novos campos de interesse;
4. Divulgar as tendências de áreas emergentes, fornecendo aos cientistas ideia da relevância de seu trabalho;
5. Testar a confiabilidade de novos conhecimentos, diante da possibilidade de testemunhas e verificações;
6. Redirecionar ou ampliar o rol de interesse dos cientistas;
7. Fornecer feedback para aperfeiçoamento da produção do cientista.

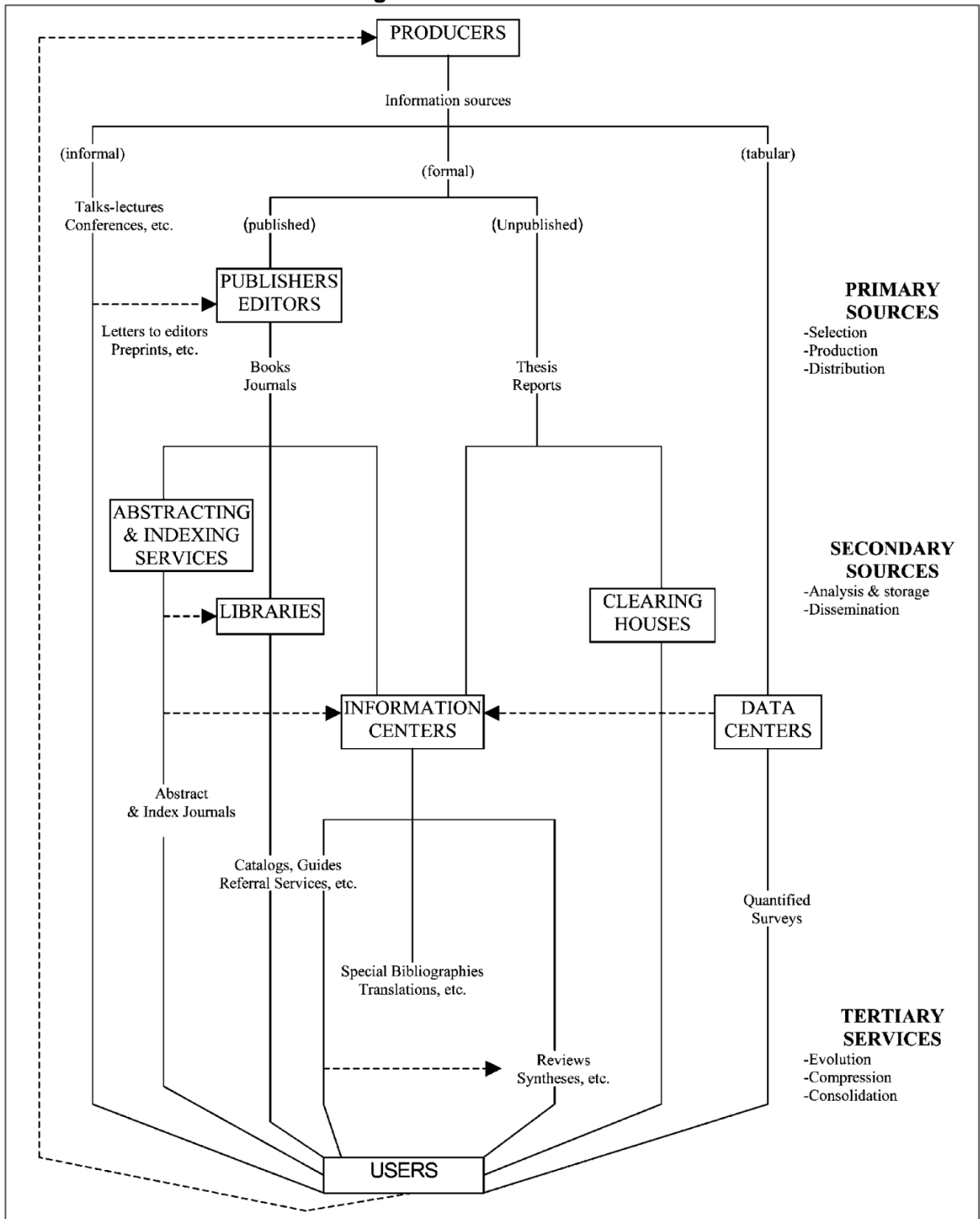
Com a revolução de Gutenberg, um novo cenário surgiu para a ciência, segundo Andrade (2015, p. 268), “levando ao aumento da circulação do conhecimento científico e seu reconhecimento social”. Nas últimas décadas do século XX, as técnicas e meios de comunicação da informação no contexto científico passaram por mais uma evolução, feito esse provocado pelas Tecnologias de Informação e Comunicação (TICs) que acarretou em um aumento expressivo da produção da literatura científica. Le Coadic (2004, p. 84) versa sobre as TICs, apresentando os objetos dessa tecnologia “a concepção de produtos, sistemas e serviços que permitem a construção, comunicação, armazenamento e uso da informação”.

Entre as comunidades científicas, esses novos meios de transmissão do conhecimento foram bem vistos, já que aceleraram a disseminação e a reutilização desses conteúdos em novos estudos entre os pares, otimizando assim os avanços científicos. Como abordado por Drexler (1991), as comunidades científicas

desenvolveram tradições e instituições que fomentam a replicação, transformação, e seleção efetiva de ideias, desenvolvendo o conhecimento com eficiência.

A comunicação científica envolve vários elementos em seu percurso entre produtores e usuários. Diferentes autores criaram modelos de comunicação científica para explicar como essas interações ocorrem (GARVEY; GRIFFITH, 1979; HURD, 2000; PINTO, COSTA, 2018). Um famoso modelo de comunicação científica chamado UNISIST (Figura 6), desenvolvido pela Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco) em parceria com o Conselho Internacional de Uniões Científicas (ICSU), foi publicado em 1971 e traz como fontes de informação os canais formais, canais informais e canais tabulares. A apresentação de canais tabulares ainda na década de 1970 salienta que os dados de pesquisa já se mostravam como insumo importante na comunicação científica.

Figura 6 - Modelo Unisist.



Fonte: Sondergaard et al. (2003)

Com a publicação do modelo UNISIST, o fluxo da comunicação científica vem se alterando em razão dos canais de comunicação apoiados pela *internet*. Do ponto de vista do uso da *internet*, Leite (2011) apresenta os canais de comunicação, sendo os canais formais divididos em documentos publicados e não publicados; os canais

informais em trocas de informações, correspondências, indicação de leitura entre outros; e os canais tabulares, comunicam dados técnicos científicos em formato de tabelas. Diferente dos dois canais anteriores, os canais tabulares não seguem um fluxo linear de texto falado ou escrito (LEITE, 2011).

Os canais formais são os mais tradicionais, pois é por meio deles que os pesquisadores recebem o aval dos seus pares, legitimando assim as suas pesquisas. Contudo, as recentes tecnologias oferecem novas possibilidades, segundo Costa (2017). Além das possibilidades proporcionadas pelas TICs, existe um movimento de Acesso Aberto e Ciência Aberta ao conhecimento científico. O Acesso Aberto foi lançado através do Projeto Gutenberg em 4 de julho de 1971, em que os livros eram digitalizados e distribuídos gratuitamente. A partir do Projeto Gutenberg, as discussões sobre a disponibilização do acesso ao conteúdo científico foram sendo discutidas no cenário mundial.

Em 2016, o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), publicou um documento chamado Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã, que trazia algumas recomendações gerais relacionadas ao compartilhamento, reprodução, avaliação, reutilização dos dados de pesquisa destinados às universidades e institutos de pesquisa, às sociedades científicas, às agências de fomento à pesquisa e aos editores de periódicos científicos.

O acesso aberto à ciência, pode ser realizado de diferentes formas, mas inicialmente o acesso era feito de duas maneiras, através da Via dourada e a Via verde. Alves (2008, p. 133) descreve as duas vias iniciais como:

1. Via verde (green road), que trata do arquivamento que poderá ser realizado pelos próprios autores de artigos científicos já publicados ou aceitos para publicação, obtendo autorização (sinal verde) dos editores que os aceitaram para que possam disponibilizar em um servidor de arquivo aberto.
2. Via dourada (golden road), que abrange os periódicos científicos eletrônicos cujo acesso aberto a seus conteúdos é garantido pelos próprios editores. Sendo assim, a publicação em ambiente de acesso aberto está assegurada no próprio periódico.

Já a Ciência Aberta é um modelo de prática científica em consonância com o desenvolvimento da cultura digital. É um movimento que vem ganhando cada vez mais seguidores fora e dentro da academia, propondo uma nova forma de fazer ciência, de maneira colaborativa, partilhada e pública (BIAZON; MARIN, 2016). Conforme

Carvalho (2018, p. 48), no cenário da Ciência Aberta “[...] a pesquisa científica produz conhecimentos que são agregados à própria ciência e constituem insumos para a geração de novas pesquisas”. Desse modo, o movimento Ciência Aberta produz no campo científico diferentes discussões em volta dos pilares que a compõe.

3.2 Ciência aberta

A Ciência Aberta é um movimento para tornar a pesquisa científica acessível à comunidade. A abertura do conhecimento promovida por ela oferece não apenas aumento da produção científica e inovação, mas como levantado por Albagli (2015), existe o retorno social, a promoção da circulação imediata da informação e as novas formas da produção colaborativa.

Um dos pressupostos da Ciência Aberta é exatamente a disponibilização do conhecimento científico para que quaisquer pessoas, independente da sua formação, “possam usá-lo, reusá-lo e distribuí-lo sem constrangimentos tecnológicos, econômicos, sociais ou legais” (SAYÃO; SALES, 2014, p. 77).

Por ser um processo discutido entre diferentes grupos (editoras, pesquisadores, instituições de pesquisa e governo), Albagli (2015) aponta duas vertentes possíveis de se olhar para a Ciência aberta. Por um lado, a socialização do conhecimento e por outro a sua privatização. De acordo com a autora, esse conflito é gerado devido à expansão iniciada na década de 1980 de mecanismos de apropriação intelectual, os altos custos dos periódicos e as políticas restritivas de acesso e uso da informação. Em contrapartida, o movimento de acesso livre a publicações científicas facilitadas pelo uso de sistemas eletrônicos ganhava força aumentando os embates sobre a democratização do conhecimento.

Apesar de o movimento existir há algumas décadas, as mudanças ocorrem lentamente. Delfanti e Pitrelli (2015) versam sobre a evolução científica ocorrida até então através da Ciência Aberta. Conforme os autores, “[...] a ciência aberta não é necessária, mas, sim, uma entre muitas evoluções possíveis dependentes de vários fatores que incluem (e vão além da) evolução e a adoção tecnológica, e até mesmo mudança cultural” (DELFANTI; PITRELLI, 2015, p. 62). Dessa forma, a Ciência Aberta deve ser vista como um processo condicional que envolve diferentes atores em

diferentes contextos sociais e institucionais com valores, motivações, objetivos diversos, tendo como propósito a colaboração em rede (SHRUM *et al.*, 2007).

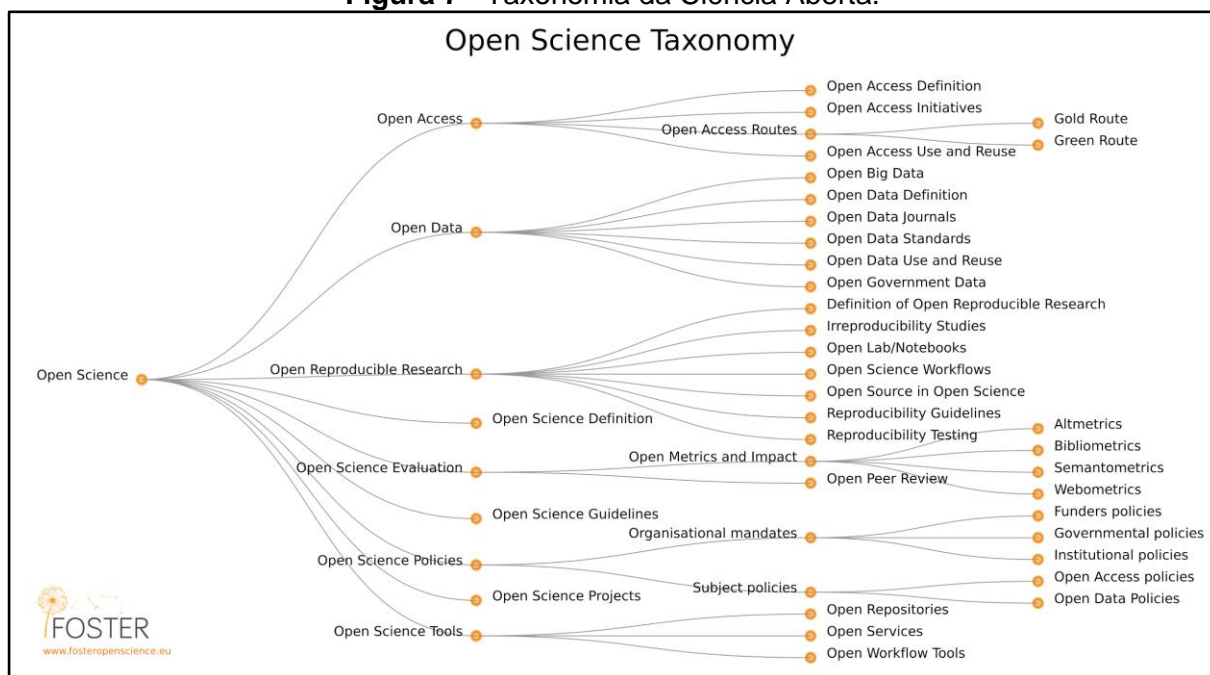
Para Bartling e Friesike (2014) a Ciência Aberta se desenvolve por meio da *web* e suas ferramentas, não sendo limitada aos grupos tradicionais de pesquisa, contando com a colaboração de cientistas autônomos e o uso de *hardwares* e *softwares* de código aberto. Essa iniciativa provoca mudanças no fazer ciência que se conhecia até então. A *European Commission* (EC) (2019, on-line, tradução nossa): descreve a Ciência Aberta como um movimento que

[...] trata da maneira como a pesquisa é realizada, disseminada, implantada e transformada por ferramentas e redes digitais. Ele se baseia nos efeitos combinados do desenvolvimento tecnológico e da mudança cultural no sentido de colaboração e abertura na pesquisa. Ao fornecer acesso aberto, ilimitado e sem barreiras aos resultados da pesquisa, a ciência aberta torna os processos científicos mais eficientes, transparentes e responsivos aos desafios da sociedade. Oferece novas ferramentas para colaboração científica, experimentos e análises e torna o conhecimento científico mais facilmente acessível.

Conforme avança, a Ciência Aberta agrega novos elementos à sua agenda de ações. Albagli, Clinio e Raychtock (2014) relatam que a Ciência Aberta é um termo guarda-chuva que envolve não só o livre acesso a publicações científicas, mas cobre outros princípios, como dados de pesquisa, ferramentas científicas abertas, *hardware* científico aberto, citações abertas, cadernos científicos abertos, e wikipesquisa, ciência cidadã, educação aberta.

O projeto *Facilitate Open Science Training For European Research* (FOSTER) aponta seis categorias para se refletir a ciência aberta (Figura 7), como o: *Open Access*, *Open Data*, *Open Reproducible Research*, *Open Science Evaluation*, *Open Science Policies*, *Open Science Tools*.

Figura 7 - Taxonomia da Ciência Aberta.



Fonte: Foster (2015)

Costa (2017) discute as seis categorias sinalizadas pelo projeto Foster:

1. *Open Access*: trata do Movimento de Acesso Aberto e as Vias verde e dourada, tendo em vista que as publicações científicas devem estar em acesso aberto, por meio de periódicos científicos ou repositórios.
2. *Open Data*: trata dos dados de pesquisa em suas diferentes categorias, indo de dados de pesquisa até dados governamentais. Os dados devem fazer parte da publicação, citada no *Open Access*.
3. *Open Reproducible Research*: Trata sobre a reprodutividade da pesquisa.
4. *Open Science Evaluation*: trata das formas de se avaliar a ciência, levando em consideração métricas abertas e o impacto amplo. São formas de avaliação: bibliometria, semantometria, webmotria e métricas alternativas.
5. *Open Science Políticas*: Trata das políticas para ciência aberta.
6. *Open Science Tools*: Trata das ferramentas usadas para a pesquisa, que devem ser preferencialmente desenvolvidas em *software* e *hardwares* abertos.

Sob esse ponto de vista, Fecher e Friesike (2014) apontam que o termo “ciência aberta” evoca diferentes entendimentos, abrindo uma infinidade de cenários para discussão, que vão do direito democrático de acesso ao conhecimento científico financiado (acesso aberto a publicações), passando pela inclusão da sociedade na

pesquisa (ciência cidadã) até desenvolvimento de ferramentas gratuitas para colaboração (mídias sociais para cientistas).

Para Gezelter (2009) a Ciência Aberta vai muito além da abertura de códigos, dados e Bloco de Notas Aberto. Esse movimento abarca quatro objetivos fundamentais para a ciência: a transparência na metodologia (observação e coleta dos dados); disponibilidade e reutilização dos dados científicos; acessibilidade e transparência da comunicação científica e utilização de ferramentas baseadas na *web* para viabilizar a colaboração científica.

Nessa mesma linha, Lopez (2015) qualificou os seguintes elementos como os cinco princípios da Ciência Aberta: metodologia aberta (documentos relevantes, processos, métodos, etc); *software* e *hardware* livres (códigos abertos); dados abertos (disponível para reutilização); acesso aberto (livre e gratuito) e recursos educacionais abertos (*Massive Open on-line Course* (MOOC)). Tanto o que foi apresentado por Gezelter (2009) quanto por Lopez (2015), mostram que toda a estrutura da Ciência Aberta precisa estar pautada por uma metodologia aberta, em que são aplicados processos relevantes cujas ferramentas tecnológicas empregadas sejam construídas em códigos fonte abertos, os dados coletados nas pesquisas estejam disponíveis de forma gratuita e que sejam capazes de serem reutilizados, bem como as publicações científicas possam estar acessíveis na *web* com o mínimo de restrição possível de acesso.

Costa (2017, p. 36) enumera seis categorias de como se pensar a Ciência Aberta:

1. Acesso aberto às publicações científicas – disponibilização do artigo científico na Internet, sem custos e restrições para o acesso. O acesso aberto pode ser proporcionado pelo próprio periódico ou por uma base de dados interoperável no contexto dos arquivos abertos, ou seja, repositórios digitais.
2. Dados de pesquisa abertos – os dados coletados e utilizados pela pesquisa devem ser publicados como meio de validar a pesquisa comunicada, permitir a refutação do argumento proposto e promover o desenvolvimento do tema estudado. Sua publicação deve ser feita em uma base de dados específica para a área do conhecimento ou segundo as orientações dos editores científicos. O conjunto de dados deverá receber um identificador persistente, que será relacionado ao artigo do periódico.
3. Metodologias abertas – a pesquisa deve ser desenvolvida a partir de métodos e processos consistentes. Isso permitirá que ela possa ser reproduzida ou analisada corretamente.

4. Avaliação aberta - tanto o artigo do periódico quanto os outros produtos da pesquisa deverão considerar avaliações alternativas, além das tradicionais. Nas novas formas de avaliação, destacam-se as métricas alternativas e a webmetria.
5. Ferramentas abertas – a pesquisa deverá apoiar-se em tecnologias abertas, tanto softwares como hardwares livres. Além das cinco categorias comuns às duas propostas, considerou-se que a sexta categoria sugerida por FOSTER (2015) é relevante para se pensar a ciência aberta.
6. Políticas abertas – são os documentos das instituições ou agências fomentadoras de pesquisa que promovem as iniciativas de acesso aberto nas pesquisas que financiam.

De acordo com Costa (2017), essas seis categorias não são critérios para se avaliar a produção científica conforme os princípios da Ciência Aberta, mas devem ser vistas como condições elementares para se analisar as práticas de pesquisa realizadas e os preceitos fundamentais discutidos para possibilitar o acesso aos resultados de pesquisa por todos.

A Ciência Aberta possui esses preceitos de como fazer ciência a partir da *web* e suas ferramentas, apresenta ainda uma divisão em escolas, de como se pensar a Ciência Aberta. Fecher e Friesike (2014) indicam 5 escolas teóricas de ciências abertas, sobre o pensar científico e as discussões envolvidas, sendo elas: democrática, pragmática, infraestrutura, pública e medição. Na Tabela 3 é apresentada a teoria central, grupos envolvidos, objetivo central e ferramentas e métodos de cada uma das escolas.

Tabela 3 - Cinco escolas de pensamento de ciências abertas.

Escola de Pensamento	Teoria Central	Grupos Envolvidos	Objetivo Central	Ferramentas e Métodos
Democrática	O acesso ao conhecimento é desigualmente distribuído	Cientistas, políticos, e cidadão	Tornar o conhecimento livremente disponível para todos	Acesso aberto, propriedade intelectual direitos, dados abertos, códigos abertos
Pragmática	A criação de conhecimento poderia ser mais eficiente se os cientistas trabalhassem juntos.	Cientistas	Abrir o processo de criação de conhecimento.	Sabedoria das multidões, rede efeitos, dados abertos, código aberto
Infraestrutura	Uma pesquisa	Cientistas e	Criar plataformas	Plataformas e

	eficiente depende da ferramentas e aplicativos disponíveis.	plataforma fornecedores	abertas, ferramentas e serviços para cientistas.	ferramentas de colaboração
Pública	A ciência precisa estar acessível para o público.	Cientistas e cidadãos	Tornar a ciência acessível a cidadãos.	Cidadão, ciência, relações públicas da ciência, blogs científicos
Medição	As contribuições científicas hoje precisam medições de impacto alternativas.	Cientistas e políticos	Desenvolver um sistema alternativo de métrica para impacto científico.	Altmetrics, revisão por pares, citação, fatores de impacto

Fonte: Adaptado de Fecher e Friesike (2014)

Mesmo que a Ciência Aberta seja vista como um conceito que engloba várias frentes, o que acaba a caracterizando como um movimento amplo, como pode ser observado neste levantamento feito por Fecher e Friesike (2014), a Ciência Aberta evoca entendimentos distintos, sendo abordada por diferentes aspectos em cada escola, apontando a complexibilidade que envolve o termo Ciência Aberta.

No contexto da Ciência Aberta, os dados de pesquisa ganham um papel fundamental no desenvolvimento da ciência, vide o Programa Horizon 2020, que foi o maior programa da Comissão Europeia em pesquisa e inovação. No Horizon 2020, um dos critérios para o financiamento de pesquisa era que os resultados dos estudos deveriam ter acesso aberto, juntamente aos dados que acompanham esses resultados.

Com o fim do Horizon 2020, o programa Europe Horizon, está dando continuidade às ações relacionadas à Ciência Aberta como um dos seus ideais e princípios. Duas das nove vias de impacto principais, são focadas na Ciência Aberta, sendo elas: “a via «Promover a difusão de conhecimentos e a ciência aberta», orientada para o impacto científico, e a via «Reforçar a adoção da I&I na sociedade», orientada para o impacto na sociedade” (HORIZON EUROPE, 2021, online

Como apontado por Pontika *et al.* (2015), outros feitos que demonstram a importância dos dados de pesquisa nesse movimento de Ciência Aberta é a quantidade de ações líderes de promoção aos dados abertos. Duas dessas iniciativas

começaram nas décadas de 1950 e 1960 e se mantêm ativas até os dias de hoje. Em 1957 foi fundado o *Word Data Center* (WDC) durante o Ano Geofísico Internacional, para o arquivamento e distribuição dos dados coletados dos programas de observação do *International Council of Science*¹ (ICSU). Quase uma década mais tarde, em 1966, foi fundado o *Committee on Data for Science and Technology* (CODATA) como um comitê interdisciplinar do ICSU, conhecido anteriormente como *International Council of Scientific Unions*, onde busca melhorar a avaliação, seleção, armazenamento e recuperação dos dados de importância à ciência e tecnologia.

Mais recentemente, outras duas iniciativas foram criadas para ampliar o compartilhamento e visibilidade dos dados. A primeira foi o *Re3data*², financiado pela *German Research Foundation* (DFG), e começou a funcionar em 2012. A segunda foi o *Databib*³, gerido pela *Purdue University*. Ambas eram respectivamente plataformas de registro de repositórios de dados de pesquisa que abrangiam diferentes disciplinas acadêmicas. Em 2013, *Re3data* e *Databib* anunciaram a fusão dos dois diretórios a fim de reduzir a duplicação de esforços e oferecer aos pesquisadores um registro único de repositórios de dados de pesquisa. A *Simmons University* desenvolveu um diretório sobre acesso aberto, o *Open Access Directory*⁴ (OAD), e nele são disponibilizadas listas de referências atualizadas sobre acesso aberto, onde os usuários podem contribuir para a construção das listas.

A abertura dos dados de pesquisa destina-se a promover condições melhores para avaliação, reuso, reprodutibilidade de métodos e colaborativismo na produção e disseminação do conhecimento científico. Conforme Costa (2020, p. 28) os debates em torno

[...] da abertura dos dados de pesquisa têm se mostrado como um marco relevante da ampliação e do amadurecimento dos projetos de Ciência Aberta. Sua relevância é estabelecida a partir da sua articulação em uma rede complexa de elementos e atores, especialmente as instituições produtoras de conhecimento científico, as agências de fomento e os editores.

Entretanto, de acordo com Pontika *et al.* (2015), outros fatores estão por trás do compartilhamento dos dados de pesquisa, fatores esses que podem ser

¹ Em julho de 2018, o ICSU uniu-se ao *International Social Science Council* (ISSC) para formar o *International Science Council* (ISC).

² Disponível em: <http://re3data.org/>

³ Disponível em: <http://databib.org/>

⁴ Disponível em: http://oad.simmons.edu/oadwiki/Main_Page

operacionais (gestão de dados de pesquisa), legais e éticos. Torino, Roa-Martínez e Vidotti (2020) levantam outra questão na dificuldade do compartilhamento dos dados de pesquisa, a falta de conhecimento técnico e especializado nos processos de curadoria dos dados pelos pesquisadores. Com todas as problemáticas existentes em meio à necessidade de disponibilizar os dados de pesquisa e a falta de clareza por parte dos pesquisadores de como descrevê-los e onde fazer o seu depósito para que os dados se tornem acessíveis a terceiros, criam-se barreiras na comunicação dos dados e conseqüentemente o seu uso por outros pesquisadores. Visto que para compreender esses obstáculos enfrentados pelos pesquisadores, é necessário entender aspectos relacionados à comunicação dos dados de pesquisa.

3.3 Dados de pesquisa

Os dados de pesquisa vêm ganhando notória visibilidade no movimento pela Ciência Aberta e questões relacionadas a sua disponibilização para se realizar uma Ciência com transparência e integridade são discutidas globalmente.

Mesmo que os dados de pesquisa seja um insumo importante para o desenvolvimento da pesquisa, sua definição se mostra complexa e essa diversidade vem sendo construída

Pelas especificidades de cada disciplina, suas condicionantes metodológicas, protocolos e seus objetivos, e se torna um desafio, mesmo para o pesquisador, pelo alto grau de contextualização necessário, definir precisamente o que é dado de pesquisa de uma forma transversal aos diversos domínios disciplinares (SAYÃO; SALES, 2016, p. 94)

Na ciência de forma geral, os dados são utilizados em diferentes contextos e têm seu significado alterado de acordo com sua natureza e o campo de conhecimento em que é empregado. A *National Academy of Science* (NAS) (1999, p. 15) em seu relatório, define dados de pesquisa como “[...] são fatos, números, letras e símbolos que descrevem um objeto, ideia, condição, situação ou outros fatores”. Para Borgman (2015), os dados de pesquisa são representações de objetos, observações ou outras entidades que são utilizadas como evidências de um fenômeno para os objetivos de uma pesquisa.

Veiga (2019) define dados de pesquisa como qualquer dado que seja utilizado com a finalidade de pesquisa. Sayão e Salles (2014, p. 77-78) discorrem de forma

sucinta o que é o dado de pesquisa, sendo “uma sequência genômica, a velocidade de partículas subatômicas, as respostas de levantamento social, a frequência de substantivos num corpus de textos, as imagens de satélites de outros planetas”. A Universidade de Boston (2015, on-line) trata os dados de pesquisa como sendo:

[...] informações distintas, geralmente formatadas de maneira especial [...] são coletados, observados ou criados para fins de análise para produzir dados originais resultados da pesquisa [...] podem ser gerados para diferentes finalidades e através de diferentes processos, e podem ser divididos em diferentes categorias.

Beagrie, Lavoie e Woollard (2010, p. 23) adota a seguinte definição para dados de pesquisa,

[...] coleções de dados digitais estruturados de quaisquer disciplinas ou fontes que possam ser usadas por pesquisadores acadêmicos para realizar suas pesquisas ou fornece um registro comprobatório de suas pesquisas. Os dados da pesquisa podem ser criados em vários contextos: por razões totalmente não relacionadas à pesquisa acadêmica; para pesquisa acadêmica ou como subproduto de pesquisa acadêmica.

A Fundação Oswaldo Cruz (Fiocruz) (2019) em seu curso “O que é Ciência Aberta?”, refere-se a dados de pesquisa como qualquer registro usado, obtido ou produzido ao longo das etapas de uma pesquisa que possam ser utilizados para validar os resultados de um estudo, podendo ser encontrados em diferentes formas.

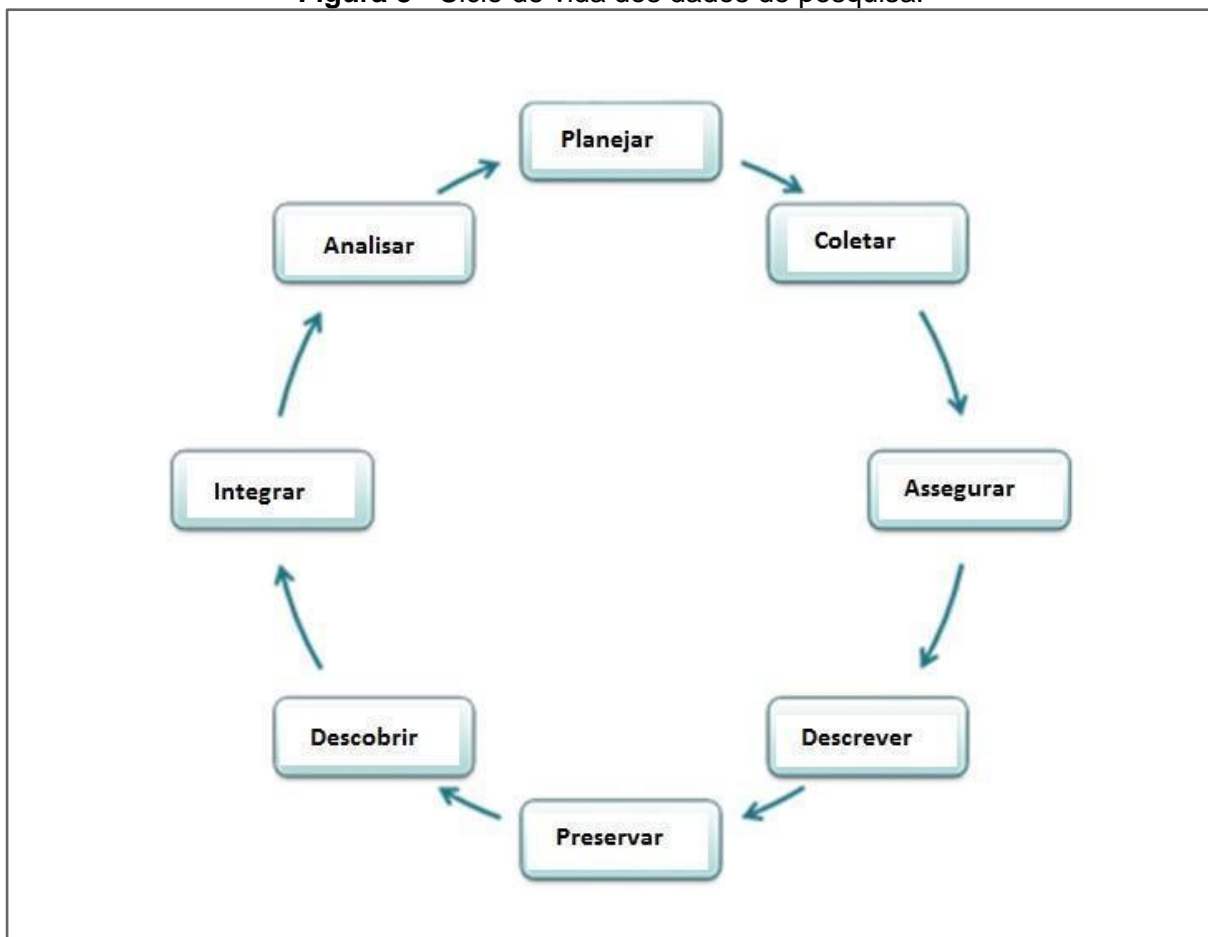
Como pode ser visto neste breve levantamento de conceitos, os dados de pesquisa possuem tipos e formatos distintos. A *National Science Foundation* (NSF) classifica os dados de acordo com a sua natureza (formato) e sua origem. Dessa maneira, os dados são gerados e processados de modos diversos. Para este estudo, os dados serão compreendidos como matéria prima produzidos ou coletados ao longo da pesquisa, utilizados como insumo comprobatório dos resultados apresentados.

3.3.1 *Ciclo de vida dos dados de pesquisa*

Os dados de pesquisa têm uma duração maior do que os projetos que os originaram, por isso eles possuem um ciclo de vida. Wigmore (2017) descreve o ciclo de vida como uma sequência de processos que estão presentes desde a origem do dado até o seu arquivamento.

Existem vários modelos de ciclo de vida de dados de pesquisa, um dos mais conhecidos é o criado pela *DataOne*, que possui 8 estágios pelos quais os dados de pesquisa passam: planejar, coletar, assegurar, descrever, preservar, descobrir, integrar e analisar (Figura 8).

Figura 8 - Ciclo de vida dos dados de pesquisa.



Fonte: Traduzido de DataOne (2013)

Cada etapa explica quais os procedimentos cujo os dados de pesquisa devem passar para garantir que possam ser preservados e utilizados (Tabela 4).

Tabela 4 - Etapas da curadoria de dados de pesquisa.

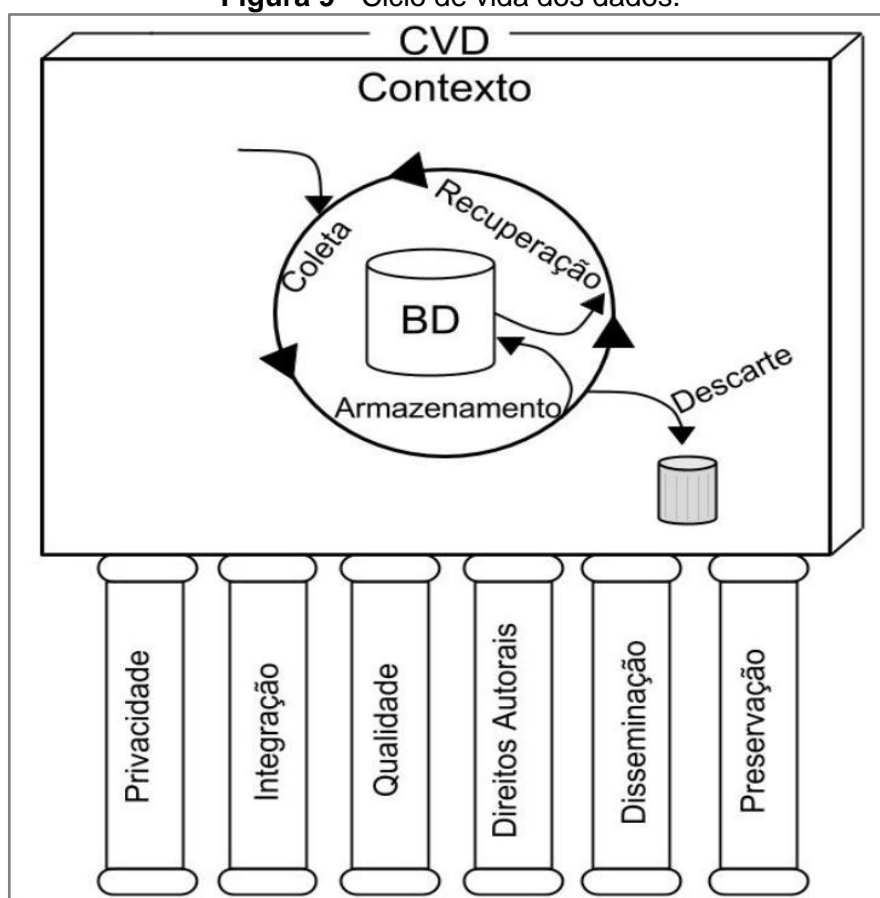
Planejar	Processo relacionado ao gerenciamento de dados. Nessa etapa é levado em conta todos os aspectos que devem ser mapeados durante o projeto, tais como políticas, planos de sustentabilidade e gerenciamento dos dados. As decisões realizadas nesse estágio são documentadas e implantadas posteriormente no plano de gestão dos dados de pesquisa.
-----------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Coleta	Define como os dados serão organizados, guardados e obtidos (observação humana, utilização de sensores ou instrumentos). Ao final desse processo, o modelo reproduz de maneira clara como os dados serão estruturados, processados, o seu formato e quais são os dados.
Assegurar	Estabelecimento de critérios para avaliação e controle de qualidade dos dados. Aplica procedimentos de identificação e correção de erros que possam prejudicar o uso pretendido dos dados.
Descrever	Descrição dos dados por meio dos metadados. Os metadados são utilizados para registrar informações que facilitem a recuperação, interpretação, uso e reuso dos dados futuramente. Informações como: organização geral dos dados, formato, nomes dos arquivos relacionados, extensão temporal, explicações sobre o processamento dos dados, técnicas de medição, detalhamento das variáveis usadas são exemplos de metadados de boa qualidade, que asseguram os armazenamentos com precisão adequada.
Preservar	Armazenamento dos conjuntos de dados que compreende duas abordagens, de curto e longo prazo. A preservação dos dados em curto prazo diz respeito ao uso de mecanismos como <i>backups</i> manuais e automatizados (disco rígido externo), quanto a de longo prazo depende da tecnologia e infraestrutura disponíveis, os principais tipos de locais para sua preservação são <i>data centers</i> e repositórios institucionais. O planejamento de guarda dos dados pode assegurar a integridade e acessibilidade dos <i>backups</i> dos dados.
Descobrir	Os dados precisam ser detectáveis na <i>web</i> após seu armazenamento. Estratégias para maximizar a visibilidade dos dados podem abarcar ações como registrar o projeto em diretórios, depositar e descrever os metadados do conjunto de dados em um repositório aberto.
Integrar	Combinação de diferentes fontes de dados que facilitam novas análises. Essas integrações podem ser feitas através de dados de vários projetos, dados espalhados (registros históricos, estatísticos, de pesquisadores, modelos, entre outros) e dados tradicionais. Como boas práticas de integração, a ativação da descoberta na <i>web</i> , ferramentas que identifiquem os dados e verificação dos dados.
Analisar	Fase onde os dados coletados/produzidos em um projeto são analisados. Nessa etapa são levados em conta os objetivos do projeto, o público (pesquisadores, profissionais, participantes e outros que possam ter interesse) e os objetivos dos dados. Diferentes <i>softwares</i> podem ser usados para a visualização, exploração e análise dos dados coletados.

Fonte: Elaboração própria (2020)

Como aponta Sant'Ana (2019), os processos que ocorrem dentro do ciclo de vida dos dados estão intimamente ligados, ao ponto que para a realização do planejamento de cada uma, tenha-se que levar em consideração características contextuais dos demais. Ainda conforme o autor, outros fatores devem estar presentes em todas as fases do ciclo de vida, são eles: privacidade, integração, qualidade, direitos autorais, disseminação e preservação (Figura 9).

Figura 9 - Ciclo de vida dos dados.



Fonte: Sant'Ana (2016, p. 20)

Todo o ciclo de vida dos dados juntamente ao uso de ferramentas tecnológicas serve para preparar os dados de pesquisa para que possam ser armazenados a longo prazo e reutilizados por outros pesquisadores em novos estudos. Para tanto, eles precisam possuir estrutura e organização e nesse ponto, entra a gestão dos dados de pesquisa, como um alicerce para as boas práticas de pesquisa.

3.3.2 Gestão dos dados de pesquisa

Ao longo dos últimos anos, a gestão de dados de pesquisa tem sido discutida em diferentes campos científicos, devido aos benefícios que traz, como o aumento de citações e o seu reuso (PIWOWAR; DAY; FRIDSMA, 2007). A gestão de dados diz respeito à organização dos dados, estando presente desde a elaboração da pesquisa até o arquivamento de seus resultados. Todo esse processo visa garantir uma consulta confiável dos resultados e seu uso para novas investigações (WHYTE; TEDDS, 2011).

Essas ações que envolvem a gestão dos dados de pesquisa, vem se consolidando “[...] em consonância com padrões de ampla aceitação, condicionantes de domínios disciplinares, requisitos estabelecidos pelos pesquisadores e, quando existentes, políticas institucionais e diretrizes de alcance nacional e internacional.” (SAYÃO; SALES, 2016)

Sales, Costa e Shintaku (2020) conceituam gestão de dados de pesquisa sob diferentes olhares, do ponto de visto dos profissionais de informação,

[...] gestão de dados de pesquisa é entendida como um conjunto de atividades gerenciais e tecnológicas, apoiadas por políticas gerais e específicas destinadas a garantir arquivamento, curadoria, preservação e oferta de acesso contínuo aos dados de pesquisa (SALES; COSTA; SHINTAKU, 2020, p. 16).

Já os pesquisadores entendem a gestão de dados “[...] como um conjunto de atividades que envolve a geração, o manuseio, o armazenamento e a descrição dos dados” (SALES; COSTA; SHINTAKU, 2020, p. 16). E as agências de fomento percebem a gestão de dados de pesquisa sendo “[...] um processo de acompanhamento da pesquisa que está sendo desenvolvida com recursos públicos por meio da supervisão dos dados que estão sendo gerados pela pesquisa” (SALES; COSTA; SHINTAKU, 2020, p. 16).

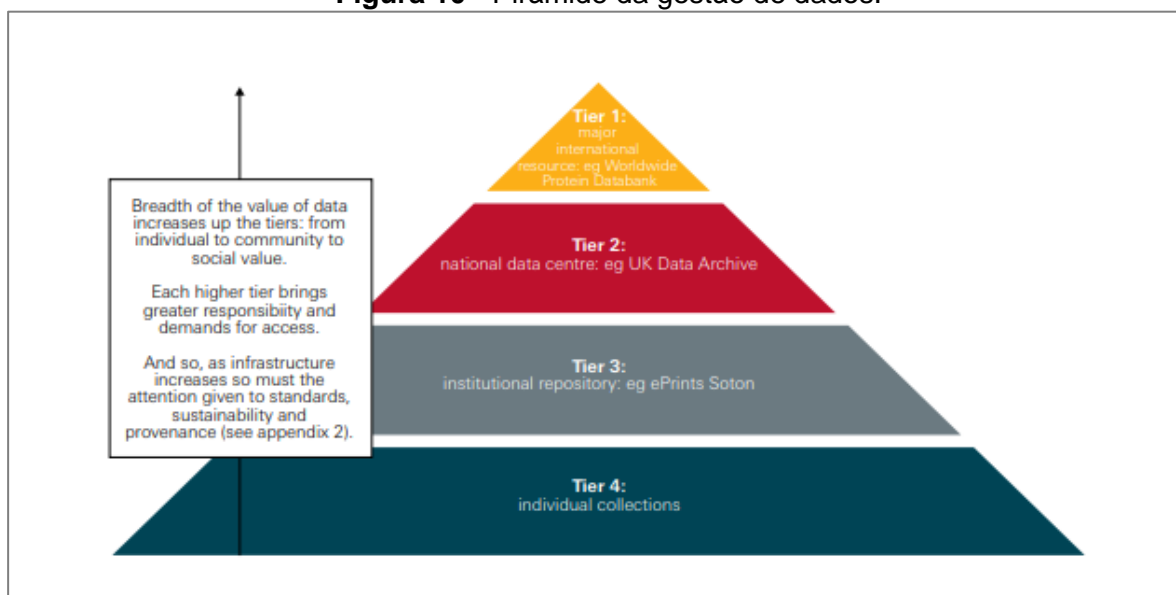
A gestão de dados pode ser vista por diferentes aspectos, por um lado ela é feita buscando uma maior visibilidade dos resultados dos dados de pesquisa, por outro lado, ela é vista como parte do ambiente *e-science* “[...] muito completo, em que as atividades do dia a dia são registradas, os dados são gerados e explorados de forma colaborativa e os resultados de investigação são preparados” (RIBEIRO *et al.*, 2016, p. 258). Corroborando com o que foi exposto pelos autores supracitados, a curadoria

dos dados de pesquisa na e-science permite traçar conexões de variados produtos obtidos em um projeto (SALES; SAYÃO, 2012)

Ainda é possível ver a gestão de dados por dois diferentes universos de pesquisa, Borgman (2012) apresenta as “big sciences” como a Física e Astronomia, campos produtores e consumidores de grande quantidade de dados e os pequenos laboratórios (universidades, pesquisadores individuais, pequenas equipes de pesquisa) que produzem uma quantidade crescente de dados, mesmo sem os aportes financeiros e tecnológicos dos grandes campos. No ambiente acadêmico é de comum entendimento de que o “[...] compartilhamento de dados pode ser complexo e custoso e precisa ser reconfigurado por estimativas realísticas de demandas e de usabilidade que justifiquem a permanência e a curadoria das coleções de dados de pesquisa” (SAYÃO; SALES, 2019, p. 36).

A *The Royal Society* (2012) propôs uma hierarquia de gerenciamento de dados composta por quatro camadas, “cada camada exige diferentes aportes financeiros e infraestruturais” (SAYÃO; SALES, 2019, p. 36). Na Figura 10, é apresentada as camadas da pirâmide da gestão de dados.

Figura 10 - Pirâmide da gestão de dados.



Fonte: The Royal Society (2012, p. 60).

Conforme apresentado no documento da *The Royal Society* (2012), cada nível é caracterizado da seguinte forma:

- **Nível 1:** dados gerados pelos principais programas internacionais como o *Large Hadron Collider*⁵ e os agregadores de fontes como o *Worldwide Protein Data Bank*⁶, ambos se caracterizam como produtores de grande quantidade de dados no contexto da *big science*.
- **Nível 2:** dados gerados por centros de dados (*data center*) financiados por órgãos nacionais.
- **Nível 3:** dados gerados por instituições individuais, universidades e institutos de pesquisa produzidos pelos seus programas de pesquisa. Os dados produzidos nesse nível possuem tratamento variado.
- **Nível 4:** dados gerados por pesquisadores individuais ou pequenos grupos de pesquisa onde a gestão dos dados é feita por eles mesmo.

Assim, Bertin, Visoli e Drucker (2017, p. 39) destacam que o planejamento é a parte com maior relevância na elaboração do plano de gestão de dados, em que assuntos relacionados aos formatos de dados, métodos utilizados para a coleta, questões ligadas à ética e aos direitos intelectuais, níveis de permissões de acesso, compartilhamento de dados e o seu reuso, depósito e preservação de curto e longo prazo precisam ser considerados e descritos.

Quando a gestão de dados é feita com eficácia

[...] reduz a quantidade de trabalho necessária para a interpretação e compilação de informações obtidas no final de um projeto de pesquisa quando se encontra bem documentado, assim a investigação em curso não precisará ser reconstruída em uma data posterior (CORRÊA, 2016, p. 388).

A gestão de dados de pesquisa é uma das atividades presentes na comunicação dos dados, pois está de forma ativa nas etapas de produção, compartilhamento e reuso de dados de pesquisa, contribuindo de maneira efetiva na publicação de dados com qualidades de forma a possibilitar o reuso por outros pesquisadores.

3.3.3 Comunicação de dados

O uso de dados em pesquisa está cada vez mais presente nas investigações científicas modernas. Nesse sentido, os dados de pesquisa fazem parte da

⁵ Disponível em: <https://home.cern/topics/large-hadron-collider>

⁶ Disponível em: <http://www wwpdb.org/>

comunicação científica, que de acordo com Garvey e Griffith (1979) é compreendida pelas atividades de produção, disseminação e uso da informação. A comunicação dos dados de pesquisa faz referência a essas práticas elencadas por Garvey e Griffith, exercendo atividades próprias na produção, compartilhamento e reuso dos dados.

A comunicação dos dados apresenta características intrínsecas a cada disciplina do conhecimento, como aponta Gomes (2013, p. 10) às comunidades científicas apresentam formas próprias de

[...] velocidades e processos díspares de pesquisa, partilha, recolha, utilização e preservação das informações, dados e do próprio conhecimento que são muitas vezes, também, acentuados por diferenças regionais e institucionais/acadêmicas.

Consequentemente, a busca pela informação, o seu acesso e uso é influenciado pelo contexto de atuação dos pesquisadores (GORRAIZ *et al.* 2009; CHOO, 1998; HUANG; CHANG, 2008). Como discorre Leite e Costa (2016, online) “[...] físicos possuem comportamento de busca, acesso e uso da informação diferente de cientistas sociais e humanistas”, essa mesma ideia pode ser aplicada a Genética e suas subáreas. As práticas culturais de descoberta da informação científica em torno da Ciência se diferem de uma disciplina para outra, essas diferenças podem ocorrer dentro de uma mesma área do conhecimento (subdisciplinas) (COSTA, 2000; SIDLER, 2014; KNORR-CETINA, 1999).

No contexto da comunicação dos dados de pesquisa, os pesquisadores costumam ter pouco conhecimento ou experiência no gerenciamento dos dados (MAYERNIK, 2011). Outra dificuldade detectada na realização da comunicação dos dados de pesquisa é a sua heterogeneidade que varia drasticamente entre as disciplinas, até mesmo entre grupos de pesquisas que trabalham com temas afins (GÓMEZ; MÉNDEZ; HERNÁNDEZ-PÉREZON, 2016).

Neste sentido, Dallmeier-Tiessen *et al.* (2014) tratam a cultura acadêmica, que são princípios de comportamento de uma área do conhecimento, refletidas na conduta do pesquisador pertencente a ela. Devida a esses princípios comportamentais, algumas áreas vão ser mais propícias ao compartilhamento dos dados de pesquisa do que outra, logo, esses comportamentos são espelhados nos pesquisadores. Além da cultura acadêmica, os autores Dallmeier-Tiessen *et al.* (2014) relatam que a

comunicação dos dados de pesquisa é mais rotineira entre pesquisadores mais jovens devido a sua facilidade em utilizar sistema de compartilhamento.

3.4 Produção de dados de pesquisa

A obtenção de dados de pesquisa pode ser feita de diferentes maneiras, dependendo das características do estudo, da natureza do fenômeno analisado, do método escolhido para a coleta de dados, os resultados que se busca e/ou entre outros aspectos. A forma de produção de dados também varia de acordo com os grupos de pesquisa que os produzem e os processos utilizados para extração e análise dos dados (SAYÃO; SALES, 2014; BORGMAN, 2015). Torres-Salinas; Robinson-Garcia; Cabezas-Clavijo (2012) discorrem que os campos de conhecimento são as responsáveis por determinar uma especificação para os dados de pesquisa, levando em conta os métodos, práticas e normas inerentes de cada disciplina.

Quanto à natureza dos dados, a NSF (2007) descreve como materiais diversificados compostos por *software*, vídeos, áudios, imagens e números. Os dados também podem ser sua natureza classificada como observacional, computacional e experimental (Tabela 5).

Tabela 5 - Natureza dos dados de pesquisa.

Tipo	Característica
Observacional	São dados obtidos por meio de observação direta, que pode ser de lugares, fenômenos e/ou tempo específico. Esse tipo de dado assume uma característica de registro histórico, visto que não é possível obter novos dados de algo que já aconteceu. Um exemplo de dado observacional é o registro de um eclipse solar em uma data específica.
Computacional	São dados obtidos através do resultado de simulação ou execução de um modelo computacional. Esse tipo de dado pode ser replicado ao longo do tempo. A criação de ambientes virtuais educacionais é um exemplo de uso de dados computacionais.
Experimental	São dados provenientes de experimentos realizados em ambientes controlados, como em laboratórios. Esses dados podem ser reproduzidos com precisão. Registros do melhoramento genético de uma espécie de planta é um dado experimental.

Fonte: National Science Board (2005); Sayão e Sales (2014); Costa e Leite (2018).

Outro aspecto interessante dos dados de pesquisa é a forma em que são organizados, como estruturados e não estruturados. Os dados estruturados

normalmente são armazenados em um conjunto de tabelas que contém relação entre si, já os dados não estruturados são aqueles apresentados em forma de texto, imagens, sons entre outros. Nas palavras de Kelleher e Tierney (2018, p. 49, tradução nossa) os

[...] dados não estruturados são muito mais comuns do que os dados estruturados. Por exemplo, as coleções de texto humano (e-mails, tweets, mensagens de texto, postagens, romances, etc.) podem ser consideradas dados não estruturados, assim como coleções de arquivos de som, imagem, música, vídeo e multimídia. [...] Muitas vezes podemos extrair dados estruturados de dados não estruturados.

Enquanto isso, os dados estruturados são apontados por Inmon e Linstedt (2014) como elementos gerenciados por um sistema de banco de dados, consistindo em: chaves, atributos, índices e registros. Os dados estruturados, conforme os autores, são bem definidos, possuindo um formato padronizado, o que possibilita uma localização correta, como é o caso de dados gerados em banco de dados, apesar dos dados estruturados dispor de uma estrutura organizada é mais comum encontrar conjuntos de dados não estruturados na *web*.

Um outro ponto que pode ser levantado sobre os dados de pesquisa é o seu nível de tratamento. Como mostra Costa (2017), os dados podem ser classificados como brutos e derivados.

A primeira é a dos dados brutos (raw data) ou formato primário. Esses dados são passíveis de ser submetidos a diversos estágios de processamento e podem ser utilizados por diversas pesquisas, que não participaram de seu processo de produção ou coleta. Essa é a categoria de dados de maior interesse em ser compartilhada pela sua capacidade de ser reutilizada facilmente por outras pesquisas. [...] A segunda categoria é denominada de dados derivados, que são versões de dados brutos após processamento específico (COSTA, 2017, p. 40).

No sistema científico moderno, os dados de pesquisa podem nascer digitalmente ou passarem por uma conversão de formato, do físico para o digital. Sales e Sayão (2020) versam sobre os formatos em que os dados de pesquisa são encontrados, onde por muito tempo eles se apresentavam em artigos de periódicos, teses, livros e outros canais tradicionais da comunicação acadêmica, mas hoje esses dados se manifestam em *softwares*, modelos, multimídias, base de dados, entre outros. Essa flexibilidade de formato facilita o gerenciamento dos dados em suportes

tecnológicos. Os dados de pesquisa podem ser obtidos em diferentes formatos e suportes, como observado na Tabela 6 abaixo.

Tabela 6 - Tipos de dados de pesquisa.

Formato	Suporte
Texto	MS Word, Portable Document Format (PDF), Rich Text Format (RTF), Hyper-Text Markup Language (HTML), Extensible Markup Language (XML), etc.
Numérico	SPSS, Stata, MS Excel, SAS, arquivos hierárquicos, etc.
Multimídia	JPEG, TIFF, GIF, MPEG, Quicktime, Bitmap, PNG, etc.
Modelo	3D, estatístico, macroeconômico, causal, de similaridade, etc.
Software	Java, C, Perl, Python, Ruby, PHP
Disciplina-específico	Crystallographic Information File (CIF)
Instrumento-específico	Carl Zeiss Digital Microscopic Image Format (ZVI).
Espacial	Shapefile

Fonte: adaptado de Bertin, Visoli e Drucker (2017, p. 37)

A produção de dados de pesquisa pode ser feita por procedimentos variados, dependendo muito em que área se desenvolve a pesquisa e os métodos necessários para aquisição de tais dados para o desenvolvimento do estudo. Costa e Leite (2018) discorrem sobre os dados serem utilizados para a promoção dos pesquisadores em suas carreiras acadêmicas, visto que os dados são necessários para o desenvolvimento de pesquisa e o resultado dela é necessário para a progressão profissional dos pesquisadores, entretanto, nem sempre a recompensa vem da comunicação somente dos dados, mas sim da pesquisa como um todo.

Deste modo, as características inerentes à produção dos dados se mostram fatores importantes na atividade de compartilhamento, visto que o compartilhamento dos dados quando padronizados e legíveis por máquinas e humanos têm uma representação significativa no ecossistema científico, pois esses dados que são gerados através de pesquisas são capazes de contribuir em novos estudos, otimizando o tempo de desenvolvimentos dessas pesquisas.

3.5 Compartilhamento de dados de pesquisa

O compartilhamento de dados contribui de forma expressiva para o avanço da ciência e maximiza os investimentos em pesquisas.

De acordo com Sayão e Sales (2015) o compartilhamento de dados proporcionam três ações positivas para a ciência, a primeira é relacionada ao aumento do volume de produtividade das atividades científicas, a segunda é referente ao crescimento do número de publicações que têm como base os dados de pesquisa que são compartilhados, e a terceira é relativa às citações (pesquisadores que disponibilizam seus dados, são mais citados).

Além dos benefícios citados, um outro motivo que influencia o aumento do compartilhamento pelos pesquisadores está relacionado aos critérios de financiamento de pesquisas das agências de fomento, como levantado pela Nassi-Calò (2019, on-line)

O número de conjuntos de dados disponibilizados nos últimos dez anos vem crescendo de forma exponencial, tendo atingido a marca de cerca de 10.000 em 2018, motivado principalmente por mandatos institucionais ou de agências de fomento.

Sayão e Sales (2015, p. 53) enumeram algumas vantagens do ponto de vista do pesquisador sobre o compartilhamento de dados de pesquisa.

1. Aumento na eficiência da pesquisa, por exemplo, evitando a duplicação de esforços através da pronta disponibilidade de ferramentas de pesquisa, protocolos e exemplos de boas práticas, pela redução do custo de formação de coleção de dados e pela promoção e adoção de padrões abertos;
2. Incentiva a investigação e o debate científico;
3. Promove a inovação e potencializa novos usos para os dados;
4. Possibilita novas formas de colaboração entre usuários e criadores de dados;
5. Maximiza a transparência e a prestação de contas;
6. Permite uma avaliação mais precisa das descobertas científicas;
7. Colabora para o aperfeiçoamento e a validação dos métodos científicos;
8. Evita o custo da duplicação de coleções de dados;
9. Aumenta a visibilidade do impacto e a visibilidade dos resultados de pesquisa;
10. Contribui para dar os devidos créditos ao criador dos dados;
11. Oferece um recurso importante para a educação e a capacitação

Por outro lado, como apresentado por Walport e Brest (2011), o hábito de compartilhamento de dados ainda não foi totalmente adotado pela comunidade

científica. Ainda de acordo com os autores, os dados são tratados pelos pesquisadores que os coletam como propriedades particulares. Borgman (2012) explana que o compartilhamento de dados de pesquisa pode ser prejudicado por problemas relacionados a questões éticas e de privacidade e receios dos pesquisadores de como compartilhar seus dados. Ainda conforme a autora, a apreensão por parte dos pesquisadores para realizarem o compartilhamento dos dados pode estar ligado há várias razões, tais como: ausência de experiência em compartilhar, falta de incentivos ou recursos para o compartilhamento, os dados não se encontrarem compatíveis com sistemas de guarda e compartilhamento de dados, receio de não receberem os créditos pelos dados, entre outros motivos.

Apesar das questões abordadas acima, o compartilhamento é uma das formas de potencializar a difusão da informação (HALL, 2001), essa ideia também se aplica ao compartilhamento dos dados de pesquisa, já que uma vez compartilhados, podem ser utilizados por outros pesquisadores.

No contexto do compartilhamento de dados de pesquisa, existem as boas práticas que visam “[...]” tornar a pesquisa mais eficiente, eficaz e satisfatória para os pesquisadores (BAYNES, 2017, on-line). Essas boas práticas têm como objetivo que os dados de pesquisa não sejam só compartilhados, mas que possam também serem reutilizados por outras pessoas. Para tornar o compartilhamento de dados uma ação mais adequada, recentemente foram criados os Princípios FAIR, que no inglês significa *Findable, Accessible, Interoperable e Reusable*, um conjunto de orientações para o tratamento dos dados. Mons *et al.* (2017, p. 51-52, tradução nossa) descreve os princípios FAIR como um guia que

[...] fala sobre a necessidade de descrever um processo - mecanizado ou manual - para acessar dados localizáveis; um requisito para descrever abertamente e ricamente o contexto em que esses dados foram gerados, para permitir a avaliação de sua utilidade; definir explicitamente as condições sob as quais eles podem ser reutilizados; e fornecer instruções claras sobre como eles devem ser citados quando reutilizados. Nenhum desses princípios exige que os dados sejam “abertos” ou “livres”. Eles, no entanto, exigem clareza e transparência em torno das condições que regem o acesso e a reutilização. Como tal, embora os dados do FAIR não precisem estar abertos, a fim de cumprir com a condição de reusabilidade, os dados do FAIR precisam ter uma licença clara, de preferência legível por máquina.

Assim, os princípios FAIR oferecem uma descrição rica dos dados de pesquisa e seus metadados, garantindo o compartilhamento e o uso/reuso dos dados de pesquisa. Os elementos dos princípios FAIR estão correlacionados, porém, podem ser utilizados de formas independentes (JORGE, 2018). Cada módulo dos princípios FAIR apresenta um conjunto de instruções de acordo com a Figura 11.

Figura 11 - Princípios FAIR.



Fonte: Fiocruz (2019)

Os princípios FAIR apresentam formas de o pesquisador produzir dados eficientes, visto que apenas a disponibilização deles na internet não é garantia de que sejam reutilizados. Para que sejam de fato reutilizados por outras pessoas, é necessário que o pesquisador disponibilize a documentação dos dados, com informações que descrevam os processos aplicados para a produção e/ou coleta. Sayão e Sales (2015) dialogam sobre a importância da documentação para que o compartilhamento seja efetivo.

Sem uma descrição minuciosa do contexto tecnológico dos arquivos de dados, do contexto no qual os dados foram criados ou coletados, das medidas que foram feitas, dos detalhes espaciais e temporais, dos

instrumentos usados, dos parâmetros e unidades e da qualidade dos dados e da sua proveniência, é improvável que os dados possam ser descobertos, interpretados, gerenciados e efetivamente usados e reusados. (SAYÃO; SALES, 2015, p. 20)

Autores como Riley (2017) e Méndez Rodríguez (2002) versam sobre os metadados e os seus descritores, buscando garantir que a documentação de um dado seja completa. Wittenburg, Hellström e Zwölf (2017) tratam da necessidade dos dados terem identificadores persistentes. Tais necessidades levantadas pelos autores supracitados são pautadas nos princípios FAIR, dada a importância da etapa de compartilhamento na comunicação dos dados.

Como levantado por Jorge (2018, p. 59), os princípios FAIR “[...] têm sido reconhecidos por organismos internacionais como uma estrutura útil para pensar o compartilhamento e a abertura de dados, por exemplo a Comissão Europeia”. Diferentes iniciativas vêm procurando soluções que eliminem ou diminuam as barreiras do compartilhamento de dados e garantam o seu uso/reuso eficiente. Assim, o uso/reuso de dados que elevam a produtividade da pesquisa está diretamente ligado com o compartilhamento de dados (PIENTA; ALTER; LYLE, 2010).

3.6 Reuso de dados de pesquisa

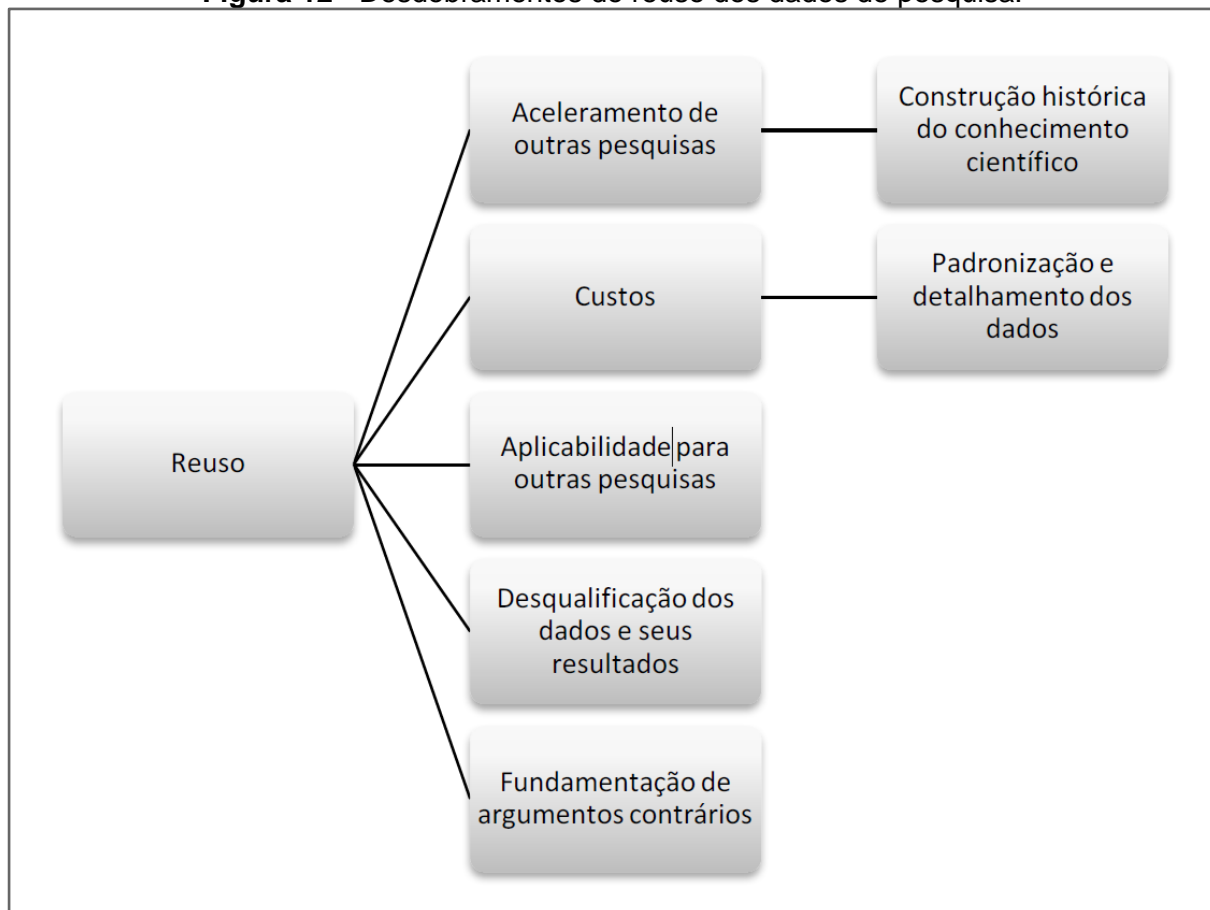
O reuso de dados de pesquisa é uma característica inerente da Ciência Aberta, pois garante a retroalimentação científica (SANTAREM SEGUNDO, 2020). Entretanto, como apontam os autores Park e Wolfram (2017), questões relacionadas ao uso/reuso de dados são pouco entendidas entre os pesquisadores. Apesar disso, temas que envolvem a utilização de licenças que permitam o uso dos conjuntos de dados por terceiros, desde que respeitem os critérios estabelecidos, resguardam os pesquisadores produtores. Outro ponto debatido também é a citação dos dados, que tem como finalidade o desenvolvimento da prática de citação, onde os dados são compreendidos pelos seres humanos e operados pelas máquinas.

3.6.1 Características do uso/reuso dos dados de pesquisa

Costa (2017) destaca os desdobramentos que o reuso dos dados de pesquisa podem ter. De acordo com a autora, o reuso dos dados assume características distintas, ora promotores da comunicação e em outras, inibidores. Essa dicotomia

ocorre devido aos vários aspectos presentes nessa etapa da comunicação dos dados, como apresentado na Figura 12.

Figura 12 - Desdobramentos do reuso dos dados de pesquisa.



Fonte: Costas (2017)

Para que o uso/reuso dos dados de pesquisa por outros pesquisadores seja potencializado, questões relacionadas à estruturação das informações dos dados e as condições de seu uso devem ser claras. A *Australian National Data Service* (ANDS) (2018) elenca ações que favorecem o uso/reuso de dados, elas são relacionadas ao: planejamento, metadados, licenciamento de dados, vocabulário, permissões e citação. Essas ações listadas pela ANDS vão de encontro com os princípios FAIR, visto que assim como o FAIR, são atividades que visam às boas práticas no compartilhamento e reuso dos dados. Uma vez que os dados atendam às orientações da ANDS, uma parte considerável dos princípios FAIR relacionados à acessibilidade, interoperabilidade e reutilização, são contemplados.

No que tange o planejamento dos dados de pesquisa, envolve políticas e procedimentos institucionais para a reutilização dos dados por outros pesquisadores,

pode-se citar como exemplo de política de uso de dados, o *Australian Code for Responsible Conduct of Research*, que é a política de gerenciamento de dados de pesquisa da *National Health and Medical Research Council* (NHMRC) (2018, p. 2, tradução nossa), que traz essa etapa como procedimentos que trata “[...] da propriedade de materiais e dados de pesquisa, seu armazenamento, retenção após o final do projeto e acesso adequado a eles pela comunidade de pesquisa”.

Os metadados “[...] são informações sobre um objeto ou recurso que descrevem características desse objeto, como conteúdo, qualidade, formato, localização e direitos de acesso” (ANDS, 2016, tradução nossa). O registro de dados de pesquisa por meio de metadados bem descritos, ou seja, metadados ricos, proporciona que conjuntos de dados sejam detectados, citados, reutilizados e acessados ao longo do tempo.

3.6.2 Licenças abertas

A aplicação do licenciamento aberto aos dados de pesquisa permite que eles possam ser utilizados por outras pessoas e especifica as formas permitidas de uso possível de acordo com os aspectos legais. As licenças concedem permissões que variam os graus de reuso. Ball (2014, p. 4, tradução nossa) elenca 3 condições comuns que as licenças têm:

1. **Atribuição:** significa que o licenciante deve receber o devido crédito pela obra quando ela é distribuída, exibida, executada ou usada para derivar uma nova obra;
2. **Copyleft:** significa que quaisquer novos trabalhos derivados do licenciado devem ser liberados sob a mesma licença e somente essa licença; e
3. **Não comercial:** impede que o licenciado explore o trabalho comercialmente.

Embora o uso de licenças personalizadas seja pertinente para atender a situações específicas, as licenças padrões podem atender de maneira satisfatória aos projetos. A *Creative Commons* tem um conjunto de licenças que permitem os dados serem utilizados desde que respeitem o direito de propriedade até as questões relacionadas ao uso comercial. Guanaes (2020, p. 82-83) fez um levantamento das

licenças da *Creative Commons* (Tabela 7) que permitem o reuso de dados de pesquisa com bases legais, sendo:

Tabela 7 - Licenças para o reuso de dados de pesquisa.

LICENÇA (OPEN DATA COMMONS)	SIGLA	CARACTERÍSTICAS
Licença e Dedicção ao Domínio Público	PDDL	Dedica o banco de dados e seu conteúdo ao domínio público, gratuitamente, para que todos usem como acharem melhor.
Licença de Atribuição	ODC-By	Os usuários são livres para usar o banco de dados e seu conteúdo de maneiras novas e diferentes, desde que forneçam a atribuição à fonte dos dados e/ou do banco de dados.
Licença de Banco de Dados Aberta	ODC-ODbL	Uso subsequente do banco de dados deve fornecer atribuição de autoria, uma versão irrestrita do novo produto deve estar sempre acessível e quaisquer novos produtos feitos com material ODbL devem ser distribuídos usando os mesmos termos. É a mais restritiva de todas as licenças ODC.
Open Government License.	OGLv3	Esta licença aplica-se a provedores de informação do setor público do Reino Unido. Protege, originalmente, direitos autorais e de banco de dados ante o licenciamento de informações da Coroa britânica.

Fonte: Guanaes (2020, p. 82-83)

O uso de vocabulário é capaz de oferecer maneiras mais claras de comunicação, pois oferecem um conjunto de termos (palavras, códigos, ícones, entre outros) estruturados que se relacionam entre si, utilizados em um grupo específico para representar os seus conceitos (ANDS, 2017). O vocabulário auxilia no entendimento correto dos dados, permitindo que proprietários e usuários tenham uma compreensão comum dos significados desses dados.

O reuso dos dados abrange também a temas relacionados com as permissões de uso dos dados de pesquisa e questões legais. Como aponta a ANDS (2018) as permissões de uso de dados devem fornecer informações relacionadas a legislações; licenciamento; *De-identification*; consentimento informado e compartilhamento de dados; diretrizes e requisitos do financiador, caso possua; controle de acesso e entre

outras orientações, que representam as necessidades de uma determinada comunidade científica.

3.6.3 Critérios para citação de dados

As citações de dados de pesquisa têm como objetivos transformar esse insumo em parte integrante de registros acadêmicos. Assim, em 2014, a Force11 em parceria com repositórios de dados, instituições de pesquisa e editoras, elaboraram os Princípios de Citação de Dados que elenca oito critérios ligados às funções e atributos das citações (Tabela 8).

Tabela 8 - Princípios para citação de dados.

Princípios	Características
Importância	Os dados devem ser considerados produtos legítimos e citáveis da pesquisa. A mesma importância deve ser atribuída às citações de dados em registros acadêmicos e às citações de outros objetos de pesquisa, como publicações.
Crédito e atribuição	As citações de dados devem facilitar a designação do crédito acadêmico e a atribuição normativa e legal de todos os colaboradores dos dados, reconhecendo que um único estilo ou mecanismo de atribuição pode não se aplicar a todos os dados.
Evidência	Na literatura acadêmica, sempre e onde quer que uma alegação for baseada em dados, os dados correspondentes devem ser citados.
Identificação única	Uma citação de dados deve incluir um método persistente de identificação que possa ser acionado por máquina, seja único em todo o mundo e amplamente usado pela comunidade.
Acesso	As citações aos dados devem facilitar o acesso aos próprios dados e aos metadados associados à documentação, ao código e a outros materiais, conforme houver necessidade, para que tanto seres humanos quanto máquinas usem os dados de referência adequadamente.
Persistência	Os identificadores únicos, e os metadados que descrevem os dados, e sua disposição, devem ser retidos – mesmo após o período de validade dos dados que forem descritos.
Especificidade e possibilidade de verificação	As citações devem facilitar a identificação, o acesso e a verificação de dados específicos usados como base em uma alegação. As citações ou metadados de citações devem incluir informações sobre a proveniência e

	imutabilidade suficientes para facilitar a verificação de que o específico período de tempo, a versão e/ou a porção granular dos dados recuperados subsequentemente sejam os mesmos da citação original.
Interoperabilidade e flexibilidade	Os métodos de citação de dados devem ser suficientemente flexíveis para acomodar a variação de práticas entre as comunidades, mas não devem diferir a ponto de comprometer a interoperabilidade das práticas de citação de dados entre comunidades.

Fonte: Force11 (2014)

Os Princípios de Citação de Dados oferecem uma síntese de como eles devem ser descritos para ser citáveis e consequentemente reutilizados, abrangendo características quanto a sua função, atributos e propósito. Por não serem princípios rígidos, podem ser adaptados conforme as práticas científicas de cada comunidade e as tecnologias empregadas.

Como pode ser percebido, o uso/reuso dos dados de pesquisa por outros estudiosos aceleram o desenvolvimento das pesquisas, evitando assim o retrabalho e gastos de verbas entre outros benefícios. Estudos que abordam o reuso dos dados, a sua produção e compartilhamento e ao mesmo tempo questões relativas à ética, curadoria, diretrizes, licenciamentos, padronização e transparência da pesquisa no contexto da comunicação científica estão presentes em diferentes disciplinas, se mostrando como uma preocupação em comum das diferentes áreas do conhecimento.

3.6.4 Estudos relacionados à abertura dos dados de pesquisa

A abertura dos dados, juntamente às técnicas de produção, compartilhamento e seu uso/reuso vem sendo abordados no mundo todo por diferentes organizações. Nas últimas décadas, iniciativas como a *Budapest Open Access Initiative* (BOIA), ganharam destaque por promover ações para a abertura dos dados de pesquisa. Desde então, diferentes autores discutem sobre os dados de pesquisa em seus diferentes contextos disciplinares.

Os dados de pesquisa têm seu significado modificado dependendo do contexto ao qual estão ligados (BORGMAN, 2007), assim “[...] para sismólogos, os dados podem ser as sequências dos bits de sensores sísmicos; para geomorfólogos, podem ser fragmentos de rochas; e, para os sociólogos, podem ser as gravações de uma conversa” (CARVALHO, 2018, p. 80).

Corroborando com Borgman (2007), autores como Gómez, Méndez e Hernández-Pérez (2016) e Murray-Rust (2008) versam sobre as interpretações de cada domínio científico acerca dos dados de pesquisa conforme a sua natureza, coleta, compartilhamento e reuso. Práticas de compartilhamento e reuso de dados de pesquisa são mais comuns em áreas como a Física, Astrologia e Genética que possuem uma cultura científica colaborativa (SALES; SAYÃO, 2020; NATURE COMMUNICATIONS, 2018).

Devido aos diferentes olhares que os dados de pesquisa recebem de acordo com as práticas científicas de cada campo, serão apresentados alguns estudos que explanam sobre a produção, compartilhamento e uso/reuso dos dados gerados por pesquisas.

Em 2010, a *Key Perspectives Ltd* elaborou uma revisão comparativa utilizando dezesseis estudos de caso para identificar as diferenças disciplinares relacionadas à forma de produzir, depositar, compartilhar, reutilizar e preservar os dados de pesquisa por diferentes campos de estudo. O documento foi dividido em quatro grupos, são eles: Artes e Humanidades, Ciências Sociais, Ciências da Vida e Ciências Exatas.

A abordagem adotada para a seleção de literatura de cada área foi a combinação de métodos de estudos de caso com levantamento de literaturas sobre curadoria digital consideradas relevantes em cada domínio. O objetivo deste estudo foi identificar as boas práticas referentes às disciplinas.

O estudo se mostra interessante, pois além de permitir observar como os pesquisadores veem o exercício de compartilhar seus dados para que outros os reutilizarem, é possível perceber também as atitudes e comportamentos dos pesquisadores, que estão diretamente ligadas à disciplina em que atuam. Ao longo da revisão, é apresentada a partir do estudo desses dezesseis casos a percepção dos pesquisadores sobre os dados de pesquisa e suas facetas. Nas disciplinas ligadas a Artes e Humanidades, a produção de dados de pesquisa em grande quantidade e o seu compartilhamento não era muito comum em 2010, quando o estudo foi feito, visto que as pesquisas são elaboradas e desenvolvidas em forma de monografias. Apesar do compartilhamento de dados não ser uma atividade corriqueira entre esse grupo, até março de 2008 o *Arts and Humanities Research Council (AHRC)* manteve o *Arts and Humanities Data Service (AHDS)*, um serviço de banco de dados onde os

pesquisadores podiam depositar e baixar os conjuntos de dados, coleções multimídia de artes, textos literários e outros materiais.

Nas Ciências Sociais, o processo de descoberta/coleta de conjuntos de dados e o seu compartilhamento é mais habitual entre os pesquisadores, principalmente aqueles que recebem financiamentos públicos. Serviços como o *Economic and Social Data Service* (ESDS) oferecem uma ampla gama de dados de pesquisa que passaram por curadoria e estão disponíveis para uso. Apesar desses esforços de compartilhamento de dados de pesquisa, muitos pesquisadores relatam dificuldades de acesso a conjuntos de dados, tais como: privacidade e direitos à confiabilidade, acesso a dados limitados a apenas as equipes ou organizações que os coletaram, receios relacionados ao uso irrestrito e deturpação dos dados, disponibilização de dados mediante o pagamento de taxas elevadas, acesso aos dados restrito a máquinas físicas, pois o mesmo não se encontra em ambiente digital, entre outras dificuldades.

Nas Ciências da Vida, a prática de produzir, compartilhar e reutilizar dados de pesquisa não é uma novidade entre os pesquisadores. O acesso, coleta e reutilização de dados genômicos retirados de banco de dados públicos, por exemplo, é um processo relativamente fácil para eles, além dos campos de metadados estarem bem estruturados e descritos, os pesquisadores têm familiaridade com os bancos de dados que são relevantes para o desenvolvimento de seus trabalhos.

Nas Ciências Exatas a produção, compartilhamento e reuso dos dados estão muito ligados às características de cada disciplina. As Engenharias são apontadas no estudo da *Key Perspectives Ltd* (2010) como exemplo de disciplina onde o compartilhamento de dados não é tão comum entre seus pares. O projeto piloto chamado *Pilot Engineering Repository Xsearch* (PERX) tinha como objetivo de fornecer um serviço de descobertas de recursos em vários repositórios de interesse de estudantes em engenharia e comunidades de pesquisa, como conclusão do projeto, foi notada a inexistência de suporte da comunidade para abordagem baseada em temas para a descoberta de recursos, diferenças disciplinares dentro das engenharias e que as informações nas engenharias são complexas, dificultando a tarefa de fornecer serviços entre repositórios. Como é apresentado no estudo elaborado da *Key Perspectives Ltd* (2010), diferentemente das Engenharias, as ciências que trabalham com o clima, por exemplo, são melhores atendidas nesse

aspecto, possuindo *data centers* que oferecem uma vasta quantidade de conjuntos de dados para reuso.

Esse estudo realizado pela *Key Perspectives Ltd* possibilitou uma maior compreensão sobre as diferenças disciplinares e como os dados de pesquisa são vistos na primeira década do século XXI, uma vez que cada campo do conhecimento apresenta suas peculiaridades e maneiras específicas de produzir, compartilhar e reutilizados os dados.

Veiga *et al.* (2019) realizou um estudo parcial para compreender a percepção dos pesquisadores sobre o compartilhamento de dados de pesquisa, os pesquisadores são ligados à Fiocruz, que é uma instituição de pesquisa com foco nas áreas de ciência, tecnologia e saúde pública no Brasil. Os autores, para a coleta de dados, utilizaram questionário on-line criado pelo *Google Forms* e a aplicação do questionário se deu de 22 de dezembro de 2018 a 7 de junho de 2019, obtendo 94 respostas. Para análise dos dados, os autores selecionaram seis questões para investigação e discussão.

A primeira análise foi referente à compreensão dos pesquisadores da Fiocruz sobre a atitude e valores relativos à abertura de dados de pesquisa na comunicação científica, foi possível observar que:

1. 67% concordaram totalmente que a abertura dos dados traz transparência para a pesquisa;
2. 66% concordaram totalmente que a abertura dos dados acelera o progresso científico, uma vez que esses dados podem ser utilizados por outros pesquisadores;
3. 62% concordam totalmente que a abertura dos dados favorece a reprodutibilidade da pesquisa; e
4. 53% concordam totalmente que a abertura dos dados de pesquisa contribui para a ampliação da credibilidade e integridade da pesquisa.

A segunda questão analisada foi relacionada aos motivos para a abertura dos dados de pesquisa. Entre os fatores observados, 51,1% dos pesquisadores se sentem motivados a compartilhar os seus dados para ajudar a melhorar a qualidade da pesquisa; 48,9% se sentem motivados pela visibilidade que a sua pesquisa vai ter; 47,9% dos pesquisadores veem motivação na abertura dos seus dados porque

acreditam que os leitores têm boas intenções no uso dos dados; 45,5% dos pesquisadores abrem seus dados para obter maior confiabilidade em seus dados, visto que os repositórios oferecem maior credibilidade aos dados depositados, 41,5% abrem seus dados motivados pela possibilidade de obterem mais financiamento para a pesquisa.

O terceiro ponto abordado na pesquisa de Veiga *et al.* (2019) foram os principais motivos para não realizarem o compartilhamento de dados de pesquisa em plataformas de dados. Os cinco principais motivos são: 47,5% dos pesquisadores entrevistados não compartilham os dados por falta de conhecimento de como o fazer e pela necessidade de auxílio para executar a ação; 47,5% não sabem se podem abrir os seus dados de pesquisa, pois contém informações sigilosas; 45,5% dos pesquisadores tem receio que seus dados sejam mal usados e/ou interpretados por outros pesquisadores, por isso não os compartilham; 31,1% alegam não abrir seus dados por não terem conhecimento de nenhum repositório.

A quarta análise foi feita em relação aos tipos de dados que os pesquisadores utilizam. 52,1% fazem uso de dados pessoais⁷ e dados pessoais sensíveis⁸; 47,9% usam dados não sigilosos; 47,9% fazem uso de dados potencialmente sigilosos, porém que ainda não passaram por análise de confidencialidade; 43% utilizam dados em parcerias interinstitucionais; e 27% dos pesquisadores usam os dados que eles mesmo geram ou que podem gerar patentes.

O volume total dos conjuntos de dados também foi analisado. Em ordem crescente: 8% dos pesquisadores informaram que seus conjuntos de dados são maiores que 100 GB e menor que 1TB; 12% responderam que é menor que 1GB; 12% menor que 10GB; 26% disseram ser menor que 1GB; e 30% dos pesquisados se quer sabem o volume de seus dados.

O sexto ponto discutido no estudo foi sobre os interesses/necessidades dos pesquisadores em serviços institucionais de boas práticas na gestão de dados em todo o ciclo de vida da pesquisa. Em ordem decrescente as respostas foram:

⁷ De acordo com Lei Geral de Proteção de Dados (LGPD) (2019), dados pessoais são informações que possibilita identificar a identidade do indivíduo.

⁸ De acordo com Lei Geral de Proteção de Dados (LGPD) (2019), dados pessoais sensíveis são informações que revelam questões Genéticas, origem racial ou étnica, convicções religiosas, opiniões políticas entre outros.

1. 73% responderam necessitarem de um serviço de preservação e armazenamento de dados para garantir o acesso futuro;
2. 72% de um sistema de *backup* de arquivos
3. 67% apoio técnico na criação de um plano de gestão de dados
4. 52% interesse em serviços de acesso e compartilhamento de dados;
5. 50% formação e consultoria na gestão de dados;
6. 44% informações referentes a repositórios temáticos; e
7. 36% interesses em identificação de normas e esquemas disciplinares de descrição de dados.

Apesar de quase 10 anos terem se passado desde o estudo da Key Perspectives Ltd até o da Veiga *et al.* (2019), algumas questões continuam presentes no debate sobre os dados de pesquisa. A falta de conhecimento de como compartilhar os dados é uma delas, a maioria dos pesquisadores acham importante o compartilhamento e preservação dos dados, porém, não sabem se podem fazer isso de maneira legal e nem onde realizar o depósito. O que pode demonstrar que a gestão de dados não é tão disseminada entre os grupos que trabalham com dados.

Chawinga e Zinn (2019) realizaram um estudo sobre o compartilhamento de dados de pesquisa sob a perspectiva global. O trabalho se propôs a contribuir com a ciência de duas formas, a primeira enriquecendo a literatura oferecendo *insights* sobre o compartilhamento de dados em diferentes níveis (individual, institucional e internacional) e a segunda propor aos interessados em pesquisa (editores, agências dos governos, bibliotecários, agências de fomento e cientistas) formas de melhorar o compartilhamento de dados.

Para a elaboração do estudo, os autores Chawinga e Zinn (2019) utilizam como metodologia a revisão sistemática com abordagem quantitativa. Após a aplicação dos critérios de seleção de material estabelecidos na metodologia, foram coletadas e analisadas 105 publicações.

Os resultados da pesquisa realizada por Chawinga e Zinn (2019) mostram que diferentes fatores dificultam o compartilhamento de dados de pesquisa. No nível individual, falta de tempo e o apoderamento de forma indevida dos dados pelos pesquisadores. Em nível institucional, a falta de treinamento de como compartilhar os dados, políticas internas desfavoráveis e escassez de remuneração contribuem para

o não compartilhamento dos dados de pesquisa. Em nível internacional, normas éticas e legais, políticas fracas, problemas com a infraestrutura dos dados e a sua interoperabilidade limitam o compartilhamento.

O estudo de Carvalho (2018) tinha como propósito identificar e analisar as práticas de comunicação de dados de pesquisa na Química, Antropologia e Educação. A autora destacou que existem diferenças em relação às práticas de produção/obtenção de dados de pesquisa entre os campos do conhecimento, na Química os dados quantitativos são predominantes, já na Antropologia são os dados qualitativos, enquanto na Educação, o tipo de dado varia conforme a natureza do fenômeno estudado.

Em relação ao compartilhamento de dados de pesquisa, os pesquisadores das três disciplinas analisadas por Carvalho (2018) se mostraram abertos ao compartilhamento. Essa prática ocorre principalmente entre os membros do próprio grupo de pesquisa, docentes, discentes e normalmente após a conclusão do estudo são disponibilizados os dados em livros, artigos, capítulos de livros e outros. Sobre a exigência do compartilhamento dos dados de pesquisa por parte das agências de fomento, editores de periódicos científicos e pelas instituições em que atuam, Carvalho (2018) observou diferentes reações dos pesquisadores em cada área, que vão desde não veem problema algum até não aceitarem/utilizarem os recursos.

No tocante do reuso de dados de pesquisa, Carvalho (2018) traz como resultado que na Química a prática de reuso dos dados não é comum e que entre os pesquisadores entrevistados, não existe intenção para tal prática, exceto na química bioinorgânica por realizar estudos colaborativos. Na Antropologia, o reuso dos dados ocorre entre pesquisadores de áreas similares, os entrevistados demonstram interesse em reusar dados, porém faltam incentivos financeiros do Estado e de agências de fomento à pesquisa. Na educação, os entrevistados acham importante o reuso dos dados de pesquisa, entretanto não se mostraram abertos para tal prática, exceto em relação ao próprio grupo de pesquisa.

O estudo de Shen (2015) visou analisar as práticas de compartilhamento e reutilização de dados de pesquisa por pesquisadores acadêmicos da *Virginia Tech* para compreender e desenvolver sistemas de infraestrutura, gerenciamento, preservação e curadoria dos dados de pesquisa e ainda a contribuição das bibliotecas

acadêmicas nesse movimento, com serviços para avaliação e compartilhamento dos dados.

Como procedimento metodológico, Shen (2015) usou um questionário com 32 perguntas. O questionário foi enviado em novembro de 2014 para 2.532 docentes e pesquisadores da *Virginia Tech*, ao qual obteve 423 respostas válidas.

Os resultados da pesquisa realizada por Shen (2015) mostram que:

1. 57% dos professores consideram que seus dados tem valor de longo prazo;
2. 45% indicam que seus dados não podem ser reproduzidos ou recoletados;
3. 44% relatam que seus dados podem ser reutilizados por pesquisadores da mesma área;
4. 40% consideram seus dados complexos e com inter-relações com outros conjuntos de dados;
5. 29% consideram que seus dados têm implicações importante para a política, sociedade ou negócios; e
6. 28% indicam que seus dados podem ser utilizados por pesquisadores de outras disciplinas.

Shen (2015) identificou em seu estudo que os dados têm sua reutilização muito limitada, existindo a troca de conjuntos de dados entre colaboradores ou contatos pessoais. Mais da metade dos pesquisadores acadêmicos não reutilizam dados de outros pesquisados ou produzidos em outras disciplinas.

Como se pode observar, mesmo os pesquisadores considerando que seus dados são aptos para serem reutilizados por outros investigadores, a prática em si do compartilhamento dos dados é baixa. Assim como Shen (2015) relata essa dificuldade dos pesquisadores de tornarem acessível os seus dados a outros pesquisadores, percepções parecidas são vistas nos trabalhos de Veiga *et al.* (2019) e Chawinga e Zinn (2019).

Questões relacionadas à falta de documentação e à falta de metadados descritivos dos dados são levantadas como barreiras para o compartilhamento e o reuso dos dados, mostrando-se necessário o gerenciamento dos dados e assistência aos pesquisadores para conseguirem passar de uma abordagem micro para uma macro de compartilhamento e reuso de dados. Como é apresentado por Sayão e

Sales (2019, p. 45), os dados de pesquisa normalmente são “[...] acompanhados de poucos metadados ou de nenhum e apresentam documentação pobre ou inexistente, que limita ainda mais o compartilhamento aos pesquisadores mais próximos e constitui um obstáculo decisivo para o reuso”

Nos estudos abordados neste tópico, os autores tratam de assuntos relacionados ao compartilhamento e reuso de dados em diferentes disciplinas, mostrando que apesar dos pesquisadores verem com bons olhos o reuso dos dados por terceiros, o ato de compartilhar não é uma prática comum, por vários motivos. Dessa forma, fatores levantados nos estudos acima como as práticas realizadas para compartilhar e reusar os dados de pesquisa juntamente com a percepção dos pesquisadores sobre essas rotinas, serão elementos fundamentais para compreender os costumes de produção, compartilhamento e reuso dos dados pelos pesquisadores na Genética.

4 METODOLOGIA

A metodologia é uma forma de se pensar e estudar a realidade social, como definem Strauss e Corbin (2008). Os métodos precisam ser organizados sistematicamente para que seja viável a realização e compreensão da pesquisa. Assim, este tópico apresenta os métodos e os procedimentos metodológicos que serão utilizados ao longo do trabalho com o propósito de responder o objetivo geral que é analisar as práticas de comunicação dos dados de pesquisa entre os pesquisadores brasileiros da Genética.

4.1 Caracterização da pesquisa

Para se desenvolver uma pesquisa, é necessário pensar nas estratégias de investigação para se alcançar os objetivos do trabalho. A pesquisa científica pode ter diferentes concepções filosóficas, Creswell (2010) em seu livro, apresenta as quatro posições de alegação do conhecimento, as quais são: pós-positivismo, construtivismo, reivindicatória/participatória e pragmatismo.

Segundo Creswell (2010) o construtivismo social é uma abordagem predominantemente qualitativa. Tem como principal alegação que os indivíduos têm entendimentos e experiências subjetivas sobre o mundo, propiciando significados

diferentes para as coisas, cabendo ao pesquisador decifrar esses pensamentos baseados no contexto em que estão inseridos.

Com base nas alegações do conhecimento de Creswell (2010) e no objetivo geral desta pesquisa “analisar as práticas de comunicação dos dados de pesquisa entre os pesquisadores brasileiros da Genética”, foi adotado o construtivismo social. Isso porque o construtivismo busca analisar o entendimento do sujeito sobre ele mesmo e o seu meio, baseado nas construções sociais, que é justamente o que a pesquisa busca ao compreender a visão dos pesquisadores sobre a comunicação dos dados de pesquisa.

Além disso, uma pesquisa pode ter aspectos exploratórios, descritivos e/ou explicativos. Levando em conta que o objetivo da pesquisa é identificar e apresentar a percepção dos pesquisadores da Genética sobre a produção, compartilhamento e uso/reuso dos dados de pesquisa, o estudo apresenta características descritivas. Como explica Creswell (2010) um estudo tem características descritivas quando apresenta uma nova visão de um assunto já conhecido. De acordo com Kauark, Manhães e Medeiros (2010) os estudos que têm características descritivas costumam assumir a forma de levantamento. Corroborando com os autores supracitados, Walliman e Baiche (2001) elucidam que a pesquisa descritiva pode ser observada por diferentes meios, como entrevistas, percepções visuais, aplicação de questionário entre outras formas de coleta de dados. Deste modo, este estudo analisa e descreve os fenômenos da comunicação dos dados de pesquisa na Genética.

A natureza dessa pesquisa é qualitativa, que conforme Creswell (2010, p. 26), corresponde a “um meio para explorar e para entender o significado que os indivíduos ou grupos atribuem a um problema social ou humano”. A pesquisa qualitativa se utiliza de várias características próprias, usa texto ao invés de números, como é o caso da pesquisa quantitativa, (FLICK, 2009). As abordagens qualitativas podem se dar de diferentes formas, visando ver o mundo por diferentes ângulos. Esse tipo de pesquisa

[...] consiste em um conjunto de práticas interpretativas e materiais que tornam o mundo visível. Essas práticas transformam o mundo, fazendo dele uma série de representações, incluindo notas de campo, entrevistas, conversas, fotografias, gravações e anotações pessoais (DENZIN; LINCOLN, 2005, p. 3, tradução nossa).

Nesse sentido, de acordo com Strauss e Corbin (2008, p. 24), a pesquisa qualitativa tem como característica analisar os dados por meio do "processo não matemático de interpretação, feito com o objetivo de descobrir conceitos e relações nos dados brutos e de organizar esses conceitos e relações em um esquema exploratório teórico". Assim, a natureza da pesquisa que é qualitativa está de acordo com as discussões metodológica do trabalho, uma vez que se busca compreender o fenômeno através da percepção dos pesquisadores "[...] por meio da precisão dos detalhes, ela fornecerá informações contextuais que poderão servir de base para pesquisas explicativas mais desenvolvidas" (DESLAURIERS; KÉRISIT, 2014, p. 130).

Assim, a pesquisa tal como foi apresentada aqui visa atender ao objetivo geral da pesquisa que é "analisar as práticas de comunicação dos dados de pesquisa entre os pesquisadores brasileiros da Genética", entende-se que essas três etapas, produção, compartilhamento e reuso de dados, apresentados extensivamente na revisão de literatura, são os três macros processos que compõem a comunicação dos dados de pesquisa, objeto de estudo desse trabalho.

Neste sentido, entende-se a produção de dados de pesquisa como qualquer registro produzido/obtido por meio de experimento, observação ou sistema de computador que seja utilizado com a finalidade de pesquisa (NATIONAL SCIENCE BOARD, 2005; SAYÃO; SALES, 2014; COSTA; LEITE, 2018; VEIGA, 2019). Já o compartilhamento de dados é compreendido como insumo que pode estar em formato bruto ou processado disponibilizado em banco de dados ou compartilhado entre os pares (SAYÃO; SALES, 2019; FIOCRUZ, 2019). Por fim, o reuso de dados de pesquisa é entendido aqui como a reutilização de conjuntos de dados produzidos/obtidos por outros pesquisados para o desenvolvimento de novos estudos (SANTAREM SEGUNDO, 2020; MEDEIROS; CAREGNATO, 2012).

4.2 Procedimentos metodológicos

Para possibilitar que a questão de pesquisa seja respondida e que o objetivo geral seja alcançado, torna-se necessária a descrição do passo a passo dos procedimentos que irão conduzir este estudo. Nesta seção serão abordados a amostragem da pesquisa, o método, a fonte dos dados, a técnica de coleta e a análise dos dados.

4.2.1 Amostragem da pesquisa

Com a finalidade de cumprir o objetivo geral da pesquisa, que é “analisar as práticas de comunicação dos dados de pesquisa entre os pesquisadores brasileiros da Genética”, delineou-se que o universo de análise será constituído por bolsistas de produtividade em pesquisa do CNPq cadastrados na Plataforma Carlos Chagas⁹. A escolha da Plataforma Carlos Chagas se deu por permitir filtrar os pesquisadores por região brasileira e por diferentes níveis de bolsa de produtividade para ter uma maior cobertura da população.

Dessa forma, a amostragem é composta por 12 Bolsistas de Produtividade da Genética cadastrados na Plataforma Carlos Chagas, por entender que são os pesquisadores que mais comunicam seus dados.

A Plataforma Carlos Chagas registra atualmente 267 pesquisadores¹⁰ do campo da Genética como bolsistas de alta produtividade no Brasil, sendo divididos em níveis de produtividade em que pesquisadores 1A são considerados os mais produtivos. Desse total de pesquisadores 26 são pesquisadores 1A, 29 são pesquisadores 1B, 41 pesquisadores são 1C, 38 são pesquisadores 1D, 128 pesquisadores são nível 2 e 5 pesquisadores encontram-se com suas bolsas interrompidas por motivo de estágio ou pós-doutorado no exterior.

A maior quantidade de pesquisadores se concentra na região Sudeste (159), seguido das regiões Sul (56), Nordeste (21), Norte (18) e Centro-Oeste (13). Sendo que desse total 203 são docentes/pesquisadores de Faculdades/Universidades Públicas, 41 são docentes/pesquisadores de Instituto/Centros de pesquisa, 17 são docentes/pesquisadores de Universidades/Faculdades Particulares e 6 são pesquisadores em Hospitais. A Tabela 9 sintetiza as informações aqui apresentadas.

Tabela 9 - Quantidade de pesquisadores por estado e local de atuação.

Quantidade de pesquisador por região	Norte	Nordeste	Sul	Sudeste	Centro-Oeste
	18	21	56	159	13
Docente/Pesquisador Universidade pública	203				

⁹ Disponível em: <http://carloschagas.cnpq.br/>

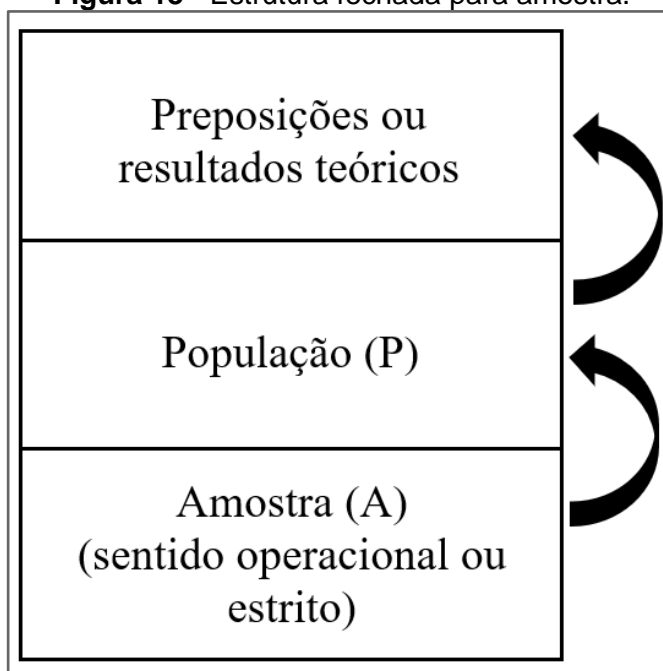
¹⁰ Levantamento feito em 16 de setembro de 2020.

Docente/Pesquisador Instituto/Centro de pesquisa	41
Docente/Pesquisador Faculdade/universidade particular	17
Pesquisador em Hospital	6

Fonte: Elaboração própria.

A seleção da amostra será caracterizada por uma estrutura fechada ou como também é conhecida, convencional. Nesta estrutura o pesquisador retira uma amostra bem definida, já que é complexo alcançar todos os integrantes que compõem a população (PIRES, 2014). Dessa forma, “a população também pode ser relacionada a um conjunto de objetos ou informações. [...] Amostra diz respeito a um subconjunto da população, fração ou uma parte do grupo” (SILVA, 2018, on-line). Pires (2014) Hierarquiza a estrutura fechada para a escolha da amostra conforme apresentado a Figura 13.

Figura 13 - Estrutura fechada para amostra.



Fonte: Pires (2014, p. 162)

A abordagem adotada neste trabalho é a da amostragem intencional não probabilística, devido às suas características qualitativas, enfatizando “[...] mais as relações entre a amostra e o objeto, do que as regras técnicas de amostragem” (PIRES, 2014, p. 157). De acordo com Deslauriers e Kèrisit (2014), a amostragem não probabilística, também conhecida como teórica, permite que o pesquisador tenha

acesso as histórias de vida dos participantes que vão além do fenômeno observado e que podem influenciar suas ações, propiciando um aprofundamento nos processos envolvidos na comunicação dos dados de pesquisa. Os autores destacam ainda que

Se a regularidade e a dimensão da amostra probabilística nos possibilitam conhecer aspectos gerais da realidade social, o caráter exemplar e único da amostra não probabilística nos dá acesso a um conhecimento detalhado e circunstancial da vida social. É, pois, em relação aos resultados que ela acarreta, bem como à sua pertinência, que a amostra não probabilística se justifica. (DESLAURIERS; KÉRISIT, 2014, p. 139).

Na Plataforma Carlos Chagas é possível realizar a busca por pesquisadores por diferentes parâmetros, um deles é a busca por “estado”. Atualmente, os 267 bolsistas de alta produtividade do campo da Genética estão distribuídos em 16 estados brasileiros conforme apresentado na plataforma, para a composição da amostragem foram selecionados docentes/pesquisadores da Genética que são bolsistas de produtividade do CNPQ por serem mais reconhecidamente envolvidos com as pesquisas, portanto de acordo com as características da amostragem não probabilística, os bolsistas de produtividade são os principais sujeitos para debater esse fenômeno.

Para discutir a comunicação dos dados de pesquisa à luz da Ciência Aberta, a amostra tem como intenção de observar as variadas percepções dos pesquisadores em diferentes contextos de trabalho de acordo com o tipo de instituição em que atua e suas práticas de produção, compartilhamento e reuso dos dados. Além disso, acredita-se que com a seleção dos bolsistas de produtividade do CNPQ, será possível identificar as práticas de comunicação de dados de pesquisa na Genética, visto que eles são os que mais produzem no país.

Portanto, a amostra também será do tipo estratificada. Flick (2010) descreve a amostra estratificada como um procedimento que tem como propósito representar subgrupos de forma mais focada e a partir disso extrair semelhanças entre os participantes. Campbell (1986) chama essa simetria dos participantes de “similaridade proximal”, em conformidade com as características da amostra estratificada será levado em conta a área de concentração da sua pesquisa e a região de atuação do docente/pesquisador. A obtenção desses dados ocorreu por meio do Currículo Lattes dos pesquisadores. De acordo com requisitos apresentados, foram convidados inicialmente para a entrevista 1 pesquisador de cada estado brasileiro que estão

contemplados na Plataforma Carlos Chagas, totalizando 16 pesquisadores. A seleção dos pesquisadores ocorreu de forma aleatória.

No dia 4 de junho de 2021 foi enviado por e-mail o convite aos 16 pesquisadores selecionados inicialmente convidando-os a participarem da entrevista. Desses 16 convites enviados via e-mail, apenas um respondeu positivamente. Durante o período de 7 a 22 de junho foram feitas novas seleções de pesquisadores conforme o registro na Plataforma Carlos Chagas. Foram ao todo enviado 80 convites, uma média de 16 convites por estado. Desse total, 8 responderam negativamente, 15 aceitaram participar, porém três pesquisadores não compareceram na data escolhida para a entrevista e não retornaram aos e-mails seguintes.

Após a aplicação do procedimento descrito acima, temos uma amostra composta por 1 pesquisador da região Centro-Oeste, 4 pesquisadores da região Sudeste e 7 pesquisadores da Região Sul. É importante ressaltar que o quantitativo de pesquisadores referente a região Sul é maior do que as demais regiões, pelo fato dos pesquisadores da região Sul terem aceitado participar da entrevista durante o período escolhido para a coleta de dados.

Dessa forma, os dados coletados para a realização dessa pesquisa foram levantados durante o dia 16 de junho a 17 de julho de 2021. O tempo médio das entrevistas foi de 47 minutos, sendo que a menor entrevista durou 27 minutos e a maior 1 hora e 26 minutos. Buscou-se nas entrevistas compreender as práticas de comunicação dos dados de pesquisa a partir das percepções dos bolsistas de produtividade do CNPq. Todas as informações coletadas foram utilizadas para fins de pesquisa e as identidades dos entrevistados foram preservadas.

A Tabela 10 apresenta de forma sucinta a descrição dos pesquisadores entrevistados neste estudo, são apontados os seguintes aspectos: a área de concentração da pesquisa, o estado em que reside; o tipo de instituição em que atua e o extrato ao qual o pesquisador pertence de acordo com a Plataforma Carlos Chagas. As informações expostas abaixo têm o intuito de mostrar a cobertura das entrevistas.

Tabela 10 - Descrição dos pesquisadores.

Pesquisadores	Área de concentração	Estados	Instituições	Nível
----------------------	-----------------------------	----------------	---------------------	--------------

Pesquisador 1	Biologia molecular de baculovírus	Centro-Oeste	Instituição de ensino pública	1C
Pesquisador 2	Genética Evolutiva Molecular	Sul	Instituição de ensino privada	1C
Pesquisadora 3	Biologia molecular do desenvolvimento - pesquisa básica	Sudeste	Hospital	1A
Pesquisador 4	Análise em larga escala de genomas, transcriptomas e proteomas	Sul	Instituto de pesquisa	2
Pesquisadora 5	Genética Vegetal	Sul	Instituição de ensino pública	2
Pesquisadora 6	Mutagênese ambiental e humana	Sul	Instituição de ensino pública	1C
Pesquisador 7	Neurogenética	Sul	Hospital	2
Pesquisador 8	Desenvolvimento e aplicação de Microarranjos de DNA	Sudeste	Instituição de ensino pública	2
Pesquisador 9	Genética de populações de moscas	Sudeste	Instituição pública	1B
Pesquisadora 10	Metabolismo de DNA em plantas	Sudeste	Instituição de ensino pública	1B
Pesquisador 11	Genética Evolutiva Molecular	Sul	Instituição de ensino privada	1C
Pesquisadora 12	Risco Ocupacional	Sul	Instituição de ensino pública	1D

Fonte: Elaboração própria.

O recorte amostral apresentado na Tabela 10 pode ser justificado pela busca da compreensão das atividades de produção, compartilhamento e uso/reuso dos dados de pesquisa através das experiências e práticas vivenciadas pelos pesquisadores ao longo dos anos na pesquisa. Além disso, como aponta Flick (2009, p. 47), a amostra deve ser representativa, porém não necessariamente de forma estatística, mas “representar a relevância do fenômeno que queremos estudar em termos de experiência e envolvimento dos participantes de nossa pesquisa com esses fenômenos”.

4.2.2 Métodos e técnicas para coleta e análise dos dados

Na pesquisa qualitativa, Zelditch (1969) aponta dois principais critérios para julgar e validar os instrumentos utilizados para coletar os dados. O primeiro é sobre a capacidade dos instrumentos escolhidos trazerem as informações pretendidas pelo pesquisador, ou seja, instrumentos que irão possibilitar coletar a maior quantidade possível de informações sobre o tema pesquisado. O segundo é a efetividade desses instrumentos.

Neste trabalho, por ter um aspecto descritivo, o método adotado foi o de levantamento (*survey*). Para Bell (2007) o método de levantamento tem como objetivo apresentar informações sobre uma amostra considerável da população a fim de conseguir tirar conclusões que representem a população integralmente. Esse método pode ser aplicado a diferentes assuntos e combinado com outros métodos de pesquisa (BABBIE, 2003).

No método levantamento, o pesquisador leva em conta as características de sua amostra e a natureza dos dados que serão coletados por ele. Para Tomanik (2004) os dados podem ser obtidos de inúmeras formas, por exemplo, por meio de questionários, observações, entrevistas, entre outros. Dentre as opções supracitadas, Kvale (1996) define que uma das técnicas mais utilizadas para coletar dados é o de questionários e entrevistas.

Neste estudo é adotada a entrevista semiestruturada como técnica de coleta de dados da pesquisa, visto que, como expõe Poupart (2008), as entrevistas são aplicadas aos indivíduos com o objetivo de estudar o ponto de vista deles sobre determinado contexto. Por esse motivo, essa técnica se mostra a mais adequada à medida que se procura obter informações de um determinado grupo.

Para a aplicação das entrevistas semiestruturadas, neste estudo, foi utilizado como base o roteiro de entrevista que foi construído pela Érika Rayanne Silva de Carvalho (Anexo 1) e aplicado em sua pesquisa de mestrado em 2018. A utilização do roteiro de entrevista de Carvalho (2018) justifica-se na uniformidade das perguntas para a coleta de dados e assim, permitir futuramente a realização de estudos comparativos relacionados as práticas de comunicação de dados de pesquisa nas diferentes áreas do conhecimento brasileiras.

Carvalho em sua pesquisa buscou compreender as atividades de produção, compartilhamento e uso/reuso de dados de pesquisa dos pesquisadores das áreas das Ciências Exatas, Ciências Sociais e Humanidades da Universidade de Brasília (UnB). Para que essas três áreas fossem contempladas, Carvalho (2018) entrevistou pesquisadores das disciplinas de Química, Antropologia e Educação da Universidade.

Portanto, o uso do mesmo instrumento de coleta de dados aplicado por Carvalho (2018) em seu estudo será valioso a esta pesquisa, pois possibilita observar o mesmo fenômeno em outra área do conhecimento que é a Genética, apresentando mais informações pertinentes que viabiliza comparar os estudos e perceber as semelhanças e diferenças entre as áreas nos dois estudos em relação a tais práticas, além de enriquecer o debate sobre a produção, compartilhamento e uso/reuso de dados entre os vários campos do conhecimento a luz da Ciência Aberta.

Creswell (2010) salienta a possibilidade de extrair significados dos dados por meio do processo analítico contínuo, que acontece desde a organização dos dados até a sua observação. A análise dos dados obtidos a partir da coleta é uma etapa essencial de uma pesquisa. Para Creswell (2010) e Flick (2009), a melhor maneira de se analisar dados provenientes de entrevistas é por meio da codificação e da categorização. Rossman e Rallis (1998, p. 171, tradução nossa) explicam que “a codificação é o processo de organização do material em blocos ou segmentos de texto antes de atribuir significado às informações”.

A codificação “[...] é aqui entendida como representação das operações pelas quais os dados são fragmentados, conceitualizados e reintegrados de novas maneiras” (CRESWELL, 2010, p. 277). A codificação possui categorias de informação, sendo elas aberta, axial e seletiva (FLICK, 2009; STRAUSS; CORBIN, 2008; CRESWELL, 2010).

A codificação aberta tem o propósito de separar os dados em grupos de palavras ou expressões curtas, as quais têm conceitos associados a eles. Strauss e Corbin (2008, p. 103) descrevem a codificação aberta como “[...] processo analítico por meio do qual os conceitos são identificados e suas propriedades e suas dimensões são descobertas nos dados”.

Logo após a coleta dos dados, foi utilizado o software *Transcribe* para a transformação dos áudios em textos, em seguida, para auxiliar nas etapas da

codificação e da categorização, optou-se pelo uso do software NVivo (Starter) para análise de dados qualitativa. A coleta e análise dos dados coletados seguiu os procedimentos descritos na Tabela 11.

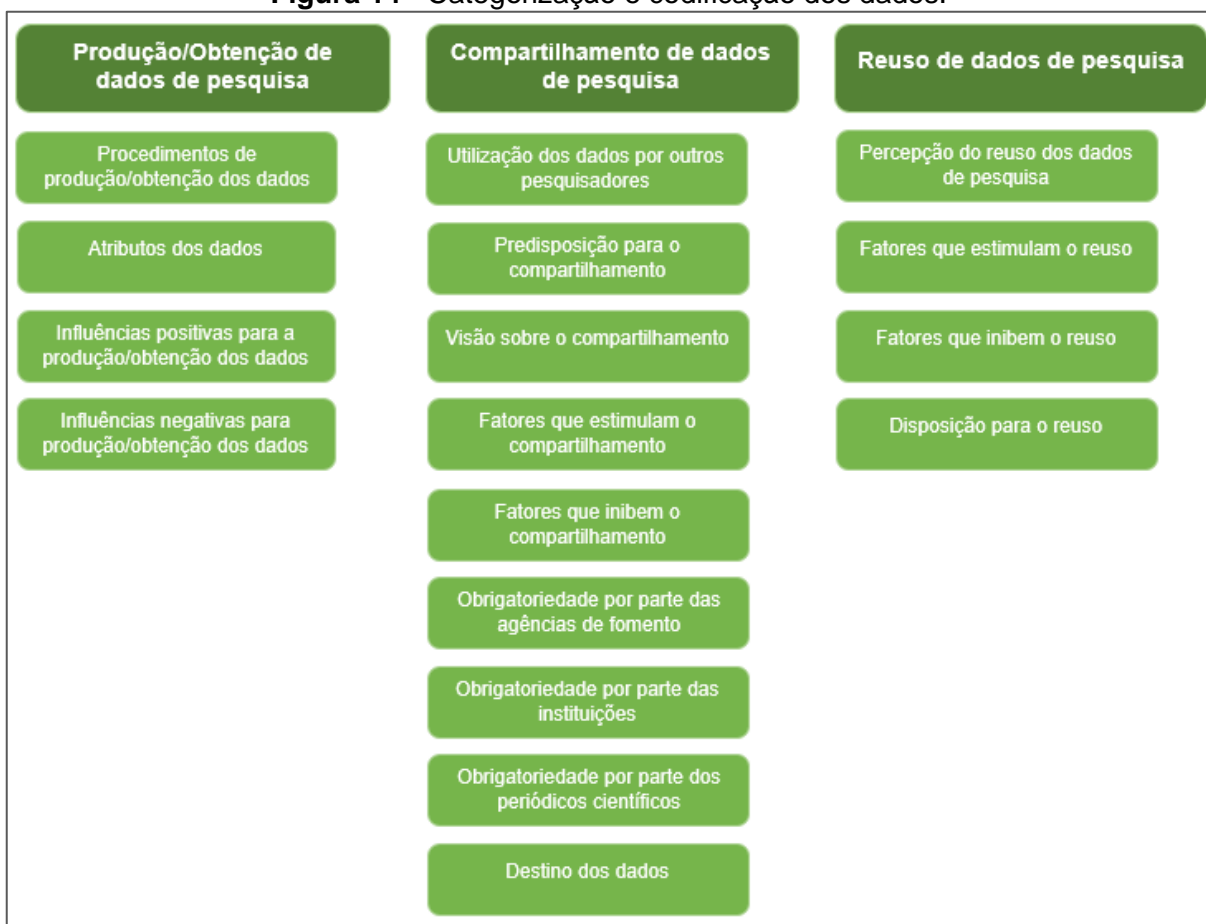
Tabela 11 - Procedimentos de coleta de dados.

Coleta de dados	Degração	Codificação e categorização
Entrevistas	Software <i>Transcribe</i>	Software NVivo

Fonte: Elaboração própria.

Uma vez que as entrevistas foram transcritas, foi possível realizar a codificação das falas dos pesquisadores conforme os assuntos abordados durante a entrevista semiestruturada (Anexo 2). Como pode ser observado abaixo (Figura 14), encontram-se os temas que foram identificados.

Figura 14 - Categorização e codificação dos dados.

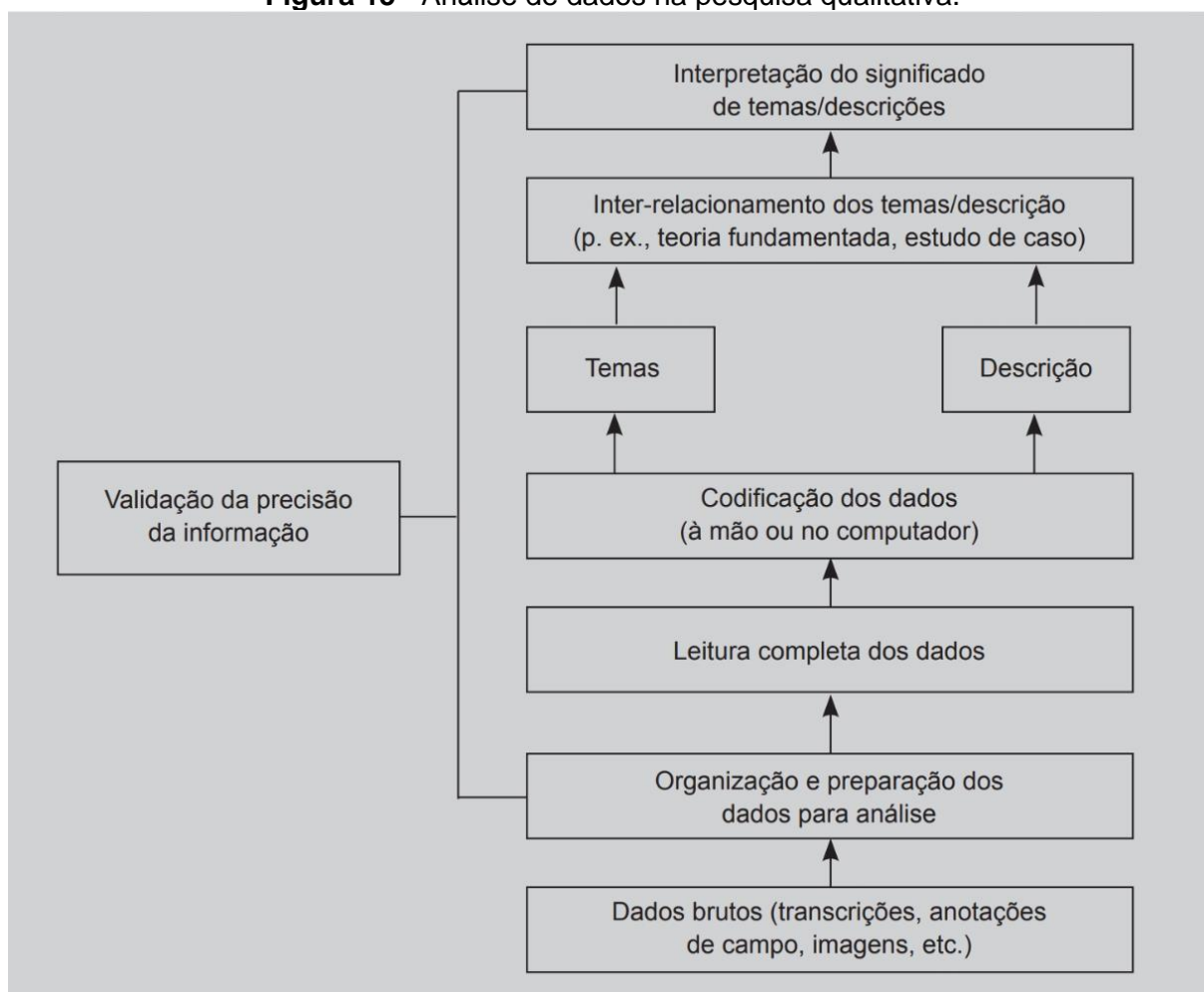


Fonte: Elaboração própria

Juntamente às fases descritas acima, a análise de dados é um ponto muito relevante para o processo analítico na pesquisa qualitativa, como Creswell (2010)

apresenta as etapas aplicadas a análise de dados na pesquisa qualitativa, a qual segue uma estrutura hierárquica em sentido crescente, ou seja, de baixo para cima (Figura 15). De acordo com o próprio autor, esse modelo nem sempre conserva essa estrutura. Levando em conta que cada pesquisa é única, esse processo pode sofrer variações.

Figura 15 - Análise de dados na pesquisa qualitativa.



Fonte: Creswell (2010, p. 218)

Com os dados coletados, a análise ocorreu seguindo os seguintes passos (CRESWELL, 2010):

1. Organização e preparação dos dados para a análise: organização e estruturação dos dados coletados.
2. Leitura dos dados: compreensão geral das informações (ideias, perspectivas, reflexões) por meio dos dados.

3. codificação dos dados: processo que reúne as informações em blocos com o intuito de atribuir significados a eles. Se utiliza normalmente da linguagem natural.
4. Uso da codificação para gerar descrição e categorias: exposição detalhada de local, lugares ou pessoas. O uso da codificação gera categorias que devem abarcar os diferentes entendimentos sobre os participantes.
5. Apontar como a descrição e as categorias serão apresentadas narrativa: é o modo de apresentação dos resultados. Normalmente se utiliza o modo narrativo na discussão, expondo os resultados de forma cronológica conforme os acontecimentos ou interconectando os temas.
6. Compreensão dos dados: envolve a extração de significado dos dados pelo pesquisador. Neste caso, existe a comparação dos resultados encontrados com a literatura científica ou teoria.

Creswell (2010) ressalta a legitimação da informação, que acontece nas fases de organização, preparação e interpretação dos dados. Dado o cunho analítico da pesquisa, reitera a relevância da categorização e codificação dos dados. Uma vez que as entrevistas estão categorizadas e codificadas, é possível tornar textos longos em categorias teóricas, possibilitando ter uma visão mais abrangente do fenômeno em observação.

A Tabela 12 exibe os objetivos apresentados ao longo desta pesquisa, mantendo relação direta com os procedimentos metodológicos adotados, sendo eles: o método, a amostra, os instrumentos, a técnica de coleta e a técnica de análise.

Tabela 12 - Relação entre os objetivos e procedimentos metodológicos adotados.

Objetivos específicos	Identificar práticas de produção de dados de pesquisa no campo da Genética	Identificar práticas de compartilhamento de dados de pesquisa no campo da Genética	Identificar práticas de reuso de dados de pesquisa no campo da Genética
Método	Levantamento		
Fonte dos dados	Pesquisadores da Genética		
Técnica de coleta	Entrevista semiestruturada		
Técnica de análise	Categorização e codificação		

Fonte: Elaboração própria.

5 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Esta seção apresenta a análise dos dados e os resultados da pesquisa realizada através das entrevistas com os bolsistas de produtividade da área da Genética do CNPq. Serão tratados aqui questões relacionadas às práticas de produção, compartilhamento e reuso dos dados de pesquisa, contemplando toda a comunicação dos dados.

5.1 Levantamento

De acordo com o descrito na seção de metodologia, este trabalho se utiliza da técnica de levantamento, esta modalidade de método tem como objetivo a interrogação direta de um grupo previamente selecionado ao qual se deseja conhecer. Para alcançar o objetivo geral da pesquisa, será executado o roteiro de entrevista elaborado por Carvalho (2018).

Para isso, as perguntas contemplam os três principais aspectos da comunicação dos dados de pesquisa: a produção, o compartilhamento e o reuso, que são o foco da pesquisa. O roteiro da entrevista semiestruturada encontra-se no Anexo 1.

5.1.1 Produção/obtenção de dados de pesquisa

Esta seção aborda questões relativas às técnicas de produção/obtenção de dados de pesquisa, tais como métodos utilizados para produção/obtenção de dados, as características desses dados e os aspectos que influenciam os pesquisadores na produção/obtenção de dados da Genética.

Realização da pesquisa científica na Genética

A Genética é um guarda-chuva que engloba diferentes disciplinas que possuem características distintas na realização de pesquisa em seu campo, podendo ainda assim encontrar ramificações dentro delas. Dessa forma, entender como a pesquisa ocorre na Genética e em algumas de suas ramificações é de suma importância para o desenvolvimento desta pesquisa. Assim, observou-se que os pesquisadores da área de Genética podem trabalhar em diferentes linhas de pesquisa, como: bioinformática (campo interdisciplinar que trabalha com grandes quantidades de dados biológicos e se utiliza da informática para realizar as análises),

pesquisa experimental (se utiliza de ambiente controlado para a realização de testes em organismos e assim coletar o máximo de informações sobre ele), pesquisa básica (aplicação do conhecimento pelo conhecimento, sem necessariamente ter uma finalidade imediata), pesquisa clínica (estuda novos tratamentos e exames, além de avaliar seus efeitos na saúde humana ou animal). As declarações dos pesquisadores mostram que o campo da Genética é fértil possuindo diferentes linhas de pesquisa para se atuar.

É praticamente uma pesquisa científica laboratorial, então eu tenho que produzir dados científicos no laboratório por experimentos (Pesquisador 1)

Nós fazemos pesquisa em laboratório né, pesquisa básica e em alguns casos pesquisa pré-clínica ou desenvolvimento inicial de algum produto, mas basicamente, pesquisa laboratorial (Pesquisador 2)

Eu trabalho em uma área de pesquisa que se chama bioinformática. (...) Então no meu grupo de pesquisa a gente trabalha fundamentalmente com dados do ser humano, então pode ser tanto dado de genoma tanto dado de transcriptoma ou dado de proteoma. (Pesquisador 4)

A gente atua em pesquisa clínica. (...) Uma porção menor do que eu faço, mas tem outros colegas que fazem mais é o que a gente chama de pesquisa básica em animais ou em modelos celulares. (Pesquisador 7)

Toda pesquisa aqui é empírica né, a gente levanta uma hipótese, a partir dessa hipótese traçamos possíveis objetivos e aí dentro desses objetivos a gente investiga empiricamente para tentar chegar em uma resposta. (Pesquisador 8)

A pesquisa na Genética se dá de diferentes formas, como pode ser observado na fala dos entrevistados, os estudos podem ser desenvolvidos em laboratórios, em sistemas de computadores, de forma empírica ou em nível básico. Esse resultado vai ao encontro dos estudos de Sidler (2014), de acordo com o autor, quanto mais uma disciplina se aproxima das áreas exatas ou rígidas, os insumos para o desenvolvimento de pesquisas são predominantemente obtidos por meio de laboratórios e/ou computadores.

Como pode ser observado nos resultados, existem especificidades na realização da pesquisa dentro da própria Genética. Neste sentido, Knorr-Cetina (1999) oferece um olhar das culturas epistêmicas, analisando as práticas presentes no processo de produção do conhecimento científico. Que tal diversidade cultural causa divisões dentro das ciências, que se refletem nos costumes das disciplinas e entre os grupos de pesquisa ao qual os pesquisadores fazem parte. A autora ainda relata que essas culturas epistemológicas são responsáveis pela reconfiguração dos

pesquisadores em suas áreas, moldando os em relação aos tipos de procedimentos, tecnologias e métodos praticados.

Produção/obtenção de dados de pesquisa

As práticas de produção/obtenção de dados dependem da linha de pesquisa a qual o pesquisador faz parte. Entre os entrevistados, a maioria produz os seus dados em laboratório.

A nossa pesquisa é muito de bancada mesmo tá, então tem toda uma parte de extração de ácidos nucleicos, de sequenciamento de proteômica, muita abordagem em larga escala como a gente chama e a gente obtém uma quantidade de dados muito grandes e esses dados são normalmente são analisados em programas específicos, em softwares específicos de análise tá. (Pesquisadora 3)

Bem, a produção de dados é mesmo a partir da experimentação né, no laboratório, tudo no laboratório. (Pesquisadora 6)

Em laboratório, totalmente em laboratório. A gente tem biotério próprio, produtos animais, os trabalhos são submetidos ao [nome do comitê], comitê de ética no uso de animais para aprovação e aí fica tudo dentro da Universidade. São experimentos de bancada mesmo de tratamento, modelos animais e aí se coleta amostras de sangue e tecidos para fazer análises moleculares. (Pesquisadora 12)

Nesse mesmo contexto da produção/obtenção de dados, foi possível observar que mesmo que a grande parte dos dados sejam produzidos em laboratório, os pesquisadores costumam obter dados de pesquisa associados a outros métodos de coleta, através de pesquisa de campo e por meio de softwares.

Trabalho laboratorial, in vitro... A gente faz uma parte também de informática, atrelados a dados em larga escala também. (Pesquisador 2)

A produção dos dados... os dados brutos eles são feitos ou através do sequenciamento de nova geração (...) Você faz o preparo em bancada molhada e coloca esse material preparado dentro do sequenciador e o resultado é um arquivo um monte, milhões de sequências de trechos de sequenciamento da sequência de DNA que você colocou pra fazer o sequenciamento, e aí vem uma parte de análise bioinformática. (Pesquisador 4)

A gente faz saída de campo, a gente trabalha em laboratório, então a gente faz coletas em campo, fazemos atividades no laboratório e também bastante atividade via computador né, bioinformática também. (Pesquisadora 5)

Então, eu trabalho nos três segmentos, né. Quando precisa a gente faz campo. Aí nesses trabalhos de campo a gente coleta as amostras que serão processadas em laboratório. Aí uma vez coletada a gente vai pro laboratório e tenta validar algumas questões que foram hipotesadas e objetivadas, e muitas das vezes antes de fazer esse trabalho, esse esforço, a

gente submete a uma análise preliminar do ponto de vista computacional para tentar entender se tudo isso é viável. (Pesquisador 8)

A produção dos dados pode ocorrer através de diferentes métodos, corroborando com essa ideia, Sayão e Sales (2014) discorrem sobre as formas de obtenção de dados, de acordo com os autores, os processos utilizados para a produção dos dados varia de grupo de pesquisa para outro. Porém, Sayão e Sales (2014) pontuam, que independente dos métodos utilizados para a produção/obtenção do dado de pesquisa, o seu verdadeiro valor está nas oportunidades geradas pelo seu reuso e em novas interpretações em áreas diferentes da que gerou os dados.

No caso da Genética, a natureza dos dados abrange os três tipos que é amplamente citado na literatura científica, sendo elas a observacional, computacional e experimental, dependendo da disciplina, há dados das três naturezas. É possível inferir por meio dos resultados que o tipo de técnica utilizada para a obtenção dos dados está ligado às práticas de produção de cada grupo. Neste sentido, Borgman (2012) reforça o entendimento sobre a produção de dados nas ciências da vida, em que a sua maioria é coletada pelos próprios pesquisadores através da observação, experimento e/ou modelos.

Os pesquisadores relataram que nem sempre são eles que produzem os dados de pesquisa. Em algumas situações, o material é enviado para uma empresa terceirizada para realizar a produção dos dados.

Nós temos aqui máquinas que fazem o sequenciamento, no caso, para sequenciamento de ácido nucléico. Algumas vezes eu encaminho o ácido nucleico para uma empresa privada fazer o sequenciamento. Eles mandam os dados brutos, aí os dados brutos nós mexemos, nos dados brutos, e analisamos os dados ou eu posso até pagar para empresa fazer análise dos dados e já dar os dados prontos analisados por eles. Mas nós normalmente recebemos dados brutos e fazemos análises. (Pesquisador 1)

Tudo tem relação com custo/benefício. Então olha só, se tem um determinado equipamento que a gente tem aqui no nosso laboratório, pra gente fazer uma análise vai custar 3 mil reais, e essa mesma análise é feita por uma empresa a mil reais, certamente a gente vai pagar para essa empresa fazer. Aí você vai falar assim “mais como? O equipamento está lá disponível”. A questão é que eles estão cobrando muito menos do que a gente teria para fazer o equipamento funcionar, entende? Aí como o nosso investimento é escasso, qualquer economia que aparecer no processo a gente apoia. (Pesquisador 8)

O sequenciamento é terceirizado, só. Até a coleta a gente faz... o sequenciamento é terceirizado. E aí volta a sequência e aí a gente continua a análise. (Pesquisador 9)

A produção dos dados de pesquisa, como foi relatado por alguns pesquisadores, pode ser feito por empresas terceirizadas, em outros casos até mesmo a análise dos dados. Como apontado pelo Pesquisador 8, essa prática tem relação com o custo/benefício na realização da atividade, levando em conta o gasto dos recursos escassos.

Pode-se supor que essa busca de economia de recursos de acordo com o que foi exposto pode estar ligada com questões relacionadas ao fomento à pesquisa que vem diminuindo gradativamente ao longo dos anos, principalmente nos últimos dois anos, devido a pandemia causada pelo Coronavírus (FAPESP, 2021; UNESCO, 2020; POLO; TAVARES, 2019; IPEA, 2020). Portanto, o montante de recursos para a pesquisa influencia, também, nos modos de produção/obtenção de dados de pesquisa.

Características dos dados produzidos/obtidos na pesquisa

Os dados produzidos/obtidos na Genética possuem características diferentes conforme a linha de pesquisa ou projeto em que o pesquisador está envolvido. Nesse contexto os dados podem ser arquivos computacionais, fotografia, imagens, figuras, planilhas em formato TXT, vídeos e texto, dependendo do método utilizado.

Eu faço agora o sequenciamento em uma máquina chamada de sequenciamento de última geração que ela vai agora sequenciar todos os ácidos nucleicos e vai produzir informações de milhares, milhões de fragmentos de ácido nucleico isso é guardado como programa como arquivos e esses arquivos agora são tratados em programas de computador (...) Esse é um tipo de dados gerados que é um dado computacional (...) Outros dados são dados de microscopia (...) Tudo isso é fotografado e eu posso fazer ensaios bioquímicos pra saber o que está sendo produzido. Também tem vários tipos de dados que você tira, dados genômicos, dados bioquímicos e dados morfológicos. São vários dados diferentes (Pesquisador 1)

Todo tipo. Tabela de excel, imagens, microscopia... Bom depois dados de escala, aí são arquivos especiais (Pesquisador 2)

A grande maioria são planilhas de Excel né. Os dados clínicos patológicos dos pacientes são planilhas de Excel que a gente faz tabelas com médias né, que a gente usa para botar de cada paciente. Depois os dados brutos eles são analisados, cada um em um tipo de programa, eles geram figuras, não é fotos, é figuras. (Pesquisadora 3)

Em termos de bioinformática a gente recebe as informações sempre no formato texto, a partir do formato texto a gente faz o processamento. (Pesquisador 4)

Depende muito, porque a linha ela é ampla, mas assim, basicamente se usa bastante Excel pra ver se os dados são quantitativos. (...) Então a gente usa o Excel e aí transforma essas planilhas dependendo do programa que a gente vai utilizar para fazer análise né num outro formato, mas eles são tabuladas no Excel e depois de transformados em outros formatos dependendo do programa que a gente usa. Mas também fotos né, então a gente é. e figuras né, basicamente isso. Programas que lidam com a questão das figuras, Excel, Word. (Pesquisadora 5)

Então depende dos dados, quando a gente fala de metadado quase sempre são planilhas, agora se a gente vai fazer alguma análise de tecido ou de perfil crescente de microbiano aí geralmente são imagens ou vídeos. (Pesquisador 8)

Como observado, os dados têm características distintas, o que resulta em diferentes formatos e suportes na utilização para a sua guarda. Entre as respostas dos pesquisadores, em suas áreas, os dados costumam se apresentar em 5 formatos principais como pode ser visto na Tabela 13. Através dos resultados, foi possível identificar uma predominância de dados numéricos, multimídia e softwares.

Tabela 13 - Características dos dados.

Formato	Suporte
Texto	Docs, PDF
Numérico	XLSX, SPSS
Multimídia	JPEG, PNG
Softwares	Perl, Python, Fasta, R
Instrumento-Específico	Digital Microscopic

Fonte: elaboração própria.

Podemos observar também que os dados produzidos são dados estruturados e não estruturados. Como é colocado por Kelleher e Tierney (2018) e Inmon e Linstedt (2014), os dados estruturados podem ser planilhas que apresentam relação entre si e são gerenciados por atributos, chaves e registros. Já os dados não estruturados são textos, imagens, vídeos, entre outros. Como discorre Kelleher e Tierney (2018), é possível retirar dados estruturados de dados não estruturados.

Outra análise que podemos fazer a partir da fala da pesquisadora 3 é a questão do tratamento do dado, bruto e derivados. Os dados brutos são considerados os dados passíveis de diversos níveis de tratamento, já os dados derivados são versões processadas dos dados brutos (COSTA, 2017).

Aspectos que influenciam positivamente na produção/obtenção de dados de pesquisa

Os pesquisadores destacaram como ponto positivo para a produção/obtenção dos dados de pesquisa a qualidade da equipe (Pesquisador 2), a facilidade de ter acesso às informações necessárias para a produção dos dados (Pesquisadora 3), o ambiente de trabalho juntamente com os colegas (Pesquisadora 5), ter uma infraestrutura adequada, financiamento aos projetos e apoio técnico (Pesquisadora 10), a flexibilidade de gerar os dados em outros locais e a colaboração de pesquisadores de outros países (Pesquisador 11) e o apoio da instituição para a realização da pesquisa (Pesquisadora 12).

No nosso caso é basicamente a equipe que é muito boa tá, a equipe é fundamental, sempre tá, a equipe humana. (Pesquisador 2)

Primeiro aspecto é que a gente está dentro do [nome da instituição], então a gente tem acesso direto aos prontuários médicos, as análises histopatológicas dos pacientes. Então todos esses dados já estão computadorizados e a gente tem acesso tá. Então facilita você escolher pacientes do tipo que você quer e achar todos os dados que você precisa deles, e também os médicos estão aqui ao lado e eles participam nas pesquisas, ou seja, a gente conversa com eles, o médico que atendeu o paciente. Então eles sabem todo e qualquer detalhe que a gente precisa saber, esse primeiro ponto. (...) Segundo ponto, é que o [nome da instituição] tem um [nome do banco de dados]. (Pesquisadora 3)

Eu acho que uma coisa muito positiva no lugar é o ambiente de trabalho, assim os colegas dentro do meu núcleo de pesquisa né. A gente... somos seis professoras dentro da Genética vegetal, então o departamento de Genética da [nome da instituição] ao qual eu pertencço ele é bem amplo, porque tem desde Genética médica, Genética de animais, Genética imuno Genética... várias áreas. Aí tem a Genética vegetal e dentro da Genética vegetal nós somos seis professoras e a gente compartilha alguns laboratórios e esse ambiente, essa troca eu acho que é algo muito positiva. Ambiente muito bom de trabalho, então acho que essa seria a parte positiva. (Pesquisadora 5)

Então, o projeto só está em andamento porque a gente tem a infraestrutura necessária para realizar esse projeto. Então a gente tem um conjunto de pesquisadores engajados, a gente tem um conjunto de alunos de Pós-Graduação e graduação que são engajados, então pessoal. Também temos financiamento ao projeto de pesquisa, significa que a parte de material e serviço de terceiros a gente consegue financiar, e o grupo conta com alguns apoios técnicos assim né, o pessoal técnico é bastante qualificado pra realização dessa pesquisa. Então a gente está no local que permite a realização do projeto de pesquisa, se não eu não estaria fazendo. (Pesquisadora 10)

Então poder gerar os dados em vários locais diferentes do mundo significa que num local der algum problema eu posso mandar fazer em outro, em outro serviço, em outra empresa que faz. (...) A outra coisa que eu posso imaginar é que, no meu caso, eu faço muitos estudos em

que as amostras vêm do mundo inteiro, não só amostras brasileiras. Então, em alguns estudos são espécies que nem ocorrem no Brasil, a gente faz coisas do mundo inteiro, de lobos marinhos do mundo inteiro ou baleias do mundo inteiro. Então assim, a gente consegue ter colaboradores do mundo inteiro que nos mandam as amostras e aí a gente gera os dados e faz as análises em cima dessas amostras de outros locais. Então poder fazer online e remoto as reuniões e o pessoal mandar via correio é bom. (Pesquisador 11)

Eu tenho total apoio da instituição. (...) E aqui a gente tem vários editais internos, tem apoio... A gente tem taxa de bancada do próprio programa de Pós-Graduação. Meu programa é [nota do programa], então a gente tem [nome da agência de fomento] por ser um dos programas de excelência do Brasil né. Então a gente tem tanto apoio da Universidade quanto apoio da [nome da agência de fomento] pelo nível de excelência do programa. (Pesquisadora 12)

Entre os resultados obtidos na pesquisa, podemos destacar que para os pesquisadores o ambiente e a interação com os colegas são importantes para a produção de dados. Assim como ocorre na comunicação científica, para que a comunicação dos dados de pesquisa aconteça com maior qualidade, é importante que ocorra o debate científico entre os pares para o desenvolvimento de pesquisas e em consequência disso, o avanço da ciência.

Também podemos ressaltar que a interação com outros profissionais durante a pesquisa, como é o caso da Pesquisadora 3, é um fator positivo, além da instituição contar com o seu próprio banco de dados, o que facilitar o acesso as informações dos pacientes. Em síntese, os comentários feitos pela Pesquisadora 3, vão ao encontro com algumas das atribuições da comunicação científica elencadas por Menzel, Somers e Glaser (1958) que é a possibilidade de verificar e testar a confiabilidade de novos conhecimentos, permitir alcançar respostas para problemas específicos, promover novas descobertas e a troca de feedbacks que viabiliza o aprimoramento da pesquisa.

De modo geral, essa interação vista pelos pesquisadores como relevante para a produção de dados, independente da sua área de estudo, pode ser compreendida como tradições desenvolvidas ao longo dos séculos pelas comunidades científicas para fomentar novos estudos e ideias entre os pesquisadores para uma comunicação mais eficiente (DREXLER, 1991).

Aspectos que influenciam negativamente na produção/obtenção de dados de pesquisa

Os pesquisadores foram questionados sobre quais pontos eles identificam como negativos para a produção/obtenção de seus dados. O Pesquisador 1 relatou as dificuldades que um pesquisador novo enfrenta para realização de pesquisa na Genética por ser uma área cara.

A pesquisa científica não é algo barato, então quando você tá no início de carreira é muito difícil alguém que trabalha na seara de Genética, de Genética molecular, de organismos conseguir fazer experimentos complexos sozinho. Ele tem que tá já no laboratório que já tem alguma história, tem recursos, tem equipamento para conseguir produzir alguma coisa relevante. Se ele não tem, ele realmente não vai conseguir fazer isso. Então existe essa dificuldade de você ter equipamentos e reagentes porque são realmente muito caros e você só consegue isso ao longo do tempo. Se associa a um pesquisador, começa a fazer trabalho em conjunto né, submete projeto de pesquisa. Quando você aprova o projeto de pesquisa, se compra equipamento, compra reagente, começa orientar um aluno de graduação, depois de pós-graduação, publica seus artigos e submete outros projetos. É um processo lento até você ganhar corpo e conseguir ter o seu próprio laboratório com toda infraestrutura necessária para fazer uma pesquisa de qualidade. (Pesquisador 1)

Os pesquisadores destacaram também a falta de verbas provenientes de editais de financiamento a pesquisa que estão cada vez mais escassos para a compra de reagentes, equipamentos e materiais para o desenvolvimento de seus projetos. Nesse sentido, a Pesquisadora 5 evidenciou a necessidade de ter que usar o próprio dinheiro para o desenvolvimento de sua pesquisa, além da frustração e desgaste gerado pela concorrência elevada pela verba cada vez mais restrita disponível para a pesquisa.

Tem o limitante grande sim, que é o custo dos reagentes, principalmente quando você trabalha com sequenciamento em larga escala, tanto genoma, como que a gente faz análise da expressão do tipo transcriptoma, aí sim são custos altos por cada. (...) mas esse é um limitante no entanto. Um limitante porque são reagentes caros, cotados em dólar, importados né, então tem toda essa problemática de custo, de demora, de obtenção desses reagentes... Esse talvez é o maior limitante que a gente tem. (Pesquisador 2)

Tem a questão financeira né, a questão negativa de você produzir dados no Brasil é que você precisa ter dinheiro para ter o equipamento, para ter o contrato de manutenção do equipamento, você ter dinheiro pra pagar os reagentes pra fazer o sequenciamento ou fazer espectrômetro de massa, no caso de proteômica. (...) Eu não vejo muito um aspecto negativo, eu vejo como dificuldade, o financiamento de pesquisas, tá. Todo mundo sabe, o [nome da agência de fomento] divulga essas informações, a [nome da agência de fomento], a [nome da agência de fomento]. Todas as grandes agências de fomento a pesquisa elas divulgam, assim,

que a cada ano que passa tá, tem menos dinheiro disponível para pesquisa, então isso dificulta muito. (Pesquisador 4)

No momento não dá para fugir que é conseguir verba para isso né. A verba é limitada que a gente tem, então muitas vezes não é raro a gente tem que usar dinheiro pessoal né. (...) A gente que trabalha com biologia molecular é muito caro né, bastante, então assim a gente estima, claro vai depender, mas com menos de 20 e 30 mil não dá para fazer cada tese né, mais ou menos, muito por cima. Claro que vai variar de acordo com qual é a área exata da pesquisa. Mas então essa questão da gente ter que ir atrás do dinheiro com a concorrência que tem e com a verba cada vez menor né, o montante cada vez mais reduzido isso é muito desgastante. (Pesquisadora 5)

No momento porque é terrível o que a gente tá vivendo é falta de dinheiro né, as fontes secaram e sobretudo a falta de bolsa. (...) uma boa parte da pesquisa ou quase toda a parte prática é levada a cabo pelos alunos de iniciação, mestrado e doutorado. São eles que ao mesmo tempo que se formam, eles também produzem os dados e produzem... enfim, as descobertas que serão feitas. Só que os alunos precisam ser sustentados e se sustentarem, e pra isso eles precisam de bolsa. E agora, agora eu digo, sei lá, tem quatro anos, seis anos que fechou tudo né, não tem, não tem jeito. Começou a fechar claramente foi na administração [nome do presidente] e com administração [nome do presidente] fechou mais ainda. Agora quando um aluno vem me pedir estágio, a primeira coisa que eu pergunto é “você tem dinheiro? Você consegue pagar sua passagem? Você consegue vir pagar sua passagem e pagar sua refeição?”, se ele me disser que tudo bem, que consegue tudo, aí dá. Se falar não, aí não tem como começar. (Pesquisador 9)

Como citado anteriormente, desde 2012 a verba destinada ao MCTI para pesquisa vem sofrendo gradativamente uma diminuição. Esse corte orçamentário afeta também as agências de fomento, como é o caso da Fapesp, uma das principais Faps brasileiras. Em consequência disso, existe uma carência de editais para o financiamento de projetos de pesquisa, podemos citar como exemplo, os Editais Universais que ao longo dos últimos 10 anos só dispôs de 6 editais publicados.

Em relação a fala da pesquisadora 5, a concorrência pela verba nos poucos editais publicados é cada vez maior. É evidente que não houve só redução de editais, mas também na disponibilização de bolsas de pesquisa, no caso da Genética, entre 2014 a 2020, os anos que tiveram a menor quantidade de pesquisadores contemplados com bolsa de pesquisa foram 2019 com 94.706 e 2020 com 94.883 (GEOCAPES, 2021).

É possível analisar a percepção dos pesquisadores em relação a perda de recursos por outra perspectiva, essa questão de ausência de verba para o desenvolvimento de pesquisa pode estar atrelada também a região de atuação dos

pesquisadores, como abordam alguns autores, as regiões Sul e Sudeste costumam acumular recursos (SIDONE; HADDAD; MENA-CHALCO, 2016; MELO; SANTANA; SILVA, 2018; SOUZA et al., 2020; MCTI, 2011), o que pode ter levado pesquisadores de outros estados sentirem mais fortemente a concorrência e a redução de fomento para as suas pesquisas.

Os resultados apresentados pelas falas dos pesquisadores podem ser relacionados ao período em que se deu a pesquisa. Desde o início de 2020 o mundo passa por uma pandemia, muitos dos editais publicados em 2020 foram direcionados para o enfrentamento da Covid-19, logo, pesquisadores que não trabalhavam em alguma das grandes áreas temáticas, não puderam concorrer aos editais.

Quanto ao relato dos Pesquisadores 3, 6 e 8, infere-se que as dificuldades enfrentadas na produção/obtenção de seus dados é a falta de apoio administrativo e/ou institucional, o que acarreta a uma sobrecarga de atividades administrativas aos pesquisadores durante um tempo que poderia ser utilizado na produção dos dados.

Olha, eu diria assim que nós não temos problemas para gerar os dados. O que a gente tem é que nós somos um laboratório de assistência e pesquisa, portanto a gente tem uma boa carga administrativa, de logística né, de material e tudo mais e de fazermos os diagnósticos. Então a gente passa bastante do nosso tempo né em trabalhos que não são de pesquisa né. Logicamente a gente tem alunos pós docs que aí ajudam a gente a andar com a pesquisa, porque a gente mesmo nosso tempo não é 100% do tempo dedicado à pesquisa. (Pesquisadora 3)

Eu nunca tive um técnico de laboratório, então nós fazemos tudo mesmo, desde lavar vidraria e esterilizar. Então nós perdemos né. Eu falo que os meninos perdem muito tempo nesse trabalho braçal. (...) O negativo é não ter um apoio maior né institucional. Nós infelizmente... Nós não temos técnicos especializados no nosso laboratório, nós precisaríamos desse apoio para dar mais resultados. Eu acho que se nós tivéssemos uma retaguarda técnica nós teríamos mais tempo pra nos debruçarmos sobre os dados e pra trabalharmos até mais e produzir mais do que a gente já produz. (Pesquisadora 6)

Apoio administrativo não existe. Eu não sei o que os outros estão respondendo, mas pelo menos na realidade que eu vivo aqui, toda a parte administrativa de todo o projeto é meu. Sou eu que faço a contabilidade, sou eu que faço, que faço administração de recursos, sou eu que faço o levantamento de preço, entende? (...) A gente não tem recursos para insumos, não tem também para contratar terceiros para fazer esses serviços para a gente, entende? Então tudo isso passa a ser obrigatoriamente uma das funções dos pesquisadores. Então além de orientar, além de pesquisar e gerar conhecimento, a gente tem que fazer toda a questão administrativa. (Pesquisador 8)

As respostas dos Pesquisadores 3, 6 e 8 corroboram com o estudo dos autores Souza *et al.* (2020) onde eles elencam os principais desafios que os pesquisadores brasileiros vivenciam no desenvolvimento de suas pesquisas, entre os resultados apresentados, questões associadas com o acúmulo de funções, falta de uma equipe de apoio é sentida também pelos pesquisadores da Genética.

O Pesquisador 11 vê como algo negativo a produção dos dados a falta de contato com os alunos no laboratório, o que ocasiona uma troca menor de pensamento e de resolução de problemas ao seu ver.

Claro, não poder estar presencialmente no laboratório junto com os alunos, digamos. O fluxo de ideias é menor né. A gente tava o dia inteiro lá com os alunos na bancada, qualquer problema na hora rapidinho se resolve, eu olho lá e vejo o que tá acontecendo, posso ajudar a resolver. (Pesquisador 11)

A resposta do Pesquisador 11 vai de encontro com questões levantadas no tópico sobre os aspectos que influenciam positivamente a produção de dados de pesquisa. A interação entre os pesquisadores no desenvolvimento de uma pesquisa proporciona o descobrimento de outras indagações que até então não tinham sido pensadas, ou até mesmo apontamentos mais adequados sobre uma pergunta que não tinha ocorrido inicialmente (DESLAURIERSE; KÉRISIT, 2008).

Reconhecendo a importância da interação dos pesquisadores no processo da pesquisa, Carvalho (2018) expõe que a produção dos dados de pesquisa acontece sobretudo no ambiente de convívio dos pesquisadores. Ainda de acordo com a autora, a produção/obtenção de dados muitas vezes advém de observações e conversas ligadas a diferentes contextos.

5.1.2 Compartilhamento de dados de pesquisa

Na seção de compartilhamento de dados de pesquisa serão apresentados os aspectos da prática de compartilhamento realizada na Genética, tendo como enfoque a utilidade dos dados por outros pesquisadores, interesse no compartilhamento dos dados, experiência no compartilhamento deles, pontos que motivam e inibem o compartilhamento, obrigatoriedade do compartilhamento e o destino dos dados de pesquisa.

Utilidade dos dados de pesquisa por outros pesquisadores

Foi possível observar que os pesquisadores acreditam que seus dados podem ser utilizados por outros pesquisadores fora do seu grupo de pesquisa, e se mostram abertos ao compartilhamento. A única ressalva apresentada ao compartilhamento dos dados é caso a solicitação de compartilhamento ocorresse antes da publicação dos resultados (Pesquisadores 1, 6 e 7).

Sim, com certeza. Com certeza estaria disposto sim. A não ser que eu não tenho ainda publicados, se é um dado sensível que eu não publiquei ainda e a pessoa que está solicitando é uma pessoa que eu desconheço, eu não conheço. Então assim, aí é mais complicado, mas de uma forma geral dados que não são dados que vão gerar publicações que eu ainda não escrevi eu compartilho. (Pesquisador 1)

Eu acho que esses dados que a gente tá produzindo são extremamente úteis. Tanto os meus porque tem muita pouca gente trabalhando. Então como a área de conservação, os dados que o nosso grupo gera eu acredito que eles são muito importantes, porque o que a gente diz para os outros né, a gente faz como se fosse assim o diagnóstico da viabilidade das populações no sentido genérico assim. (...) Então acho que eles têm bastante valor nesse sentido. (Pesquisadora 5)

Sim, a gente tem compartilhado em congresso né, e também em revistas científicas. E a gente tem visto bastante interesse né nos nossos resultados até em congressos internacionais. Nós temos sido convidados para divulgar os dados, então eu acho que sim, que tem interesse sim. Até porque né trabalhar com câncer e trabalhar com fitoterápico tem... acaba gerando muito interesse nas pessoas. (...) É meio complicado você compartilhar antes de ser publicado, complicado né. Porque os dados para você publicar têm que ser originais, as revistas pedem isso, então depende do objetivo do pesquisador que queira alguma informação nossa. Dependendo do objetivo dele, da informação que a pessoa quer... Sim a gente pode compartilhar, mas normalmente nós não compartilhamos antes de publicar. (Pesquisadora 6)

Eu acho que, eu diria médio ou pouco. Porque a minha área de atuação é bem restrita tá, mas para os pesquisadores da área restrita que eu atuo, eventualmente esses dados podem ser bem importantes. (...) Mediante uma solicitação adequada eu estaria disposto. Não é aberta completamente, mas se eu não coloquei, se a pessoa me dá uma justificativa que é plausível, se é uma coisa que não tem competição com o que eu tô fazendo naquele momento. (Pesquisador 7)

Todos os dados que a gente tem gerado nos últimos 20 anos eles são depositados nesse [nome do banco de dados] e eles são livremente acessados por outros pesquisadores. Os dados são usados e a gente tem vários trabalhos (...) Então eles são úteis, eles são usados, nem sempre eles são reconhecidos. (Pesquisadora 10)

Total. Porque eu também dependo direto de dados gerados por outros pesquisadores do mundo inteiro. Então nessa nossa área a gente usa, dependendo do projeto é claro, do tipo de organismo a gente usa direto dados produzidos pelos outros pesquisadores ao longo dos anos todos. (...) Então assim, os nossos dados são usados e são sempre depositados nesses bancos e serão usados um monte de vezes. Ao longo dos anos já foram usados nos diferentes projetos, pra projetos de pessoas no mundo inteiro, que usam pra comparar, que pegam e

fazem uma análise mais ampla com os nossos dados e com o conjunto que estão tão gerando. (Pesquisador 11)

Mesmo boa parte dos respondentes demonstrando serem abertos ao compartilhamento dos seus dados de pesquisa, essa abertura é percebida mais de acordo com as respostas dos pesquisados que trabalham com algum ramo da Genética Básica, por ser uma área ampla. No caso do Pesquisador 7, em que seu campo de atuação é mais restrito, a Neurogenética, os dados por ele produzidos teriam pouco ou média utilização para pesquisadores de fora e por isso não costuma compartilhá-los. Corroborando com o exposto, Dallmeier-Tiessen *et al.* (2014) discute de forma mais profunda a questão da cultura acadêmica, onde os pesquisadores refletem os comportamentos sobre o compartilhamento dos dados de pesquisa de acordo com as normas comportamentais da área do conhecimento em que fazem parte.

Em comparativo com o estudo realizado por Carvalho (2018), os pesquisadores das disciplinas de Educação, Antropologia e Química acreditam que seus dados são úteis para outros pesquisadores e estão dispostos a compartilhá-los. No entanto, existem também nessas disciplinas ressalvas na realização do compartilhamento. No caso da Educação o compartilhamento só seria feito com o consentimento dos participantes da pesquisa, já na Antropologia os pesquisadores acreditam que dados relativos à intimidade do pesquisador não deveriam ser revelados.

Em geral, as ressalvas feitas pelos pesquisadores da Genética em relação às solicitações de compartilhamento de dados têm a ver com o fato do ineditismo dos dados e das publicações, pois o seu compartilhamento antes da publicação pode comprometer as pesquisas que estão em andamento. O que pode virar um problema em momentos de crises de saúde, como a da Zika e a mais recentemente, Covid-19. Neste sentido, iniciativas como a Wellcome Trust criada em 2016, ações de cooperação científica da ONU e declarações de editoras científicas internacionais que visam garantir o acesso aberto a dados de pesquisa nessas ocasiões de crise, podem gerar uma corrida por publicação e pesquisadores que trabalham com assuntos correlatos segurarem mais ainda seus dados por medo de perderem suas pesquisas em desenvolvimento.

Percepção sobre o compartilhamento dos dados

Ao serem questionados sobre o compartilhamento dos dados de pesquisa em suas disciplinas, os pesquisadores demonstraram familiaridade com essa atividade, e salientaram que o compartilhamento é importante para a pesquisa em suas áreas. São diferentes os motivos que encoraja a realização do compartilhamento dos dados nas áreas que os pesquisadores atuam: o reconhecimento recebido pela divulgação (Pesquisador 1), o custo elevado da pesquisa no campo da Genética (Pesquisadora 3), por ser uma prática já estabelecida na área (Pesquisador 4) ou uma exigência dos periódicos científicos para a publicação dos artigos (Pesquisadores 9 e 11).

Ela é recorrente, ela é muito, ela acontece sempre. (...) Na minha área é muito importante divulgar, por isso, que você se torna conhecido também. Mas você só não divulga se você ainda não publicou. Tem muita gente que divulga porque acho que não vai publicar e já divulga, descobriu alguma coisa, divulga, e aí deixa o pessoal usar o dado do jeito que quiser. Mas algumas vezes eu faço isso, outras não, quando eu acho que é interessante eu seguro por um tempo, mas depois eventualmente eu divulgo. (Pesquisador 1)

Olha, eu acho que o compartilhamento é essencial, em primeiro lugar porque essa pesquisa, essa área de Genética não é uma área nem simples e nem barata né, é tudo é muito caro e aí portanto sempre que você puder compartilhar é importante. (Pesquisadora 3)

Nessa área genômica, transcriptômica, proteômica e bioinformática é meio que padrão todo mundo compartilha e sabe da importância de compartilhar, depois de publicado. (Pesquisador 4)

Eu acho que cada vez mais vai ser maior. Porque houve muito problema... Houve algum problema das pessoas repetirem o experimento dos outros e não dar certo. Aí tendo os dados publicados dá para você verificar se tem alguma coisa errada ou não. Então agora as revistas estão pedindo quase que direto, quase que todas as revistas estão pedindo que coloque os dados à disposição. (Pesquisador 9)

Uma boa parte das melhores revistas científicas né, onde a gente publica os artigos é obrigatório. Pra eles aceitarem um artigo, aprovarem um artigo tu tem que indicar onde estão e qual é o acesso livre para todos os dados que foram usados no artigo, inclusive os dados brutos. (Pesquisador 11)

O compartilhamento de dados de pesquisa já foi abordado em outros estudos, como por exemplo, o elaborado pela *Key Perspectives Ltd* em 2010, no estudo, foi constatado que nas Ciências da Vida o compartilhamento de dados de pesquisa é comum entre os pesquisadores.

A percepção sobre o compartilhamento dos dados é crescente nas subdisciplinas da Genética como pode ser observado nas falas dos pesquisadores.

Os motivos para o compartilhamento são diversos. Na área da Pesquisadora 1, é através do compartilhamento que o pesquisador será reconhecido entre seus pares. Nas áreas dos Pesquisadores 9 e 11 o compartilhamento dos dados é feito principalmente por ser uma exigência dos periódicos científicos. Essa condição imposta pelos editores de periódicos vai de encontro com a promoção da Ciência Aberta que tem como princípio manter a transparência da comunicação científica (GEZELTER, 2009) e evitando problemas, como os relatados pelo Pesquisador 9.

Apesar da maioria dos pesquisadores reconhecerem a prática do compartilhamento de dados como algo importante e crescente em sua área e de realizarem o compartilhamento dos dados que produzem durante as suas pesquisas, em algumas subdisciplinas da Genética, essa atividade não é sentida de forma habitual em seus campos, como relatado pelos Pesquisadores 6 e 7. Considerando isso, o Pesquisador 8 apresenta como ponto para o não compartilhamento dos dados de pesquisa em sua área os benefícios ou produtos que os dados podem gerar para a instituição, agências que fomentam sua pesquisa e/ou a União, pois a divulgação desses dados poderia ser utilizada por outros países.

O Pesquisador 7 traz um outro dado interessante relacionado ao fato de pesquisadores mais jovens serem mais abertos ao compartilhamento de dados de pesquisa, de acordo com os autores Zuiderwijk e Spiers (2019), Jarolímková, Drobíková e Souček (2018) e Dallmeier-Tiessen *et al.* (2014), as práticas de compartilhamento de dados é mais comum entre pesquisadores mais novos do que entre os pesquisadores de faixas etárias mais elevadas. Uma justificativa para esse fato, de acordo com os autores, é a domínio de mecanismos de compartilhamento de dados de pesquisa pelos pesquisadores mais jovens.

Eu acho que não, que de modo geral não existe. O pessoal compartilha após, como eu te disse, após a publicação né. O que nós fazemos, eu faço nas minhas disciplinas, e até em palestra, a gente mostra os nossos dados, mas a gente não fornece isso para as pessoas. Mostra para discutir para saber opinião, pra receber críticas e sugestões e tal e então isso sim. Nas disciplinas a gente tem compartilhado, mas não é entregue né. Não sei se você tá entendendo, mas compartilhar no sentido de abrir pra discussão sim, mas não pra fornecer o dado pra ser usado por outro antes de ser publicado. (Pesquisadora 6)

Eu acho que ainda é pouco. Cada vez mais, mas acho que ainda é pouco. Os artigos que eu vejo dos meus colegas, eu não vejo tanto disposição de dados, mas assim, os mais jovens eu vou dizer que tem uma tendência a fazer isso um pouco mais tá, mas é feito, acho que ainda pouco. O que eu enxergo é um aumento. (Pesquisador 7)

Hoje existem países mais desenvolvidos que estão super ligados ali esperando uma nova informação científica interessante para transformar em produto, correto? Então o que acontece, toda vez que a gente divulga essa informação, é óbvio que a gente divulga pensando no bem estar da ciência. Mas como eu disse, se os resultados que a gente gerou no laboratório são passíveis de gerar algum benefício para a universidade, para a União, pra quem fomentou nossas pesquisas, a gente vai segurar esses dados até que todas as comprovações científicas sobre a eficiência daquilo sejam comprovadas. Pra que a partir daí gere um produto comercializável e que possa ser transferido para a população de uma maneira geral, entende? Então se esse produto tem essa função, essa possibilidade a gente não compartilha o dado em hipótese alguma. Porque no momento que a gente compartilha esse dado, que ele torna-se público, países mais competentes em ciências, com mais recursos financeiros e que tem estrutura para fazer esse tipo de análise, vai fazer em curto prazo de tempo comparado com a gente, e aí a gente perde o bônus. A gente perde aquele momento oportuno de transformar ideia em produto. (Pesquisador 8)

Como aponta Walport e Brest (2011) e Pontika *et al.* (2015), o compartilhamento dos dados de pesquisa não é tão comum em uma boa parte das comunidades científicas, isso pode ser percebido dentro da própria Genética que é vista como um campo aberto a ideia do compartilhamento, mas que na prática, em algumas das suas ramificações de estudos existe uma resistência por parte dos pesquisadores, seja por receio da perda do ineditismo, por causa da alta competitividade ou questões relacionadas à geração a qual o pesquisador faz parte.

Além disso, a ausência de locais para o depósito de dados pode ser vista como um aspecto negativo. No Re3data, estão registrados apenas 12 repositórios brasileiros distribuídos nos estados de São Paulo, Rio de Janeiro, Paraná, Brasília, Rio Grande do Sul, Goiás. Mesmo o Brasil sendo o país na América do Sul com o maior quantitativo de repositório de dados abertos, quando comparado com países da América do Norte, como os Estados Unidos (1135 repositórios), com o Reino Unido (303 repositórios) na Europa ou com a China (48 repositórios) na Ásia, o Brasil tem poucas alternativas para os pesquisadores, principalmente quando se é levado em conta as temáticas dos repositórios brasileiros para compartilhamento de dados de pesquisa.

No entanto, a compreensão sobre o compartilhamento dos dados para um futuro reuso por outros pesquisadores não é percebida em algumas disciplinas da Genética de acordo com as respostas dos pesquisadores. Ao questionar a Pesquisadora 12 sobre o compartilhamento dos dados para o reuso em novos

estudos, ele demonstrou surpresa sobre essa prática, relatando que em sua área, os dados que foram publicados não são usados em novos estudos.

Não sei se eu sei o que que é isso, reutilizar dados brutos para outro trabalho tipo... Eu não, não sei... Eu não, eu não sei se isso existe. Eu realmente fiquei surpresa, não sei que isso poderia... Porque na minha formação, o que eu sempre aprendi é uma vez que eu já publiquei esse trabalho, é publicado. Então o que eu tenho que fazer é citar o trabalho. (...) No próximo eu não posso usar esses dados, esses dados tem que ser citados que “no artigo publicado de fulano de tal o modelo foi caracterizado, e neste artigo a gente tá indo um passo adiante que é dosar as proteínas relativas a não sei o que”. Mas o modelo já foi caracterizado e publicado em outro. É assim que eu trabalho, e pra mim isso é o que é ou deveria ser. Na minha formação com os meus anteriores é... Supervisores de mestrado e doutorado pra mim era isso. Eu desconheço que se possa reaproveitar dados. (Pesquisadora 12)

O relato da Pesquisadora 12 vai ao encontro com os resultados do estudo da *Key Perspectives Ltd* (2010), que mostra que o compartilhamento de dados está diretamente ligado às práticas realizadas na disciplina, ou seja, mesmo em áreas como as das Ciências da Vida que se tem familiaridade com o compartilhamento de dados de pesquisa e o uso de sistemas, a complexidade dos dados que são produzidos e/ou as metodologias utilizadas impossibilita o seu compartilhamento. Essa mesma observação foi feita pela *Key Perspectives Ltd* (2010), nas Ciências Sociais e nas Ciências Exatas, em relação as práticas de compartilhamento de dados, devida aos métodos específicos empregados para a coleta de dados e a cultura acadêmica das disciplinas de cada área, o compartilhamento acaba não sendo uma atividade recorrente entre os pesquisadores.

Do mesmo modo, Murray-Rust (2008) explana sobre as diferenças de práticas de compartilhamento dos dados de pesquisa entre as disciplinas, onde algumas, por tradição exigem que pesquisas que foram financiadas com verbas públicas tenham seus dados posteriormente disponibilizados em bancos de dados, já outras, disponibilizam seus dados totalmente.

Experiência em compartilhamento de dados

A experiência em compartilhamento de dados foi abordada nesta pesquisa. Nesse sentido, os pesquisadores afirmam já terem realizado o compartilhamento dos seus dados de pesquisa, por uma demanda induzida pelas revistas, em repositórios públicos e internacionais, quando solicitado por outros pesquisadores ou no próprio núcleo, entre os colegas.

Todos os dados de sequenciamento vão para repositórios públicos, todos eles. (Pesquisador 2)

Em revistas. Então nas publicações isso é um compartilhamento, e dentro do próprio núcleo onde tá também. A gente compartilha bastante mesmo, até no cafezinho, quando a gente tá tomando o cafezinho que a gente troca né as ideias e “ah, eu consegui isso”. Então em todos os âmbitos assim eu acho que a gente compartilha. (Pesquisadora 5)

Qualquer trabalho científico que a gente publica é um compartilhamento de dados de pesquisa. Na área da Genética quem trabalha com dados muito volumosos, hoje, praticamente é obrigado a depositar em repositórios. As revistas científicas inclusive não aceitam trabalhos se os dados não estiverem depositados. Então a própria ciência está se cobrando em relação a isso. (Pesquisador 8)

Saiu um artigo, acho que foi em 2012 (...) Eles [pesquisadores de outro país] pediram os dados para mim e eu mandei. (...) Eu não me lembro se tava numa base de dados, mas eu mandei uma planilha Excel, mandei pra eles, eles analisaram e aí eles publicaram um artigo criticando o nosso artigo. (...) Aí a gente escreveu uma resposta pras críticas que eles tinham feito. E aí também é claro, que a gente fez uma resposta criticando o trabalho deles. Mas não tem problema, não nesse sentido, é até bom porque você publica um artigo e amplia a tua audiência. (Pesquisador 9)

No meu caso, é tudo em banco internacional, praticamente tudo internacional. Até porque é um tipo de dado muito simples né, quer dizer, relativamente. São sequências de DNA que são aquelas quatro letras né. (Pesquisador 11)

Observou-se a partir das falas dos pesquisadores que as experiências relacionadas ao compartilhamento de dados de pesquisa acontecem mais facilmente entre os próprios grupos de pesquisa e/ou em imposição por parte das revistas em que os pesquisadores submetem seus trabalhos para publicação. Sayão e Sales (2019) salientam que o compartilhamento frequentemente acontece entre os pares mais próximos ou nos sites de suas instituições.

Igualmente foi observado que mesmo que seja comum a prática de compartilhar dados de pesquisa entre os pares, os pesquisadores demonstraram existir frequência na disponibilização de seus dados em bancos de dados como o *The Cancer Genome Atlas* (TCGA), *Genotype-Tissue Expression* (GTEx), *Sequence Read Archive* (SRA), *ProteomeXchange*, Genbank, *Universal Protein Resource* (Uniprot) e *Kyoto Encyclopedia of Genes and Genomes* (KEGG), principalmente para o cumprimento de exigência dos periódicos científicos e por serem os bancos de dados mais adequados aos *datasets* produzidos.

Experiências parecidas sobre compartilhamento de dados por pesquisadores de disciplinas das grandes áreas de Ciências Exatas, Ciências Sociais e Humanidades são descritas por Carvalho (2018), de acordo com a autora, mais de 50% dos pesquisadores entrevistados afirmaram que o compartilhamento de seus dados é realizado entre os membros do grupo ou do projeto de pesquisa em que fazem parte. Entretanto, o compartilhamento de dados apenas entre os pares não é suficiente para garantir o rápido progresso científico e o alinhamento com as práticas defendidas pela Ciência Aberta.

A Ciência Aberta trata o compartilhamento de dados de pesquisa como uma ação que deve ser praticada por pesquisadores, instituições de ensino e pesquisa, agências de fomento e governo para assegurar a construção colaborativa e transparente da ciência, permitindo que os dados produzidos sejam livremente acessados e reutilizados em outras pesquisas.

Motivos que estimulam o compartilhamento de dados

Quando perguntado aos pesquisadores quais motivos que os estimulam a compartilharem os dados em seus respectivos campos de atuação, os pesquisadores mencionaram cinco causas principais que impulsionam a realização do compartilhamento:

- A cooperação científica entre os pesquisadores e as discussões que elevam a qualidade da pesquisa;
- A escassez de recursos financeiros para a produção dos dados;
- O desejo de contribuir para o avanço da ciência;
- O repasse do conhecimento a população que é gerado dentro de instituições públicas;
- Registrar a ciência por meio da publicação.

O que estimula é a cooperação científica, a partir do momento que você compartilha, você vai cooperar com mais pessoas, você abre um leque de possibilidades para sua pesquisa. (...) Então quem é muito produtivo nessa área que eu atuo é porque faz muita colaboração, porque sozinho você não faz nada, porque é tudo muito caro. (...) Então é super comum esse compartilhamento e só eleva né a pesquisa. Acho que melhora muito a cooperação unindo o trabalho, porque você tem o input né. A geração de dados e a discussão com vários pesquisadores diferentes. (Pesquisador 1)

Pra nós é muito importante, fora isso porque, como tava te falando, nós temos limitantes econômicos principalmente, nós temos um limite de produção de dados. Tem pesquisa que a gente faz com três, quatro, cinco amostras, o ideal era fazer com 20, 50, 100. Nós não temos capacidade de fazer isso, então vai ser econômica tá, de processamento de dados, então por isso mesmo a gente utiliza muito dado já publicado tá. A gente puxa os dados das bases públicas pra reforçar os nossos. (Pesquisador 2)

A vontade de contribuir para o avanço da ciência. Se todo mundo começar do zero a gente nunca vai passar de um certo ponto né. Então se um começou do zero parar no 200, o outro começou do 150, usou aqueles dados e foi a 400 e assim por diante (Pesquisadora 3)

Eu vejo como muito importante, quanto mais dados forem gerados e as pessoas puderam ter acesso. (...) Eu acho que é bem, bem importante, assim, esse compartilhamento para que todos possam usufruir dos dados gerados dentro de uma universidade que é pública e portanto né a gente deve isso a população. (Pesquisadora 5)

A gente publica porque primeiro quer fazer ciência e, não publicar, não tem validade nenhuma. Então tipo, o trabalho não terminou enquanto ele não for publicado, pra mim isso é condição (...) o trabalho só termina quando ele de fato tá publicado. (Pesquisadora 12)

A partir dos resultados obtidos é possível aferir que o principal fator que estimula o compartilhamento dos dados pelos pesquisadores é o desejo de contribuir com a qualidade e o avanço da ciência. Veiga *et al.* (2019) reforça essa visão dos pesquisadores em seu trabalho. Conforme apresentado pela autora, 51,1% dos pesquisadores entrevistados se sentem motivados a compartilhar os seus dados tendo em vista a melhoria da pesquisa.

Outro aspecto que podemos abordar como motivo que influencia o compartilhamento de dados é as condições impostas por agências de fomento e instituições para financiar pesquisa (NASSI-CALÒ, 2019). Ao compartilhar os seus dados, os pesquisadores são beneficiados, uma vez que ampliam a visibilidade de suas pesquisas em razão que os principais repositórios em nível internacional do campo da Genética possuem identificadores persistente e metadados bem descritos que facilita a citação do produtor dos dados.

Motivos que inibem o compartilhamento de dados

Em relação aos fatores que podem inibir o compartilhamento dos dados está a exigência do ineditismo dos dados. O principal ponto exposto pelos pesquisadores é o receio de compartilhar os seus dados que são importantes e/ou que ainda estão em processamento serem utilizados por outros pesquisadores ou laboratórios de outros países que têm um poder de produção maior e apresentar resultados ou

produtos antes dos pesquisadores que geraram os dados. Como discorre o Pesquisador 7, a competitividade é vista como inibidor para o compartilhamento de dados, principalmente quando essa competitividade acontece em nível internacional.

Uma coisa que inibe é isso, o ineditismo. Se você tem um dado que é muito importante, você quer segurar aquele dado. Você não quer compartilhar antes de publicar aquele dado, então um probleminha que inibe seria isso o ineditismo. Então as pessoas ficam preocupadas de outras usarem os dados que você gerou na sua frente. Você gerou o dado, gastou dinheiro, liberou no banco de dados e vai alguém né e publica. Aí a pessoa fica inibida. (...) Acho que esse que é o principal, que é o medo do dado ser usado, que você gerou por outros pesquisadores. (Pesquisador 1)

Pra nós, por exemplo, disponibilizar um dado que a gente está processando é de alto risco digamos. (...) Nós sabemos que temos uma velocidade de trabalho que não é comparável com laboratórios mais fortes, digamos, laboratórios de primeiro mundo tá. Então, isso que corre o risco de perder o ineditismo da pesquisa nessa publicação de dados, tá. Então esse é o único que inibe a abertura total desse dado. (Pesquisador 2)

Eu acho que a competitividade, receio de perder nicho de atuação e etc. Acho que é muito mais assim, a palavra é receios, o que inibe são os receios. Competitividade, desde tu for competir com outros países, talvez seja ainda mais difícil de fazer esse nível de competição. (Pesquisador 7)

Observou-se entre as respostas dos pesquisadores que questões relacionadas aos direitos do uso dos dados (patente e propriedade intelectual) e dados sensíveis impedem o seu compartilhamento.

É só quando tem patente, propriedade intelectual, aí realmente fica mais... aí não tem como compartilhar mesmo (Pesquisador 4)

Eu acho que tem dados que são dados confidenciais. Então acho que em qualquer ambiente a gente tem dados que são confidenciais e talvez esses sejam dados sensíveis. (Pesquisadora 10)

Neste sentido, o Pesquisador 8 expressa que o compartilhamento dos dados a qualquer custo antes de esgotar todas as possibilidades de uso acabou favorecendo outros países. Além disso, como discorre o Pesquisador 8, o compartilhamento não deve ser feito de qualquer jeito, mas sim ser planejado, visando o retorno do conhecimento para a sociedade.

Esse é um dos motivos, compartilhar e perder a oportunidade de gerar alguma coisa satisfatória (...) eu tenho sido menos favorável ao compartilhamento da informação até que essa informação não possa ser mais utilizada. (...) Mas é uma forma de garantir que a herança daquele conhecimento permaneça nas nossas mãos, entende? Então aquela ideia de compartilhar o conhecimento a tudo e a qualquer custo. Tornar ele público porque ele foi um

investimento público, durante muito tempo defendido por muitas pessoas, certamente enriqueceu outros países que não necessariamente fizeram. Então eu acho que tem que compartilhar, mas no momento certo. Não compartilhar a qualquer custo, a qualquer momento sem ter esse planejamento de como a gente pode retornar isso para a sociedade em termos de produto. Geração de conhecimento na forma de produto. (Pesquisador 8)

Os resultados aqui obtidos demonstram que a Genética assim como outras áreas do conhecimento possui limitações relacionadas ao compartilhamento dos dados de pesquisa. Podemos inferir que a alta competitividade em algumas subáreas da Genética, o nível de confidencialidade e a propriedade intelectual por trás do dado são as principais responsáveis por impedir o seu compartilhamento.

Nesse sentido, Chawinga e Zinn (2019) mostram que diferentes razões podem dificultar o compartilhamento dos dados de pesquisa. Para os autores, os motivos se encontram em três níveis. O nível individual tem relação com as formas de aquisição dos dados por parte dos pesquisadores; em nível institucional com a ausência de políticas internas adequadas e em nível internacional políticas fracas e normas éticas e legais não transparentes.

As falas dos pesquisadores vão de encontro com pontos abordados no trabalho dos autores supracitados, como o apoderamento indevido dos dados por terceiros, ausência de licenças que respeitem e resguardem os participantes da pesquisa, políticas sólidas que assegurem os direitos aos pesquisadores que o produziram. No Brasil não tem nenhuma política em nível estadual ou nacional relacionada ao acesso aberto do conhecimento, apenas em nível institucional, como é o caso da Fiocruz. Grande parte das razões que inibem o compartilhamento dos dados de pesquisa elencada pelos pesquisadores poderiam ser resolvidas se tivesse um envolvimento mais efetivo de atores políticos e sociais e políticas específicas.

Exigência de compartilhamento dos dados de pesquisa por parte das agências de fomento

É muito comum as instituições receberem verbas de agências de fomento para o desenvolvimento de suas pesquisas. Sobre este aspecto, os pesquisadores foram perguntados como se sentiriam caso os seus dados de pesquisa tivessem que ser compartilhados para obtenção de financiamento para os seus estudos. Sob esta perspectiva, a maior parte dos pesquisadores entrevistados expressaram de forma

positiva a obrigatoriedade do compartilhamento dos seus dados para obtenção de financiamento das agências de fomento.

Tranquilo, já faço isso, não tem problema nenhum. Você se adapta né, você se adapta, “sim eu posso compartilhar, mas agora deixa eu publicar e vamos discutir”. Mas no final das contas você compartilha. É algo que eu já faço. (Pesquisador 1)

Sem problema, eu já compartilho eles todos. Eu não consigo fazer uma patente porque tudo é publicado imediatamente, então não tem problema. (Pesquisadora 3)

Já acontece, na minha área é assim, você tem o poder sobre o dado que você produz até a publicação, depois esse dado tem que ser público, porque por exemplo, quem financia as minhas pesquisas ou é o CNPQ ou Ministério da Saúde ou a própria Fiocruz... É dinheiro público, dinheiro público, não é meu. (Pesquisador 4)

Acho que normal, porque a gente compartilha mesmo. Então não tem problema né. (Pesquisadora 6)

Não, pra mim não tem problema não. Não teria problema, pra mim não. (Pesquisador 9)

Pra mim sempre foi uma condição de financiamento, porque eu sempre atuei tendo que disponibilizar os meus dados. O que eu acho que é grave é se essa imposição vier sem respeitar o período de publicação, aí eu acho que é uma coisa grave porque você torna ineficiente o sistema enquanto reconhecimento da ideia da elaboração do projeto, da coleta de dados e da análise primária dos dados. Então eu acredito que essa disponibilização ela faz parte do processo, mas ela deve respeitar a criação né, a criação da pergunta, a execução do trabalho, a execução da interpretação e da qualidade dos dados pra serem disponibilizados porque senão a gente vira só um produtor de dados e tanto faz se eu pensei ou não pensei no dado. (Pesquisadora 10)

No Brasil, algumas das agências de fomento à pesquisa possuem políticas que promovem a abertura dos dados de pesquisa. Desde 2017 a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) exige dos pesquisadores a submissão de um plano de gestão dos dados junto com o projeto para conseguirem financiamento. Nas Diretrizes para Planos de Gestão de Dados, entre as informações que devem constar no plano de gestão de dados é se haverá embargo, e se sim, de quanto tempo para a abertura dos dados produzidos.

A Fiocruz possui a Política de Acesso Aberto ao Conhecimento da Fiocruz, que visa garantir o acesso gratuito a toda produção intelectual produzida pela Instituição. De caráter mandatória, a Política refere-se ao depósito no Repositório Institucional Arca de teses e dissertações desenvolvidos e defendidos nos Programas de Pós-graduação da Fiocruz, além de artigos submetidos a periódicos científicos.

Neste documento é garantida o respeito ao tempo de embargo dos periódicos, só disponibilizando o acesso ao artigo somente após o cumprimento do prazo. Como incentivo ao cumprimento do depósito por parte dos pesquisadores, a Fiocruz oferece vantagens como a priorização na concorrência de editais interno.

No entanto, essa obrigatoriedade do compartilhamento de dados de pesquisa por parte das agências de fomento foi vista como positiva desde que os pesquisadores não tivessem suas publicações prejudicadas. E mesmo que os dados fossem disponibilizados, eles tivessem um tempo de embargo, não permitindo o uso por outros pesquisadores por um tempo determinado.

Eu não teria problema se isso não interferisse na publicação. Para mim não teria problema nenhum. Então que nem eu falei né, como a gente precisa da publicação para tudo, para conseguir a verba tudo, se isso não tivesse influência alguma nessa questão de impedir que isso seja publicado depois, eu não teria problema nenhum. (Pesquisadora 5)

Então, eu acho que assim, eu não vejo problema desde que... Desde que tenha essa salvaguarda para quem gerou o dado, quem gastou o dinheiro do projeto pra gerar o dado tenha as salvaguarda de ninguém publique antes. Pelo menos num tempo razoável né, antes de quem gerou o dado e colocou ali. (...) Tem que ter algum tipo de salvaguarda porque senão não é bom pra ninguém, ou seja, nem pro próprio [nome da agência de fomento]. (Pesquisador 11)

Ainda neste contexto da exigência do compartilhamento dos dados de pesquisa por parte das agências de fomento, os Pesquisadores 2 e 8 manifestaram contrariedade a essa obrigatoriedade, expondo que deixariam de pedir a verba.

Dependendo do que ele chama de compartilhamento de dados, tá. Falo porque é uma discussão que se deu, que está se dando dentro da [nome da fundação]. Um compartilhamento assim em tempo real pra nós não é viável tá, a gente não aceitaria. Porque colocaria o grupo e a nossa pesquisa em uma situação de fragilidade muito grande. (Pesquisador 2)

Eu provavelmente pararia de pedir fomento para aquela agência. Nesses termos eu pararia de pedir. Uma boa parte dessas agências já fazem isso. Se é possível fazer sem essa obrigatoriedade, sem disponibilizar a informação, eu acho melhor. Se a agência exigir que esse compartilhamento seja feito, é possível inclusive que muitos pesquisadores, entre eles eu, não entrem mais com pedido naquele fomento específico. (Pesquisador 8)

De um modo geral, os pesquisadores não veem problemas em cumprir a exigência de compartilhar os dados de pesquisa imposta pelos periódicos desde que alguns aspectos sejam assegurados, como é apresentado na fala dos Pesquisadores 5 e 11. Outros já veem essa medida como complexa ao ponto de abrirem mão do

recurso para o desenvolvimento dos seus projetos de pesquisa. O fato de ainda uma parcela considerável das agências de fomento brasileira não possuírem políticas ou iniciativas claras relacionadas a obrigatoriedade ou recomendações sobre a disponibilização dos dados de pesquisa e nem garantias aos pesquisadores que os produziram, como tempo mínimo de embargo e/ou que não comprometesse a publicação da pesquisa é um obstáculo para que os pesquisadores compartilhem seus dados de maneira mais efetiva.

Mesmo a Genética historicamente sendo conhecida como um campo aberto à partilha de conhecimento, é possível observar que essa abertura está associada às características do grupo de pesquisa e seus pares do que propriamente ao campo da Genética de forma generalizada.

Carvalho (2018) em sua pesquisa levantou a questão sobre a exigência do compartilhamento de dados de pesquisa por parte de editores de periódicos, instituição e agência de fomento, em resposta, os pesquisadores das áreas de Química, Antropologia e Educação apresentaram diferentes reações que foram desde aceitarem disponibilizar os seus dados de pesquisa até recusarem os recursos. As respostas dos pesquisadores no estudo de Carvalho (2018) são refletidas de maneira semelhante neste estudo.

Os resultados revelam que assim como em outras áreas do conhecimento, na Genética também é percebida uma resistência dos pesquisadores em abrir seus dados, mesmo que acarrete a perda de recursos. Também podemos observar que essa obrigatoriedade influencia diretamente nas práticas entre os pesquisadores.

Exigência de compartilhamento dos dados de pesquisa por parte da instituição

Os pesquisadores foram questionados como reagiriam se a instituição em que atuam exigisse o compartilhamento dos dados de suas pesquisas. Em resposta, o Pesquisador 1 demonstrou não ver problema nessa premissa por parte das instituições, porém, indicou que os pesquisadores poderiam omitir informações relacionadas a suas pesquisas.

Os Pesquisadores 5, 9 e 12 não veem problema em compartilhar os dados uma vez que esses já estão publicados. Já a Pesquisadora 6 concorda com o

compartilhamento dos dados de pesquisa desde que a instituição respeite o tempo que se leva para publicar os resultados gerados pelos dados.

É também, tranquilo. Mas assim é algo difícil de controlar. Como a instituição vai controlar o compartilhamento de dados que você está gerando no seu laboratório escondido com seus alunos ali? Quem é que tá olhando o que você tá fazendo? Só vai descobrir o que você fez depois que você já publicou, aí já tá publicado. (...) As instituições podem exigir, por exemplo, que todo projeto que tá fazendo no laboratório seja passado primeiro pela Universidade para ser aprovado, tenha relatórios e tudo mais, mas mesmo assim as pessoas podem omitir várias coisas. (Pesquisador 1)

Vou te dizer que eu não tenho muita simpatia assim pela palavra “obrigatório”, assim, não é uma coisa muito simpática né, mas eu por minha própria.... pelo meu jeito de ver, eu não me importaria. Porque eu acho que a ciência é para ser compartilhada, é a gente é um grupo né (...) Claro isso é uma pergunta que tá me fazendo de forma bem ampla né, mas de forma geral eu tenho simpatia pelo compartilhamento. (Pesquisadora 5)

Eu acho que isso, dentro da nossa profissão, acho que é uma condição mesmo. Ninguém faz pesquisa pra guardar, você faz pesquisa pra divulgar né. Então eu acho que não tem problema com tanto que eles respeitem né nosso tempo pra divulgação. Mas eu acho que é nossa obrigação divulgar os resultados de pesquisas né, então acho que não tem problema. (Pesquisadora 6)

Não teria problema não. Uma vez que o dado tá obtido, tá obtido. Não tem problema nenhum. (Pesquisador 9)

Eu acho que isso é tranquilo depois que isso já tá publicado numa revista internacional. (Pesquisadora 12)

Todavia, essa exigência por parte das instituições não foi vista com bons olhos por todos os pesquisadores entrevistados. Os Pesquisadores 7 e 8 demonstraram não ser favoráveis a essa condição caso fosse imposta pela instituição apresentando como justificativa a presença de dados sensíveis que podem ser produzidos ao longo dos projetos e a competitividade da área, principalmente quando se comparado com outros países.

Daí eu já não sei se eu sou tão a favor da instituição. Porque diante do comitê de ética, por exemplo, o comitê de ética vai avaliar e vai dizer que tem que ser aberta, eu não sei. Daí nesse caso a instituição... Eu acho que teria que ser facultativo, eventualmente fazer alguma sugestão de política institucional que os dados sejam abertos etc, que seja colocado. Mas acho que não tem que tornar obrigatório, como eu falei, vai ter projetos que são sensíveis, daí tornar obrigatório essa informação eu acho que não seria bom. (Pesquisador 7)

Ficaria bastante insatisfeito com essa obrigatoriedade (...) Só colocando aí em termos pandêmicos, eu tenho ali no laboratório a melhor vacina dentre todas essas que estão aí, só que eu preciso de mais dois anos de experimentos, correto. Seis meses se tiver mais

investimentos. “A não, você é obrigado a compartilhar”, no momento que eu compartilhar essa informação, todos os países desenvolvidos transformam essa minha vacina realmente na melhor vacina do mundo em um curto espaço de tempo. Então acho isso uma tremenda de uma bobagem essa exigência. (Pesquisador 8)

Como exposto pelo Pesquisador 8, sobre as diversidades enfrentadas para se produzir ciência, levando em consideração as práticas científicas defendidas pela Ciência Aberta, um dos objetivos desse movimento é a luta contra as desigualdades estruturais e ao acesso ao desenvolvimento científico independente da região ou país (AGUIA, 2021). Mesmo o movimento Ciência Aberta reconhecendo que estes problemas relatados pelo Pesquisador 8 não tem solução, eles são reconhecidos e discutidos também em outras estruturas científicas que compõem a Ciência Aberta, como em cadernos abertos de laboratório.

De modo geral, os pesquisadores não se oporiam ao compartilhamento dos dados caso a instituição exigisse. No entanto, assim como apresentado no tópico anterior, os pesquisadores expõem ressalvas para realizar esse compartilhamento, como o embargo temporário dos dados. Outro ponto levantado pelo Pesquisador 1 é a omissão das informações pelos pesquisadores a fim de preservar suas pesquisas.

Nesse sentido, políticas de gestão de dados poderiam auxiliar na quebra dessas barreiras. No Manifesto do IbiCT é recomendado a elaboração de políticas institucionais de caráter mandatório garantindo que os dados de pesquisa produzidos pelos pesquisadores sejam adequadamente gerenciados, porém paralelamente estimular iniciativas de Ciência Aberta e incentivar discussões sobre a importância do compartilhamento e uso dos dados de pesquisa.

Exigência de compartilhamento dos dados de pesquisa por parte dos editores de periódicos

Quando perguntado aos pesquisadores sobre o que achariam se os editores de revistas exigissem que os dados da pesquisa fossem disponibilizados junto à publicação, a resposta dos entrevistados foi favorável, uma vez que essa prática já é recorrente por parte dos editores de periódicos. A única ressalva feita pelos Pesquisadores 2, 5 e 6 foi referente a disponibilização dos dados no momento da submissão do artigo, antes do trabalho ser aceito para publicação.

Algumas revistas já fazem isso. Esses dados de sequenciamento, muitas revistas elas pedem para você depositar os dados brutos num repositório de dados brutos, já outras não. Outras falam simplesmente coloco na opção “eu tenho os dados e disponibilizarei os dados se alguém pedir”. Mas já outras obrigam, se você quiser publicar lá, você tem que disponibilizar os dados brutos para alguém poder usar de novo. Então eu acho tranquilo, já acontece. (Pesquisador 1)

Não, isso já é normal, isso acontece o tempo todo, isso não é problema. Talvez no momento da submissão também a gente precisaria avaliar um pouco. (...) Eles solicitam que seja depositado no repositório, mas não precisa tá liberado. Quando são trabalhos que tem um componente de informática muito forte o que algumas revistas solicitam e que você libere o acesso pros revisores tá, para eles conseguirem ver o dado cru, mas isso é o máximo. Em geral nenhuma revista solicita que o dado esteja aberto totalmente na hora da submissão. (Pesquisador 2)

Mas isso é obrigatório. A gente não consegue nem passar do editor se não mandar os dados. Eles perguntam “cadê os dados? Esses dados depositam em tal lugar, esse dado deposita em tal lugar”. Você já pra submeter o trabalho tem que ter depositado os dados, se não eles nem aceitam. Assim, também os revisores podem olhar os dados brutos, ver se você fez aquilo que você está falando. (Pesquisadora 3)

Também não vejo problema, porque já vai tá sendo publicado né. E aí para as pessoas eu acho que acessarem e ver como é que foi feito também não, não, não vejo problema. Na submissão... eu tava pensando assim o seguinte publicou e aí fica como sei lá, o material complementar com os dados brutos... pra mim é diferente. Agora no momento da submissão não me agrada muito não. (Pesquisadora 5)

Durante a submissão eu acho meio complicado né, acho que daí teríamos que dar uma conversada com eles, mostrar que não é uma boa ideia compartilhar antes deles aceitarem para publicar né. Acho que não, acho que daí eu teria que dar uma discutida. (Pesquisadora 6)

Isso eu concordo, ou seja, se você chegou no momento que está submetendo um artigo, tá tentando publicar aquele artigo, se você quer transformar aquela informação em conhecimento público aí eu acho que você tem que escancarar todos os seus dados. Com atenção, são perspectivas diferentes tá, chegou na conclusão que o artigo científico tem que ser publicado, então você tem que abrir a janela de todos os seus dados. Porque aí você está transparecendo a ciência que você fez para chegar naqueles resultados. (Pesquisador 8)

De um modo geral, todas as revistas na área, na área que eu atuo, elas pedem que os dados já estejam depositados no [nome do banco de dados] e que a gente tenha um código de acesso. Esse código de acesso ele normalmente não é liberado no momento da submissão da publicação, ele só é liberado uma vez o artigo sendo aceito. (Pesquisadora 10)

Ah, isso não tem problema, se eles quiserem. Isso é, isso é até meio... não é rotineiro, mas isso acontece eles pedirem as planilhas, pedirem os dados brutos para refazer estatística. Quanto a isso não tem nem um problema assim, eu mesmo como referee de artigos se eu achei muito estranho eu peço, eu peço pra refazer estatística, eu peço pra mandar os dados. (Pesquisadora 12)

Em relação aos editores de periódicos exigirem o compartilhamento dos dados, todos os pesquisadores respondentes reagiram com naturalidade, visto que já é uma prática corriqueira em suas respectivas áreas, tendo como única ressalva o fato da exigência ser feita no momento da submissão. Essa aceitação também é percebida na pesquisa de Carvalho (2018), os pesquisadores das áreas de Química e Educação tiveram algumas ressalvas em relação à obrigatoriedade, mas se mostraram bem abertos a fazê-la, diferentemente dos pesquisadores da Antropologia que reagiram com anormalidade.

Destino dos dados produzidos após a finalização da pesquisa

Em relação ao destino dos dados após a finalização da pesquisa, os pesquisadores apontaram diferentes produções, sendo elas teses, dissertações, trabalhos de conclusão de curso, artigos, capítulos de livros e apresentações em congressos. Os dados “brutos” costumam ser mantidos nos bancos de dados dos grupos, em HDs, drive, em bancos de dados internacionais.

Publicação. Artigos científicos, teses, capítulos de livro, artigos de revisão, relatórios de pesquisa. Mas eu tento publicar artigos científicos, na minha área é muito importante artigo publicado. Não livro, livro não. (Pesquisador 1)

Geralmente publicação, teses, dissertações, artigos. (Pesquisador 2)

Olha, geralmente os meus alunos, eles saem geralmente com uns dois artigos do doutorado, com alguma sorte um artigo do mestrado. (Pesquisador 4)

É ou uma monografia de graduação ou uma dissertação ou uma tese. Esse é o primeiro destino e depois a publicação em periódicos científicos mesmo. (Pesquisadora 6)

Eles ficam em bancos de dados do grupo, acaba ficando uma coisa mais fechada, restrita. Ou no drive ou em arquivo físico HDS eu diria. (Pesquisador 7)

Quase sempre, têm a tese, a dissertação associada, porque quem faz a pesquisa no fundo são os alunos. (...) aí torcendo pra que algum produto saia daí, um artigo científico ou uma patente consiga está relacionada com o desenvolvimento dessa pesquisa. (Pesquisador 8)

Ficam arquivados, ficam arquivados. No banco de dados e no laboratório, em geral é assim. (Pesquisador 9)

Então eu tenho um sistema de armazenamento de todos os alunos que passaram pelo laboratório e eu tenho registro de caderno, ata de cada aluno que passou pelo laboratório. (Pesquisadora 10)

Então, nessa área de Genética, eles realmente são mantidos nesses bancos todos né, então assim, o dado principal, tá todos nesses bancos. (Pesquisador 11)

Apresentar em congressos e submeter pra revistas internacionais pra publicação. (Pesquisadora 12)

Outros aspectos citados pelos pesquisadores foram a discussão sobre o acesso livre aos dados, o seu uso sem citar a fonte e o medo de perder publicação/produtos.

Como discorre o Pesquisador 2, o compartilhamento é uma preocupação entre os pesquisadores, pois os países de primeiro mundo tem um poder de processamento que não pode ser acompanhado devida a burocracia de compra dos materiais e o altos custos, além desse fator, o uso dos dados sem citar os pesquisadores que o produziram também foi relatado pelos Pesquisadores 2 e 10. O Pesquisador 8 expressou acreditar que os colegas de sua área não irão divulgar os seus dados integralmente por temerem perder para outras publicações ou oportunidades de gerar algum produto que pode ser retornado à sociedade.

Acho que a grande discussão que hoje tá tendo é isso é sobre o acesso livre a dados. O que nós como pesquisadores ficamos muito preocupados com essa ideia de compartilhar dado quase que em tempo real digamos né, que tá sendo produzido, isso realmente é uma questão que entendemos que isso fragiliza. Isso acabou de acontecer, houve dados relacionados com o Genoma Covid checados por grupos do Brasil que foram utilizados em uma publicação por grupos europeus, publicada obviamente antes que o grupo do Brasil e os grupos locais não foram nem citados, nem agradecidos, nem em agradecimento. Então é uma situação que é real, é real isso. Novamente, volto a falar, há uma velocidade de processamento, de capacidade de trabalho do primeiro mundo que a gente hoje não consegue equiparar, não por falta de qualidade das pessoas, como falei, por causa da logística, pelos custos. (Pesquisador 2)

O que me chateia às vezes, é que as pessoas não citam o trabalho de origem, então isso, não dão crédito ao trabalho de origem. Isso eu acho que é uma coisa que volta e meia acontece e que talvez a comunidade científica precise ganhar clareza de como citar esses dados. (Pesquisadora 10)

Eu acho que a maioria dos pesquisadores vai recuar tá. Não vai querer divulgar totalmente a informação, mas eu acho que com interesses distintos. Alguns não vão divulgar com medo de perder a publicação e outros, como eu, não vão divulgar com medo de perder a chance de gerar um produto que possa ser disponibilizado para sociedade em forma de produto, entende? Mas os dois vão recuar, os dois vão fazer de tudo para não divulgar aquela informação, sabendo de todas as dificuldades que é gerar o conhecimento para que outros transformem em produto. (Pesquisador 8)

Os resultados mostram que os dados de pesquisa têm diferentes destinos após a finalização de uma pesquisa, mas a principal é publicar os resultados em algum periódico. Essa cultura nas Ciências da Vida é discutida no estudo da Key Perspectives Ltd (2010), normalmente os dados viram artigos, mas também podem ser disponibilizados em sites dos grupos ou em blogs e/ou nas wikis dos pesquisadores. Existem situações como é abordado no trabalho Key Perspectives Ltd (2010) em que os dados são tão específicos que não existem bancos de dados para eles, o que faz com que a maior parte desse tipo de dado seja mantido em sistemas locais (computadores, drives, programas, entre outros).

Ao longo da discussão sobre o compartilhamento dos dados de pesquisa, em diferentes momentos foi levantada pelos pesquisadores receios sobre o uso indevido dos dados, a perda de ineditismo da pesquisa, a falta de citação do trabalho ou do pesquisador que deu origem àqueles dados. Isso mostra que não existem boas práticas relacionadas à gestão dos dados de pesquisa, isso pode se dar por vários motivos: as complexidades de cada subdisciplina da Genética, falta de conhecimento dos pesquisadores de tais instrumentos, falta de incentivo das instituições e assim por diante.

O desenvolvimento de boas práticas principalmente em um campo como a Genética, traz mais eficiência e transparência e conseqüentemente mais segurança aos pesquisadores. Problemáticas aqui citadas poderiam ser sanadas por meio de aplicação de medidas simples, como a dos Princípios FAIR, que garantem que os dados sejam localizáveis, acessíveis, interoperáveis e reutilizáveis através de uma descrição rica dos seus metadados.

5.1.3 Reuso dos dados de pesquisa

O enfoque desta seção são os aspectos ligados ao reuso dos dados de pesquisa. Pontos como a experiência em reutilizar dados de pesquisa, a percepção dos pesquisadores sobre o tema, a utilização de dados produzidos por outros pesquisadores, aspectos que favorecem ou dificultam o reuso de dados e a disposição para o reaproveitamento deles pelos pesquisadores.

Experiência na reutilização de dados produzidos por outros pesquisadores

Os participantes foram questionados sobre a sua experiência em reutilizar dados de pesquisas gerados por outros pesquisadores, neste ponto foi possível perceber que os pesquisadores não tinham clareza sobre a prática de reuso de dados, mesmo assim, afirmaram estarem acostumados com essa prática.

A pesquisadora 3 declarou que reutiliza dados retirados de repositórios, mas expôs a dificuldade de integração de dados coletados de bancos de dados diferentes devido às variadas metodologias existentes. Foi possível observar também que os participantes entendem que qualquer interação com dados gerados por pesquisadores de fora do seu grupo é um tipo de reuso. Vale a pena ressaltar que a falta de conhecimento por parte dos pesquisadores sobre a reutilização dos dados dificulta a análise neste ponto.

Sim, de trabalho científico que tem esses dados de patogenicidade e de bancos de dados.
(Pesquisadora 1)

Olha, a gente trabalha muito com os dados do [nome do banco de dados], muito tá, então são retirados do repositório. Agora existem outros, então não é muito simples você integrar dados de diferentes plataformas né. Então coisas que foram feitas separadas e que foram feitas com metodologias das vezes diferentes, é muito difícil você ter dados exatos ou que possam ser comparados. Normalmente você tem que fazer normatizações, aproximações, então isso torna as coisas um pouquinho mais difíceis. (Pesquisadora 3)

Já tive experiência, eu não sei se chamaria de reutilizar, de fazer metanálise. (Pesquisador 7)

A gente trabalha com outros bancos de dados, por exemplo, se eu quero estudar uma característica (...) Existem bancos de dados que são especializados, então a gente tem o [nome do banco de dados] que é um grande repositório. Mas a gente tem outros bancos de dados, por exemplo, como o [nome do banco de dados] que é um banco de dados de referência de proteínas de domínios proteicos que são aspectos relevantes às vezes do conjunto de moléculas que a gente tá estudando, então acesso ele. (Pesquisadora 10)

Entre os pesquisadores que relataram já terem feito reutilização de dados de pesquisa, foi constatado que os bancos de dados e repositórios são as principais fontes utilizadas para a busca dos conjuntos de dados para o desenvolvimento das suas pesquisas. A seleção do banco de dados e/ou repositório mais adequado ocorre conforme as características do estudo em desenvolvimento, como exposto pela Pesquisadora 10.

A predileção por utilização de bancos de dados e repositórios pelos pesquisadores da Genética podem ser vistos por dois enfoques. O primeiro está

relacionado às vantagens que esse tipo de sistema oferece aos pesquisadores, em particular a recuperação mais ágil e acurada dos datasets (SØNDERGAARD, ANDERSEN E HJØRLAND, 2003). O segundo é o reconhecimento e relevância de um determinado banco de dados em uma área específica, como por exemplo, o Genbank que é um banco de dados de nucleotídeos reconhecido mundialmente. Bancos como o Genbank possuem seus metadados bem descritos e estruturados e como relatado no Key Perspectives Ltd (2010), pesquisadores das Ciências da Vida que utilizam esses sistemas possuem familiaridade com seus descritores.

Percepção do reuso de dados de pesquisa produzidos por outros pesquisadores

O reuso de dados produzidos por outros pesquisadores é percebido pelos pesquisadores como algo natural. De acordo com as respostas dos pesquisadores, a reutilização de dados gerados em outras pesquisas tem utilidade fundamental para a realização de análises comparativas de dados.

Quando eu faço as análises... que eu faço de vírus novos, eu tenho obrigatoriamente de analisar os dados dos outros, de quem já publicou aquele momento. E esses dados estão nesses repositórios, entendeu. (...) Então é muito comum você utilizar dados de outros pesquisadores do passado para você publicar o seu trabalho (...) A parte teórica né, do que os outros já fizeram, os dados das outras para comparar com seu dado. (Pesquisador 1)

Sem dúvida nenhuma né, a gente já reutiliza hoje em dia pra principalmente pra validar os nossos dados. Mas à medida que aumenta a quantidade de dados disponíveis né, cada vez mais o que precisa ser feito é a geração de algoritmos ou programas que nos permitam a encontrar o que já existe né, e nesse aspecto a gente está engatinhando tentando assim pensar em machine learning. (Pesquisador 2)

A gente faz isso rotineiramente. Eu acho que, eu só tinha um aluno que não ia utilizar dados brutos produzidos por outros grupos de pesquisa, mas veio a pandemia e a gente teve que mudar o projeto dele. (Pesquisador 4)

Olha, a gente não tem essa prática né, de usar os dados, mas sim de utilizá-los para discutir né, de confrontar com os nossos, é nesse sentido né. Então a gente tem muito isso, toda vida buscar grupos de pesquisa que trabalhem com os mesmos compostos ou com o mesmo problema nosso e tentar confrontar e ver se a gente corrobora os dados deles ou não. É nesse sentido sim, nós fazemos isso frequentemente né. E utilizar dados nossos mesmo. (...) Agora de fora, eu acho que é mais para comparar mesmo, pra discutir, pra tentar contribuir pra ciência de modo geral. A gente não costuma utilizar de outros pesquisadores não. (Pesquisadora 6)

Eu acho que é uma das práticas bem do momento até eu diria, porque a pandemia atrapalhou a aquisição de dados. Então acho que o pessoal tá olhando pra os dados já gerados pra conseguir... Acho até que isso tenha acelerado suas práticas. (Pesquisador 7)

Sim, de novo é a prática corrente. Então isso é uma coisa que a gente faz também, direto né. (...) Então eu tô falando de uma área que eu tenho bastante conforto porque faz a prática assim tá, de uso e reuso. (Pesquisadora 10)

Na nossa área é total. (...) essa necessidade de tornar público os dados originais né, inclusive os dados já trabalhados e acessível a qualquer um, acelerou isso mais ainda e deixou mais claro do quão importante isso é necessária pra a área. (...) Cada novo conjunto de dados acrescenta mais, ou seja, você já tá fazendo análise de grandes grupos, sei lá, pegar um exemplo, eu tô fazendo uma análise de um grupo lá de baleias, espécies de baleias, tem várias outras espécies de baleias se eu... Eu não tenho acesso a todas e nem tem dinheiro pra produzir informação sobre todas elas, mas o tipo de informação é o mesmo. Então se eu puder comparar o meu conjunto com várias outras espécies, produzidos em vários artigos diferentes, por vários autores diferentes e fazer uma comparação mais completa, a qualidade ganha enormemente né e sem precisar repetir tudo de novo desnecessariamente. (Pesquisador 11)

Conforme a maioria das respostas dos pesquisadores, observou-se que a percepção do reuso dos dados de pesquisa é comum em suas áreas de atuação dentro da Genética. Entre os respondentes, a Pesquisadora 6 foi a única a relatar não perceber esse reuso dos dados para além da comparação e discussão entro os dados do seu grupo e dados de outros grupos. A pandemia foi vista pelo Pesquisador 7 como agente motivador para o aumento da prática de reuso de dados entre eles. Mais uma vez foi possível evidenciar pelas falas dos entrevistados a falta de clareza sobre o reuso de dados de pesquisa.

Nas falas dos pesquisadores foi notado os benefícios que o reuso dos dados podem proporcionar para a Ciência, como a economia de recursos, permitir novas discussões por meio da comparação dos dados de pesquisa, maior quantidade de dados disponíveis, acelerar a pesquisa e entre outras vantagens. Esses aspectos indicados pelos pesquisadores corroboram com as discussões levantadas pelos autores Pienta, Alter e Lyle (2010), Santarem Segundo (2020) e Sielemann, Hafner e Pucker (2020) relacionadas aos benefícios gerados pela reutilização de dados de pesquisa. Conforme Costas (2017) discorre, o reuso dos dados tem inúmeros desdobramentos, tais como: celeridade a pesquisa, redução de custos, reprodutibilidade de métodos, discussões acerca de conjuntos de dados existentes, aplicação em outras pesquisas e validação de dados e seus resultados.

No que tange a localização dos dados, o Pesquisador 2 trata esse fato como facilmente sanado pela utilização de programas e machine learning. Com tudo, para que os conjuntos sejam descobertos são necessárias outras ações além da utilização de tecnologias, os dados precisam ter uma curadoria, ou seja, possuir um conjunto de requisitos previamente padronizados e ser bem documentados, assim será possível localizar, acessar, processar e conectar os conjuntos de dados de pesquisa a outros (TORINO; TREVISAN; VIDOTTI, 2019).

A aplicação dessas práticas amplia o reuso de dados, mas o efetivo reuso só é possível se os pesquisadores compartilharem seus dados amplamente e reutilizar dados produzidos por outros pesquisadores. Conforme Shen (2015) discorre sobre as práticas de compartilhamento e reutilização de dados, entre os 423 respondentes, 44% indicam que seus dados podem ser reutilizados por pesquisadores da mesma disciplina e 28% relataram que os seus dados podem ser reutilizados por pesquisadores de outras áreas, contudo, mais da metade dos entrevistados não reutilizam dados de outras pesquisas ou campos do conhecimento.

Características similares podem ser percebidas na pesquisa de Carvalho (2018), onde o reuso de dados da Química não é uma prática comum entre os pesquisadores e nem há disposição para o reuso. Na Antropologia o reuso se concentra em áreas semelhantes. Na Educação a autora observou que o reuso de dados de pesquisa é importante, contudo não existe disposição para o reuso de dados gerados por outros pesquisadores, a não ser que os dados tenham sido obtidos por integrantes do mesmo grupo de pesquisa. Traços análogos podem ser vistos na Genética, como é relatado pela Pesquisadora 6.

Fatores que contribuem para a reutilização de dados de pesquisa

Em relação aos elementos que contribuem para o reuso dos dados pelos pesquisadores, foi visto que diferentes fatores os encorajam a reutilizar dados já produzidos. Os principais motivos observados foram:

- Comparar seus dados com outros produzidos anteriormente para discussão de resultados;
- Validar modelos em desenvolvimento;
- Praticidade para a pesquisa;

- Custos elevados e falta de verba;
- Poder explorar e compreender o funcionamento dos bancos de dados;
- A qualidade dos dados que está sendo utilizado

O principal fator é o fator de comparação. Como é que eu vou propor alguma coisa se eu não analisei o que os outros já fizeram? Então eu tenho que discutir os meus resultados com base no que os outros já fizeram. (Pesquisador 1)

Você confirmar dados que você tá tendo, aumentar o universo da tua amostragem, confirmar em outros modelos ou situações que você vê no modelo que você tá trabalhando. (Pesquisador 2)

Primeiro lugar facilidade, segundo lugar, custo né. Então, você quer ver o perfil de expressão gênica de mil tumores, você fazer o RNA SEQ de mil tumores é caro pra chuchu, é muito caro. Então assim, cem tumores RNA SEQ dá 50 mil reais, mil tumores dá 500 mil reais, então é caro, é muito caro. Se ele já existe, se esse dado já existe nas plataformas né, nos repositórios né, os bancos de dados porque não utilizá-los? (Pesquisadora 3)

Falta de verba para fazer trabalho de bancada. Então a gente tá tendo que ser bastante criativo nesse momento tão assim, vamos pensar “que fazer?”. Então assim a gente tá tendo que ser bem criativo né... (...) Então a gente tá particularmente numa época que tá se fazendo muito isso, porque a gente tem que ter alternativas e a ciência não pode parar, então vamos trabalhar com que a gente tem né. (Pesquisadora 5)

É aprender melhor sobre os bancos públicos, conhecer melhor, quais são que estão disponíveis, dominar um pouco melhor as ferramentas de análise desses dados, saber explorar esses bancos de dados. Acho que seria isso aí. Eu tenho a intenção de usar mais esses dados, eu ainda não aprendi a fazer isso. (Pesquisador 7)

Quase sempre é a qualidade do trabalho esboçado em grandes periódicos de circulação internacional. Quer dizer, a gente acredita que quanto melhor o periódico, mais confiável é o trabalho daquele pesquisador e obrigatoriamente maior a chance de você ter sucesso no uso dele durante a sua pesquisa. (Pesquisador 8)

É o controle de que o comportamento do dado ele é coerente. Então acho que isso é muito importante na grande área biológica, é você ter claro que o que você tá analisando não é um fenômeno particular do experimento que você desenhou. Então o fato de você ter respostas macros que são consistentes entre diferentes experimentos é relevante. E o outro aspecto é você ter um, que eu também mencionei, é você poder ter uma inspiração. (Pesquisadora 10)

Em relação aos fatores que estimulam o reuso dos dados de pesquisa, os resultados obtidos mostram que os pesquisadores da Genética reutilizam os dados produzidos anteriormente para comparar, validar e discutir os seus próprios dados. O reuso dos dados também é visto pelos Pesquisadores 2 e 10 como um meio de legitimar os processos utilizados em seus estudos e garantir que os seus dados tenham qualidades.

Da mesma forma como citado nos tópicos de produção e compartilhamento de dados de pesquisa, o desenvolvimento de pesquisa na Genética é custoso e o reuso é uma forma de economizar recursos que são escassos. Apesar do reuso de dados de pesquisa trazerem benefícios para aqueles que os reutilizam e para os que o produziram, diante das falas dos pesquisadores, o reuso dos dados não é claramente entendido por eles, pois só o uso comparativo do dado não garante a sua realização propriamente dita do dado.

Corroborando com essa afirmativa, Park e Wolfram (2017) salientam que campos como a Genética e Hereditariedade possuem um grande quantitativo de registros no *Data Citation Index* (DCI), ainda assim, o reuso de dados é pouco praticado entre os pesquisadores. Além disso, a pouca literatura científica focada nos processos relacionados ao reuso dos dados e a sua gestão sob a ótica dos pesquisadores favorecem que esse tema continue nebuloso no entendimento dos pesquisadores (DIAS, ANJOS E ARAÚJO, 2019).

Fatores que inibem a reutilização de dados de pesquisa

A principal razão do não reuso dos dados, de acordo com os pesquisadores respondentes, é a falta de qualidade apresentada pelos dados. Foram vários aspectos relacionados com a qualidades dos dados que os pesquisadores mencionaram, tais como:

- A estatística utilizada nos trabalhos não ser pertinente;
- O volume de dados não ser adequado;
- A metodologia não ser clara;
- Possuir viés na coleta dos dados;
- O meio de divulgação e/ou quem gerou os dados (encontros, periódicos, laboratórios, grupo de pesquisas) não serem de confiança;
- Falta de informação na documentação dos dados.

Só a qualidade dos dados. Se a estatística não está relevante, se não é um número mínimo necessário de amostra, como foi gerado o dado. (Pesquisador 2)

Baixa qualidade e um volume pequeno, aí não adianta... Não tem o que fazer. (Pesquisador 4)

Não me lembro especificamente de algum caso, mas quando a gente vai pegar esses dados a gente sempre dá uma checada, vê qual revista tá publicada, se é uma revista rigorosa né, se tem um método adequado. (Pesquisadora 5)

Talvez a gente não ter certeza se as metodologias utilizadas por aquele grupo foram as mais adequadas. Se a gente não tiver acesso né a forma como foram obtidos os dados. (...) Não dá para pegar qualquer dado da literatura e utilizar, a não ser que você tenha bastante confiança no que você tá usando né. E também, é lógico, que eu acho que é uma questão ética aí né tem que ter a concordância lógico, você não vai utilizar um dado de outro grupo sem eles concordarem com isso. (Pesquisadora 6)

Eu acho que a baixa confiança. Se eu tiver alguma dúvida em relação a qualidade dos dados, eu não usaria se o dado não é confiável. Se tem viés ali em relação a coleta. Acho que é mais a questão da confiabilidade da informação. (Pesquisador 7)

Ele [dado] estando divulgado em encontros duvidosos, como os papers, a gente procura evitar, “há mais só tem aqueles dados”, aí a gente até tenta refazer os dados, mas a reputação da qualidade dos periódicos é que determina se usa ou se não usa. (Pesquisador 8)

Se eu tiver suspeitas de que os dados não são bons. Se eu achar que... Porque às vezes a gente conhece o laboratório que tá gerando o dado, e a gente tem uma opinião ruim, aí nesse caso eu não uso. Não uso nada do laboratório. Se eu achar que não é sério e aí eu me reservo o direito (...) Mas se eu achar que um determinado laboratório não tá fazendo, enfim, tá inventando, aí não vou usar o dado deles. (Pesquisador 9)

Quando os dados não estão claramente descritos, quando eu não sei informações básicas do experimento eu não posso usar, assim, não dá pra a gente usar. Então existem alguns datasets que os alunos usam e que se a gente identifica que informações relevantes não constam para uma interpretação que justifique uma interpretação de resultados comparativa a gente não usa. (Pesquisadora 10)

A qualidade e o nível do periódico com certeza. A gente sabe hoje, aí as revistas predatórias. Hoje em dia artigo publicado não quer dizer artigo bom né, e também não é só porque a revista é paga. Tem revista paga que é boa e tem revista paga que é péssima. Então isso eu acho que é muito importante a qualidade, o nível da publicação, talvez o histórico do grupo, quem é essa pessoa, até onde eu sei se ela é confiável ou não. Se eu fosse usar eu acho que seriam coisas que eu levaria em consideração, sabe. (Pesquisadora 12)

Os resultados mostram que quase todos os fatores inibidores do reuso dos dados de pesquisa listados pelos pesquisadores têm relação com a falta de curadoria dos dados que têm a ver com a sua gestão. Problemas análogos são notados entre os pesquisadores de Química, Antropologia, Educação e Engenharias (CARVALHO, 2018; KEY PERSPECTIVES LTD, 2010).

Neste contexto, infere-se que dados não acurados pelos seus produtores dificultam o seu reuso em novas pesquisas. Segundo Sales e Sayão (2012) a

curadoria dos dados proporciona tratamento técnico e avaliação de qualidade, além disso, como expõe os autores, a curadoria facilita o acesso persistente aos dados através das tecnologias, estratégias e metodologias aplicadas assegurando o uso dos dados no futuro.

O DataOne (2013) em seu modelo de ciclo de vida de dados de pesquisa, lista 8 etapas com procedimento para garantir a curadoria dos dados, além da sua preservação e reutilização futura. Com base nas respostas obtidas e no modelo do DataOne é possível sugerir algumas ações que promovam a reutilização dos dados na Genética.

- **Planejamento:** percebe-se que não existe uma clareza de gerenciamento sobre a produção dos dados. O planejamento permite levantar todos os processos realizados no decorrer do projeto, permitindo definir políticas, planos de sustentabilidade dos dados e seu gerenciamento. Tal atividade é importante pois impacta no compartilhamento, preservação e reutilização dos dados a longo prazo.
- **Assegurar qualidade dos dados:** foram verificados elementos relacionados à qualidade dos dados, como por exemplo, dados com vieses e/ou meio de divulgação não confiável. Sendo assim, definir critérios para avaliar e controlar a qualidade dos dados é essencial para impulsionar o reuso por pesquisadores.
- **Documentação dos dados:** Entre as respostas, observou-se a importância da documentação dos dados. Metadados descritos de forma rica favorecem a recuperação e o reuso dos dados em novas pesquisas. Além de propiciar um armazenamento com precisão adequada.

Disposição para o reuso pelos pesquisadores

Quando perguntados sobre a disposição de reutilizarem os dados produzidos por outros pesquisadores, as respostas foram positivas. Destaca-se, especialmente, a fala do Pesquisador 1 em relação ao reuso dos dados, onde o pesquisador que produziu o dado e disponibiliza a outros pesquisadores exige que o seu nome apareça nos trabalhos por ver o dado como propriedade particular.

O reuso dos dados pelos pesquisadores é entendido pela Pesquisadora 6, como não sendo algo claro para ela, mas acredita que entre os pesquisadores de sua

área existiria uma resistência à reutilização de dados gerados por outros pesquisadores. A Pesquisadora 10 também destaca que o reuso de dados está ligado principalmente à área de atuação do pesquisador.

Sim, isso é algo muito comum. Muito comum mesmo, você utilizar dados de outros pesquisadores nas suas pesquisas. A grande maioria não, tem gente que segura, como eu falei para você, porque ele quer ser o primeiro a publicar. Se ele já publicou alguma coisa aí ele compartilha o dado, aí deixa você usar. Mas tem muita gente que se acha dono do dado também né, ele vai te dar, mas quer o nome dele em qualquer trabalho que a partir daquele momento que ele te deu aquele dado, ele tem que aparecer nos trabalhos. Mas isso não é justo né, porque você só utilizou o dado, não usou todo seu intelecto para fazer uma outra análise, mas tem gente que mais ou menos exige que você acrescente o nome da pessoa quando ele te dá um dado. (Pesquisador 1)

É o que a gente faz né, a gente parte de um trabalho publicado, de vários trabalhos publicados para fazer as nossas hipóteses e testar. Então a gente na verdade está se baseando em dados que foram obtidos por outros pesquisadores, mesmo que você não esteja usando eles propriamente dito, você está se baseando neles. Nas pesquisas a gente faz reuso dos nossos próprios dados porque a quantidade de dados é tão grande que você usou ele para responder uma pergunta, mas tem várias outras que você poderia usar esses mesmos dados para responder e normalmente você deu aquela pergunta para um aluno, o aluno fez lá o mestrado, o doutorado dele e foi embora e os dados ficam aí jogados. (Pesquisadora 3)

São tranquilos, não veem problema em utilizar outros dados que já foram produzidos anteriormente. Só tem que ver de qual repositório você vai pegar. Você pode pegar de um repositório que não tenha um controle de métricas ou você pode pegar também e fazer seu próprio controle, falou assim “eu só vou querer trabalhar com amostras que tem um milhão de volume de informação lá do dado bruto”, é um critério, mas eu acho que a dificuldade é saber filtrar. (Pesquisador 4)

Eu acho que, como eu falei, tá tendo bastante (...) Dentro da Genética, o pessoal apresenta os resultados... todo mundo indo mais para esse lado em função que a falta de verba despertou isso né da gente buscar esse caminho, então eu acho que é bem frequente. (Pesquisadora 5)

Não sei, eu nunca discuti muito isso, mas eu acho que existe na minha área uma certa resistência, eu não vejo muito essa utilização. (Pesquisadora 6)

Acho que a tendência é cada vez mais de usar dados, assim não é usar só esses dados, mas cada vez mais utilizar dados produzidos por outros pesquisadores. Esses grupos mais jovens que eu te falei, principalmente no contexto da pandemia, vários utilizaram os bancos que mencionei pra análise e vários artigos publicados com base nesses dados gerados pelos outros grupos. Isso cresceu bastante. Pesquisador 7)

Há com muita frequência. Se usa dados de outros pesquisadores com muita frequência. De novo, seja para corroborar com os dados que você acabou de gerar ou pra refuta (Pesquisador 8)

Acho que sim. Na área molecular tem que usar. Quando você vai identificar uma espécie usando ferramenta molecular você vai jogar num banco de dados que tem milhares de pesquisadores. E aí você vai ver aquela sequência que você tem é igual a sequência de quem você acha lá no banco de dados que tem milhares ou milhões de consequências, vê qual sequência é homóloga. E aí quando você acha, você usou inevitavelmente dados de outros pesquisadores, e não tem outro jeito. Eu acho que isso não é problema não. Acho que em Genética isso é rotina. (Pesquisador 9)

Acho que depende muito da área que a pessoa tá atuando, assim como eu tenho uma vivência de diferentes abordagens eu acho que isso é muito afeito a abordagem que as pessoas usam. Eu acho que tem algumas áreas na grande área das ciências da vida que se vale de algumas metodologias cuja sensibilidade e o controle precisam tomar muito cuidado na interpretação, e essa área é uma área que o compartilhamento de dado ele é praticamente impossível de reprodução, assim de você garantir a reprodutibilidade. (Pesquisadora 10)

De modo geral, os pesquisadores acreditam que existe uma disposição para a prática de reutilização de dados de pesquisa pelos pesquisadores. Contudo foram observados indícios que apontam que o reuso não é homogêneo dentro da Genética, essa diversificação pode ser explicada por meio do comportamento de produção, busca, acesso e reuso da informação que se dá de formas diferentes dentro de um mesmo campo do conhecimento (LEITE; COSTA, 2016; COSTA 2000; SALES; SAYÃO, 2012).

Na literatura são vastamente discutidos os benefícios que o reuso dos dados pode oferecer para a elevação da produtividade científica e o avanço da Ciência, contudo, na prática, é constatada uma realidade diferente. Autores como Curty (2017) e Tenopir *et al.* (2011) tratam sobre os receios dos pesquisadores em relação ao reuso de dados de pesquisa e que essa prática está longe de se consolidar no meio científico. Mesmo quatro anos após o estudo de Curty, na Genética o reuso de dados não parece ser amplamente aceito e praticado pelos pesquisadores.

A prática do reuso de dados de pesquisa é fomentada pela Ciência Aberta e pela web e suas ferramentas. No contexto brasileiro, a Ciência Aberta ainda é uma ideia recente, com poucas iniciativas que estimulam a criação de políticas e diretrizes para as práticas adequadas de produção, compartilhamento e reuso dos dados utilizando. A Genética ainda é uma ciência tradicional e competitiva, a disponibilização dos dados pode retirar uma grande parcela das vantagens competitivas dos pesquisadores.

Por causa disso, a criação de políticas institucionais e nacionais se mostram necessárias para criar o hábito de atividades de gestão e curadoria dos dados, juntamente com esses elementos, a aplicação dos princípios FAIR permite que os dados sejam encontrados tanto por humanos quanto por máquinas, proporcionando uma ciência de qualidade, transparência e com reprodutibilidade.

6 SÍNTESE DOS RESULTADOS DA PESQUISA

Esta seção apresenta a síntese dos dados e os resultados da pesquisa apresentados no tópico anterior.

6.1 Produção/obtenção de dados de pesquisa

Realização da pesquisa científica na Genética

Conforme as respostas dos pesquisadores, na Genética as pesquisas ocorrem em diferentes linhas. A partir dos dados coletados, as principais áreas dos pesquisadores respondentes são a:

- **Bioinformática:** com a produção e análises de grandes conjuntos de dados.
- **Experimental:** realizada em laboratório.
- **Pesquisa Básica:** voltada para a parte teórica, melhoramento de teorias.
- **Pesquisa Clínica:** realiza o desenvolvimento de medicamentos para o tratamento de uma dada patologia.

Tabela 14 - Realização da pesquisa científica na Genética.

Bioinformática	Utilização de softwares para análises de grandes quantidades de dados de pesquisa.
Experimental	Pesquisa realizada em ambiente controlado (laboratório) com organismos in vitro ou vivos para testar uma teoria.
Pesquisa Básica	A pesquisa caracteriza-se no melhoramento de teorias científicas da Genética e na compreensão de seus fenômenos.
Pesquisa Clínica	São desenvolvidos produtos e/ou tratamentos para uso humano ou animal visando o tratamento de uma determinada doença.

Fonte: Elaboração própria.

Produção dos dados de pesquisa e suas características

Dentro das linhas de pesquisa apresentadas anteriormente, a produção dos dados, a partir dos resultados encontrados, podem ocorrer em laboratório, por observação e por aplicação de programas específicos de computadores.

Na literatura científica amplamente disseminada, a geração dos dados pode ser observacional; experimental ou aplicação de programas de computador. E na Genética é possível observar conforme é colocado por Borgman (2012) que a produção dos dados nas Ciências da Vida, em sua maioria são coletados através de observações, experimentos ou por modelos/programas de computador.

Além disso, os dados gerados apresentam diferentes características que vão desde textos até instrumentos específicos, o que exige o uso de variados tipos de suporte para garantir o seu armazenamento e uma gestão mais apurada dos dados.

Tabela 15 - Produção de dados de pesquisa na Genética.

Natureza dos dados	Características dos dados
Laboratoriais, Observacionais e Computacionais	Texto, Números, Multimídia, Softwares, Instrumento-específico

Fonte: Elaboração própria.

Aspectos que influenciam positivamente e negativamente na produção de dados de pesquisa

Sobre os aspectos que favorecem a produção dos dados de pesquisa, os pesquisadores pontuaram a qualidade da equipe por ser um fator fundamental para o desenvolvimento da pesquisa, juntamente com o ambiente de trabalho e a sua infraestrutura.

O apoio financeiro, técnico e institucional também foram citados como fatores positivos pelos Pesquisadores 2, 5, 10 e 12. Esses mesmos pontos foram colocados pelos Pesquisadores 3, 6 e 8, como fatores inibidores, uma vez que eles relataram não receberem apoio, principalmente técnico e institucional. O que corrobora com o estudo de Sousa (2020) que elenca os principais desafios enfrentados pelos pesquisadores brasileiros, entre eles, a falta de equipe e apoio institucional.

O custo elevado dos reagentes e equipamentos é outro aspecto negativo que foi bastante enfatizado pelos pesquisadores, devido a compra desses materiais serem cotados em dólar. Também foi citado como fator inibidor, a concorrência elevada pela verba disponibilizada nos poucos editais publicados nos últimos anos, sendo colocada pela Pesquisadora 5 como uma atividade desgastante e que prejudica a produção dos dados.

Tabela 16 - Aspectos que influenciam a produção de dados de pesquisa.

Aspectos positivos	<ul style="list-style-type: none"> ❖ Qualidade da equipe ❖ Acesso facilitado às informações necessárias ❖ Ambiente de trabalho e infraestrutura adequada ❖ Apoio financeiro aos projetos ❖ Flexibilidade na geração dos dados ❖ Apoio técnico e institucional
Aspectos negativos	<ul style="list-style-type: none"> ❖ Custo elevado ❖ Falta de fomento à pesquisa ❖ Concorrência/competitividade

Fonte: Elaboração própria.

6.2 Compartilhamento de dados de pesquisa

Utilidade, percepção e experiência no compartilhamento dos dados de pesquisa

Quando perguntado aos pesquisadores sobre a utilidade de seus dados para outros pesquisadores, foi notado que a utilidade dos dados depende da disciplina a qual o pesquisador faz parte. Os pesquisadores de áreas mais amplas, como pesquisa básica, acreditam que seus dados podem ser utilizados por outros pesquisadores, por serem dados mais simples. Já os pesquisadores de áreas mais restritas, como o Pesquisador 7 que é da Neurogenética não vê seus dados sendo muito úteis para outros pesquisadores que não trabalhem com os mesmos temas. Mesmo assim, de maneira geral, os pesquisadores veem o compartilhamento de dados de pesquisa como uma prática comum em suas respectivas áreas.

Os pesquisadores também relataram já terem compartilhado seus dados de pesquisa, principalmente por exigência de editores de periódicos científicos, assim como através de solicitações de outros pesquisadores (desde que não afetem suas pesquisas), em repositórios, além de realizarem o compartilhamento entre seus colegas.

Tabela 17 - Compartilhamento de dados de pesquisa

Utilidade dos dados de pesquisa por outros pesquisadores	Depende do campo de pesquisa
Percepção sobre o compartilhamento dos dados	Prática recorrente
Experiência em compartilhamento de dados	Tem experiência no compartilhamento dos dados de pesquisa

Fonte: Elaboração própria.

Motivos que estimulam e inibem o compartilhamento de dados

No tocante aos aspectos que influenciam positivamente o compartilhamento dos dados de pesquisa, a maioria dos motivos relatados tem a ver com o desejo de colaborar para o avanço da ciência. Essa vontade de contribuir com o avanço da Ciência é observado em outros estudos, como o da Veiga (2019), onde mais de 50% dos pesquisadores veem o compartilhamento dos dados como benéfico, pois proporciona uma ciência com mais qualidade.

Como pontos inibidores, o compartilhamento dos dados de pesquisa causa preocupações aos pesquisadores, principalmente quando não se tem uma política clara de acesso aberto aos dados que garantam os direitos dos pesquisadores que produzem os dados e penalize o seu uso indevido por terceiros. A falta de conhecimento das práticas de compartilhamento de dados é outro fator que faz crescer os receios dos pesquisadores, inibindo ainda mais o compartilhamento dos dados.

Os autores Chawinga e Zinn (2019) tratam em seu estudo as razões que podem dificultar o compartilhamento dos dados de pesquisa pelos pesquisadores. De acordo com os autores, esses obstáculos podem se encontrar em 3 níveis: o nível individual (a falta de conhecimento), o nível institucional (ausência de políticas) e em nível internacional (normas éticas e legais pouco claras).

Tabela 18 - Aspectos que estimulam e inibem o compartilhamento de dados de pesquisa.

Aspectos que favorecem o compartilhamento dos dados de pesquisa	<ul style="list-style-type: none"> ❖ A cooperação científica entre os pesquisadores e as discussões que elevam a qualidade da pesquisa; ❖ A escassez de recursos financeiros para a produção dos dados; ❖ O desejo de contribuir para o avanço da ciência; ❖ O repasse do conhecimento a população que é gerado dentro de instituições públicas; ❖ Registrar a ciência por meio da publicação.
Aspectos que inibem o compartilhamento dos dados de pesquisa	<ul style="list-style-type: none"> ❖ Receio de seus dados serem utilizados por outros pesquisadores/laboratórios com um maior poder de processamento; ❖ O alto nível de competitividade em nível internacional; ❖ Questões ligadas à propriedade intelectual e/ou patentes ❖ Compartilhamento de dados sem planejamento

Fonte: Elaboração própria.

Exigência de compartilhamento dos dados de pesquisa

Quando a exigência é feita pelas Agências de fomento, a maioria dos pesquisadores se mostram abertos, já que é uma condição para conseguir o financiamento, como colocado pela Pesquisadora 10. Somente os Pesquisadores 2 e 8 que se mostraram totalmente contra a essa exigência, ao ponto de até deixarem de solicitar a verba. O Pesquisador 2 que trabalha com Genética Evolutiva, colocou como justificativa a essa contrariedade a impossibilidade de compartilhar os dados rapidamente. E o Pesquisador 8, que é da Biogenética, declarou a alta competitividade da sua área, principalmente quando comparado com outros países, como justificativa.

Caso a exigência fosse feita pela instituição do pesquisador, a maioria dos respondentes são favoráveis ao compartilhamento, mas com ressalvas, desde que exista um tempo de embargo dos dados para não afetarem o andamento das pesquisas. Ou como colocado pelo Pesquisador 1, a possível omissão de informação por receio de terem suas pesquisas prejudicadas. Apenas os Pesquisadores 7 e 8 foram contra a instituição exigir o compartilhamento dos dados, como justificativa citaram a produção de dados sensíveis e a alta competitividade.

Já o compartilhamento dos dados de pesquisa imposta por editores de periódico científico é algo visto com naturalidade pelos pesquisadores, já que a publicação é considerada por eles como o momento de mostrar as descobertas feitas.

A única ressalva feita foi em relação a disponibilização dos dados na hora da submissão, pois pode ocorrer do trabalho não ser aceito.

Mas no geral, foram poucos os pesquisadores que se mostraram contrários a qualquer tipo de exigência de compartilhamento.

Tabela 19 - Exigência do compartilhamento dos dados de pesquisa.

Obrigatoriedade (agências de fomento)	Maioria dos pesquisadores são favoráveis
Obrigatoriedade (instituição)	Maioria dos pesquisadores são favoráveis Ressalva: omissão de informações, embargo temporário
Obrigatoriedade (editores de periódicos)	Reagiram com naturalidades Ressalva: disponibilização dos dados no momento da submissão

Fonte: Elaboração própria.

6.3 Reuso dos dados de pesquisa

Experiência e percepção na reutilização de dados produzidos por outros pesquisadores

Em relação ao reuso, a maioria dos pesquisadores relataram já terem reutilizado dados de pesquisa produzidos por outros pesquisadores. A Pesquisadora 3 relatou, por exemplo, a dificuldade de integrar os dados coletados de diferentes bancos devido a variedade das metodologias utilizadas. Mostrando a necessidade de uma padronização dos metadados.

A Pesquisadora 12 foi a única que mencionou de forma clara que não utiliza dados gerados por outros pesquisadores devida as características de pesquisa da sua área, que é experimental.

Mas entre os relatos de reuso de dados de pesquisa, foi possível perceber que os pesquisadores não têm muita clareza sobre as práticas de reuso. Compreendendo o reuso de dados apenas para comparação de seus conjuntos de dados com os produzidos por outros pesquisadores. O que também faz os respondentes perceberem o reuso de dados em suas áreas como recorrente já que a reutilização é basicamente para discussões de resultados e comparações.

Tabela 20 - Experiência e percepção sobre o reuso dos dados de pesquisa.

Experiência na reutilização de dados produzidos por outros pesquisadores	<ul style="list-style-type: none"> ❖ Falta de clareza sobre prática de reuso de dados ❖ Dificuldade de integrar dados ❖ Maioria afirmaram terem experiência em reuso de dados
Percepção do reuso de dados de pesquisa produzidos por outros pesquisadores	<ul style="list-style-type: none"> ❖ Percebido pelos pesquisadores como algo natural ❖ Essencial para comparação de dados e discussões

Fonte: Elaboração própria.

Fatores que contribuem e inibem a reutilização de dados de pesquisa

Como aspectos positivo ao reuso dos dados de pesquisa, os pesquisadores citaram a possibilidade de comparar seus conjuntos de dados com outros conjuntos já existentes, dar agilidade no desenvolvimento da pesquisa e a economia de recursos.

Do mesmo modo como apresentados nos resultados dos processos de produção e compartilhamento de dados de pesquisa, o desenvolvimento de pesquisa na Genética é custoso e o reuso é uma forma de economizar recursos. Mas mesmo assim, a prática de reutilização dos dados não é totalmente explorada pelos pesquisadores, se mantendo apenas no nível de debate como constatado nas falas dos Pesquisadores 1, 2 e 10.

Em relação aos aspectos que inibem o reuso dos dados, as motivações estão diretamente ligadas a qualidade e a quantidade dos dados disponibilizados, além da falta de documentação e uma curadoria adequada dos dados. Fora que o local responsável pela disponibilização dos dados já é motivo suficiente para que os pesquisadores não utilizarem o material daquele lugar devido a sua reputação, como colocado pelos Pesquisadores 5, 8 e 9.

Tabela 21 - Fatores que favorecem e inibem o reuso dos dados de pesquisa.

Aspectos que contribuem com o reuso dos dados de pesquisa	<ul style="list-style-type: none"> ❖ Comparar seus dados com outros produzidos anteriormente para discussão de resultados; ❖ Validar modelos em desenvolvimento; ❖ Praticidade para a pesquisa; ❖ Custos elevados e falta de verba; ❖ A qualidade dos dados que estão sendo utilizados.
Aspectos que impedem o reuso dos dados de pesquisa	<ul style="list-style-type: none"> ❖ O volume de dados não ser adequado; ❖ A metodologia não ser clara; ❖ Possuir viés na coleta dos dados; ❖ O meio de divulgação e/ou quem gerou os dados (encontros, periódicos, laboratórios, grupo de pesquisas) não serem de confiança; ❖ Falta de informação na documentação dos dados.

Fonte: Elaboração própria.

Como finalização deste tópico, foi possível observar que assim como em outros campos do conhecimento que tiveram suas práticas de comunicação de dados de pesquisa analisadas, a Genética mesmo sendo uma área historicamente conhecida pelo costume de compartilhar e reutilizar dados de pesquisa, também sofre com dificuldades ligadas as suas práticas de produção, compartilhamento e reuso dos dados de pesquisa, em maior ou menor grau em relação a outros campos do conhecimento.

7 CONCLUSÃO

Este estudo teve como objetivo geral analisar as práticas de comunicação dos dados de pesquisa entre os pesquisadores brasileiros da Genética. Para isso, foram estabelecidos três objetivos específicos, a saber: 1. Identificar as práticas de produção de dados de pesquisa no campo da Genética; 2. Identificar as práticas de compartilhamento de dados de pesquisa no campo da Genética; 3. Identificar as práticas de reuso de dados de pesquisa no campo da Genética.

7.1 Identificar as práticas de produção de dados de pesquisa no campo da Genética

O primeiro objetivo específico foi alcançado com os dados obtidos através do primeiro conjunto de perguntas do roteiro de entrevista semiestruturado (Anexo 1). Foi possível concluir que o campo da Genética é bastante amplo, o que possibilita a aplicação de uma vasta quantidade de métodos e processos que permitem os dados serem coletados, logo, por ser uma área heterogenia que engloba diferentes

disciplinas, os processos de produção/obtenção naturalmente variam de uma para outra.

Outro aspecto relevante é a natureza dos dados produzidos, os dados podem ser obtidos em laboratórios, por experimentos ou sistemas de computadores, dependendo do tipo de pesquisa que está sendo desenvolvida. O mais comum, são dados de laboratório, mas também ocorre a sua obtenção por meio de experimentos e sistemas de computadores, logo, a natureza do dado interfere nos formatos e suportes que serão utilizados para o armazenamento desse insumo.

Questões relacionadas as relações interpessoais, ambiente de trabalho (estrutura), falta de apoio da instituição de ensino e verba escassa podem comprometer a produção dos dados de pesquisa, seja de maneira positiva ou negativa. Atualmente a falta de verba e a escassa quantidade de editais disponibilizados para financiamento de pesquisa pelas agências de fomento brasileiras é a principal dificuldade enfrentada pelos pesquisadores.

7.2 Identificar as práticas de compartilhamento de dados de pesquisa no campo da Genética

O objetivo específico foi alcançado através dos relatos dos pesquisadores referentes as suas práticas de compartilhamento de dados de pesquisa. Foi possível identificar quais aspectos estão envolvidos no processo de compartilhamento de dados pelos pesquisadores da Genética. Podem-se observar a seguir as conclusões.

- Práticas de compartilhamento de dados de pesquisa: o compartilhamento de dados de pesquisa acontece mais facilmente entre os pesquisadores que trabalham com Genética molecular, por ser um campo mais amplo. Pesquisadores que atuam em áreas mais restritas (neurogenética) ou competitivas (biogenética) não se mostram totalmente abertos para o compartilhamento. O fator idade também foi observado como ponto relevante para o compartilhamento de dados em disciplinas pouco adeptas a essas práticas, pesquisadores mais jovens costumam disponibilizar seus dados mais facilmente.
- Estímulos recebidos para o compartilhamento de dados de pesquisa: os pesquisadores da Genética realizam o compartilhamento dos seus dados

visando contribuir com discussões entre seus pares e construir uma ciência com mais qualidade; devolver a população em forma de conhecimento os investimentos públicos recebido; minimizar os efeitos da escassez de recursos financeiros para a produção de dados; desejo de contribuir com o avanço da ciência.

- Fatores inibidores para o compartilhamento de dados de pesquisa: as alegações feitas pelos pesquisadores da Genética sobre receios enfrentados para a disponibilização de seus dados são semelhantes as apresentadas como resultados em outros estudos presentes na revisão de literatura deste trabalho. Ressalta-se a alta competitividade em nível internacional em algumas disciplinas dentro da Genética; o receio do uso indevido dos dados por pesquisadores/laboratórios; questões relacionadas a sensibilidade dos dados e os direitos de uso, como patentes e propriedade intelectual.
- Exigência de disponibilização de dados de pesquisa: a obrigatoriedade imposta pelos editores de periódicos científicos em que os pesquisadores para terem um artigo aprovado publicado na revista precisam depositar os dados do trabalho em bancos de dados ou repositórios, não é visto como um problema, uma vez que já é uma prática recorrente para a publicação. O mesmo não é visto com bons olhos quando essa obrigatoriedade é imposta por agências de fomento ou a instituição de ensino a qual o pesquisador é colaborador. Em nível institucional, a resistência é maior, principalmente entre os pesquisadores que não recebem nenhum apoio da instituição.

7.3 Identificar as práticas de reuso de dados de pesquisa no campo da Genética

O objetivo foi alcançado, uma vez que os dados obtidos em diferentes questões do roteiro possibilitaram compreender as práticas de reuso de dados de pesquisa e que o entendimento sobre reuso de dados muda de uma disciplina para outro. Quanto a isso foi possível concluir que:

- O entendimento sobre reuso de dados de pesquisa não é claro entre os pesquisadores da Genética, mesmo eles relatando já terem tido experiências anteriormente em reutilizar dados produzidos por outros pesquisadores em suas pesquisas.

- Os conjuntos de dados que são reutilizados costumam ser retirados de bancos de dados ou repositórios que possuem relevância e/ou reconhecimento nas respectivas áreas dos pesquisadores, como é o caso do Genbank, reconhecido mundialmente. A facilidade gerada na utilização desses sistemas devida a rápida recuperação dos datasets e a sua documentação são outra vantagem apontada pelos pesquisadores.
- No quesito de fatores que contribuem e inibem o reuso dos dados de pesquisa, os pesquisadores da Genética consideram diferentes condições. Como fatores positivos para o reuso, estão incluídas a qualidade dos dados disponibilizados; agilidade no desenvolvimento da pesquisa; economia de recursos; comparar dados de pesquisa. As condições que podem desmotivar o reuso dos dados incluem procedimentos metodológicos não descritos com clareza; volume de dados pequeno; documentação pobre sobre os dados; coleta de dados apresentando vieses; estatística inadequada; meio de divulgação dos dados de pesquisa ser duvidoso.
- A disposição para a reutilização dos dados gerados por outros pesquisadores, similarmente como nas práticas de produção de dados, varia de uma área para a outra, o que está associada as culturas acadêmicas de cada disciplina refletidas em seus pesquisadores.

Esta pesquisa considerou os bolsistas de produtividade cadastrados na Plataforma Carlos Chagas, por considera-los os pesquisadores que mais produzem, compartilham e reusam dados de pesquisa e assim poder exprimir suas percepções sobre essas práticas dentro das suas respectivas áreas e apresentarem um panorama sobre a comunicação dos dados na Genética.

As conclusões desta pesquisa revelam que existem diferenças das práticas de produção, compartilhamento e reuso de dados de uma disciplina para outra e até mesmo entre grupos de pesquisas similares. Logo, os resultados mostram que não é possível generalizar a Genética em uma grande área homogênea, mais sim, como um campo do conhecimento com diferentes necessidades de comunicação dos dados intrínsecas a cada disciplina que a compõe.

Embora os resultados sejam baseados apenas nas percepções de bolsistas de produtividade, eles fornecem uma significativa visão sobre as práticas de

comunicação de dados de pesquisa, e o mais importante, as barreiras enfrentadas para a realização de uma pesquisa de qualidade dentro da Genética. Esses pontos problemáticos listados ao longo da discussão dos dados, podem auxiliar as instituições, agências de fomento e governos na elaboração de mecanismos e iniciativas adequadas que resguardecam os pesquisadores ao longo do andamento de suas pesquisas.

7.4 Limitações do estudo

Apesar do objetivo geral da pesquisa de “identificar as práticas de produção/obtenção, compartilhamento e reuso de dados de pesquisa de pesquisadores da Genética” ter sido alcançado, limitações podem ser vistas:

Clareza sobre as práticas de reuso dos dados: embora as respostas dos pesquisadores terem aspectos relevantes sobre as práticas reuso, a falta de clareza em relação ao termo “reuso” não permitiu análises mais profundas a respeito desta dimensão.

Pandemia do Covid-19: devido ao momento que estamos passando, o acesso aos pesquisadores, mesmo por meio de sistemas de comunicação (e-mail e Teams) se mostrou limitado, o que acarretou uma amostra pequena.

Generalização das conclusões: apesar de todos os entrevistados serem pesquisadores da Genética, suas atividades de pesquisa são em áreas e linhas diferentes, mesmo que algumas das práticas de comunicação dos dados de pesquisa sejam semelhantes, não permite generalização.

7.5 Sugestões para futuras pesquisa

Neste estudo, evidencia-se aspectos relacionadas as práticas de comunicação dos dados de pesquisa que não foram aprofundadas por não ser o objetivo, portanto, seguem como sugestões de estudos futuros:

- Analisar a falta de literatura sobre as práticas de produção, compartilhamento e reuso de dados de pesquisa sob a percepção dos pesquisadores brasileiros;
- Investigar o comportamento de produção, compartilhamento e reuso de dados de pesquisa à luz da Ciência Aberta;

- Analisar modelos acerca do comportamento de produção, compartilhamento e reuso de dados de pesquisa;
- Investigar o atendimento por parte dos pesquisadores aos critérios exigidos por editores de periódicos científicos que seguem os princípios da Ciência Aberta.
- Analisar os conjuntos de dados de pesquisa depositados em bancos de dados e repositórios da Genética de acordo com os princípios FAIR.

REFERÊNCIA

ALBAGLI, S.; CLINIO, A.; RAYCHTOCK, S. Ciência aberta: correntes interpretativas e tipos de ação. **Liinc em Revista**, Rio de Janeiro, v. 10, n. 2, nov., 2014. Disponível em: <http://revista.ibict.br/liinc/index.php/liinc/article/view/749/512>. Acesso em 10 jun. 2015

ALBAGLI, S. Ciência aberta em questão. *In*: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Orgs.). **Ciência aberta, questões abertas**. Brasília: IBICT, 2015, p. 09-25. Disponível em: [https://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20abertas_PORTUGUES_DIGITAL%20\(5\).pdf](https://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20abertas_PORTUGUES_DIGITAL%20(5).pdf). Acesso em: 07 jul. 2020.

ALVES, V. B. A. Open archives: via verde ou via dourada?. **PontodeAcesso**, v. 2, n. 2, p. 127-137, 2008. Disponível em: <https://portalseer.ufba.br/index.php/revistaici/article/view/1780/2172>. Acesso em: 06 jul. 2020.

ANDRADE, V. T. A. Comunicação científica e ciência aberta: produção e circulação de conhecimento em redes digitais. **Comunicação & Sociedade**, v. 37, n. 1, p. 259-287, 2015. Disponível em: <https://www.metodista.br/revistas/revistas-metodista/index.php/CSO/article/view/4817>. Acesso em: 10 abr. 2020.

AUSTRALIAN NATIONAL DATA SERVICE (ANDS). **Metadata**. 2016. Disponível em: https://www.ands.org.au/_data/assets/pdf_file/0004/728041/Metadata-Workinglevel.pdf. Acesso em: 23 maio 2020.

AUSTRALIAN NATIONAL DATA SERVICE (ANDS). **Vocabularies and research data**. 2017. disponível em: https://www.ands.org.au/_data/assets/pdf_file/0003/690870/Vocabularies-and-Research-Data.pdf. Acesso em: 24 maio 2020.

AUSTRALIAN NATIONAL DATA SERVICE (ANDS). **Australian code for responsible conduct of research**. Austrália: National Health and Medical Research Council, Australian Research Council and Universities Australia, 2018. disponível em: encurtador.com.br/wKMRV. Acesso em: 23 maio 2020.

BABBIE, E. **Métodos de pesquisas de survey**. Belo Horizonte: Ed. UFMG, 2003.

BALL, A. **How to license research data**. Edinburgh: Digital Curation Centre, 2014. Disponível em: <https://www.dcc.ac.uk/guidance/how-guides/license-research-data>. Acesso em: 23 dez. 2019.

BARONE, L.; WILLIAMS, J.; MICKLOS, D. Unmet needs for analyzing biological big data: a survey of 704 nsf principal investigators. **Plos Computational Biology**, [S.l.], v. 13, n. 10, p. 1-1, 19 out., 2017. <http://dx.doi.org/10.1371/journal.pcbi.1005755>.

BARTLING, S.; FRIESIKE, S. **Opening science**. New York: Springer, 2014.

BAYNES, G. **Colaboração e cooperação são fundamentais para tornar a ciência aberta uma realidade**. 2017. Disponível em: <encurtador.com.br/eqFT5>. Acesso em: 20 maio 2020.

BEAGRIE, N.; LAVOIE, B.; WOOLLARD, M. **Keeping research data safe 2: final report**. 2010. Disponível em: <http://repository.essex.ac.uk/2147/>. Acesso em: 30 mar. 2020.

BELL, J. **Projeto de pesquisa: guia para pesquisadores iniciantes em educação, saúde e ciências sociais**. 4. ed. Porto Alegre: Artmed, 2007.

BERTIN, P. R. B.; VISOLI, M. C.; DRUCKER, D. P. A gestão de dados de pesquisa no contexto da e-science: benefícios, desafios e oportunidades para organizações de p&d. **Ponto de Acesso**, v. 11, n. 2, p. 34-48, 2017. Disponível em: <https://portalseer.ufba.br/index.php/revistaici/article/view/21449>. Acesso em: 03 abr. 2020.

BIAZON, T.; MARIN, T. **Ciência aberta: uma nova forma de fazer ciência**. Disponível em: <http://www.dicyt.com/noticia/ciencia-aberta-uma-nova-forma-de-fazer-ciencia>. Acesso em: 22 mar. 2020.

BORGMAN, C. L. **Scholarship in the digital age: information, infrastructure, and the Internet**. Massachusetts: MIT Press, 2007.

BORGMAN, C. L. **The Conundrum of Sharing Research Data**. 2012. Disponível em: <https://on-linelibrary.wiley.com/doi/pdf/10.1002/asi.22634>. Acesso em: 18 mar. 2020.

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

BORGMAN, C. L. *et al.* The durability and fragility of knowledge infrastructures: lessons learned from astronomy. **Proceedings Of The Association For Information Science And Technology**, [S.l.], v. 53, n. 1, p. 1-10, 2016. <http://dx.doi.org/10.1002/praz.2016.14505301057>.

BRASIL. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes). **GEOCAPES - Sistema de Informações Georreferenciadas**. 2021. Disponível em: <https://geocapes.capes.gov.br/geocapes/>. Acesso em: 7 set. 2021.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). **Políticas públicas para pesquisa e inovação em face da crise da COVID-19**. Nota Técnica. Brasília: Ipea, 2020. Disponível em: <http://repositorio.ipea.gov.br/handle/11058/10034>. Acesso em: 7 set. 2021

BRASIL. MINISTÉRIO DA CIÊNCIA E TECNOLOGIA E INOVAÇÃO (MCTI). **Estratégia nacional de ciência, tecnologia e inovação 2012-2015**: balanço das atividades estruturantes. 2012. Disponível em: <https://livroaberto.ibict.br/218981.pdf>. Acesso em: 7 set. 2021.

BRASIL. Ministério da Ciência, Tecnologia e Inovações (MCTI). Coordenação de Indicadores de Ciência, Tecnologia e Inovação (COICT). **Comparações internacionais - Produção científica**. 2021. Disponível em: https://livroaberto.ibict.br/218981.pdfdores/detalhe/comparacoesInternacionais/Producao_cientifica_8.3.2.html. Acesso em: 7 set. 2021.

BRASIL. Presidência da República. **Lei nº 13.853, de 8 de julho de 2019**. Altera a Lei nº 13.709, de 14 de agosto de 2018, para dispor sobre a proteção de dados pessoais e para criar a Autoridade Nacional de Proteção de Dados; e dá outras providências. Brasília: Diário oficial, 2019. Disponível em: http://www.planalto.gov.br/ccivil_03/ato2015-2018/2018/lei/L13709.htm#art65. Acesso em: 11 jul. 2020.

BRASIL. Senado Federal. Agência Senado. **Corte de verbas da ciência prejudica reação à pandemia e desenvolvimento do país**. 2020. disponível em: <https://www12.senado.leg.br/noticias/infomaterias/2020/09/corte-de-verbas-da-ciencia-prejudica-reacao-a-pandemia-e-desenvolvimento-do-pais>. Acesso em: 7 set. 2021.

BUENO, W. C. Comunicação científica e divulgação científica: aproximações e rupturas conceituais. **Informação & Informação**, v. 15, n. 1, esp, p. 1-12, 2010. Disponível em: <http://www.uel.br/seer/index.php/informacao/article/view/6585/6761>. Acesso em: 19 mar. 2020.

CARIBÉ, R. DE C. DO V. Comunicação científica: reflexões sobre o conceito. **Informação & Sociedade**: estudos, v. 25, n. 3, p. 89-104, 28 dez., 2015. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/23109>. Acesso em: 15 fev. 2021.

CAMPBELL, D. T. Relabeling internal and external validity for applied social scientists. *In*: TROCHIM, W. M. **Advances in quasi-experimental designs and analysis**. San Francisco: Jossey-Bass, 1986.

CAMPBELL E. G. *et al.* Data withholding in academic genetics: evidence from a national survey. **Jama**, v. 287, n. 4, p. 473-480, 2002. Disponível em: <https://jamanetwork.com/journals/jama/article-abstract/194592>. Acesso em: 11 jun. 2021.

CAMPBELL E. G.; BENDAVID, E. Data-sharing and data-withholding in genetics and the life sciences: Results of a national survey of technology transfer officers. **J.**

Health Care L. & Pol'y, v. 6, p. 241, 2003. Disponível em: <https://bityli.com/Cft1CP>. Acesso em: 11 jul. 2021.

CARVALHO, É. R. S. de. **Diferenças na produção, compartilhamento e (re)uso de dados**: a percepção de pesquisadores da Química, Antropologia e Educação visual. 2018. 173 f. Dissertação (Mestrado em Ciência da Informação) - Universidade de Brasília, Brasília, 2018. Disponível em: <https://repositorio.unb.br/handle/10482/32559>. Acesso em: 19 maio 2020.

CENTRO DE GESTÃO E ESTUDOS ESTRATÉGICOS (CGEE). Panorama da ciência brasileira: 2015-2020. **Boletim Anual OCTI**, Brasília, v.1, jun., 2021. Disponível em: https://www.cgEE.org.br/documents/10195/734063/CGEE_Pan_Cie_Bra_2015-20.pdf. Acesso em: 7 set. 2021.

CHAWINGA, W. D.; ZINN, S. Global perspectives of research data sharing: A systematic literature review. **Library & Information Science Research**, v. 41, n. 2, p. 109-122, 2019. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S074081881830330X>. Acesso em: 29 maio 2020.

CHOO, C. W. **Gestão da informação para a organização inteligente**: a arte de escanear o ambiente. New York: Information Today, Inc., 2002.

CORRÊA, F. C. O papel dos bibliotecários na gestão de dados científicos. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 14, n. 3, p. 387-406, 2016. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8646333>. Acesso em: 11 maio 2020.

COSTA, S. M. S. Mudanças no processo de comunicação científica: o impacto do uso de novas tecnologias. *In*: MUELLER, S. P. M.; PASSOS, E. (Orgs.). **Comunicação científica**. Brasília: Editora Universidade de Brasília, 2000.

COSTA, M. **Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da Zika, na perspectiva de pesquisadores**. 2017. 269 f. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2017. Disponível em: <http://repositorio.unb.br/handle/10482/23000>. Acesso em: 11 maio 2020.

COSTA, M. Políticas de Ciência Aberta e a abertura dos dados de pesquisa. *In*: SHINTAKU, M.; SALES, L. F; COSTA, M. (Orgs.). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 23-29. DOI: [10.21452/978-85-93910-04-3.cap2](https://doi.org/10.21452/978-85-93910-04-3.cap2).

COSTA, M.; LEITE, F. C. L. Fatores que exercem influência na comunicação dos dados de pesquisa: uma revisão sistematizada da literatura no campo da Ciência da Informação. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 19, 2018, Londrina. **Anais...** Londrina: Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual de Londrina (PPGCI/UEL),

2018. Disponível em:

<http://enancib.marilia.unesp.br/index.php/XIXENANCIB/xixenancib/paper/viewFile/1265/1869>. Acesso em: 26 abr. 2020.

CRESWELL, J. W. **Projeto de pesquisa**: métodos qualitativo, quantitativo e misto. 3. ed. Porto Alegre: Artmed, 2010.

CURTY, R. **Disseminação desigual**. Entrevista. 2017. Disponível em:

<https://revistapesquisa.fapesp.br/2018/03/20/disseminacao-desigual/>. Acesso em: 18 mar. 2020.

DATAONE. **Data life cycle**. 2013. Disponível em: <https://www.dataone.org/data-life-cycle>. Acesso em: 6 maio 2020.

DELFANTI, A.; PITRELLI, N. Ciência aberta: revolução ou continuidade. *In*: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Orgs.). **Ciência aberta, questões abertas**. Brasília: IBICT, 2015, p. 59-69.

DENZIN, N. K.; LINCOLN, Y. S. **Introduction**: the discipline and practice of qualitative research. 2005.

DESLAURIERS, J.; KÉRISIT, M. O delineamento de pesquisa qualitativa. *In*: POUPART, J. *et al.* **A pesquisa qualitativa**: enfoques epistemológicos e metodológicos. Rio de Janeiro: Vozes, 2014, p. 127-153.

DIAS, G. A.; ANJOS, R. L. dos; ARAÚJO, D. G. de. A gestão dos dados de pesquisa no âmbito da comunidade dos pesquisadores vinculados aos programas de pós-graduação brasileiros na área da Ciência da Informação: desvendando as práticas e percepções associadas ao uso e reúso de dados. **Liinc em Revista**, Rio de Janeiro, v.15, n. 2, p. 5-31, nov., 2019. Disponível em:

<http://revista.ibict.br/liinc/article/view/4683/4306>. Acesso em: 13 dez. 2019.

DREXLER, K. E. Hypertext publishing and the evolution of knowledge. **Social Intelligence**, v. 1, n. 2, p. 87-120, 1991. Disponível em:

<http://edrexler.com/d/06/00/Hypertext/HPEK0.html>. Acesso em: 21 set. 2019.

EUROPEAN COMMISSION. **Open Science**. 2019. Disponível em:

https://ec.europa.eu/info/sites/info/files/research_and_innovation/knowledge_publications_tools_and_data/documents/ec_rtd_factsheet-open-science_2019.pdf. Acesso em: 05 dez. 2020.

FANIEL, I. M.; JACOBSEN, T. E. Reusing scientific data: how earthquake engineering researchers assess the reusability of colleagues' data. **Computer Supported Cooperative Work (CSCW)**, v. 19, n. 3-4, p. 355-375, 2 ago. 2010. DOI: [10.1007/s10606-010-9117-8](https://doi.org/10.1007/s10606-010-9117-8). Disponível em:

<https://link.springer.com/article/10.1007/s10606-010-9117-8>. Acesso em: 10 mar. 2020.

FACILITATE OPEN SCIENCE TRAINING FOR EUROPEAN RESEARCH (Foster). **The future of science is Open**. 2015. Disponível em:

<https://www.fosteropenscience.eu>. Acesso em: 27 jul. 2019.

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

FEDERER, L. M. *et al.* Biomedical data sharing and reuse: attitudes and practices of clinical and scientific research staff. **Plos One**, [S.l.], v. 10, n. 6, jun., 2015. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pone.0129506>.

FERGUSON, A. R. *et al.* Big data from small data: data-sharing in the 'long tail' of neuroscience. **Nature neuroscience**, v. 17, n. 11, p. 1442-1447, 2014. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4728080/>. Acesso em: 30 jun. 2021.

FIENBERG, S. E.; MARTIN, M. E.; STRAF, M. L. (Orgs.). **Sharing research data**. Washington: National Academy Press, 1985.

FLICK, U. **Desenho da pesquisa qualitativa**. Porto Alegre: Bookman, 2009.

FLICK, U. **Introdução à metodologia de pesquisa: um guia para iniciante**. Porto Alegre: Penso, 2012.

FORCE11. **Data citation synthesis group: joint declaration of data citation principles**. 2014. Disponível em: <https://www.force11.org/datacitationprinciples>. Acesso em: 24 maio 2020.

FRANCISCO, N. A.; ZUCATTO, L. C. Estado da arte das produções referente ao financiamento e fomento da pesquisa científica no brasil: uma análise a partir do ciclo de políticas públicas. **Práticas de Administração Pública**, [S.l.], v. 2, n. 1, p. 41, jan., 2019. <http://dx.doi.org/10.5902/2526629230604>

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO (Fapesp). **Relatório de Atividades FAPESP 2020**. 2021. Disponível em: <https://fapesp.br/relatorio2020>. Acesso em: 8 set. 2021.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO (Fapesp). Agência Fapesp. **Fapesp respondeu aos desafios da pandemia sem comprometer o apoio às demais áreas de pesquisa**. 2021. Disponível em: encurtador.com.br/dexEW. Acesso em: 8 set. 2021.

FUNDAÇÃO OSWALDO CRUZ (Fiocruz). **Curso Ciência Aberta - Aula 4: FAIR: dos princípios à prática**. 2019. Disponível em: <https://mooc.campusvirtual.fiocruz.br/rea/ciencia-aberta/serie3/curso2/aula4.html>. Acesso em: 07 jul. 2020.

GARVEY, W. D.; GRIFFITH, B. C. Communication and information processing within scientific disciplines: empirical findings for psychology. *In*: GARVEY, W. D. (Ed.). **Communications, the essence of science: facilitating information exchange**

students. Oxford: Pergamon Press, 1979. p. 127–174. DOI: [10.1016/C2013-0-05843-4](https://doi.org/10.1016/C2013-0-05843-4).

GEZELTER, D. What, exactly, is open science. **The OpenScience Project**, 2009. Disponível em: <http://openscience.org/what-exactly-is-open-science/>. Acesso em: 23 mar. 2020.

GOMES, C. M. **Comunicação científica**: alicerces, transformações e tendências. 2013. Disponível em: <https://ubibliorum.ubi.pt/handle/10400.6/2127>. Acesso em: 13 set. 2021.

GÓMEZ, N.; MÉNDEZ, E.; HERNÁNDEZ-PÉREZ, T. Datos y metadatos de investigación en ciencias sociales y humanidades: una aproximación desde los repositorios temáticos de datos. **El Profesional de la Información**, v. 25, n. 4, p. 545-555, jul./ago., 2016. Disponível em: http://www.elprofesionaldelainformacion.com/contenidos/2016/jul/04_esp.pdf. Acesso em: 04 nov. 2020.

GORRAIZ, J. *et al.* research impact in social sciences: comparison of the Universities of Vienna, Zurich and Oslo. **Research Evaluation**, v. 18, n. 3, p. 221-232, 2009. Disponível em: <https://academic.oup.com/rev/article-abstract/18/3/221/1546298>. Acesso em: 13 set. 2021.

GUANAES, P. C. V. Quais direitos podem ser alegados sobre dados de pesquisa?. *In*: SHINTAKU, M.; SALES, L. F.; COSTA, M. (Orgs.). **Tópicos sobre dados abertos para editores científicos**. São Paulo: ABEC, 2020. p. 79-85. DOI: [10.21452/978-85-93910-04-3.cap7](https://doi.org/10.21452/978-85-93910-04-3.cap7).

HALL, H. Social exchange for knowledge exchange. *In*: MANAGING KNOWLEDGE: CONVERSATIONS AND CRITIQUES, 2001. **Presentation...** Leicester: University of Leicester Management Centre, 2001.

HEY, T.; TANSLEY, S.; TOLLE, K. M. Jim gray on escience: a transformed scientific method. *In*: **THE FOURTH PARADIGM**: data-intensive scientific Discovery. [S.l.]. Microsoft Research, 2009. Disponível em: <http://itre.cis.upenn.edu/myl/JimGrayOnE-Science.pdf>. Acesso em: 16 ago. 2020.

HJORLAND, B. **Data**: with big data and database semantics. 2018. Disponível em: <https://www.isko.org/cyclo/data>. Acesso em: 10 maio 2020.

HORIZON EUROPE. **Horizon Europe, open science**. 2021. Disponível em: <https://op.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/9570017e-cd82-11eb-ac72-01aa75ed71a1>. Acesso em: 13 ago. 2021.

HUANG, M.; CHANG, Y. Characteristics of research output in social sciences and humanities: from a research evaluation perspective. **Journal of the American Society for Information Science and Technology**, v. 59, n. 11, p. 1819-1828, 2008. Disponível em: <https://bityli.com/LbDb2>. Acesso em: 13 set. 2021.

HURD, J. M. The transformation of scientific communication: a model for 2020. **Journal of the American Society for Information Science**, v. 51, n. 14, p. 1279–1283, 1 jan., 2000. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/1097-4571%282000%299999%3A9999%3C%3A%3AAID-ASI1044%3E3.0.CO%3B2-1>. Acesso em: 20 mar. 2020.

INMON, W. H.; LINSTEDT. **Data Architecture: a primer for the data scientist- big data, data warehouse and data vault**. USA: Elsevier, 2014.

JAROLÍMKOVÁ, A.; DROBÍKOVÁ, B.; SOUČEK, M. Attitudes of charles university academic staff to data sharing. *In*: CONFERENCE ON GREY LITERATURE AND REPOSITORIES. 3, 2018, Prague. **Anais...** Prague: National Library of Technology, 2017, p. 57-66.

JORGE, V. de A. **Abertura e compartilhamento de dados para pesquisa nas situações de emergência em saúde pública: o caso do vírus Zika**. 2018. 263 f. Tese (Doutorado em Ciência da Informação) - Universidade Federal do Rio de Janeiro; Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2018. Disponível em: <https://www.arca.fiocruz.br/handle/icict/32604>. Acesso em: 20 maio 2020.

KAUARK, F.; MANHÃES, F. C.; MEDEIROS, C. H. **Metodologia da pesquisa: guia prático**. Itabuna: Via Litterarum, 2010.

KELLEHER, J. D.; TIERNEY, B. **Data Science**. Cambridge: The MIT Press, 2018.

KEY PERSPECTIVES LTD. **Data dimensions: disciplinary differences in research data sharing, reuse and long term viability**. 2010. Disponível em: <https://www.dcc.ac.uk/sites/default/files/documents/publications/SCARP%20SYNTHESIS.pdf>. Acesso em: 27 maio 2020.

KNORR-CETINA, K. **Epistemic cultures: how the sciences make knowledge**. Cambridge, Mass.; London: Harvard University Press, 1999.

KURATA, K.; MATSUBAYASHI, M.; MINE, S. Identifying the complex position of research data and data sharing among. **Researchers in Natural Science. Sage Open**, [S.l.], v. 7, n. 3, p. 215824401771730, jul., 2017. <http://dx.doi.org/10.1177/2158244017717301>.

KVALE, S. **Interviews: an introduction to qualitative research interviewing**. London: SAGE, 1996.

LE COADIC, Y. **A ciência da informação**. 2. ed. rev. atual. Brasília: Briquet de Lemos, 2004.

LEITE, F. C. L. **Modelo genérico de gestão da informação científica para instituições de pesquisa na perspectiva da comunicação científica e do acesso aberto**. 2011. 262 f. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Faculdade de Ciência da Informação, Programa de Pós-Graduação em

Ciência da Informação, 2011. Disponível em:

<https://repositorio.unb.br/handle/10482/9753>. Acesso em: 10 mar. 2020.

LEITE, F. C. L.; COSTA, S. M. de S. Modelo genérico de gestão da informação científica para instituições de pesquisa na perspectiva da comunicação científica e do acesso aberto. **Investigación Bibliotecológica: Archivonomía, Bibliotecología e Información**, [S.l.], v. 30, n. 69, p. 43-74, 2016.

<http://dx.doi.org/10.1016/j.ibbai.2016.04.012>.

LOPEZ, F. **Workshop de datos científicos**: introducción. Slides. 2015. Disponível em: <https://pt.slideshare.net/fernandoariellopez/workshop-de-datos-cientificos-introduccion>. Acesso em: 23 mar. 2020.

MARCUM, D. B., GEORGE, G. Introduction. *In*: ABC-CLIO (Ed.). **The data deluge: can libraries cope with e-science?**. California: Greenwood Press, 2010.

MARTÍNEZ-URIBE, L.; MACDONALD, S. **Un nuevo cometido para los bibliotecarios académicos**: data curation. 2008. Disponível em:

<http://www.elprofesionaldelainformacion.com/contenidos/2008/mayo/03.pdf>. Acesso: 18 mar. 2020

MARTINS, M. F. M. **Estudos de revisão de literatura**. Rio de Janeiro: Fiocruz, 2018. Disponível em:

https://www.arca.fiocruz.br/bitstream/icict/29213/2/Estudos_revisao.pdf. Acesso em: 20 jul. 2020.

MATTOS, P. C. **Tipos de revisão**. Universidade Estadual Paulista: Faculdade de Ciências Agrônomas, 2015. Disponível em:

<https://www.fca.unesp.br/Home/Biblioteca/tipos-de-revisao-de-literatura>. Acesso em: 18 set. 2020.

MAYERNIK, M. S. **Metadata realities for cyberinfrastructure**: data authors as metadata creators. 2011. 337 f. Dissertation (Doctor of Philosophy in Information Studies) - University of California, Los Angeles, 2011. DOI: [10.2139/ssrn.2042653](https://doi.org/10.2139/ssrn.2042653).

Disponível em: <https://ssrn.com/abstract=2042653>. Acesso em: 9 set. 2020.

MAZZA, M. **Comissão da SBPC faz balanço das ações estratégicas para o financiamento do setor**. *Jornal da Ciência*. 2021. Disponível em:

<http://www.jornaldaciencia.org.br/comissao-da-sbpc-faz-balanco-das-acoes-estrategicas-para-o-financiamento-do-setor/>. Acesso em: 10 set. 2021.

MEADOWS, A. J. **A comunicação científica**. Brasília: Brique de Lemos, 1999.

MEDEIROS, J. D. S.; CAREGNATO, S. E. Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. **Liinc em Revista**, v. 8, n. 2, 18 dez., 2012. Disponível em: <http://revista.ibict.br/liinc/article/view/3368>.

Acesso em: 10 abr. 2020.

MELO, J. N.; DE SANTANA, J. R.; SILVA, G. F. Ciência, tecnologia e inovação no Brasil: Uma análise inter-regional por meio de indicadores. **Revista Brasileira de**

Gestão e Desenvolvimento Regional, v. 15, n. 1, 2019. Disponível em: <https://rbgdr.net/revista/index.php/rbgdr/article/view/4321>. Acesso em: 7 set. 2021.

MÉNDEZ RODRÍGUEZ, E. **Metadatos y recuperación de información: estándares, problemas y aplicabilidad en bibliotecas digitales**. Gijón: Trea, 2002.

MENZEL, H, SOMERS, R.; GLASER, W. A. **The flow of information among scientists: problems, opportunities and research questions**. New York: Columbia University., Bureau Of Applied Social Research, 1958.

MILHAM, M. P. *et al.* **Assessment of the impact of shared data on the scientific literature**. 2017. Disponível em: <https://www.biorxiv.org/content/10.1101/183814v1>. Acesso em: 17 mar. 2020.

MONS, B. *et al.* Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. **Information Services & Use**, v. 37, n. 1, p. 49–56, Jan., 2017. Disponível em: <https://pdfs.semanticscholar.org/bae3/d1c60dae3654903add37aa88eb581712aae5.pdf>. Acesso em: 20 maio 2020.

MONTEIRO, G.; LUCAS, E. R. de O. Dados científicos abertos: identificando o papel das políticas de gestão e das agências de fomento. **AtoZ: novas práticas em informação e conhecimento**, v. 8, n. 1, p. 13-20, 2019. Disponível em: <https://brapci.inf.br/index.php/res/v/114899>. Acesso em: 02 out. 2020.

MURRAY-RUST, P. Open Data in Science. **Nature Precedings: pre-publication research and preliminar findings**. 2008. Disponível em: <http://precedings.nature.com/documents/1526/version/1>. Acesso em: 30 nov. 2020.

NASSI-CALÒ, L. Promovendo e acelerando o compartilhamento de dados de pesquisa [on-line]. **SciELO em Perspectiva**, 2019. Disponível em: <https://blog.scielo.org/blog/2019/06/13/promovendo-e-acelerando-o-compartilhamento-de-dados-de-pesquisa/#.YDJvDehKjIU>. Acesso em: 21 ago. 2020.

NATIONAL ACADEMY OF SCIENCES (NAS). **A question of balance: private rights and the public interest in scientific and technical databases**. Washington: National Academies Press, 1999. Disponível em: <https://www.nap.edu/download/9692>. Acesso em: 30 mar. 2020.

NATIONAL HEALTH AND MEDICAL RESEARCH COUNCIL ((NHMRC). **Australian code for responsible conduct of research**. Austrália: National Health and Medical Research Council, Australian Research Council and Universities Australia. 2018. Disponível em: <https://www.nhmrc.gov.au/about-us/publications/australian-code-responsible-conduct-research-2018#block-views-block-file-attachments-content-block-1>. Acesso em: 23 maio 2020.

NATIONAL SCIENCE BOARD. **Long-lived digital data collections: enabling research and education in the 21st century**. New Jersey: National Science Foundation. 2005. Disponível em: <https://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>. Acesso em: 13 maio 2020.

NATIONAL SCIENCE FOUNDATION (NSF). **Cyberinfrastructure vision for 21st century Discovery**. 2007. Disponível em: <https://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf>. Acesso em: 30 mar. 2020.

NATURE COMMUNICATIONS. Data sharing and the future of science. **Nature Communications**, [S.l.], v. 9, n. 1, jul., 2018. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/s41467-018-05227-z>. Disponível em: <https://www.nature.com/articles/s41467-018-05227-z#citeas>. Acesso em: 04 set. 2020.

NAVARRO, F. P.; CONEGLIAN, C. S.; SANTARÉM SEGUNDO, J. E. Big data no contexto de dados acadêmicos: o uso de machine learning na construção de sistema de organização do conhecimento. **Informação&Tecnologia (ITEC)**, João Pessoa, v.5, n.2, p.181-200, jul./dez., 2018. Disponível em: <https://periodicos.ufpb.br/index.php/itec/article/view/45807/27751>. Acesso em: 10 mar. 2020.

OLIVEIRA, J. **COVID-19 concentra recursos da pesquisa, e outras áreas importantes perdem**. Estado de Minas. 2020. Disponível em: https://www.em.com.br/app/noticia/gerais/2020/10/04/interna_gerais,1191424/covid-19-concentra-recursos-da-pesquisa-areas-importantes-perdem.shtml. Acesso em: 11 set. 2021.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). **OECD principles and guidelines for access to research data from public funding**. New York: OECD Publishing, 2007. Disponível em: <http://www.oecd.org/sti/inno/38500813.pdf>. Acesso em: 4 mar. 2020.

PARK, H.; WOLFRAM, D. An examination of research data sharing and reuse: implications for data citation practice. **Scientometrics**, v. 111, n. 1, p. 443-461, 2017.

PIENTA, A. M.; ALTER, G. C.; LYLE, J. A. **The enduring value of social science research: the use and reuse of primary research data**. 2010. Disponível em: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/78307/pienta_alter_lyle_100331.pdf?sequence=1&isAllowed=y. Acesso em: 20 maio 2020.

PINTO, C. S.; COSTA, J. L. Padrões de comunicação em diferentes comunidades científicas. *In*: COSTA, S. M. DE S.; LEITE, F. C. L.; TAVARES, R. B. (Orgs). **Comunicação da informação, gestão da informação e gestão do conhecimento**. Brasília: Ibict, 2018. p. 145-159. Disponível em: <http://livroaberto.ibict.br/handle/123456789/1071>. Acesso em: 20 mar. 2020.

PIRES, A. P. Amostragem e pesquisa qualitativa: ensaio teórico e metodológico. *In*: POUPART, J. *et al.* **A pesquisa qualitativa: enfoques epistemológicos e metodológicos**. Rio de Janeiro: Vozes, 2014.

PIWOWAR, H. A.; DAY, R. S.; FRIDSMA, D. B. Sharing detailed research data is associated with increased citation rate. **PloS one**, v. 2, n. 3, p. 308, jan., 2007. Disponível em:

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000308>. Acesso em: 20 mar. 2020.

POLO, M. C. E.; TAVARES, G. H. Políticas públicas de financiamento de pesquisas: a participação das mulheres no âmbito dos estudos do lazer e do esporte. **Revista Brasileira de Estudos do Lazer**, v. 6, n. 2, p. 56-71, 2019. Disponível em: <https://periodicos.ufmg.br/index.php/rbel/article/view/16137>. Acesso em: 7 set. 2021.

PONTIKA, N. *et al.* Fostering open science to research using a taxonomy and an eLearning portal. *In: INTERNATIONAL CONFERENCE ON KNOWLEDGE TECHNOLOGIES AND DATA-DRIVEN BUSINESS*, 15, 2015. Áustria. **Proceedings...** Áustria: Association for Computing Machinery, p. 1-8. Disponível em: <https://dl.acm.org/doi/abs/10.1145/2809563.2809571>. Acesso em: 24 mar. 2020.

POUPART, J. **A entrevista de tipo qualitativo: considerações epistemológicas, teóricas e metodológicas.** *In: POUPART, J. et al. A pesquisa qualitativa: enfoques epistemológicos e metodológicos.* Rio de Janeiro: Vozes, 2008.

RADECKI, J.; SCHONFELD, R. C. **Os impactos do COVID-19 na empresa de pesquisa: uma revisão da paisagem.** 2020. Disponível em: <https://sr.ithaka.org/publications/the-impacts-of-covid-19-on-the-research-enterprise/>. Acesso em: 11 set. 2021.

RIBEIRO, C. *et al.* Projeto TAIL - Gestão de dados de investigação da produção ao depósito e à partilha (resultados preliminares). **Cadernos BAD**, n. 2, p. 256-264, 2016. Disponível em: <https://www.bad.pt/publicacoes/index.php/cadernos/article/view/1603>. Acesso em: 11 maio 2020

RILEY, J. **Understanding metadata: what is metadata, and what is it for?** Baltimore: National Information Standards Organization (NISO), 2017. Disponível em: https://groups.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf. Acesso em: 11 maio 2020.

ROSA, F. G.; GOMES, M. J. **Comunicação científica: das restrições ao acesso livre.** 2010. Disponível em: https://repositorium.sdum.uminho.pt/bitstream/1822/11412/1/RI_FI%C3%A1via%20Rosa%20%26%20Maria%20Jo%C3%A3o%20Gomes.pdf. Acesso em: 19 mar. 2020.

ROSA, F. G.; BARROS, S. S. Comunicação científica: reflexões preliminares para o GT" Relevância dos livros acadêmicos na comunicação da pesquisa". **SciELO 20 Years epository**, p. 1-12, 2018. Disponível em: https://www.scielo20.org/redescielo/wp-content/uploads/sites/2/2018/07/ROSA-F_-BARROSS.-Comunicacao-Cientifica.pdf. Acesso em: 19 mar. 2020.

ROSSMAN, G.; RALLIS, S. E. **Learning in the field: an introduction to qualitative research.** Thousand Oaks: Sage, 1998.

RUTTENBERG, S.; RISHBETH, H. World data centres: past, present and future. **Journal of Atmospheric and Terrestrial Physics**, Oxford, v. 56, n. 7, p. 865-870, may, 1994. Disponível em: <https://ui.adsabs.harvard.edu/abs/1994JATP...56..865R/abstract>. Acesso em: 21 ago. 2020.

SALES, L. F.; SAYÃO, L. F. O impacto da curadoria digital dos dados de pesquisa na comunicação científica. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, [S.l.], v. 17, n. esp, p. 118-135, 18 dez., 2012. Universidade Federal de Santa Catarina (UFSC). <http://dx.doi.org/10.5007/1518-2924.2012v17nesp2p118>. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17nesp2p118>. Acesso em: 13 set. 2021.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva partir da Ciência da Informação. **Informação & informação**, v. 21, n. 2, p. 116-142, 2016. Disponível em: <https://www.readcube.com/articles/10.5433%2F1981-8920.2016v21n2p116>. Acesso em: 16 set. 2020.

SANT'ANA, R. C. G. Campo informacional resultante da interação de ciclos de vida dos dados. *In*: DIAS, G. A.; OLIVEIRA, B. M. J. F. de. **Dados científicos: perspectivas e desafios**. João Pessoa: Ufpb, 2019. Disponível em: <http://www.editora.ufpb.br/sistema/press5/index.php/UFPB/catalog/download/359/508/2949-1?inline=1>. Acesso em: 16 set. 2020.

SANTAREM SEGUNDO, J. E. Web semântica e linked data como aliados na consolidação da Ciência Aberta. *In*: SHINTAKU, M.; SALES, L. F.; COSTA, M. (Orgs.). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 65-78. DOI: [10.21452/978-85-93910-04-3.cap6](https://doi.org/10.21452/978-85-93910-04-3.cap6).

SANTIN, D. M.; VANZ, S. A. de S.; STUMPF, I. R. C. Produção científica em Ciências Biológicas da UFRGS: tendências temáticas no período 2000-2011. **Perspectivas em Ciência da Informação**, v. 20, p. 3-21, 2015. Disponível em: <https://www.scielo.br/j/pci/a/bx6JMF8THGhBSff98CqJFyM/?lang=pt>. Acesso em: 5 ago. 2021.

SAYÃO, L. F.; SALES, L. F. **Dados abertos de pesquisa: ampliando o conceito de acesso livre**. 2014. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/viewFile/611/1252>. Acesso em: 13 maio 2020.

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.aben.com.br/noticias/guia-de-gestao-de-dados-de-pesquisa-para-bibliotecarios-e-pesquisadores>. Acesso em: 15 mar. 2020.

SAYÃO, L. F.; SALES, L. F. Curadoria digital e dados de pesquisa. **AtoZ**, v. 5, n. 2, p. 67, jan., 2016. Disponível em: <https://revistas.ufpr.br/atoz/article/view/49708>. Acesso em: 18 mar. 2020.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, v. 21, n. 2, p. 90-115, 2016.

Disponível em:

<http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939>. Acesso em: 19 ago. 2021.

SAYÃO, L. F.; SALES, L. F. A ciência invisível: os dados da cauda longa da pesquisa científica. *In*: DIAS, G. A.; OLIVEIRA, B. M. J. F. de. **Dados científicos: perspectivas e desafios**. João Pessoa: Ufpb, 2019. Disponível em:

<http://www.editora.ufpb.br/sistema/press5/index.php/UFPB/catalog/download/359/508/2949-1?inline=1>. Acesso em: 16 set. 2020.

SALES, L. F.; SAYÃO, L. F. Ampliando as fronteiras da editoração científica: o papel dos repositórios de dados. *In*: SHINTAKU, M.; SALES, L. F.; COSTA, M. (Orgs.). **Tópicos sobre dados abertos para editores científicos**. Botucatu: ABEC, 2020. p. 39-45. DOI: [10.21452/978-85-93910-04-3.cap4](https://doi.org/10.21452/978-85-93910-04-3.cap4).

SALES, L. F.; COSTA, M.; SHINTAKU, M. Ciência aberta, gestão de dados de pesquisa e novas possibilidades para a editoração científica. *In*: SHINTAKU, M.; SALES, L. F.; COSTA, M. (Orgs.). **Tópicos sobre dados abertos para editores científicos**. Botucatu: ABEC, 2020. p. 13-21. DOI: [10.21452/978-85-93910-04-3.cap1](https://doi.org/10.21452/978-85-93910-04-3.cap1).

SCHAUDER, D. Electronic publishing of professional articles: attitudes of academics and implications for the scholarly communication industry. **Journal of the American society for information science**, v. 45, n. 2, p. 73-100, 1994. Disponível em: <https://asistdl.on-linelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-4571%28199403%2945%3A2%3C73%3A%3AAID-ASI2%3E3.0.CO%3B2-5>. Acesso em: 19 mar. 2020.

SCIMAGO INSTITUTIONS RANKINGS. **Ranking**. 2021. Disponível em: <https://www.scimagoir.com/rankings.php>. Acesso em: 21 ago. 2021.

SHEN, Y. Research data sharing and reuse practices of academic faculty researchers: a study of the Virginia Tech data landscape. **International Journal of Digital Curation**, v. 10, n. 2, p. 157-175, 2015. Disponível em: <http://www.ijdc.net/article/view/10.2.157>. Acesso em: 1 jun. 2020.

SHRUM, W.; GENUTH, J.; CHOMPALOV, I. **Structures of scientific collaboration**. Cambridge: MIT Press, 2007.

SIDONE, O. J. G.; HADDAD, E. A.; MENA-CHALCO, J. P. A ciência nas regiões brasileiras: evolução da produção e das redes de colaboração científica. **Transinformação**, v. 28, p. 15-32, 2016. Disponível em: <https://www.scielo.br/j/tinf/a/tvBDyptMBFSxRSt3VngySRC/abstract/?lang=pt>. Acesso em: 7 set. 2021.

SIDLER, M. Open science and the three cultures: expanding open science domains of knowledge creation. *In*: BARTLING, S.; FRIESIKE, S. (Eds.). **Opening Science**:

the evolving guide on how the internet is changing research, collaboration and scholarly publishing. Heidelberg: Springer Open, 2014.

Sielemann, K., HAFNER, A., PUCKER, B. The reuse of public datasets in the life sciences: potential risks and rewards. **PeerJ**, v. 8, n. 9954, 2020. Disponível em: <https://peerj.com/articles/9954/>. Acesso em: 2 ago. 2021.

SILVA, M. N. P. da. **População e amostras**. 2018. Rio de Janeiro: Brasil Escola. Disponível em: <https://brasilecola.uol.com.br/matematica/populacao-amostras.htm>. Acesso em: 21 fev. 2020.

SILVA, R. M. *et al.* Perfil e financiamento da pesquisa em saúde desencadeada pela pandemia da COVID-19 no Brasil. **Vigilância Sanitária em Debate: Sociedade, Ciência & Tecnologia**, v. 8, n. 2, p. 28-38, 2020. Disponível em: <https://visaemdebate.incqs.fiocruz.br/index.php/visaemdebate/article/view/1579>. Acesso em: 7 set. 2021.

SOCIEDADE BRASILEIRA PARA O PROGRESSO DA CIÊNCIA (SBPC). **Em plena pandemia, orçamento da CT&I sofre cortes**. 2021. Disponível em: <https://portal.sbpcnet.org.br/noticias/em-plena-pandemia-orcamento-da-cti-sofre-cortes/>. Acesso em: 8 set. 2021.

SØNDERGAARD, T. F. *et al.* Documents and the communication of scientific and scholarly information: revising and updating the UNISIST model. **Journal of Documentation**, London, v. 59, n. 3, p. 278-320, 2003. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/00220410310472509/full/html>. Acesso em: 16 ago. 2021.

SOUSA, R. P. M. de; DIAS, G. A.; SOUSA, M. R. F. de. Análise sobre dados abertos e regulação autoral no contexto da editoria científica. *In*: SHINTAKU, M.; SALES, L. F.; COSTA, M. (Orgs.). **Tópicos sobre dados abertos para editores científicos**. São Paulo: ABEC, 2020a. p. 119-135. DOI: [10.21452/978-85-93910-04-3.cap10](https://doi.org/10.21452/978-85-93910-04-3.cap10).

SOUZA, D. L. *et al.* A perspectiva dos pesquisadores sobre os desafios da pesquisa no Brasil 1. **Educação e Pesquisa**, v. 46, p. e221628, 2020. Disponível em: <https://www.scielo.br/ijep/a/WgdZnSMrX49LLTJMffmsqNK/?format=html&lang=pt>. Acesso em: 2 set. 2021.

STRAUSS, A.; CORBIN, J. **Pesquisa Qualitativa: técnicas e procedimentos para o desenvolvimento de teoria fundamentada**. Porto Alegre: Artmed. 2008.

SUBRAMANYA, S. H.; LAMA, B.; ACHARYA, K. P. Impact of COVID-19 pandemic on the scientific community. **Qatar Medical Journal**, [S.l.], v. 2020, n. 1, p. 1-3, mar., 2020. Hamad bin Khalifa University Press (HBKU Press). <http://dx.doi.org/10.5339/qmj.2020.21>. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7372729/>. Acesso em: 11 set. 2021.

TARGINO, M. das G. M. **Comunicação científica: o artigo de periódico nas atividades de ensino e pesquisa do docente universitário brasileiro na pós-**

graduação. 1998. 387 f. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 1998. Disponível em: <https://repositorio.unb.br/handle/10482/34362>. Acesso em: 19 mar. 2020.

TENOPIR, C. *et al.* Data sharing by scientists: practices and perceptions. **Plos One**, [S.l.], v. 6, n. 6, 2011. <http://dx.doi.org/10.1371/journal.pone.0021101>.

TENOPIR, C. *et al.* Changes in data sharing and data reuse practices and perceptions among scientists worldwide. **Plos One**, [S.l.], v. 10, n. 8, ago., 2015. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pone.0134826>.

TENOPIR, C. *et al.* Data management education from the perspective of science educators. **International Journal Of Digital Curation**, [S.l.], v. 11, n. 1, p. 232-251, nov., 2016. Edinburgh University Library. <http://dx.doi.org/10.2218/ijdc.v11i1.389>.

THE ROYAL SOCIETY. **Science as an open enterprise**. London: The Royal Society Science Policy Centre, 2012. Disponível em: <https://royalsociety.org/~media/policy/projects/sape/2012-06-20-saoe.pdf>. Acesso em: 04 out 2020.

TOMANIK, E. A. **O olhar no espelho: “conversas” sobre a pesquisa em Ciências Sociais**. Maringá: EDUEM, 1994.

TORRES-SALINAS, D.; ROBINSON-GARCÍA, N.; CABEZAS-CLAVIJO, A. Compartir los datos de investigación em ciência: introducción al data sharing. **El profesional de la información**, v. 21, n. 2, p. 173-184, mar./abr., 2012. Disponível em: <http://eprints.rclis.org/16786/>. Acesso em: 10 jan. 2021.

TORINO, E.; TREVISAN, G. L.; VIDOTTI, S. A. B. G. Dados abertos CAPES: um olhar à luz dos desafios para publicação de dados na web. **Ciência da Informação**, v. 48, n. 3, 20 mar., 2019. Disponível em: <http://revista.ibict.br/ciinf/article/view/4866>. Acesso em: 4 out. 2020.

TORINO, E.; ROA-MARTÍNEZ, S. M.; VIDOTTI, S. A. B. G. Dados de pesquisa: disponibilização ou publicação?. *In*: SHINTAKU, M.; SALES, L. F; COSTA, M. (Orgs.). **Tópicos sobre dados abertos para editores científicos**. São Paulo: ABEC, 2020. p. 183-201. DOI: [10.21452/978-85-93910-04-3.cap15](https://doi.org/10.21452/978-85-93910-04-3.cap15).

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: 17 mar. 2020.

VEIGA, V. S. de O. **Gestão de dados de pesquisa FAIR: dando um jump em seus dados**. 2019. Disponível em: https://www.arca.fiocruz.br/bitstream/icict/33343/2/ve_Veiga_Viviane_ICICT_I_Encontro_Sudeste_RIAA_2019.pdf. Acesso em: 07 maio 2020.

VEIGA, V. S. de O. *et al.* Compartilhamento de dados de pesquisa na Fiocruz: diagnóstico e percepção do pesquisador. **Ciência da Informação**, v. 48, n. 3, 2019. Disponível em: <https://www.arca.fiocruz.br/handle/icict/31461>. Acesso em: 10 nov. 2020.

VIEIRA, E. **Sociologia da educação**: reproduzir e transformar. São Paulo: FDT, 1996, p. 12-41.

UNIVERSIDADE DE SÃO PAULO. Jornal da USP. **Dados mostram que ciência brasileira é resiliente, mas está no limite**. 2021. Disponível em: <https://jornal.usp.br/universidade/politicas-cientificas/dados-mostram-que-ciencia-brasileira-e-resiliente-mas-esta-no-limite/>. Acesso em: 1 set. 2021.

UNIVERSIDADE DE BOSTON. **Helping researchers with their data**. 2015. Disponível em: <https://www.bu.edu/data/>. Acesso em: 30 mar. 2020.

ZELDITCH, M. Some methodological problems of field studies. *In*: McCALL, G. M.; SIMMONS, J. L. (Orgs.). **Issues in participant observation**: a text and a reader. Reading: Addison-Wesley, 1969, p. 5-19.

ZUIDERWIJK, A.; SPIERS, H. Sharing and re-using open data: a case study of motivations in astrophysics. **International Journal Of Information Management**, [S.l.], v. 49, p. 228-241, dez., 2019. Elsevier BV. <http://dx.doi.org/10.1016/j.ijinfomgt.2019.05.024>.

WHYTE, A.; TEDDS, J. **Making the case for research data management**. 2011. Disponível em: <http://www.dcc.ac.uk/resources/briefing-papers/making-case-rdm>. Acesso em: 06 maio 2020.

WIGMORE, I. **Data life cycle**. 2017. Disponível em: <https://whatis.techtarget.com/definition/data-life-cycle>. Acesso em: 06 maio. 2020.

WALLIMAN, N., BAICHE, B. **Your research project**: a step-by-step guide for the first-time researcher. London: Sage Publications, 2001.

WALPORT, M.; BREST, P. Sharing research data to improve public health. **The Lancet**, v. 377, n. 9765, p. 537-539, 2011. Disponível em: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(10\)62234-9/fulltext?_eventId=login](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(10)62234-9/fulltext?_eventId=login). Acesso em: 18 maio 2020.

WITTENBURG, P.; HELLSTRÖM, M.; ZWÖLF, C. M. (Eds.). **Persistent identifiers**: consolidated assertions. 2017. Disponível em: <https://dx.doi.org/10.15497/RDA00027>. Acesso em: 17 mar. 2020.

ANEXO 1

Para a coleta de dados desta pesquisa foi utilizado de forma integral o questionário semiestruturado desenvolvido pela Érika Rayanne Silva de Carvalho em sua pesquisa de mestrado. Ao total foram elaboradas 22 perguntas, divididas em produção/obtenção de dados de pesquisa (6 perguntas), compartilhamento de dados de pesquisa (10 perguntas) e reuso de dados de pesquisa (6 perguntas)

PERGUNTA DE ABERTURA

1. Em linhas gerais, como é e como se dá a realização da pesquisa científica em seu campo?

PRODUÇÃO/OBTENÇÃO DE DADOS DE PESQUISA

2. Como se dá a produção/obtenção de dados que permitirão as descobertas científicas em suas pesquisas?

3. Quais são as características dos dados produzidos/obtidos durante a pesquisa que realiza (tipos, formatos, estrutura, volume, formas de análise, tecnologias envolvidas)?

4. Considerando a disciplina em que atua, quais aspectos você percebe como influência positiva na produção/obtenção de dados de pesquisa?

5. Considerando a disciplina em que atua, quais aspectos você percebe como influência negativa na produção/obtenção de dados de pesquisa?

6. Há qualquer outro aspecto que você queira mencionar sobre a prática de produção/obtenção de dados em sua disciplina?

COMPARTILHAMENTO DE DADOS DE PESQUISA

7. Quão úteis os dados que você produz/obtem durante a realização de suas pesquisas seriam para outros pesquisadores (fora de seu grupo de pesquisa)? Você estaria disposto a compartilhá-los? Por quê? Se sim, com quem?

8. O compartilhamento de dados de pesquisa, de modo que outros pesquisadores possam reutilizá-los em novos ciclos de produção do conhecimento, tem sido uma prática em algumas disciplinas. Como você percebe essa questão na disciplina em que você atua?

9. Você já teve alguma experiência de compartilhamento de dados de pesquisa? Se sim, como se deu? Se não, por quê?

10. Considerando a disciplina em que atua, quais motivos estimulariam o compartilhamento de dados de pesquisa?

11. Considerando a disciplina em que atua, quais motivos inibiriam o compartilhamento de dados de pesquisa?

12. Como você reagiria caso a agência que financia seus projetos tornasse o compartilhamento de dados de suas pesquisas uma condição para o financiamento?

13. Como você reagiria se sua instituição tornasse obrigatório o compartilhamento de dados de suas pesquisas?

14. Como você reagiria caso editores dos periódicos nos quais você publica exigissem que o artigo submetido ou aceito para publicação fosse acompanhado dos dados da pesquisa

que gerou?

15. Quando você finaliza determina pesquisa, qual é o destino dos dados que foram produzidos?

16. Há qualquer outro aspecto que você queira mencionar sobre a prática ou a possibilidade/viabilidade (ou não) de compartilhamento de dados em sua disciplina?

(RE)USO DE DADOS DE PESQUISA

17. É crescente a percepção de que o avanço da ciência pode ser acelerado e potencializado caso pesquisadores reutilizassem dados de pesquisa gerados/obtidos anteriormente por outros pesquisadores. Como você percebe essa questão? Você reutilizaria dados gerados/obtidos anteriormente por outros pesquisadores?

18. Você já teve alguma experiência de reutilização de dados que foram gerados por pesquisas conduzidas por outros pesquisadores? Se sim, como se deu? Se não, por quais motivos?

19. Quais fatores contribuiriam para que você reutilizasse em suas pesquisas dados gerados por outros pesquisadores?

20. Quais fatores contribuiriam para que você não reutilizasse em suas pesquisas dados gerados por outros pesquisadores?

21. Em sua disciplina, haveria disposição dos pesquisadores para reutilizar dados gerados por pesquisas de outros pesquisadores? Por quê?

22. Há qualquer outro aspecto que você queira mencionar sobre a prática ou a possibilidade/viabilidade (ou não) de reutilização de dados em sua disciplina?