

Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz



Instituto de Comunicação e Informação
Científica e Tecnológica em Saúde

**INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E
TECNOLÓGICA EM SAÚDE – ICICT
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMAÇÃO E COMUNICAÇÃO EM
SAÚDE – PPGICS**

LUISE MARIA COSTA DE OLIVEIRA

**GESTÃO DE DADOS DE PESQUISA EM SAÚDE NO CONTEXTO DA
COLABORAÇÃO CIENTÍFICA INTERNACIONAL: um olhar a partir das políticas dos
financiadores**

RIO DE JANEIRO

2021

LUISE MARIA COSTA DE OLIVEIRA

**GESTÃO DE DADOS DE PESQUISA EM SAÚDE NO CONTEXTO DA
COLABORAÇÃO CIENTÍFICA INTERNACIONAL: um olhar a partir das políticas dos
financiadores**

Dissertação apresentada ao Programa de Pós-Graduação em Informação e Comunicação em Saúde do Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICICT) para obtenção do grau de Mestre em Ciências.

Orientadora: Dr^a. Cícera Henrique da Silva
Coorientadora: Dr^a. Viviane Santos de Oliveira
Veiga

RIO DE JANEIRO
2021

Oliveira, Luisi Maria Costa de.

Gestão de dados de pesquisa em saúde no contexto da colaboração científica internacional: um olhar a partir das políticas dos financiadores / Luisi Maria Costa de Oliveira. - Rio de Janeiro, 2021.

150 f.; il.

Dissertação (Mestrado) – Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Pós-Graduação em Informação e Comunicação em Saúde, 2021.

Orientadora: Cícera Henrique da Silva.

Coorientadora: Viviane Santos de Oliveira Veiga.

Bibliografia: f. 122-143

1. Acordos de Cooperação Científica e Tecnológica. 2. Colaboração Científica Internacional. 3. Dados de Pesquisa em Saúde. 4. Políticas de Gestão de Dados de Pesquisa. I. Título.

LUISE MARIA COSTA DE OLIVEIRA

**GESTÃO DE DADOS DE PESQUISA EM SAÚDE NO CONTEXTO DA
COLABORAÇÃO CIENTÍFICA INTERNACIONAL: um olhar a partir das políticas dos
financiadores**

Aprovado em: 23 de fevereiro de 2021.

Banca examinadora:

Dr^a. Cícera Henrique da Silva – Orientadora
PPGICS/ICICT/Fiocruz

Dr^a. Viviane Santos de Oliveira Veiga – Coorientadora
PPGICS/ICICT/Fiocruz

Dr^a. Maria Cristina Soares Guimarães – Titular interno
PPGICS/ICICT/Fiocruz

Dr^a. Patrícia Corrêa Henning – Titular externo
PPGSTEH/UNIRIO

Dedico este trabalho à minha mãe e à minha avó
(*in memoriam*), que sempre incentivaram meus
sonhos.

AGRADECIMENTOS

Primeiramente, agradeço a Deus por ter me dado a oportunidade de vivenciar algo tão brilhante e desafiador que é a vida da pós-graduação e por ter me dado saúde e força nos momentos de maiores dificuldades para o desenvolvimento da pesquisa.

À minha família que soube compreender minhas ausências e foi uma rede de apoio e afeto que me sustentaram nesse percurso, em especial, à minha mãe que sempre me incentiva e encoraja a trilhar novos caminhos e vibra com as minhas conquistas. Você é a razão de tudo.

Ao meu amor e melhor amigo, Matheus Fernandes, que com sua paciência, carinho e incontestável cumplicidade e companheirismo me fortaleceu para concluir essa pesquisa. É um privilégio partilhar a vida com você! Estendo meus agradecimentos à sua família que me acolheu com muito amor e torceu por mim durante essa caminhada.

Às minhas amigas queridas Brenda Souza e Mariana Dias por compreenderem meu tempo limitado e pela cumplicidade e torcida em todos esses anos de amizade. Gostaria também de agradecer à Ana Rosa por toda energia positiva enviada para eu terminar a pesquisa.

Às minhas orientadoras, Cícera Henrique da Silva e Viviane Veiga, por me conduzirem nesse caminho da pesquisa com profissionalismo, entusiasmo e compartilhando conhecimentos. Agradeço pelo acolhimento com muito carinho, empatia e paciência.

À minha eterna mestre Dayanne Prudencio, quem me incentivou a entrar no mestrado ainda na graduação e quem me deu suporte em todos os momentos desse percurso.

Aos colegas da turma de 2019 que tornaram o primeiro ano de mestrado mais leve, guardarei os momentos de afeto e acolhimento para sempre. Aos professores do Programa de Pós-graduação em Informação e Comunicação em Saúde (PPGICS) com os quais pude compreender melhor que os direitos à comunicação e à informação são indissociáveis do direito à saúde. À equipe da secretária acadêmica que sempre foi muito gentil, solícita e eficiente no atendimento.

Aos professores e membros da banca, Cristina Guimarães, Josué Laguardia, Luana Sales e Patrícia Henning que prontamente aceitaram o convite para compor a banca de defesa e me deixaram muito honrada com o aceite.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa concedida para que eu pudesse me dedicar integralmente à pesquisa.

*“Todo caminho da gente é resvaloso.
Mas também, cair não prejudica demais - a gente
levanta, a gente sobe, a gente volta!
O correr da vida embrulha tudo, a vida é assim:
esquenta e esfria, aperta e daí afrouxa, sossega e
depois desinquieta.
O que ela quer da gente é coragem”.*

(Guimarães Rosa, 2005)

OLIVEIRA, Luisi Maria Costa de. **Gestão de dados de pesquisa em saúde no contexto da colaboração científica internacional**: um olhar a partir das políticas dos financiadores. 2021. 150 f. Dissertação (Mestrado em Informação e Comunicação em Saúde) – Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, 2021.

RESUMO

O presente estudo teve como objetivo investigar as orientações dos financiadores da pesquisa para gestão de dados produzidos em contexto de colaboração científica internacional no campo da saúde. A metodologia consistiu na pesquisa bibliográfica em fontes de informação nacionais e internacionais que forneceu as bases teóricas da pesquisa e permitiu identificar o National Institutes of Health e a European Commission como os principais financiadores mundiais da pesquisa em saúde. Utilizou a pesquisa documental a fim mapear e analisar os documentos normativos emitidos pelos financiadores com políticas para gestão de dados de pesquisa em saúde. A análise dos documentos baseou-se em uma estrutura de avaliação de políticas de dados que contemplou as categorias grau de aplicação do documento e cobertura de critérios. Identificou-se que ambos financiadores possuem políticas recomendatórias para gestão de dados como também compulsórias que se não atendidas podem impedir a concessão de financiamento. Verificou-se que dentre os critérios contemplados na estrutura de avaliação, o NIH e a EC abordaram totalmente os critérios de escopo, monitoramento de conformidade e sanções, suporte, repositórios e centros de dados, custo, compartilhamento e acesso e questões éticas relativas aos dados de pesquisa. Constatou-se que os financiadores reconhecem que algumas questões relativas à gestão dos dados requerem negociação e construção de acordos entre os parceiros antes do início de um projeto colaborativo. Verificou-se a existência de orientações para gestão de dados produzidos em contextos colaborativos, sem, no entanto, ser possível verificar se eram aplicadas igualmente em colaborações internacionais. Concluiu-se que a lacuna de normativas para gestão de dados produzidos nesses contextos pode provocar confrontos normativos, uma vez que possivelmente os pesquisadores partícipes dessas colaborações resgatarão suas normas nacionais e institucionais para sustentar a realização da pesquisa. Constatou-se ainda que na perspectiva das políticas dos financiadores a maior parte das responsabilidades com as questões relacionadas à gestão dos dados gerados na investigação centraliza-se na figura do coordenador do projeto, mesmo a pesquisa sendo realizada pelo trabalho colaborativo entre pesquisadores.

Palavras-chaves: Acordos de Cooperação Científica e Tecnológica. Colaboração Científica Internacional. Dados de Pesquisa em Saúde. Políticas de Gestão de Dados de Pesquisa.

OLIVEIRA, Luisi Maria Costa de. **Gestão de dados de pesquisa em saúde no contexto da colaboração científica internacional**: um olhar a partir das políticas dos financiadores. 2021. 150 f. Dissertação (Mestrado em Informação e Comunicação em Saúde) – Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, 2021.

ABSTRACT

This study aimed to investigate the guidelines of research funders for data management produced in the context of international scientific collaboration in the field of health. The methodology consisted of bibliographic research on national and international sources of information that provided the theoretical basis of the research and allowed identifying the National Institutes of Health and the European Commission as the main global funders of health research. It used the documentary research in order to map and analyze the normative documents issued by funders with policies for health research data management. The analysis of the documents was based on a data policy evaluation framework that contemplated the degree of document application and criteria coverage. It was identified that both funders have recommendatory policies for data management as well as compulsory policies that if not met can prevent the granting of funding. It was verified that among the criteria contemplated in the evaluation structure, the NIH and the CE fully addressed the scope criteria, compliance monitoring and sanctions, support, repositories and data centers, cost, sharing and access and ethical issues related to the research data. It was found that funders recognize that some data management issues require negotiation and agreement building among partners before a collaborative project can begin. Guidelines for data management produced in collaborative settings were found to exist, but it was not possible to verify that they were also applied in international collaborations. It was concluded that the lack of norms for the data management produced in these contexts can provoke normative confrontations, since possibly the researchers participating in these collaborations will rescue their national and institutional norms to support the realization of the research. It was also found that in the perspective of the policies of the funders most of the responsibilities with the issues related to the management of the data generated in the research are centered in the figure of the project coordinator, even though the research is carried out by collaborative work among researchers.

Keywords: Scientific and Technological Cooperation Agreements. International Scientific Collaboration. Health Research Data. Research Data Management Policies.

LISTA DE ILUSTRAÇÕES

| | | |
|-------------|---|---------|
| Quadro 1 – | Termos selecionados para a busca bibliográfica..... | 67-68 |
| Quadro 2 – | Principais financiadores da pesquisa em saúde..... | 69 |
| Quadro 3 – | Estrutura de avaliação de políticas de dados de pesquisa..... | 71 |
| Quadro 4 – | Descrição dos elementos-chaves da categoria “cobertura de critérios” presentes na estrutura de avaliação de políticas de gestão de dados de pesquisa..... | 72 |
| Quadro 5 – | Documentos normativos do NIH com políticas para gestão de dados de pesquisa..... | 79 |
| Quadro 6 – | Definições concebidas pelo NIH para alguns tipos de dados de pesquisa..... | 80 |
| Quadro 7 – | Métodos previstos pelo NIH para o compartilhamento de dados de pesquisa..... | 81 |
| Figura 1 – | Repositório de dados genômicos humanos e não humanos recomendados pelo NIH..... | 83 |
| Quadro 8 – | Documentos normativos do H2020 com políticas para gestão de dados de pesquisa..... | 88 |
| Figura 2 – | Fluxos para o acesso aberto a publicações científicas e dados de pesquisa no contexto mais amplo de divulgação e exploração..... | 90 |
| Quadro 9 – | Avaliação geral das políticas para gestão de dados de pesquisa do NIH..... | 95 |
| Quadro 10 – | Grau de aplicação das principais políticas para gestão de dados no NIH..... | 96 |
| Gráfico 1 – | Porcentagem de cobertura da subcategoria “aspectos gerais” por critério (NIH)..... | 98 |
| Gráfico 2 – | Porcentagem de cobertura da subcategoria “planejamento e infraestrutura” por critério (NIH)..... | 99 |
| Gráfico 3 – | Porcentagem de cobertura da subcategoria “disponibilização e reuso” por critério (NIH)..... | 101 |
| Quadro 11 – | Avaliação geral das políticas para gestão de dados de pesquisa do H2020..... | 103 |
| Quadro 12 – | Grau de aplicação das principais políticas para gestão de dados de pesquisas envolvidas em ações de saúde no H2020..... | 104 |
| Quadro 13 – | Razões para a não concessão de acesso aberto aos dados de pesquisas financiadas pelo programa europeu H2020..... | 105 |
| Gráfico 4 – | Porcentagem de cobertura da subcategoria “aspectos gerais” por critério (EC)..... | 106 |
| Gráfico 5 – | Porcentagem de cobertura da subcategoria “Planejamento e infraestrutura” por critério (EC)..... | 107 |
| Gráfico 6 – | Porcentagem de cobertura da subcategoria “Disponibilização e Reuso” por critério (EC)..... | 109 |
| Figura 3 – | Distribuição das principais responsabilidades identificadas para a gestão de dados de pesquisa no âmbito dos consórcios de pesquisa (EC)..... | 115 |
| Quadro 14 – | Recomendações para a elaboração de PGD de projetos colaborativos no âmbito do H2020..... | 115-116 |

LISTA DE SIGLAS E ABREVIATURAS

| | |
|---------|--|
| ANDS | <i>Australian National Data Service</i> |
| ARC | <i>Australian Research Council</i> |
| BBSRC | <i>Biological and Biomedical Sciences Research Council</i> |
| BD2K | <i>Big Data to Knowledge</i> |
| BDTD | Biblioteca Digital Brasileira de Teses e Dissertações |
| BRAPCI | Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação |
| caBIG | <i>Cancer Biomedical Informatics Grid</i> |
| CARL | <i>Canadian Association of Research Libraries</i> |
| CERN | <i>European Organization for Nuclear Research</i> |
| C&T | Ciência & Tecnologia |
| CGHub | <i>Cancer Genome Hub</i> |
| CIHR | <i>Canadian Institutes of Health Research</i> |
| CODATA | <i>Committee on Data of the International Science Council</i> |
| CRIS | Centro de Relações Internacionais em Saúde |
| DACs | <i>Data Access Committees</i> |
| DataNet | <i>Sustainable Digital Data Preservation and Access Network Partners</i> |
| DataOne | <i>DataNet Observation Network for Earth</i> |
| dbGaP | <i>Database of Genotypes and Phenotypes (The)</i> |
| DCC | <i>Digital Curation Centre</i> |
| DDBJ | <i>DNA Data Bank of Japan</i> |
| DeCS | Descritores em Ciências da Saúde |
| DFG | <i>Deutsche Forschungsgemeinschaft</i> |
| DOI | <i>Digital Object Identifier</i> |
| DSPC | <i>NIH Data Science Policy Council</i> |
| EC | <i>European Commission</i> |
| EDPB | <i>European Data Protection Board</i> |
| ENA | <i>European Nucleotide Archive</i> |
| EOSC | <i>European Open Science Cloud</i> |
| EPSRC | <i>Engineering and Physical Sciences Research Council</i> |
| ERC | <i>European Research Council</i> |

| | |
|-----------|---|
| ERCEA | <i>European Research Council Executive Agency</i> |
| ESRC | <i>Economic and Social Research Council</i> |
| EU | <i>European Union</i> |
| EUA | Estados Unidos da América |
| EUDAT CDI | <i>EUDAT Collaborative Data Infrastructure</i> |
| EVS | <i>Enterprise Vocabulary Services</i> |
| FAIR | <i>Findable, Accessible, Interoperable e Reusable</i> |
| FAPESP | Fundação de Amparo à Pesquisa do Estado de São Paulo |
| Fiocruz | Fundação Oswaldo Cruz |
| FOIA | <i>Freedom of Information Act</i> |
| GDPR | <i>General Data Protection Regulation</i> |
| GDS | <i>Genomic Data Sharing</i> |
| GEO | <i>Gene Expression Omnibus</i> |
| GWAS | <i>Genome-Wide Association Studies</i> |
| H2020 | <i>Horizon 2020</i> |
| HHMI | <i>Howard Hughes Medical Institute</i> |
| HHS | <i>US Department of Health and Human Services</i> |
| IBICT | Instituto Brasileiro de Informação em Ciência e Tecnologia |
| ICICT | Instituto de Comunicação e Informação Científica e Tecnológica em Saúde |
| ICs | Institutos e Centros |
| ICSU | <i>International Council of Science</i> |
| Inserm | <i>Institut National de la Santé et de la Recherche Médicale</i> |
| JISC | <i>Joint Information Systems Committee</i> |
| LC | <i>Library of Congress</i> |
| LGPD | Lei Geral de Proteção de Dados Pessoais |
| MCTIC | Ministério da Ciência, Tecnologia e Inovações |
| Medline | <i>Medical Literature Analysis and Retrieval System Online</i> |
| MGI | <i>Mouse Genome Informatics</i> |
| MRC | <i>UK Medical Research Council</i> |
| MS | Ministério da Saúde |
| MSDWG | <i>Metadata Standards Directory Working Group</i> |
| NCI | <i>National Cancer Institute</i> |
| NDAR | <i>National Database for Autism Research (The)</i> |

| | |
|-------------|---|
| NHMRC | <i>Australia's National Health and Medical Research Council</i> |
| NIAGADS | <i>National Institute on Aging Genetics of Alzheimer's Disease Data Storage (The)</i> |
| NIH | <i>National Institutes of Health</i> |
| NINDS | <i>National Institutes of Neurological Disorders and Stroke</i> |
| NIST | <i>National Institute of Standards and Technology</i> |
| NLM | <i>National Library of Medicine</i> |
| NSERC | <i>Natural Sciences and Engineering Research Council of Canada</i> |
| NSF | <i>National Science Foundation</i> |
| NIHGPS | <i>NIH Grants Policy Statement</i> |
| OCDE | <i>Organização para a Cooperação e Desenvolvimento Econômico</i> |
| OD | <i>Office of the Director</i> |
| ODSS | <i>Office of Data Science Strategy</i> |
| OECD | <i>Organisation for Economic Co-operation and Development</i> |
| OER | <i>Office of Extramural Research</i> |
| OMB | <i>Office of Management and Budget</i> |
| OMS | <i>Organização Mundial da Saúde</i> |
| OpenAIRE | <i>Open Access Infrastructure for Research in Europe</i> |
| ORD | <i>Open Research Data Pilot</i> |
| OSC | <i>Office of Strategic Coordination</i> |
| OSP | <i>Office of Science Policy</i> |
| OSTP | <i>Office of Science and Technology Policy</i> |
| P&D | <i>Pesquisa e Desenvolvimento</i> |
| PCD | <i>Plano de Compartilhamento de Dados</i> |
| PCDG | <i>Plano de Compartilhamento de Dados Genômicos</i> |
| PGD | <i>Plano de Gestão de Dados</i> |
| PMC | <i>PubMed Central</i> |
| PPGICS | <i>Programa de Pós-Graduação em Informação e Comunicação em Saúde</i> |
| RCUK | <i>Research Council UK</i> |
| RDA | <i>Research Data Alliance</i> |
| RDM | <i>Research Data Management and Sharing Policies</i> |
| Re3data.org | <i>Registry of Research Data Repositories</i> |
| SNP | <i>Single Nucleotide Polymorphisms</i> |

| | |
|--------|--|
| SRA | <i>Sequence Read Archive</i> |
| SSHRC | <i>Social Sciences and Humanities Research Council</i> |
| TIC | Tecnologias de Informação e Comunicação |
| US DoD | <i>United States Department of Defense</i> |
| VODAN | <i>Virus Outbreak Data Network</i> |
| WDC | <i>World Data Center</i> |
| ZFIN | <i>Zebrafish Model Organism Database</i> |

SUMÁRIO

| | | |
|----------|---|-----------|
| 1 | INTRODUÇÃO..... | 16 |
| 1.1 | OBJETIVOS..... | 21 |
| 1.2 | JUSTIFICATIVA..... | 21 |
| 2 | A CIÊNCIA E O TRABALHO COLABORATIVO..... | 23 |
| 2.1 | COOPERAÇÃO E COLABORAÇÃO CIENTÍFICA INTERNACIONAL..... | 24 |
| 2.2 | COLABORAÇÃO CIENTÍFICA NA PESQUISA EM SAÚDE..... | 28 |
| 2.3 | DESAFIOS NAS COLABORAÇÕES CIENTÍFICAS INTERNACIONAIS..... | 30 |
| 3 | ESTRUTURA NORMATIVA DA CIÊNCIA..... | 32 |
| 3.1 | O <i>ETHOS</i> DA CIÊNCIA REVISITADO..... | 35 |
| 3.2 | CONSIDERAÇÕES EM TORNO DA CONDUTA E INTEGRIDADE DA PESQUISA..... | 37 |
| 4 | GESTÃO DE DADOS DE PESQUISA: EMERGÊNCIA DE UMA AGENDA DE PESQUISA..... | 41 |
| 4.1 | POLÍTICAS DE FINANCIADORES INTERNACIONAIS E NACIONAIS..... | 48 |
| 4.2 | DADOS DE PESQUISA NO CONTEXTO DA PESQUISA EM SAÚDE..... | 55 |
| 4.3 | GESTÃO, CURADORIA E PRESERVAÇÃO DE DADOS DE PESQUISA..... | 57 |
| 4.4 | REPOSITÓRIOS DE DADOS, PADRÕES DE METADADOS E OS PRINCÍPIOS FAIR..... | 61 |
| 5 | PERCURSO METODOLÓGICO..... | 65 |
| 5.1 | CARACTERIZAÇÃO DA PESQUISA..... | 65 |
| 5.2 | DEFINIÇÃO DE CRITÉRIOS PARA COLETA E ANÁLISE DE DADOS..... | 66 |
| 6 | APRESENTAÇÃO DOS RESULTADOS..... | 74 |
| 6.1 | NATIONAL INSTITUTES OF HEALTH (NIH)..... | 74 |
| 6.2 | EUROPEAN COMMISSION (EC)..... | 85 |
| 7 | ANÁLISE DOS RESULTADOS..... | 95 |
| 7.1 | ANÁLISE DAS POLÍTICAS PARA GESTÃO DE DADOS DE PESQUISA DO NIH..... | 95 |
| 7.2 | ANÁLISE DAS POLÍTICAS PARA GESTÃO DE DADOS DE PESQUISA NO ÂMBITO DO PROGRAMA HORIZON 2020..... | 103 |
| 7.3 | CONSIDERAÇÕES ACERCA DA GESTÃO DE DADOS DE PESQUISA EM SAÚDE NO CONTEXTO DA PESQUISA COLABORATIVA | |

| | | |
|----------|---|------------|
| | INTERNACIONAL..... | 111 |
| 8 | CONSIDERAÇÕES FINAIS..... | 118 |
| | REFERÊNCIAS..... | 122 |
| | APÊNDICE A – Plano de Gestão de Dados..... | 144 |
| | ANEXO A – Componentes do PGD FAIR para projetos do Horizon 2020..... | 148 |
| | ANEXO B – Componentes do PGD para projetos financiados pelo ERC no âmbito do Horizon 2020..... | 150 |

1 INTRODUÇÃO

A pesquisa científica envolve cada vez mais esforços interdisciplinares, interinstitucionais e internacionais a fim de ampliar o conhecimento científico e solucionar os problemas sociais. Na emergência de novos desafios globais como as alterações climáticas, biodiversidade, alimentação, segurança energética e hídrica e saúde global, a Royal Society (2011, p. 72, tradução nossa¹) considera ser imperativa a “cooperação internacional em larga escala devido à natureza e magnitude das potenciais consequências desses problemas. Nenhum país ou disciplina científica será capaz de oferecer soluções completas”.

Em relatório emitido pela Organização Mundial da Saúde (OMS) é relatado o aumento da publicação da pesquisa em saúde como resultado da colaboração internacional. Ainda que os países de alta renda continuem a desempenhar um papel de destaque na maioria dos estudos, os cientistas de países de renda média e baixa estão cada vez mais envolvidos em parcerias internacionais, sendo a parcela de crescimento de coautorias internacionais com pesquisadores brasileiros ainda vagoroso (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2014). Estudos mais recentes apontaram que o Brasil ocupava a quarta posição mundial de países que mais colaboraram internacionalmente na pesquisa em resposta à emergência de saúde pública da infecção pelo vírus Zika em 2015, sendo os Estados Unidos da América (EUA) o principal colaborador, seguido pela Inglaterra, França e Alemanha (OLIVEIRA *et al.*, 2020).

As colaborações em pesquisa entre países abrangem um escopo complexo de considerações legais e regulatórias que precisam ser reconhecidas e negociadas, a fim de evitar choques normativos e garantir um horizonte normativo que sustente a realização dos compromissos e responsabilidades assumidos, conforme afirmam Bohnhorst *et al.* (2011) e Jesus (2019). Os desafios mais proeminentes em colaborações entre países decorrem da necessidade de pesquisadores e instituições atenderem a demandas de diferentes sistemas nacionais de pesquisa que variam em quatro principais dimensões: organização e financiamento das pesquisas; ambientes legais e normativos; regulamentação e supervisão e; educação e treinamento (ANDERSON, 2011b). Variações internacionais nessas áreas podem levar a expectativas diferentes sobre a condução da pesquisa, comprometendo a qualidade e integridade da pesquisa (ANDERSON, 2011b).

¹ No original: “[...] *international co-operation on a large scale because of the nature and magnitude of the potential consequences of these problems. No one country or scientific discipline will be able to offer complete solutions*” (ROYAL SOCIETY, 2011, p. 72).

Direcionando a discussão para dimensão normativa, Jesus (2019, p. iv) explica que atores envolvidos em colaborações internacionais estão, por vezes, sujeitos a uma “pluralidade normativa e possíveis diferenças e/ou variações normativas em torno de questões que potencialmente afetam a integridade da pesquisa”, como a concepção do projeto de pesquisa, gestão de dados de pesquisa e materiais de pesquisa, análise e proteção de dados, propriedade intelectual, conflitos de interesse, pesquisas envolvendo seres humanos e animais, entre outras. Entre essas questões, discute-se nesse estudo as políticas para a gestão de dados de pesquisa gerados no contexto da colaboração científica internacional, ou seja, o conjunto de recomendações, normas, diretrizes, regras, entre outros instrumentos normativos.

Os dados de pesquisa têm assumido protagonismo central no processo de comunicação na ciência, dado o seu potencial de apoiar a transparência, reprodutibilidade na pesquisa e possível reuso para nova produção de conhecimento. Há pelo menos uma década, instituições acadêmicas, financiadores da pesquisa em saúde e os principais periódicos da área emitiram declarações nas quais ratificaram que as estratégias de gestão de dados são preponderantes para garantir o compartilhamento, acesso e reuso de dados de pesquisa, de modo a maximizar o conhecimento e os seus benefícios para a saúde (WELLCOME TRUST, 2012; 2016).

Gestão de dados de pesquisa é um termo genérico que envolve os processos de coleta, seleção, estruturação, nomeação, interpretação, análise, *backup*, armazenamento, preservação, compartilhamento, acesso, abertura e todas demais ações que garantam a segurança e reuso dos dados e contribuam para a integridade da pesquisa (GIFFELS; VOLLMER; BIRD, 2010; SCHÖPFEL *et al.*, 2018). Segundo Sayão e Sales (2015, p. 16), “é uma das áreas essenciais na conduta responsável da pesquisa nos ambientes científicos atuais”, sendo associada a benefícios como garantia da confiabilidade, integridade e replicabilidade dos dados; proteção de investimentos financeiros empregados na coleta dos dados; suporte à disponibilização dos dados da pesquisa para o escrutínio e/ou verificação por outros pesquisadores; diminuição de perdas ou esforços duplicados para recriar dados perdidos (DOUCETTE; FYFE, 2013).

Nessa perspectiva, nos últimos anos, financiadores internacionais da pesquisa, expressivamente os dos Estados Unidos e do Reino Unido e, mais recentemente em nível nacional, passaram a estabelecer recomendações e exigências mandatórias em políticas de dados com o propósito de “reordenar as condutas dos pesquisadores e das instituições de pesquisa em relação à gestão e à disponibilidade dos dados” (SAYÃO; SALES, 2018, p. 4188). Como condição para concessão de financiamento, pesquisadores e instituições precisam atender a requisitos, cada vez mais compulsórios, relacionados à gestão dos dados, como a elaboração do Plano de Gestão de Dados (PGD), de acordo com Veiga *et al.* (2019).

Os dados de pesquisa são o cerne das colaborações internacionais em pesquisa. O uso desses dados requer o conhecimento acerca das implicações do acesso e compartilhamento dos dados por diferentes atores e das preocupações culturais, legais, políticas e técnicas determinantes para manutenção da privacidade, confidencialidade e segurança dos dados (NATIONAL ACADEMIES OF SCIENCES, ENGINEERING, AND MEDICINE, 2018).

A necessidade de harmonização das práticas de gestão de dados no âmbito dos acordos estabelecidos entre parceiros individuais e institucionais envolvidos em colaborações internacionais foi reconhecida em documentos deliberativos como a “Declaração de Montreal sobre a integridade da pesquisa em colaborações de pesquisa entre fronteiras”, de 2013. O documento “*Practical guide to the international alignment of research data management*”² emitido pela Science Europe (2018), também aborda a necessidade de identificar a coordenação de responsabilidades com a gestão dos dados entre os parceiros envolvidos em projetos colaborativos. Nesse contexto, um número crescente de projetos de pesquisa, principalmente realizados em contextos internacionais, está adotando políticas de gestão e compartilhamento de dados para chegar a uma abordagem comum entre os países envolvidos (SHEARER, 2015).

Não obstante, os principais financiadores da pesquisa científica reconhecem que a adesão dos pesquisadores às políticas de dados é um grande desafio em virtude das diferenças significativas de infraestruturas disponíveis para viabilizar a gestão e compartilhamento dos dados e das práticas diversificadas para o tratamento dos dados que estão enraizadas na cultura de cada disciplina (SHEARER, 2015).

Sobre a questão, Pisani *et al.* (2018) citam como exemplo a área genômica, na qual a gestão de dados é uma prática consolidada, pois financiadores além de incentivarem o compartilhamento de dados, investiram fortemente no desenvolvimento de padrões e outras formas de infraestrutura. No entanto, na pesquisa em saúde, Pisani e Abouzahr (2010) verificam que a gestão de dados é uma atividade subvalorizada e subfinanciada, oferecendo prejuízos à pesquisa. Além da carência de infraestruturas, particularidades como a proteção de dados sensíveis e confidenciais, preocupações éticas quanto à integridade e privacidade dos sujeitos da pesquisa, restrições legais e regulatórias específicas dos países, interesses comerciais e contratuais podem se tornar obstáculos determinantes para o compartilhamento de dados de pesquisa em saúde (PISANI; ABOUZAHAR, 2010; SANTOS *et al.*, 2017).

²Nova versão publicada em 27 de janeiro de 2021. Disponível em: <https://www.scienceeurope.org/our-resources/practical-guide-to-the-international-alignment-of-research-data-management/>. Acesso em: 27 jan. 2021.

As colaborações cada vez mais “inter” – interdisciplinares, interinstitucionais e internacionais – podem tornar a gestão dos dados produzidos nesses contextos ainda mais complexa. Wang, Göpfert e Stark (2016) problematizam os possíveis desacordos em projetos interdisciplinares, visto que a gestão dos dados ainda é caracterizada a partir de uma perspectiva disciplinar, no qual cada domínio possui seus próprios conjuntos de metadados. Nas colaborações interinstitucionais, a gestão de dados pode tornar-se problemática uma vez que as instituições envolvidas geralmente possuem sua própria organização e sistemas técnicos, além de seus pesquisadores terem entendimentos diferentes do que realmente são “dados de pesquisa” (WANG; GÖPFERT; STARK, 2016). No contexto da colaboração internacional, a conformidade das atividades de gestão de dados pode ser dificultada, pois as políticas podem ou não existir de forma abrangente (BRINEY; GOBEN; ZILINSKI, 2015).

Finkel *et al.* (2020) afirmam que a gestão de dados de pesquisa não é uma atividade necessária apenas para atender ao objetivo abrangente da ciência aberta de garantir o acesso amplo aos resultados de pesquisa produzidos pelo financiamento público, mas também desempenha papel crucial para o sucesso das pesquisas individuais e colaborativas. Particularmente na pesquisa colaborativa, é fundamental uma infraestrutura que permita o compartilhamento de dados entre os pesquisadores envolvidos, tornando a gestão dos dados uma atividade chave na conduta responsável da pesquisa (GIFFELS; VOLLMER; BIRD, 2010; CURDT; HOFFMEISTER, 2015).

No entanto, em sua pesquisa Jesus (2019) verifica certo desconhecimento dos pesquisadores envolvidos em projetos colaborativos internacionais quanto às normas relativas à gestão de dados de suas instituições. Giffels, Vollmer e Bird (2010) explicam que embora a adoção de boas práticas de gestão de dados de pesquisa pareça ser um elemento óbvio e essencial para qualquer investigação científica, em muitos casos, o conjunto total ou parcial dessas práticas é negligenciado. Segundo Wallis (2012), existem muitas visões sobre a gestão de dados de pesquisa a partir de uma perspectiva de “cima para baixo”, ou seja, a partir das expectativas dos financiadores de pesquisa e curadores de dados. Todavia, as funções e responsabilidades com a gestão de dados, na maioria das vezes, não são claras para os pesquisadores e produtores de fontes primárias para pesquisa (WALLIS, 2012).

Para Sallans e Lake (2014), uma das barreiras para as práticas de gestão de dados é a abrangência das informações oferecidas nos documentos normativos dos financiadores da pesquisa referentes à maneira como o pesquisador deve realizar as atividades relacionadas à gestão e curadoria de dados. Terry, Littler e Olliaro (2018) convalidam essa percepção na pesquisa em saúde ao ressaltarem que ainda existe pouca conformidade nas políticas das

agências e instituições de fomento, devido às orientações genéricas e aos recursos limitados oferecidos pelos financiadores para apoiar os pesquisadores à implementação das políticas no cotidiano do trabalho científico.

Por outro lado, a saúde pública, considerada um direito do cidadão, necessita de novas soluções para o acesso e reuso de dados de pesquisa para “subsidiar conhecimentos mais integrados sobre os determinantes e os condicionantes da saúde, bem como o enfrentamento de emergências sanitárias” (SANTOS *et al.*, 2017, p. 14). Sendo assim, Pisani e Abouzahr (2010) declaram que o estabelecimento ou melhorias de práticas de gestão de dados de pesquisa para subsidiar o processo de compartilhamento é o primeiro passo para reduzir a chances de perda de dados, que na área da saúde pública, pode traduzir-se em expectativas de tempo de vida menores ou vidas menos saudáveis.

Nessa perspectiva, por um lado, observa-se que embora o número de políticas para gestão de dados de pesquisa emitidas principalmente por financiadores da pesquisa tenha crescido nos últimos anos, ainda existe uma baixa adesão dos pesquisadores em atividades relacionadas à gestão de dados de pesquisa no campo da saúde. Por outro lado, observa-se que os pesquisadores envolvidos em colaboração científica internacional podem enfrentar desafios em virtude das variações e/ou diferenças normativas em torno de questões relativas à gestão de dados de pesquisa, que podem resultar em efeitos deletérios sobre a integridade da pesquisa. Diante dessas considerações iniciais, a pergunta que norteia a pesquisa é: como os financiadores da pesquisa em saúde orientam a gestão de dados de pesquisa aos pesquisadores que estão em programas ou projetos de pesquisas colaborativas internacionais?

A expectativa dessa dissertação foi de fornecer um panorama das recomendações e exigências relacionadas à gestão de dados de pesquisa estabelecidas pelos financiadores de pesquisa em saúde, contribuindo para um maior conhecimento a respeito das políticas e/ou normativas que precisam ser atendidas para garantir a realização dos compromissos e responsabilidades com o tratamento dos dados gerados no âmbito da colaboração científica internacional.

Olhar para as políticas dos financiadores pode ser um primeiro passo para observar a existência de lacunas de normativas ou recomendações que pode tornar a gestão dos dados um desafio nesse tipo de empreendimento científico, no qual possivelmente pesquisadores de diferentes países precisarão equacionar divergências de entendimentos e equalizar práticas sobre aspectos da gestão dos dados da pesquisa que não foram abordados previamente no âmbito dos instrumentos normativos e estruturas regulatórias fornecidas pelos financiadores como condição para realização da pesquisa.

1.1 OBJETIVOS

Para responder à questão da pesquisa, este estudo teve como objetivo principal investigar as orientações dos financiadores da pesquisa para gestão de dados produzidos em contexto de colaboração científica internacional no campo da saúde. Para atingir esse objetivo geral, os objetivos específicos da pesquisa foram:

- a) Caracterizar os principais financiadores da pesquisa em saúde;
- b) Identificar e analisar as orientações para a gestão de dados nos documentos normativos dos financiadores da pesquisa em saúde;
- c) Mapear e analisar as responsabilidades com a gestão de dados produzidos em pesquisas colaborativas envolvendo pesquisadores de diferentes países;

1.2 JUSTIFICATIVA

Como observado, os pesquisadores e instituições envolvidos em colaborações internacionais de pesquisa devem estar especialmente atentos, dentre outros aspectos, às questões normativas que podem afetar a integridade da pesquisa. Dentre as questões que precisam ser reconhecidas e negociadas por esses pesquisadores e instituições, discute-se nesta pesquisa a gestão de dados de pesquisa, uma temática que desencadeia uma série de desafios para a seara científica no que se refere à coleta, organização, análise, preservação, compartilhamento, acesso, abertura e reuso dos dados de pesquisa.

A proposta de realizar um estudo que analise como a gestão de dados de pesquisa é orientada por financiadores da pesquisa em saúde em contextos de pesquisas colaborativas internacionais tem a intenção de contribuir para o campo teórico da área, fazendo avançar o conhecimento acerca das recomendações e exigências endereçadas a pesquisadores quando trabalham nesse tipo de empreendimento científico.

A discussão no contexto institucional é propícia em um momento em que a Fundação Oswaldo Cruz (Fiocruz), a maior instituição de Ciência & Tecnologia (C&T) em Saúde do Brasil, está preocupada em adaptar-se às novas práticas de produção científica desencadeadas pelo movimento do acesso aberto e Ciência Aberta e cada vez mais se torna uma instituição estratégica do Estado brasileiro expressiva no cenário global da saúde, envolvida em ações de cooperação internacional.

A temática ainda encontra aderência na relevância social, uma vez que as estratégias de gestão de dados de pesquisa oferecem maior segurança aos investimentos financeiros

públicos ao mitigar a possibilidade de perda de dados e a necessidade de duplicação de esforços para recriá-los quando perdidos, visto que embora o valor real dos dados de pesquisa não seja determinado pelo custo empreendido no processo de coleta, ele se torna evidente na economia de recursos despendidos em pesquisas subsequentes. Além disso, as estratégias de gestão de dados são fundamentais para tornar os dados de pesquisa em saúde localizáveis, acessíveis, interoperáveis e reutilizáveis para pesquisas voltadas para a qualidade de vida da população brasileira e potenciais subsídios para o aperfeiçoamento de políticas públicas. As estratégias de gestão de dados de pesquisa também podem otimizar a resposta da ciência para sociedade, acelerando descobertas científicas como vacinas e medicamentos.

Ainda importa mencionar que pelas questões levantadas, esta pesquisa está alinhada ao Programa de Pós-Graduação em Informação e Comunicação em Saúde (PPGICS) do Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICICT/FIOCRUZ), especificamente como uma produção da linha 1 – “Produção, Organização e Uso da Informação em Saúde”, no eixo “regimes de produção, regulação e novas dinâmicas de pesquisa científica em saúde”.

Para o desenvolvimento do estudo, foram construídos oito capítulos. O **primeiro capítulo**, de carácter introdutório, apresenta o tema, questão de pesquisa, objetivos e justificativa. O **segundo capítulo** se dedica ao tema da colaboração científica internacional, seu papel na pesquisa em saúde e algumas considerações acerca dos desafios inerentes a esse tipo de empreendimento. O **terceiro capítulo** aborda as normas sociais da ciência e discute, brevemente, os temas da conduta responsável na pesquisa e integridade da pesquisa. O **quarto capítulo** versa a respeito da gestão de dados de pesquisa, abordando os principais fatores que impulsionaram o tema, as principais iniciativas governamentais e políticas de dados de financiadores da pesquisa científica, seguida de uma breve discussão acerca do que são dados de pesquisa, gestão, curadoria e preservação de dados, repositórios de dados, padrões de metadados e os princípios FAIR, um acrônimo para *Findable, Accessible, Interoperable e Reusable*. O **quinto capítulo** descreve os procedimentos metodológicos adotados para atingir os objetivos desta pesquisa, que teve como trabalho empírico a análise dos documentos normativos dos financiadores da pesquisa em saúde com políticas para a gestão dos dados de pesquisa. O **capítulo seis e sete** apresentam, respectivamente, os resultados obtidos e a análise. Por fim, o **capítulo oito** apresenta as considerações finais acerca do tema estudado, as limitações da pesquisa e possíveis caminhos para novos estudos.

2 A CIÊNCIA E O TRABALHO COLABORATIVO

O trabalho colaborativo científico é caracterizado pela cooperação entre cientistas que compartilham entre si a investigação científica. Segundo Ziman (1979), a “ciência é conhecimento público”, que passa por análise crítica e avaliação dos cientistas que determinam sua validade enquanto científico e universalmente aceito. O autor caracteriza a ciência como uma atividade social, sendo permeada pela interação entre cientistas que se organizam de formas específicas, compartilham informações entre si e cumpre sua função dentro do sistema que transforma o conhecimento em propriedade pública. Desse modo, a ciência não é realizada por um único indivíduo, mas por um grupo de cientistas que colaboram mutuamente e fiscalizam as contribuições de cada um.

Na visão de Meadows (1999), a colaboração científica pode ser vista desde os primórdios da pesquisa. Beaver e Rosen (1978) identificam que as primeiras colaborações foram desenvolvidas em resposta à profissionalização da ciência durante o século XVII, constituindo-se um novo estilo de pesquisa associado às transformações ocorridas durante a Revolução Científica. Para os autores, a “profissionalização” provocou mudanças estruturais dentro da ciência, redefinindo como a mesma era feita, quem a fazia, onde era feita, quem a financiava, quais eram as pretensões de seus participantes e como se tornavam cientistas.

Anteriormente à Segunda Guerra Mundial, a ciência era feita por indivíduos isolados em seus laboratórios de pesquisa que, embora recorressem a seus pares, em geral realizavam suas pesquisas em isolamento pessoal (MEADOWS, 1999). A partir do “Projeto Manhathan” que desenvolveu a bomba atômica, a ciência, nas palavras de Schwartzman (1976, sem paginação) “parece dar um salto para a *‘big science’*, caracterizada pelos grandes orçamentos e pela complexidade das atividades da pesquisa” envolvendo o trabalho coordenado de milhares de pessoas. Como afirma Price (1986), a mudança ocorrida na forma de colaboração entre os cientistas caracterizou a passagem da *“little science”* para a *“big science”*.

Para Price (1986), a estrutura e dinâmica do trabalho científico são identificadas em um amplo espectro de colaboração que ocorre em colégios invisíveis. O autor afirma que um colégio invisível é constituído por um grupo de cientistas de uma área de interesse em comum que, em níveis locais, nacionais e internacionais, contribuem materialmente para pesquisas sobre determinado assunto, compartilham informações e outros materiais de pesquisa de forma informal e formal com outros pesquisadores, encontram-se em eventos científicos, podendo também administrar fundos de pesquisa, controlar o prestígio pessoal e o destino de novas ideias científicas (PRICE, 1986).

Meadows (1999) afirma que a noção de trabalho em colaboração enquanto uma atividade orientada foi desenvolvida durante a primeira metade do século XX, quando surgiram os primeiros grupos de científicos formados por um pesquisador sênior, assistentes de pesquisa, estudantes de doutorado e técnicos, sendo intensificada após a Segunda Guerra Mundial. Segundo Price (1986), os esforços empregados na guerra forçaram o surgimento de núcleos de pesquisadores orientados ao trabalho colaborativo. Reconhecidas as potencialidades da ciência no pós-guerra, Velho (2011) ressalta uma transformação radical na percepção e a imagem que o público tinha sobre a ciência. A autora afirma que a ciência passa a desempenhar um papel estratégico como força produtiva, o que a tornou tema de grande relevância e interesse entre nações.

Nessa perspectiva, Lima, Velho e Faria (2007) afirmam que as colaborações científicas começam a ocorrer entre pesquisadores de uma mesma instituição e, em outros casos, por colegas de instituições diferentes de uma mesma região ou país e, na contemporaneidade, sobretudo no período entre 1988 a 2001, “a colaboração internacional cresceu muito mais rapidamente do que a colaboração entre as instituições dentro das fronteiras nacionais” (LIMA; VELHO; FARIA, 2007, p. 51).

Para Meadows (1999), as razões básicas do trabalho colaborativo encontram-se no próprio crescimento e especialização da pesquisa. Anderson (2011a) considera que atualmente, a pesquisa tornou-se, em certa medida, uma empresa global onde equipes transnacionais realizam projetos de pesquisa em conjunto, devido a uma multiplicidade de fatores econômicos, cognitivos e sociais internos ao empreendimento científico que serão brevemente abordados adiante. Segundo Luukkonen, Persson e Sivertsen (1992), esses fatores acrescidos das iniciativas financeiras governamentais para viagens e programas de ciência aprimoram a colaboração científica, especialmente a colaboração internacional, tornando-a objetivo político entre as nações.

2.1 COOPERAÇÃO E COLABORAÇÃO CIENTÍFICA INTERNACIONAL

Na medida em que a globalização e a internacionalização expandem na perspectiva de formação de uma comunidade global, diversos setores da sociedade passaram por transformações em suas estruturas e mecanismos de funcionamento. Nesse contexto, a cooperação internacional tornou-se um elemento essencial das relações internacionais, uma vez que “[...] não são apenas os mercados para produtos, serviços e recursos financeiros de um país que se abrem por meio da cooperação, mas também, e talvez até mais importante, a

cooperação possibilita o intercâmbio de pessoas, experiências e conhecimentos” (JESUS, 2019, p. 68), por meio da cooperação científica internacional.

Conforme destacado pela Royal Society (2011, p. 6, tradução nossa³), “o intercâmbio de ideias, conhecimentos e habilidades científicas estão mudando o foco da ciência do nível nacional para o global”. Para Braga (1974) e Lima, Velho e Faria (2007), a ciência é uma tarefa internacional que cresce exponencialmente em escala global, de tal modo que a cooperação científica internacional tornou-se fundamental para a garantia da qualidade da ciência moderna e para o sucesso da pesquisa científica. A disponibilização do conhecimento científico em periódicos disponíveis globalmente e o envolvimento dos pesquisadores em projetos de pesquisa com parceiros de outros países são alguns dos múltiplos aspectos que imprimem à ciência um caráter internacional e colaborativo (LIMA; VELHO; FARIA, 2007).

A cooperação científica pode envolver a participação pesquisadores de diferentes países e atores heterogêneos, como universidades, empresas, consórcios, agências de fomento, entre outros (JESUS, 2019). Segundo Georghiou (1998), a cooperação científica internacional é expressa através do intercâmbio entre pesquisadores, incluindo bolsas de estudo; oficinas ou outras reuniões; projetos ou redes cooperativas; oferta de acesso ou compartilhamento de custos de instrumentos científicos ou instalações de grande escala; parcerias entre laboratório; participação em programas nacionais do país colaborador; estabelecimento de laboratórios subsidiários no país parceiro; e patrocínio ou participação em programas nacionais.

A cooperação científica abrange tanto atividades formais representadas por grandes instalações, programas específicos de apoio e atividade regidos por acordos científicos, quanto atividades informais como a interação dos cientistas em viagens, comunicações e trocas de ideias e materiais, sem incorporar a esse relacionamento um contrato formal (GEORGHIOU, 1998). Segundo Jesus (2019, p. 69), a cooperação científica internacional “tem sido um relevante meio para alavancar a internacionalização da ciência e da pesquisa, sendo a colaboração científica internacional um dos muitos modos de realização da cooperação internacional, podendo assumir uma variedade de arranjos e atividades”. A autora ainda ressalta que o aumento da colaboração científica internacional expressa evidências da intensificação da dimensão global da ciência.

A colaboração científica pode ser entendida como o “[...] trabalho conjunto de pesquisadores para atingir um objetivo comum de produzir novos conhecimentos científicos”

³ No original: “[...] exchange of scientific insight, knowledge and skills, they are changing the focus of science from the national to the global level” (ROYAL SOCIETY, 2011, p. 6).

(KATZ; MARTIN, 1997, p. 7, tradução nossa⁴). Sonnenwald (2007, p. 645, tradução nossa⁵) complementa ao considerá-la como “a interação que se situa em um contexto social entre dois ou mais cientistas, a qual facilita a partilha do propósito e da ação de execução de tarefas, respeitando um objetivo superordenado e compartilhado mutuamente”. Esse contexto social envolve “elementos como revisão por pares, sistemas de recompensa, colégios invisíveis, paradigmas científicos e políticas científicas nacionais e internacionais, como também normas disciplinares e institucionais” (SONNENWALD, 2007, p. 645, tradução nossa⁶).

O colaborador na pesquisa é o indivíduo que trabalha cooperativamente ao longo do desenvolvimento da pesquisa ou em parte significativa do processo; aquele cujo nome consta na proposta original da pesquisa; aquele que é responsável por um ou mais elementos principais da pesquisa (KATZ; MARTIN, 1997). Em outros casos, também se consideram colaboradores aqueles responsáveis por ideias, hipóteses originais e interpretação teórica bem como o proponente do projeto inicial e/ou agenciador de fundos. Para Vanz (2009), geralmente, são excluídos aqueles que fizeram contribuições insignificantes e aqueles que não são considerados propriamente pesquisadores, como técnicos e assistentes de pesquisa.

Segundo Katz e Martin (1997), a colaboração pode acontecer em vários níveis a depender dos contextos em que é desenvolvida. Em um nível básico, são as pessoas que colaboram e não as instituições, desta forma a parceria entre dois ou mais pesquisadores é a unidade básica da colaboração científica. A colaboração também pode ocorrer em outros níveis – entre pesquisadores de um mesmo laboratório ou departamento de uma mesma instituição; entre instituições; entre campos disciplinares; e entre espaços geograficamente separados. Esses últimos níveis de colaboração são mais promovidos em vez de colaborações em nível interindividual (KATZ; MARTIN, 1997). Portanto, entende-se que as colaborações podem ocorrer em níveis disciplinares, setoriais, institucionais, nacionais e internacionais.

Importa sinalizar que a colaboração científica não é sinônima de equidade em pesquisa, devido a razões pragmáticas, ou seja, “só se procura uma colaboração quando se acredita que se tem algo a ganhar” (VELHO, 2001, p. 62). Silva (2007) explica que a colaboração em pesquisa implica a existência de um ator principal como o responsável pelo projeto ou programa de pesquisa, enquanto os outros membros da equipe são atores

⁴ No original: “[...] *the working together of researchers to achieve the common goal of producing new scientific knowledge*” (KATZ; MARTIN, 1997, p. 7).

⁵ No original: “[...] *interaction taking place within a social context among two or more scientists that facilitates the sharing of meaning and completion of tasks with respect to a mutually shared, superordinate goal*” (SONNENWALD, 2007, p. 645).

⁶ No original: “[...] *elements such as peer review, reward systems, invisible colleges, scientific paradigms, and national and international science policies, as well as disciplinary and university norms*” (SONNENWALD, 2007, p. 645).

coadjuvantes que se limitam à assistência técnico-científica e formação de recursos humanos para a pesquisa; à utilização de equipamentos e outros recursos oferecidos pelos “donos da pesquisa”; à possibilidade de oferecer ambientes em seu território para servir de observação e coleta de dados pelo pesquisador principal. Grande parte dos projetos colaborativos internacionais que envolvem países do Norte e do Sul possui esse perfil, no qual países desenvolvidos definem ou controlam a agenda de pesquisa, o financiamento, o gerenciamento do programa ou projeto, a prioridade nos resultados, entre outras questões (DODSON, 2017).

A respeito das razões para a colaboração em pesquisa, Sonnenwald (2007) destaca a crescente especialização da ciência, o elevado grau de complexidade dos instrumentos científicos e técnicos e a necessidade de abordagens transdisciplinares e multidisciplinares para resolver problemas complexos. Para Anderson (2011a), o alto custo para administração das infraestruturas complexas de pesquisa executadas na ‘*big science*’ também tem estimulado aos pesquisadores e instituições a colaborarem. Em alguns casos, as colaborações são incentivadas, pois muitos cientistas precisam trabalhar com outros parceiros para acessar os últimos desenvolvimentos em seu campo (ROYAL SOCIETY, 2011).

A colaboração científica internacional, caracterizada pela “pesquisa realizada em conjunto por cientistas cujas principais afiliações institucionais estão em diferentes países” (ANDERSON, 2011a, tradução nossa⁷) tem sido influenciada por três principais fatores: a) econômicos relacionados ao custo das pesquisas, vantagens de compartilhar e manter recursos e investimentos financeiros de pesquisa, e um sentido mais amplo, a possibilidade de aumento da produtividade; b) cognitivos ligados às necessidades dos cientistas no acesso a novos conhecimentos tanto interno como externo ao seu campo disciplinar, o que gera redes de colaboração multidisciplinares; c) sociais ligados à rede de relacionamento entre os cientistas devido a proximidades pessoais, afinidades temáticas, emocionais ou ideológicas (LUUKKONEN; PERSSON; SIVERTSEN, 1992).

Para Wagner *et al.* (2001), a proximidade geográfica, a economia, os laços históricos, a linguagem e os equipamentos de pesquisa, bancos de dados e laboratórios também são fatores que impulsionam as colaborações científicas internacionais. A premência em solucionar os desafios globais do século XXI, como a saúde global, tem estimulado o estabelecimento de parcerias entre pesquisadores, instituições de pesquisa e nações, impulsionando “[...] estratégias científicas nacionais e multinacionais e de muitos mecanismos

⁷ No original: “[...] *research pursued jointly by scientists whose primary institutional affiliations are in different countries*” (ANDERSON, 2011a, p. 1).

de financiamento” (ROYAL SOCIETY, 2011, p. 7, tradução nossa⁸). Wagner *et al.* (2001) consideram que os problemas transfronteiriços, como o controle de doenças, são um dos principais motivos para as colaborações internacionais. Nessa perspectiva, procurou-se, na seção seguinte, compreender o papel desse empreendimento científico na pesquisa em saúde.

2.2 COLABORAÇÃO CIENTÍFICA NA PESQUISA EM SAÚDE

A pesquisa científica e tecnológica em saúde refere-se a atividades de investigação científica que resultam em “produtos e processos que contribuem para promover, manter ou recuperar a saúde dos seres humanos” (GUIMARÃES; SOUZA; SANTOS, 2012, p. 241). De forma geral, Guimarães (2006) considera que a pesquisa em saúde pode ser dividida em quatro segmentos de pesquisa: clínica, biomédica, em saúde pública e a tecnológica em saúde.

Com exceção da pesquisa voltada para fins militares, o setor saúde mobiliza a maior quantidade de financiamento no mundo, contabilizando em 2005, cerca de US\$ 160,3 bilhões destinados à pesquisa em saúde, distribuídos de forma bastante heterogênea entre os países de renda alta e renda média ou baixa (GUIMARÃES; SOUZA; SANTOS, 2012). Nesse período, os EUA, Reino Unido, Japão, Alemanha e França eram responsáveis por 72,5% da produção total da pesquisa em saúde, enquanto o Brasil, China, Rússia, Turquia e África do Sul foram responsáveis por 4,4% do *output* bibliográfico mundial (GUIMARÃES, 2006).

De acordo com Buss e Leal (2009), o processo de globalização tem criado grandes disparidades, produzindo problemas sociais e na saúde, principalmente nos países à margem dos circuitos centrais da economia global. Nessa perspectiva, a partir da década de 1990 foram impulsionados empreendimentos colaborativos na pesquisa em saúde global diante das alegações de que as colaborações internacionais têm o potencial de desempenhar um papel significativo no tratamento das disparidades globais em saúde (PARKER; KINGORI, 2016).

De acordo com Buss (2013), a saúde global se constitui um tema de grande interesse internacional no século XXI. A saúde global tem como base a saúde pública com foco na saúde da coletividade, na interdisciplinaridade e ações de promoção, prevenção e recuperação da saúde humana e a saúde internacional que materializa a saúde como assunto tratado politicamente pelas nações, Estados e governos (BUSS; TOBAR, 2017; RIBEIRO, 2016).

A saúde global “envolve disciplinas além das ciências da saúde e que promove a colaboração interdisciplinar” com foco “na melhoria de condições de saúde ao redor do

⁸ No original: “[...] *national and multinational science strategies and many funding mechanisms*” (ROYAL SOCIETY, 2011, p. 7).

mundo, na redução de disparidades e proteção contra ameaças globais que não respeitam fronteiras nacionais” (RIBEIRO, 2016, p. 24). Buss e Leal (2009) assinalam que o Brasil tem participado ativamente no cenário da saúde global por meio do que se tem chamado de “diplomacia da saúde⁹”, sobretudo no contexto da cooperação sul-sul¹⁰, firmando acordos de cooperação com alguns países da América do Sul e África. No contexto brasileiro, Buss (2013) cita a Fiocruz que por meio do Centro de Relações Internacionais em Saúde (CRIS) tem dado maior organicidade à sua já ampla função de cooperação internacional em saúde.

As redes de colaboração, mecanismos estratégicos para a cooperação internacional, são de grande importância no desempenho das instituições de C&T em Saúde, uma vez que a “pesquisa científica e o desenvolvimento de inovações para a área da saúde são cada vez mais multidisciplinares e complexos, a inserção em redes de cooperação é condição essencial para o sucesso organizacional” (FONSECA; FONSECA, 2015, p. 188). Segundo a autora, as colaborações são empreendimentos promissores em respostas às situações de emergência na saúde pública, na promoção de inovações e serviços em saúde, no desenvolvimento de produtos e instrumentos de qualidade para os sistemas de saúde.

Em que pese às cooperações sul-sul tenham recebido crescente atenção dos países em desenvolvimento do hemisfério sul, como bem lembram Buss e Ferreira (2010), a maior parte das atividades de cooperação internacional em saúde ainda reflete as relações de poder do mundo globalizado, nas quais a ‘assistência’ internacional é ‘provida’ pelos países desenvolvidos e ‘recebida’ pelos países em desenvolvimento. Na pesquisa em saúde, grande atenção foi dada aos projetos de colaboração internacional norte-sul, nos quais os países desenvolvidos contribuem com recursos financeiros, conhecimentos, laboratórios de ponta e outros recursos especiais, enquanto os parceiros de países em desenvolvimento fornecem conhecimento clínico local e outros conhecimentos contextuais (LUCAS, 2005).

Na pesquisa médica, por exemplo, o financiamento e o gerenciamento de ensaios clínicos estão além do escopo e da capacidade de apenas uma nação (ANDERSON, 2011a). Particularmente para essa pesquisa, as colaborações internacionais são importantes, pois permite o agrupamento de dados, conhecimentos e recursos, promove a produção e troca de conhecimento entre países desenvolvidos e em desenvolvimento, além de padronizar medidas e reduzir a duplicação de pesquisas (VANNI *et al.*, 2014). Nessa perspectiva, Ferreira *et al.*

⁹ Trata-se de campo novo de conhecimento e prática, cujo objeto é a saúde e suas negociações internacionais (BUSS, 2013).

¹⁰ “[...] é o processo de parceria econômica, comercial, social ou de outra natureza que se estabelece (idealmente), trazendo vantagens mútuas para os países em desenvolvimento parceiros, geralmente localizados no Hemisfério Sul” (BUSS, 2011).

(2017) defendem que o enfrentamento dos desafios globais na saúde deve ser realizado por meio de atividades em que pesquisadores de diferentes países colaborem em projetos estratégicos e que assegurem benefícios acadêmicos e sociais para os países participantes.

Todavia, cabe considerar que na pesquisa colaborativa internacional em saúde corre-se o risco de haver certa desigualdade entre os parceiros envolvidos, na ocasião de colaboradores de países desenvolvidos explorarem as vulnerabilidades dos cientistas e instituições dos países em desenvolvimento, concentrando a pesquisa nos interesses das instituições patrocinadoras, ao invés das necessidades locais do país anfitrião (LUCAS, 2005).

A respeito do que se tem discutido, Parker e Kingori (2016) exemplificam por meio do exame de colaborações internacionais de pesquisa em saúde global entre pesquisadores americanos, europeus e os da África Central, que o papel mais frequente dos pesquisadores africanos era fornecer amostras e conduzir trabalho de campo, não o projeto de pesquisa. Essas questões vêm sendo discutidas no âmbito da ‘governança ética’ para a garantia de altos padrões éticos em colaborações de pesquisa em saúde global. Buss e Ferreira (2010) ainda complementam defendendo que o atual modelo de cooperação internacional dominante só será efetivamente modificado se também forem alteradas as estruturas e regras da governança global em saúde.

2.3 DESAFIOS NAS COLABORAÇÕES CIENTÍFICAS INTERNACIONAIS

Atores envolvidos em colaborações estão, por vezes, sujeitos a desafios que em casos extremos podem até impedir o progresso da ciência e da pesquisa (ANDERSON, 2011b; VASCONCELOS *et al.*, 2012). Muitos são os desafios para a realização das pesquisas internacionais, como a seleção de parceiros, áreas e prioridades de pesquisa; distribuição de responsabilidades; produtividade das equipes que lidam com demandas de diferentes estruturas administrativas e regulatórias; diferenças normativas entre os sistemas de pesquisas dos quais os pesquisadores fazem parte; padrões e mecanismos de supervisão distintos no que tange a garantia da integridade da pesquisa, entre outros (ANDERSON, 2011b; JESUS, 2019). Na pesquisa em saúde, Parker e Kingori (2016) relatam que na visão dos pesquisadores a condução de uma pesquisa colaborativa internacional constitui-se um complexo entrelaçamento de preocupações científicas, sociais, políticas e éticas.

Anderson (2011a) e Jesus (2019) consideram que o entusiasmo dos financiadores em esforços conjuntos e a ânsia em “tornar-se global” geralmente ocultam as complicações e desafios culturais, epistêmicos, linguísticos, normativos, entre outros, inerentes à pesquisa

colaborativa. Anderson e Steneck (2011, sem paginação, tradução nossa¹¹) ressaltam que as “abordagens desconhecidas para treinamento, complicações legais e regulatórias e diferenças de financiamento e administração colocam desafios à colaboração que são agravados pela necessidade de satisfazer os requisitos de diferentes sistemas de pesquisa”.

Pesquisadores oriundos de diferentes sistemas nacionais de pesquisa também podem enfrentar possíveis desacordos em face das diferenças de entendimentos e práticas no que diz respeito à determinação de autoria, custódia dos dados de pesquisa, transferência de materiais entre colaboradores, laboratórios e instituições, patentes, privacidade, propriedade intelectual, comunicação de resultados, entre outros (JESUS, 2019). Isso porque, as colaborações possivelmente envolvem pesquisadores com diferentes experiências disciplinares e institucionais e que seguem tradições e normas informais para lidar com aspectos dessas dimensões (SONNENWALD, 2007). Anderson (2011b) e Jesus (2019) consideram que formas diferentes de abordar essas questões podem levar a expectativas e posturas conflitantes na prática científica entre os atores envolvidos em empreendimentos colaborativos científicos internacionais, com efeitos indiretos ou diretos sobre a integridade da pesquisa.

Nesse contexto, Jesus (2019) chama atenção para os desafios que podem emergir a partir de questões de natureza normativa, como o direito e licença de propriedade intelectual, disseminação e gestão de dados de pesquisa e distribuição de responsabilidades com respeito às violações legais, entre outras questões que precisam ser reconhecidas e negociadas a fim de evitar conflitos. Para Jesus e González de Gómez (2018), a dimensão normativa constitui-se uma potencial zona de conflitos e “choques” normativos gerados pela presença de uma pluralidade de documentos normativos em torno de questões relacionadas à integridade da pesquisa, que por suas possíveis diferenças e variações pode não ser facilmente harmonizada. Sendo assim, dimensão normativa é “[...] um espaço em que o equacionamento, por meio de negociações, entre discórdia e harmonia poderá levar (ou não) a uma situação que possibilite a realização da colaboração científica internacional” (JESUS, 2019, p. 83).

Portanto, a colaboração em pesquisa, não exclusivamente a internacional, envolve desafios que precisam ser equacionados. Ainda cabe ressaltar que a pesquisa colaborativa não se desassocia das normas e práticas científicas tradicionais, conforme afirma Sonnenwald (2007). Nesse contexto das normas ligadas à própria atividade científica, Jesus (2019) destaca as normas sociais da ciência. Sendo assim, o próximo capítulo se ocupa em abordá-las.

¹¹ No original: “*Unfamiliar approaches to training, legal and regulatory complications, and differences in funding and administration pose challenges for collaborations that are then compounded by the need to satisfy the requirements of different research systems*” (ANDERSON; STENECK, 2011, sem paginação).

3 ESTRUTURA NORMATIVA DA CIÊNCIA

A investigação científica é um fenômeno orientado por normas e valores de caráter institucional e transnacional e não apenas por práticas individuais dos atores envolvidos nessa atividade. Conforme afirma Kalleberg (2015), o período anterior à Segunda Guerra Mundial, teve o controle institucional da má conduta na pesquisa realizado de forma informal, com base no poder de autocorreção e autorregulação da ciência que excluiria indivíduos considerados “*bad apples*”. A autora afirma que atualmente o sistema normativo da ciência é mais supervisionado e contemplado com uma série de instrumentos legais como ouvidoria de integridade em pesquisa e comitês de ética nas universidades e instituições de pesquisa.

O entendimento que se tem hoje sobre ciência é fruto de um processo cada vez mais sociológico que coloca o cientista e a ciência como partes integrantes da estrutura e do dinamismo social. No entanto, a antiga visão da ciência e dos cientistas como entidades isoladas, autônomas, independente da sociedade e com implicações neutras “em si” é para Morel (1979) fruto da própria história da ciência e da sociologia da ciência.

Após a Segunda Guerra Mundial, instalou-se um clima de tensão dentro da comunidade científica em virtude do papel da C&T na construção do aparato militar fascista no período entre guerras (VELHO, 2011). Para a autora, a ideia de ciência autônoma, ou seja, fora dos processos sociais, coloca-a vulnerável a ideologias e projetos políticos. Havia uma ansiedade presente entre os cientistas sobre as implicações morais do Projeto Manhattan, responsável por inaugurar uma nova relação entre ciência e Estado (MOREL, 1979).

Conforme aponta Velho (2011), uma parte influente da comunidade científica começa a distanciar-se da ciência aplicada e focaliza na chamada ciência básica. Essa noção “básica” e separada da sociedade tenta colocar a ciência “fora do alcance dos governos totalitários e, analogamente, de qualquer outra forma de interferência social” (VELHO, 2011, p. 136). Na pesquisa básica ocorre a “busca desinteressada pelo novo, procura de conhecimento universalmente válido, sem considerar a sua possível importância para a solução de problemas práticos” (STORER, 1975, 130).

Em uma tentativa de estabelecer um complexo de valores e normas morais para “os homens da ciência”, o trabalho “*Science and the Social Order*” (1938) do proeminente sociólogo da ciência Robert K. Merton foi o primeiro delineamento da estrutura normativa de autorregulação da ciência, considerado um marco para a história da ciência moderna. Storer (1975) considera que a estruturação da ciência determinada por um conjunto de normas e valores forneceu a estrutura para grande parte da pesquisa “básica” em ciência.

As críticas feitas à comunidade científica, visto que muitos cientistas tinham as mãos ‘sujas’ devido ao Projeto Manhattan, convidam-na a uma autocrítica: a conscientização da ciência como elemento integrante da sociedade e das obrigações e interesses correspondentes (MERTON, 1968). Diante dos ataques à integridade da ciência, Merton (1968 p. 651) assinala que os cientistas passam a reconhecer “sua dependência de certos tipos de estrutura social”. Para Merton (1968, p. 651) “uma instituição que sofre ataques tem que examinar de novo seus fundamentos, revisar seus objetivos, buscar sua explicação racional”.

Para Merton (1968) era necessário transformar o isolacionismo otimista da ciência em participação realista no conflito revolucionário das culturas, desse modo, o sociólogo reafirma o *ethos* da ciência moderna, conceituando-o como um “complexo de valores e normas afetivamente tonalizado, que se considera como constituindo uma obrigação moral para o cientista” (MERTON, 1968, p. 652). Essas normas são expressas “em forma de prescrições, proscricões, preferências e permissões, que se legitimam em relação com valores institucionais”; seriam transmitidas “pelo preceito e pelo exemplo e reforçados por sanções” e assimiladas “em graus variáveis pelo cientista, formando assim uma consciência científica” (MERTON, 1968, p. 653).

Nessa direção, Velho (2011, p. 137) afirma que “para justificar que pessoas carregadas de interesses, e sujeitas às relações sociais e influências culturais mais variadas sejam capazes de produzir conhecimento objetivo” as normas sociais de Merton são fundamentais para modelar e normatizar o comportamento dos membros da comunidade científica a fim de garantir a produção de conhecimento isento de valores e influências sociais.

Os quatro imperativos morais básicos propostos por Merton funcionam como um sistema de autorregulação interno da ciência, descritos como normas, que forneceram uma base para relações sociais da ciência, sendo elas: universalismo, comunismo ou comunalismo, desinteresse e ceticismo organizado (MERTON 1968). Conforme aponta Veiga (2017), a essas normas foram acrescentadas por Merton e seus seguidores a originalidade, individualismo, humildade, imparcialidade, independência e neutralidade emocional.

O “universalismo” estabelece que a aceitação e a rejeição de pretensões à verdade, quaisquer que sejam suas origens, devem ser submetidas a critérios impessoais preestabelecidos consoantes com a observação e com o conhecimento pré-estabelecido. Para Merton (1968, p. 654), “o imperativo de universalismo tem raízes profundas no caráter impessoal da ciência”, assim a ciência não deve depender dos atributos pessoais ou sociais do cientista.

O “comunismo” ou “comunalismo” defende que o conhecimento científico é produto da colaboração social e destinado à comunidade. O autor defende que o único direito do cientista à sua propriedade intelectual limita-se à gratidão e à estima institucional. Merton ainda sublinha a importância do cientista em comunicar suas descobertas à comunidade, uma vez que “o conceito institucional da ciência como parte do domínio público está ligado ao imperativo da comunicação dos resultados. O segredo é a antítese dessa norma; a plena e franca comunicação é o seu cumprimento” (MERTON, 1968, p. 658).

O “desinteresse” prevê que a atividade do cientista seja orientada ao interesse coletivo da ciência. Segundo Merton (1968, p. 660-661), o desinteresse “tem firme alicerce no caráter público e testável da ciência e podemos supor que esta circunstância contribui para a integridade do homem e da ciência” de modo que “os cientistas têm, mais cedo ou mais tarde, prestar contas perante aos seus colegas” (MERTON, 1968, p. 661).

O “ceticismo organizado” exige que o conhecimento seja submetido a escrutínio de critérios empíricos e lógicos (MERTON, 1968). Deste modo, para a pesquisa ser aceita no corpo de “conhecimento garantido”, é necessário que passe por uma vigilância recíproca entre os pares que obriga o cientista a receber as contribuições de seus pares no seu campo de estudo, “para tornar conhecida a sua avaliação e também para ser igualmente crítico de seu próprio trabalho antes de continuar a partilhá-lo com outros” (STORER, 1975, p. 125). Portanto, o cientista deve duvidar de suas próprias descobertas, bem como as de outros.

Todavia, ainda que a instituição da ciência fosse regida por valores e normas centrais que formam a base da atividade científica, a natureza desta energia é orientada ao tipo de meta que direciona a ciência, sendo endereçada à busca e ampliação de conhecimentos válidos (MERTON, 1968). Foi então que, em 1957, Merton salientou que o reconhecimento profissional é a única recompensa legítima pela realização da ciência e propicia a realimentação para o engajamento do cientista na pesquisa: a norma da originalidade da descoberta é de grande importância capital para a motivação dos cientistas (STORER, 1975).

A ciência com seu sistema de alocar recompensas pelo papel desempenhado pelo cientista relativo às contribuições científicas inéditas intensifica a competição no campo, o que para Merton (1957) desperta possíveis comportamentos desviantes na atividade científica. Para o autor “nessa situação de estresse, todo tipo de comportamento adaptativo é colocado em ação, sendo que alguns deles estão muito além dos *mores* da ciência” (MERTON, 1957, p. 659, tradução nossa¹²).

¹² No original: “*In this situation of stress, all manner of adaptive behaviors are called into play, some of these being far beyond the mores of Science*” (MERTON, 1957, p. 659).

Os quatro imperativos institucionais especificados por Merton estabelecem “uma base para o comportamento profissional do cientista individual, de forma a garantir não apenas o funcionamento da ciência, mas também uma base para a relação entre ciência e sociedade” (JESUS, 2019, p. 87). No entanto, alguns autores observam que nem sempre a atividade científica segue esses nobres padrões, sobretudo porque a instituição da ciência é permeada por um sistema de recompensas que envolve prestígio, autoridade e reconhecimento. Ainda que o trabalho de Merton seja referência das normas que orientam o comportamento dos cientistas e do próprio desenvolvimento do conhecimento científico, uma série de revisitações às formulações sobre o *ethos* da ciência foi realizada.

3.1 O *ETHOS* DA CIÊNCIA REVISITADO

Parece haver um consenso entre apoiadores e críticos às normas sociais da ciência em concordar que as formulações mertonianas forneceram um dos fundamentos teóricos mais significativos para a sociologia da ciência moderna. Ainda que as normas do *ethos* da ciência não tenham perdido sua função, essas foram e ainda são revisitadas por autores contemporâneos que buscam revisar o modelo de fazer ciência.

Para Stehr (1978) a maior parte das críticas às formulações mertonianas foi influenciada pelo trabalho seminal “*The structure of scientific revolutions*” de Thomas Kuhn (1962), que fornece uma teoria alternativa das relações da ciência, a partir das observações feitas durante sua estadia no *Center for Study in the Behavioral Sciences* (1958-1959). Os relatos de Kuhn sobre as divergências manifestadas entre os cientistas sobre a natureza dos problemas reputados legítimos até a escolha de métodos justificáveis foram basilares para as argumentações pós mertonianas. Diversos autores da nova sociologia da ciência propuseram interpretações conceituais a partir da revisitação às normas mertonianas. Destaca-se aqui as “contra-normas” do teórico organizacional americano Ian Mitroff (1974a; 1974b).

A análise de Mitroff em seu livro intitulado *The Subjective Side of Science* (1974a) parte da premissa que embora o *ethos* científico tenha sido fundamental para a ciência, sua aplicabilidade na atividade científica pode parecer um tanto ingênua (CUPANI, 1998). Mitroff (1974a, 1974b) questiona explicitamente a validade funcional do *ethos* científico para o avanço do conhecimento científico, uma vez que sustenta o entendimento que tais normas não representam, de maneira exclusiva, a racionalidade científica. Neste sentido, Mitroff “ênfatiza a importância de conjuntos de normas conflitantes como condições para a possibilidade de racionalidade e crescimento na ciência” (STEHR, 1978, p. 181).

Na década de 1970, Mitroff (1974b) investiga as atitudes e comportamento de quarenta e dois cientistas que participaram nas missões “Apollo” da NASA por aproximadamente quatro anos, concluindo que os cientistas seguem tanto as normas como as contra normas, a depender da natureza do problema. Nas palavras de Mitroff (1974b, p. 594, tradução nossa¹³), “enquanto que as normas convencionais da ciência são dominantes para problemas bem estruturados, as contra normas propostas aqui parecem ser dominantes para problemas mal estruturados”. Mitroff (1974b) evidencia a prevalência das contra normas entre os cientistas que trabalhavam no Projeto Apollo, especialmente o particularismo em vez do universalismo, o solitarismo contrapondo ao comunismo ou comunalismo, o interesse no lugar do desinteresse, e o dogmatismo organizado em oposição ao ceticismo organizado.

O “particularismo” demonstra que as características sociais e psicológicas do cientista são fatores importantes que influenciam como seu trabalho será julgado. O “solitarismo” prevê o sigilo como um ato moral necessário. O “interesse” considera que os cientistas esperam que seus colegas alcancem o interesse próprio na satisfação do trabalho e prestígio em servir suas comunidades especiais de interesse. Por fim, o “dogmatismo organizado” quando o cientista acredita em suas próprias descobertas com total convicção, duvidando das de outras pessoas com todo o seu valor (MITROFF, 1974b).

Para Veiga (2017, p. 32), enquanto a “análise de Merton buscava um ponto de vista da ciência institucionalizada, sem levar em conta os aspectos cognitivos, [...] a análise de Mitroff busca um ponto de vista psicológico, levando em conta o comportamento individual do cientista”. Tal comportamento do cientista é de extrema relevância quando se discute a má conduta na pesquisa. De acordo com Jesus (2019), as normas mertonianas parecem, por vezes, ser consideradas superficialmente pelos pesquisadores, em situações complexas no cotidiano da pesquisa, que possuem comportamentos não correspondentes à sua ‘versão literal’.

Desde as formulações mertonianas no período pós-guerra, a ciência passou por extensas mudanças tecnológicas, sociais e financeiras que suscitaram um amplo debate ético na pesquisa sobre novas questões para garantir que a pesquisa seja conduzida com responsabilidade (GONZÁLEZ DE GÓMEZ, 2017). Há um grande movimento dentro da comunidade científica para estabelecer novas diretrizes para garantir a integridade do processo de pesquisa. Desse modo, a seguir, serão contempladas algumas considerações envolvendo a conduta responsável na pesquisa e a integridade da pesquisa.

¹³ No original: “Whereas the conventional norms of science are dominant for well-structured problems, the counter-norms proposed here appear to be dominant for ill-structured problems” (MITROFF, 1974b, p. 594).

3.2 CONSIDERAÇÕES EM TORNO DA CONDUTA E INTEGRIDADE DA PESQUISA

Embora a ciência dispusesse de mecanismos internos de autorregulação providos por processos de validação, entre as décadas de 1970 e 1990 são relatados os primeiros casos de má conduta na pesquisa, identificados na condução de estudos envolvendo seres humanos e na fabricação de dados e falsificação de registros de pacientes, levando o Congresso Americano a realizar uma série de audiências públicas (JESUS, 2019). A autora sublinha a iniciativa como primeiro esforço para o exame sistemático sobre os tipos de comportamento desviante que afetam a integridade da pesquisa e o início da intervenção de atores externos à comunidade científica, como parlamentares e representantes da mídia.

A partir de uma série de mobilizações em torno do problema da má conduta na pesquisa, foi evidenciada uma “ruptura na concepção de que as normas sociais da ciência, implícitas e internalizadas pelos pesquisadores e cientistas, e a autorregulação da ciência seriam suficientes para garantir a integridade da pesquisa” (JESUS, 2019, p. 110), implicando em um grande movimento de conscientização entre as instituições sobre a necessidade de formalização de políticas, códigos de ética, códigos profissionais, entre outros mecanismos.

Russo (2014) afirma que a preocupação em torno da ética e a integridade na pesquisa começou a ser tornar assunto de interesse, em virtude dos alertas feitos pelas principais revistas científicas sobre casos de fraudes científicas. Segundo a autora, muitos casos foram descobertos e rapidamente publicizados, chamando a atenção não só das instituições, como também da sociedade em geral, que pleiteava por respostas das instituições de pesquisa envolvidas. Em decorrência, nos últimos anos formuladores de políticas e comunidade científica vêm pesquisando formas de rotular, estudar e responder a comportamentos de pesquisa que estão aquém da conduta responsável (RUSSO, 2014).

O termo “má conduta em pesquisa” é entendido como a “fabricação, falsificação ou plágio na proposição, realização ou revisão de pesquisas ou no relato de resultados de pesquisas” (OFFICE OF SCIENCE AND TECHNOLOGY POLICY, 2000, sem paginação, tradução nossa¹⁴). Segundo Steneck (2007), a fabricação está associada ao registro ou relato de dados ou resultados inventados; falsificação está na falta de precisão no registro da pesquisa devido à manipulação (alteração e omissão) de materiais, equipamentos ou processos de pesquisa; a apropriação de ideias, resultados, processos e resultados de pesquisa de autoria

¹⁴ No original: “*Research misconduct is defined as fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results*” (OFFICE OF SCIENCE AND TECHNOLOGY POLICY, 2000, sem paginação).

de outra pessoa sem dar crédito é considerada plágio. Assim, fabricação, falsificação e plágio compõem-se o entendimento que se tem hoje sobre fraude científica (RUSSO, 2014).

Entre os mais influentes trabalhos sobre mau comportamento do cientista na pesquisa, está o “*Scientists behaving badly*”, no qual Martinson, Anderson e De Vries (2005, p. 737) verificaram que entre 3.300 pesquisadores entrevistados 0,5% admitiram “falsificar” ou “cozinhar” dados da pesquisa, 1% usavam “ideias de outrem sem obter permissão ou dar o devido crédito”, 5,3% com falha “em apresentar dados que contradizem a própria pesquisa anterior” e 12,8% negligenciaram “o uso de dados defeituosos por outros ou interpretação questionável dos dados”. Para os autores, esses comportamentos apresentam maiores ameaças para ciência do que aquelas causadas por casos de “má conduta”, como fraude.

Segundo Kalleberg (2015), geralmente, casos de má conduta estão relacionadas com práticas questionáveis de pesquisa, no entanto essa é menos grave. Práticas questionáveis de pesquisa são “ações que violam os valores tradicionais da empresa de pesquisa e que podem ser prejudiciais ao processo de pesquisa” (NATIONAL ACADEMIES OF SCIENCE, ENGINEERING AND MEDICINE, 1992, p. 28). Nessa definição, prática questionável da pesquisa está separada de má conduta na pesquisa (fabricação, falsificação e plágio), pois “não danificam diretamente a integridade do processo de pesquisa” (NATIONAL ACADEMIES OF SCIENCE, ENGINEERING AND MEDICINE, 1992, p. 28).

Uma ampla gama de práticas está incluída nessa categoria, como a deturpação que ocorre quando o pesquisador não apresenta de maneira honesta suas contribuições ou deturpam a originalidade das publicações, publicando uma mesma informação mais de uma vez, produzindo o fenômeno de salamização¹⁵; imprecisão que decorre de inconsistências ou descuidos em informações contidas em notas de rodapé ou referências bibliográficas; e o viés evidente em situações em que o pesquisador toma decisões ou apresenta evidências que não decorreram de razões científicas ou acadêmicas (STENECK, 2006).

Segundo Steneck (2006, p. 54, tradução nossa¹⁶), a conduta responsável da pesquisa “representa o ideal que instituições e indivíduos se esforçam para encontrar”. Os termos “conduta responsável da pesquisa” e “integridade da pesquisa” são usados por instituições e formuladores de políticas para se referir a uma ampla gama de áreas da pesquisa, ética profissional e responsabilidade pessoal (CASSIMIRO, 2017, p. 37). O primeiro termo, normalmente usado pelo governo federal, refere-se à condução da pesquisa em “conformidade

¹⁵ O fenômeno da salamização ocorre quando os resultados de uma pesquisa são ‘fatiados’ em pedaços pequenos de modo a aumentar o número de publicações (KALLEBERG, 2015).

¹⁶ No original: “[...] *represents the ideal standard institutions and individuals endeavor to meet*” (STENECK, 2006, p. 54).

com as normas, códigos e diretrizes de sua profissão, [bem como] às diretrizes, políticas, regras e regulamentos de seus empregadores (universidades, instituições de pesquisa) e do governo (público)”, enquanto o segundo termo, usado com mais frequência por institutos de pesquisa para referir-se ao comportamento do pesquisador, é definido como “a qualidade de possuir e aderir firmemente a altos princípios morais e padrões profissionais” (STENECK, 2006, p. 55).

A concepção sobre integridade da pesquisa pode variar de maneira significativa, devido às experiências dos pesquisadores em pesquisa, origens nacionais, institucionais e seus níveis de hierarquia estrutural da ciência (ANDERSON, 2018). Nesse contexto, as conferências mundiais sobre integridade da pesquisa tornaram-se um importante mecanismo para alinhar dispositivos legais sobre má conduta na pesquisa e integridade da pesquisa.

A declaração elaborada como resultado da Terceira Conferência Mundial sobre Integridade da Pesquisa, realizada em Montreal, Canadá, em 2013, reconheceu que as colaborações apresentam desafios para a conduta responsável da pesquisa, uma vez que são empreendimentos que podem envolver diferenças substanciais nos sistemas regulatório e jurídico, estruturas organizacionais e de financiamento, culturas de pesquisa e abordagens ao treinamento. Dentre as responsabilidades assumidas pelos atores envolvidos em pesquisas colaborativas internacionais visando garantir da integridade da pesquisa, a declaração estabelece que “os parceiros colaboradores devem chegar a um acordo, no início e mais tarde, conforme necessário, sobre o uso, gestão, compartilhamento e propriedade de dados, propriedade intelectual e registros de pesquisa” (WORLD CONFERENCE ON RESEARCH INTEGRITY, 2013, sem paginação, tradução nossa¹⁷).

Nessa perspectiva, Jesus (2019, p. 145) considera a integridade da pesquisa como um termo guarda-chuva que engloba uma série questões inerentes à conduta ética na pesquisa como “pesquisa envolvendo seres humanos e animais, concepção do projeto de pesquisa, gestão de dados e de materiais de pesquisa, análise de dados, publicação, práticas de autoria, revisão por pares, mentoria, conflitos de interesse, entre muitos outros”. Dentre essas questões, a gestão de dados de pesquisa é de particular interesse para esta pesquisa.

Dados de pesquisa são considerados a base da inferência e da investigação científica, pois trazem subsídios que possibilitam a testagem de teorias e hipóteses; a identificação de novas áreas de investigação; entre outras possibilidades (STENECK, 2007). Sayão e Sales

¹⁷ No original: “*Collaborating partners should come to agreement, at the outset and later as needed, on the use, management, sharing and ownership of data, intellectual property, and research records*” (WORLD CONFERENCE ON RESEARCH INTEGRITY, 2013, sem paginação).

(2015) consideram a gestão desses dados uma das áreas centrais na conduta responsável da pesquisa, dessa forma, suas práticas devem ser acordadas antes da coleta dos dados, considerando questões referentes à propriedade, coleção, armazenamento e compartilhamento, como forma de garantir a integridade do processo da pesquisa (STENECK, 2007).

Essas questões vêm sendo reconhecidas por governos e órgãos de pesquisa, como o Reino Unido, que no âmbito do seu Escritório de Integridade da Pesquisa reconhece que as organizações devem oferecer suporte administrativo para auxiliar os pesquisadores na coleta e armazenamento de forma segura e acessível dos dados, que por sua vez, devem considerar como os dados serão coletados, analisados e gerenciados, e como e de que forma os dados relevantes serão disponibilizados para os outros (UK RESEARCH INTEGRITY OFFICE, 2009, p. 17, tradução nossa). A Royal Netherlands Academy of Arts and Sciences (2013), preocupada com a confiança pública na ciência, também advoga pela gestão de dados na prática científica, pois a mesma oferece suporte à disponibilização dos dados da pesquisa para o escrutínio e/ou verificação por outros pesquisadores.

Dessa forma, boas práticas de gestão de dados fornecem suporte para análise de dados posteriores; permite o reuso; possibilita a agregação de um conjunto de dados para análise; e subsidia a reprodução dos resultados de pesquisa (READ *et al.*, 2019). No entanto, Pimple (2002), analisando os elementos da conduta responsável da pesquisa, assinala que as questões relacionadas à gestão de dados são complexas e multivalentes, tornando-a uma prática negligenciada, ainda que seja essencial ao método científico. Sallans e Lake (2014) e Read *et al.* (2019) esboçaram algumas das principais razões pelas quais os pesquisadores não aderem as práticas de gestão de dados na pesquisa, tais como: falta de tempo devido às pressões de concessão e publicação; falta de recursos financeiros para realizar a atividade; a pouca ou nenhuma experiência nas atividades de gestão; as informações oferecidas em documentos normativos das agências de fomento sobre a modo como o pesquisador devem realizar as atividades relacionadas à gestão são vagas e limitadas.

Consequentemente, os serviços de dados são orientados, superficialmente, por requisitos oferecidos de forma limitada pelos financiadores, concentrando-se, principalmente, nos requisitos que subsidiam o compartilhamento e menos nos aspectos ideais das melhores práticas de gestão. Para Sallans e Lake (2014), os requisitos que focam no interesse dos financiadores para o compartilhamento de dados, levam a produtos insatisfatórios e, ao fim e ao cabo, não alteram realmente o comportamento e habilidades dos pesquisadores, tampouco melhoram a prática de gestão. Introduzida a problemática da gestão dos dados, na próxima seção analisou-se com mais detalhamento as questões em torno da temática.

4 GESTÃO DE DADOS DE PESQUISA: EMERGÊNCIA DE UMA AGENDA DE PESQUISA

Nos últimos anos, a gestão de dados de pesquisa foi introduzida na agenda de pesquisa de vários países, na qual atores do meio científico estão investigando as melhores formas para a construção de uma infraestrutura que viabilize a curadoria, preservação por longo prazo, compartilhamento, uso e reuso de dados. Os fatores que influenciaram o surgimento das demandas pela gestão de dados por parte dos governos, agências de fomento, formuladores de políticas, instituições de pesquisa e periódicos são muitos e por vezes seguem direções ligeiramente diferentes. Reconhecê-los torna-se essencial para posicionar o fazer científico do pesquisador em um contexto inevitavelmente politizado (COX; VERBAAN, 2018).

A gestão de dados não é uma nova prática que emergiu nos últimos anos, sendo observada a partir da década de 1960 em bibliotecas universitárias do Reino Unido, como afirma Lewis (2010). Todavia, a prevalência de dados em pesquisas digitais foi a responsável por abrir um novo conjunto de possibilidades para a ciência do século XXI e despertar o interesse nas atividades de curadoria e preservação de dados (LEWIS, 2010; BRINEY, 2015).

Parte expressiva da literatura sobre a emergência de uma agenda de pesquisa em torno da gestão de dados localiza como principal fator a mudança no *modus operandi* da ciência intimamente ligada às tecnologias digitais que culminou no surgimento de um novo modelo de ciência mais colaborativo, mais computacional e mais intensivo em dados (TENOPIR *et al.*, 2014; COX; VERBAAN, 2016; COX *et al.*, 2017).

Na década de 1990, surge uma nova onda de *big science* amparada por tecnologias digitais, facilitando a exploração e produção de dados sem precedências na história da ciência. O novo cenário foi caracterizado pelo aumento da informatização do processo de pesquisa, por meio de estruturas organizacionais virtuais baseadas em rede, novas ferramentas, métodos e instrumentação para coleta e análise de dados em grande escala, novos padrões de colaboração e comunicação dos resultados de pesquisa (JANKOWSKI, 2009).

Albagli, Appel e Maciel (2014) explicam que no final do século XX a incorporação de uma vertente computacional na ciência desencadeou uma produção exponencial de pesquisas baseadas no uso intensivo de dados (*data-intensive*), emergindo o conceito *e-Science*. Segundo Borgman (2012), o crescimento de dados nas disciplinas da *big science*, como genômica, astronomia e física, além de proporcionar o surgimento de novos campos de estudo, como biologia computacional, astroinformática, e humanidades digitais, desencadeou o novo modelo de ciência. A *e-Science* foi principalmente institucionalizada em iniciativas

emergentes na Europa, particularmente do Reino Unido, quando John Taylor em 1999 cunhou o termo para definir uma “ciência cada vez mais desenvolvida por meio de colaborações globalmente permitidas pela Internet, usando grandes coleções de dados, recursos computacionais em escala *tera* e visualização de alto desempenho” (TAYLOR, 2002, *apud* FRY; SCHROEDER; DEN BESTEN, 2009, p. 4, tradução nossa¹⁸).

Antecipadamente previsto por Hey e Trefethen (2003), o chamado “dilúvio de dados” gerado pela *e-Science* desencadeou novas preocupações sobre como os dados de pesquisa deveriam ser capturados, selecionados e mantidos para uso futuro. Esse dilúvio instaurou nas instituições de pesquisa a necessidade de desenvolver de políticas, infraestruturas e serviços para a gestão dos dados, com o objetivo de auxiliar os pesquisadores desde a criação até a preservação dos dados (PINFIELD; COX; SMITH, 2014).

A *e-Science*, reconhecida por Gray (2007) como o quarto paradigma científico que representa, antes, um *continuum* dos três paradigmas anteriores (ciência experimental, ciência teórica, simulação por computadores, respectivamente) trouxe para dentro da comunidade científica novas preocupações para o desenvolvimento de técnicas e tecnologias necessárias para executá-la. O novo modelo de ciência exige do cenário da pesquisa uma variedade de recursos tecnológicos para a gestão dos dados e serviços ao pesquisador visando à garantia da estabilidade, da integridade científica e da colaboração científica (SCHMILLEN, 2015).

Contudo, Borgman (2012) afirma que a alegação de um quarto paradigma parece atraente, mas tende a exceder. Wilbanks (2009, p. 210) ressalta que a *e-Science* não é uma mudança de paradigma no sentido kuhniano, visto que “os dados não estão varrendo a velha realidade”, na verdade, imputam às metodologias e práticas científicas um conjunto de novos desafios para exportar, transmitir e integrar o conhecimento. O autor explica que os dados não são considerados como um método em si, mas um recurso valioso para qualquer paradigma.

Avançando na discussão, Cox e Verbaan (2018) e Corti *et al.* (2019) ressaltam que a introdução das tecnologias digitais e da Internet, além de implicar mudanças no processo de pesquisa, também tornou viável a ideia de dados abertos ao facilitar o compartilhamento de dados. O desfecho bem-sucedido do Projeto Genoma Humano (1990–2003), a partir de esforço internacional para identificar e mapear o genoma humano mostrou como a abertura e o intercâmbio de informações – incluindo dados de pesquisa – poderia aumentar rapidamente a velocidade das descobertas científicas (RAY, 2014; CORTI *et al.*, 2019).

¹⁸ No texto original: “science increasingly carried out through distributed global collaborations enabled by the Internet, using very large data collections, terascale computing resources and high performance visualization” (TAYLOR, 2002 *apud* FRY; SCHROEDER; DEN BESTEN, 2009).

Nos primeiros dias do sequenciamento, as tentativas de patentear¹⁹ e transformar em *commodity* o genoma humano provocaram grande tensão sobre o controle da informação científica (BORGMAN, 2015). Ainda que a partir da década de 1980 tenham surgido relatórios ponderando as razões para aprimorar o compartilhamento e curadoria de dados, foi o estabelecimento de políticas para o acesso aberto aos dados do genoma humano um ponto de virada nas práticas científicas de compartilhamento de dados (BORGMAN, 2015).

Nessa perspectiva, a literatura revela que o crescente interesse na gestão de dados está fortemente imbricado ao movimento de acesso aberto e se encontra em um espectro mais amplo de discussões sobre o futuro da ciência que reverberam uma reforma da prática de pesquisa orientada pelos pressupostos da “ciência aberta” (COX; VERBAAN, 2018).

Entre as décadas de 1980 e 1990 do século passado, o sistema de comunicação científica foi abalado pela chamada crise dos periódicos impulsionada pela impossibilidade das bibliotecas universitárias (sobretudo, norte-americanas) manterem suas coleções de periódicos e atenderem à demanda crescente de seus usuários em virtude do aumento no custo das assinaturas dos periódicos (MUELLER, 2006).

Naquele contexto, a insatisfação da comunidade acadêmica frente à lucratividade das empresas editoriais na publicação dos periódicos científicos aliada ao advento da internet e às alternativas oferecidas pelas tecnologias de informação e comunicação (TIC), que quase suprimiam as editoras do fluxo de comunicação, impulsionou o movimento pelo acesso aberto ao conhecimento científico (MUELLER, 2006; CARDOSO; JACOBETTY; DUARTE, 2012).

Segundo Veiga (2017), o movimento pelo acesso aberto foi formalmente iniciado em 2002 com a “*Budapest Open Access Initiative (BOAI)*”, que o definiu como a disponibilidade livre e irrestrita à produção científica (BUDAPEST, 2002). A declaração propôs duas estratégias para viabilizar o acesso à produção científica: a via verde (*green road*) que se refere ao autoarquivamento de artigos publicados em periódicos em um repositório de acesso aberto pelo autor; e a via dourada (*golden road*), que consiste na publicação da produção científica em periódicos científicos de acesso aberto (VEIGA, 2017).

No ano seguinte, surgem duas outras declarações: a Declaração Bethesda e a Declaração de Berlim. Pampel e Dallmeier-Tiessen (2014) afirmam que a Declaração de Berlim é considerada uma referência central para as demandas de acesso e compartilhamento dos resultados de pesquisa, visto que ampliou a noção de acesso aberto para além do acesso

¹⁹ De acordo com Borgman (2015), uma das motivações para a discussão foi o aumento da mercantilização das informações associadas às mudanças nos regimes de propriedade intelectual nas décadas de 1970 e 1980.

livre à literatura científica revisada por pares, incluindo os “[...] resultados de pesquisas científicas originais, dados brutos e metadados, materiais de origem, representações digitais de materiais pictóricos e gráficos e materiais acadêmicos multimídia” (BERLIN, 2003, sem paginação, tradução nossa²⁰). Segundo Cavalcanti e Sales (2017), o objetivo do alargamento da noção de acesso aberto aos dados de pesquisa²¹ é colaborar na dinâmica da pesquisa em escala global, eliminando etapas já realizadas por outros pesquisadores, além de impactar na economia de recursos ao evitar a duplicação de pesquisas com propósitos similares.

Cabe ressaltar que uma das primeiras declarações²² em prol do acesso aberto aos dados de pesquisa foi no campo da saúde, com a publicação da Declaração de Helsinque, em 1964, que expressou princípios éticos para pesquisas médicas envolvendo sujeitos humanos, incluindo pesquisas sobre material e dados humanos identificáveis. Destacam-se também os Princípios de Bermudas (*Bermuda Principles*), firmados em 1996, que representaram um acordo entre os pesquisadores e suas agências financiadoras sobre a importância de se compartilhar de forma rápida, gratuita e aberta os dados de sequências genéticas de genes antes da publicação, conforme explicam Fitzgerald, Fitzgerald e Pappalardo (2011). Segundo Veiga (2017, p. 57), os princípios representam “uma conquista significativa da transferência do que se pensava ser privado para o bem público”.

Diante da expansão do conceito de acesso aberto à publicação e aos resultados de pesquisa que atualmente incorporam coleções de dados de pesquisa, importa assinalar que o regime de abertura da ciência não se restringe exclusivamente aos resultados de pesquisa subjacentes aos artigos científicos, antes, observa-se um alargamento da noção de abertura para todo o processo científico sob o auspício do movimento da ciência aberta (CARDOSO; JACOBETTY; DUARTE, 2015).

A ciência aberta advoga uma maior visibilidade e transparência para o processo da pesquisa visando melhorar a eficiência da ciência, defendendo o acesso aos cadernos de laboratório, códigos de *software*, *workflow*, metodologias, protocolos, estratégias, instrumentos e ferramentas de observação e coleta de dados (ALBAGLI, 2015; SAYÃO; SALES, 2018). Nesse sentido, Guimarães (2014, p. 142) salienta que “o comprometimento

²⁰ No original: “[...] *original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material*” (BERLIN, 2003, sem paginação, *on-line*).

²¹ As iniciativas para o acesso aberto aos dados têm raízes ainda mais antigas, datadas desde a década de 1950, sobretudo após as contribuições advindas da criação do World Data Center (WDC), pelo International Council of Science (ICSU), em 1957 (AVENTURIER; ALENCAR, 2016; COSTA; LEITE, 2017).

²² Disponível em: http://oad.simmons.edu/oadwiki/Declarations_in_support_of_OA. Acesso em: 12 jan. 2021.

com a abertura não é, necessariamente, um preceito ético, antes, faz parte do sistema de recompensa da ciência”.

Albagli (2015, p. 14) observa que ciência aberta “não expressa simplesmente um novo ciclo de revitalização do *ethos* mertoniano de uma ciência desinteressada, em contraposição ao endurecimento dos regimes de propriedade intelectual a partir da década de 1980”, na verdade, reflete novos modos de pensar e de exercer a cientificidade. A ciência aberta traz para o empreendimento científico uma maior preocupação em promover a transparência, produtividade e reprodutibilidade da pesquisa, de forma colaborativa e apoiada em recursos tecnológicos que permitam a colaboração *online*, simultânea, à distância e de modo visível e aberto à ampla contribuição (ALBAGLI; CLINIO; RAYCHTOCK, 2014). Sendo assim, Vicente-Saez; Martinez-Fuentes (2018, p. 7, tradução nossa²³) a conceituam como “o conhecimento transparente e acessível que é compartilhado e desenvolvido através de redes colaborativas”.

Fecher e Friesike (2013) defendem que a ciência aberta envolve uma infinidade de dinâmicas dentro do campo científico que parte desde o direito democrático de acesso ao conhecimento financiado por recursos públicos (acesso aberto às publicações e aos dados de pesquisa), passando por uma demanda por uma melhor participação da sociedade na ciência (ciência cidadã), até o desenvolvimento de ferramentas colaborativas (plataformas de mídia social para cientistas). Para os autores, os dados de pesquisa abertos possibilitam a reutilização e validação da ciência, enquanto o acesso aberto à publicação científica contrapõe-se às “políticas editoriais que buscam cercear o acesso à literatura científica pela cobrança de assinaturas com preços elevados” (ALBAGLI; CLINIO; RAYCHTOCK, 2014, p. 437).

Embora as tecnologias digitais aliadas aos pressupostos da ciência aberta tenham beneficiado a comunicação da ciência, uma parte relevante do trabalho científico realizado na chamada “cauda longa” (*long tail*) da ciência não está visível e acessível para sociedade e a comunidade científica (SAYÃO; SALES, 2018). Na cauda longa, o trabalho científico é conduzido por indivíduos ou pequenas equipes de pesquisadores que produzem grande quantidade de pequenas coleções de dados, mas com características heterogêneas devido às suas variações de conteúdo, estrutura e representações (BORGMAN, 2015), tornando a gestão desses dados desafiadora. Para Sales e Sayão (2019a), o foco de atenção de financiadores e formuladores de políticas, instituições de pesquisa e da própria sociedade direciona-se a

²³ No original: “*Open Science is transparent and accessible knowledge that is shared and developed through collaborative networks*” (VICENTE-SAEZ; MARTINEZ-FUENTES, 2018, p. 7).

segmentos da *big science*, provocando certa invisibilidade e opacidade de dados e outros produtos gerados na pequena ciência (*small science*).

Na *big science* são coletadas centenas de *gigabytes* de dados que tendem a ser homogêneos devido à coleta ser automatizada, à adesão aos procedimentos uniformes e padrões de estruturação para formato dos dados e à disponibilização em repositórios de dados específicos, como na área da genômica que dispõe do *GenBank*, facilitando a recuperação por outros cientistas (HEIDORN, 2008; BORGMAN, 2015). No entanto, na cauda longa há uma falta de padrões nos quais basear a infraestrutura compartilhada e de recursos para desenvolver e sustentar os dados compartilhados (BORGMAN, 2015). A carência de políticas institucionais e infraestruturas tecnológicas e gerenciais, a falta de interesse, visto que a organização de dados não é objeto de prestígio acadêmico e a falta de reconhecimento das atividades pelos sistemas de recompensa são alguns dos obstáculos que impedem as práticas de gestão e compartilhamento de dados nesse segmento, tornando-os subutilizados e eventualmente perdidos (SAYÃO; SALES, 2018).

Contudo, Heidorn (2008, p. 282, tradução nossa²⁴) considera o acesso a esses dados um fator crucial para o progresso da ciência, pois “parece mais provável que a ciência transformadora venha mais provavelmente da cauda do que da cabeça”. Sales e Sayão (2019a) ainda defendem que “a disponibilidade e o acesso aos dados da cauda longa é um elo essencial para o princípio da autocorreção da ciência”, pois permite a verificação, revisão e validação das pesquisas e reconhecimento dos vieses das publicações acadêmicas.

Compartilham de tal percepção Williams, Bagwell e Zozus (2017) ao afirmarem que historicamente, muitas das práticas de gestão de dados decorreram da necessidade em apoiar a reprodutibilidade da pesquisa, entendida como a “capacidade de um pesquisador duplicar os resultados de um estudo anterior usando os mesmos materiais usados pelo pesquisador original” (NATIONAL SCIENCE FOUNDATION, 2015, p. 3, tradução nossa²⁵). Cox e Verbaan (2018) constatam que uma crise na reprodutibilidade da pesquisa, refletindo na perda de confiança na integridade da pesquisa e na necessidade de uma maior transparência no processo de investigação, intensificou as motivações para a gestão de dados em disciplinas como a biomedicina e psicologia, que apresentavam casos de falsificação e fabricação de

²⁴ No original: “*It seems likely that transformative science is more likely to come from the tail than the head*” (HEIDORN, 2008, p. 282).

²⁵ No original: “*refers to the ability of a researcher to duplicate the results of a prior study using the same materials and procedures as were used by the original investigator*” (NATIONAL SCIENCE FOUNDATION, 2015, p. 3).

dados que inviabilizaram a reprodução dos estudos. Isso levou alguns periódicos a exigirem a disponibilização de dados como forma de avaliar os resultados de pesquisa (COX, 2015).

Borgman (2012) ainda afirma que a reprodução de um estudo confirma que o dinheiro público foi gasto devidamente. Segundo Costa (2017b), o papel do Estado no financiamento do setor da C&T mostrou que pesquisas desenvolvidas com recursos públicos são bens que pertencem a todos, por isso precisam ser amplamente acessíveis. Bhardwaj e Banks (2019) explicam que muitas organizações de fomento à pesquisa são federais ou possuem sua missão voltada para o público, assim, para que a organização e o público obtenham o retorno sobre o investimento, os dados devem ser adequadamente gerenciados e disponibilizados. Desse modo, Tenopir *et al.* (2011) consideram que o financiamento público é a principal motivação para a preservação dos dados, visando maximizar os recursos empreendidos para a coleta.

Nessa perspectiva, Latham (2017) salienta que uma das principais razões para a gestão de dados está relacionada ao recurso financeiro destinado às pesquisas. A pressão exercida por financiadores de pesquisa para a gestão de dados coincidiu, aproximadamente, com a recessão econômica global dos anos 2000, quando muitas instituições de pesquisa tiveram orçamentos reduzidos (BRINEY, 2015). Nesse período, em virtude da necessidade de justificar o dinheiro público investido nas pesquisas, visto que o mesmo deve ser utilizado para servir ao bem público (BORGMAN, 2012), muitas agências internacionais de fomento começaram a declarar a questão como prioridade. Briney (2015) comenta que ao exigir a gestão dos dados, os financiadores podem conter a perda de dados importantes, além de prestar contas àqueles que pagam pela pesquisa. Como benefício adicional, o autor acrescenta que a reutilização de dados significa que a mesma quantia em dinheiro resultará em mais pesquisas, uma vez que os dados custam mais para coletar do que para reutilizar.

O argumento do bem público é explícito nos princípios da Organisation for Economic Co-operation and Development (OECD)²⁶. No nível político, um fator chave para as mudanças atuais nas políticas pode ser localizado na virada do século, quando a OCDE começou a explorar os dados de pesquisa. Após iniciativas formais pelo acesso aberto com as declarações BBB (Budapeste, Bethesda e Berlim), em janeiro de 2004, países signatários da OCDE assinaram a “*Declaration on access to research data from public funding*” no qual reconheceram a importância do acesso aberto e uso irrestrito de dados para promover o progresso e contribuir para novas descobertas e inovações (SHEARER, 2015).

²⁶ Organização para a Cooperação e Desenvolvimento Econômico (OCDE).

Em 2007, com vista a reforçar o acordo de 2004, a OECD publicou um conjunto de princípios e diretrizes em “*OECD Principles and Guidelines for access to research data from public funding*” no qual declara que os dados de pesquisa financiada por recursos públicos são um bem público e devem ser disponibilizados abertamente com o mínimo de restrições possíveis, de maneira oportuna e responsável, sem prejudicar a propriedade intelectual (ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2007).

Desde a declaração da OECD, Jones (2012) e Shearer (2015) observam que o ímpeto para a gestão de dados como forma de maximizar o investimento público é aparentemente retórico na maioria das políticas de dados adotadas pelos países signatários da declaração. No cenário em que emergem essas políticas observa-se, por um lume altruísta, que a motivação dos governos e agências de fomento repousa na ampliação do acesso aos dados de pesquisa para aumentar a velocidade e reprodutibilidade da ciência. Outra motivação não muito explícita nas políticas é a de que os financiadores passam a reconhecer que dados bem gerenciados e preservados podem expandir as possibilidades de comercialização de suas pesquisas financiadas. A seguir, apresenta-se um panorama do surgimento dessas políticas.

4.1 POLÍTICAS DE FINANCIADORES INTERNACIONAIS E NACIONAIS

Nas primeiras duas décadas do século XXI, em reconhecimento à necessidade de tornar acessíveis os dados produzidos com recursos públicos, observa-se uma tendência de governos e agências de financiamento nacionais e internacionais em desenvolver arcabouços normativos para promover o compartilhamento de dados, visando obter maiores benefícios das pesquisas que financiam. As chamadas políticas de dados, também conhecidas na literatura internacional como “*Research Data Management and Sharing Policies*” (RDM) começaram a ser adotadas por governos, financiadores de pesquisa e instituições acadêmicas.

No entendimento de Pryor (2014), uma política de RDM é uma declaração pública de comprometimento dos principais *stakeholders* para a implantação de medidas para fornecer os recursos e infraestrutura necessária para habilitar a política. É um documento que reflete um quadro normativo mais amplo em âmbito governamental ou internacional, referindo-se, na maioria das vezes, a declarações públicas de governos que abordam agendas mais amplas de dados abertos (NEYLON, 2017). Segundo Shearer (2015), essas políticas têm os objetivos de melhorar a eficiência da pesquisa, apoiar o reuso de dados para novos *insights* e descobertas, fomentar a colaboração e facilitar uma maior transparência do processo de investigação.

Dependendo do contexto, elas podem ter caráter recomendatório ou obrigatório, podem sofrer variações em termos de motivações, cobertura, responsabilidades e requisitos bem como podem oferecer disposições específicas de acordo com a região, agência ou domínio no qual são emitidas (SHEARER, 2015). Portanto, Shearer (2015) sustenta que mesmo que haja requisitos comuns entre as políticas, ocasionalmente pode haver maior ênfase em alguns aspectos em detrimento de outros. Assim uma política predominantemente voltada para o compartilhamento basear-se-á em práticas para oferecer suporte ao acesso e reuso dos dados, enquanto uma política baseada na administração de dados provavelmente atear-se-á em funções e responsabilidades envolvidas na gestão dos dados (SHEARER, 2015).

Ainda para Shearer (2015), os ambientes políticos como maior maturidade de registros de políticas e outros instrumentos normativos sobre gestão e compartilhamento de dados de pesquisa são o Reino Unido e os EUA, embora a Austrália, a União Europeia (*European Union* – EU), Canadá e Alemanha tenham avançado na discussão. Recentemente, iniciativas brasileiras também começaram a ser exploradas na literatura nacional.

No Reino Unido, ainda na década de 1990, agências de fomento à pesquisa como Wellcome Trust e Economic and Social Research Council (ESRC) emitiram políticas de dados que, além de exigirem métodos adequados para a gestão de dados, esperavam que os institutos de pesquisa estabelecessem provisões para o financiamento público recebido nas pesquisas (BORGMAN, 2012; LATHAM, 2017). No entanto, a questão avançou a partir de novembro de 2000, com a implementação do principal Programa de *e-Science* (*e-Science Core Program*) destinado a estabelecer um conjunto de ferramentas – incluindo infraestrutura, *middleware* e documentação – para facilitar os projetos de *e-Science* (LEWIS, 2010).

Em 2003, em um relatório encomendado pelo Joint Information Systems Committee (JISC), ficou evidente a necessidade do desenvolvimento de uma estratégia nacional para gestão e infraestrutura dados científicos e da criação um centro de curadoria digital (LYON, 2007). Em resposta, o JISC e o Engineering and Physical Sciences Research Council (EPSRC) financiaram a criação do Digital Curation Centre (DCC), em 2004, um centro proativo na gestão e curadoria de dados de pesquisa. Os modelos para planos de gestão de dados do DCC são amplamente usados como base para muitos dos documentos de planejamento desenvolvidos em instituições de pesquisa internacionais (BORGMAN, 2012).

Jones (2012) averiguou dois principais picos no desenvolvimento de políticas de dados de financiadores do Reino Unido: o primeiro movimento em 2007, quando o UK Medical Research Council (UK MRC), Biological and Biomedical Sciences Research Council (BBSRC) e o Wellcome Trust introduziram novas políticas de dados com referências aos

princípios e diretrizes divulgados pela OECD; o segundo, em 2011, após a pesquisa financiada pelo Research Council UK (RCUK) intitulada “*Common Principles on Data Policy*”, que ratificou a importância das instituições de pesquisa desenvolverem políticas e planos de dados, além de outras medidas para preservar dados de valor reconhecido e criar metadados para facilitar sua recuperação (PRYOR, 2014). Jones (2012) também destacou a declaração de propósito conjunta sobre o compartilhamento de dados de pesquisa para melhorar a saúde pública²⁷, assinada em 2011 por grandes financiadores da pesquisa em saúde, como um instrumento fortalecedor dos requisitos de gestão e compartilhamento de dados no contexto britânico.

Atualmente, a política de dados de pesquisa do Reino Unido é estruturada pelo *Concordat on Open Research Data*, desenvolvido por um grupo de múltiplos *stakeholders* do Reino Unido. O *Concordat* estabelece dez princípios que cobrem questões como a gestão e curadoria de dados; questões legais e éticas e as restrições no acesso aos dados; habilidades e treinamentos; custos e financiamento; e a necessidade de revisão regular das políticas e outros documentos normativos (CORTI *et al.*, 2019). Para os autores, a ênfase dada a esses princípios é ligeiramente diferente nas políticas de dados dos financiadores do Reino Unido.

No contexto norte-americano, podem-se rastrear declarações públicas para a gestão de dados de pesquisa desde a década de 1990, quando o Office of Science and Technology Policy (OSTP) da Casa Branca publicou em 1991 declarações de política sobre gestão de dados para pesquisa sobre mudança global²⁸, visando facilitar o acesso amplo e aberto a dados de qualidade para pesquisas de mudanças globais.

Já no início dos anos 2000, o Congresso Americano começou a indagar as agências federais de fomento à pesquisa a respeito do alto custo gasto para a coleta de dados em diferentes projetos que, aparentemente, eram conectados ou apoiavam um ao outro (MULLINS, 2014). Nesse contexto, as principais agências de financiamento como o National Institutes of Health (NIH) e a National Science Foundation (NSF) responderam às questões com ações variadas.

O NIH foi uma das primeiras agências a emitir uma política sobre o compartilhamento de dados de pesquisa, em 2003, adicionando um requisito de Plano de Compartilhamento de Dados (PCD) para doações acima de \$500.000,00 para projetos de pesquisas submetidos aos institutos (BORGMAN, 2012). Já a NSF realizou uma série de audiências e designou uma

²⁷ Disponível em: <https://wellcome.org/what-we-do/our-work/sharing-research-data-improve-public-health-full-joint-statement-funders-health>. Acesso em: 15 jun. 2020.

²⁸ Disponível em: <http://www.gcrio.org/USGCRP/DataPolicy.html>. Acesso em: 12 jan. 2021.

equipe a comando de Daniel Atkins, então diretor da Escola de Informação da University of Michigan, para analisar quais eram os desafios na gestão de dados, especialmente na mineração e como o desperdício de dados poderiam ser evitado (MULLINS, 2014).

O seminal Relatório *Atkins*, publicado em 2003 reuniu, pela primeira vez, os desafios para a *cyberinfrastructure*: identificação, descrição, localização, preservação e compartilhamento de grandes quantidades de dados. Para atender a demanda de evitar que o dinheiro público fosse desperdiçado em duplicação de esforços para novas coletas de dados, a NSF, junto a agências federais e à Library of Congress (LC) alocaram recursos para a criação de bibliotecas digitais voltadas aos serviços de dados (MULLINS, 2014). Nessa perspectiva, Gobin e Griffin (2019) observaram que a partir da primeira década do século XXI, os bibliotecários acadêmicos foram motivados a desenvolver serviços e pesquisas sobre gestão de dados de pesquisa a fim de atender às demandas de pesquisadores envolvidos em projetos financiados por instituições de pesquisa e agências de fomento.

Pryor (2014, p. 13, tradução nossa²⁹) ainda afirmou que “o movimento de mudança nos EUA ocorreu menos como consequência de mandatos ou ordens do que de iniciativas surgidas na própria comunidade de suporte à pesquisa”. O autor explica que desde o surgimento da pesquisa intensiva em dados, bibliotecários começaram a reestruturar infraestruturas técnicas e humanas, além de assumirem lideranças em programas multidisciplinares como o Data Conservancy e DataNet Observation Network for Earth (DataONE), lançados no âmbito do programa Sustainable Digital Data Preservation and Access Network Partners (DataNet) da NSF.

Em maio de 2010, a NSF divulgou sua Política de Compartilhamento de Dados, exigindo que as propostas enviadas a partir de 18 de janeiro de 2011 incluíssem um documento suplementar com no máximo duas páginas rotuladas como “Plano de Gestão de Dados” com informações referentes aos tipos de dados coletados e disposições para o arquivamento, compartilhamento e acesso dos dados (MULLINS, 2014). Segundo Borgman (2012), o requisito NSF para o envio do PGD foi um ponto de inflexão, pelo menos nos EUA, ao tornar o documento como parte do processo de revisão por pares.

Não obstante as políticas do NIH (2003) e da NSF (2010) que se aplicavam a um grupo considerável de pesquisadores, suas diretrizes não eram sistêmicas em todo o sistema de financiamento à pesquisa dos EUA (BRINEY; GOBIN; ZILINSKI, 2017). Um marco regulatório para as políticas de dados dos financiadores estadunidenses foi o memorando

²⁹ No original: “[...] *the movement for change in the USA has come about less as a consequence of mandates or dictates than from initiatives arising within the research support community itself*” (PRYOR, 2014, p. 13).

“*Expanding Public Access to the Results of Federally Funded Research*”, divulgado pelo OSTP, em 2013.

Por meio do memorando, o OSTP determinou que as agências federais com mais de U\$ 100 milhões em despesas de pesquisa e desenvolvimento (P&D) garantissem o acesso público aos dados e publicações produzidas com financiamento do governo federal bem como exigia que seus pesquisadores elaborassem um PGD para pesquisas financiadas, com previsões para a preservação e o acesso a dados científicos (HOLDREN, 2013). Desde então, algumas as agências sujeitas à diretiva do Governo promulgaram requisitos em resposta ao memorando do OSTP, enquanto outras esboçaram planos preliminares como o desenvolvido pelo NIH, com vista ao estabelecimento de políticas e abordagens para melhorar o compartilhamento e gestão de dados de pesquisa (NATIONAL INSTITUTES OF HEALTH, 2015b).

A governo australiano também se moveu com relativo dinamismo para fornecer serviços e infraestrutura nacional para facilitar a coleta, geração, manipulação, curadoria, acesso e disseminação de dados de pesquisa, com a criação do Australian National Data Service (ANDS)³⁰, em 2007 (TRELOAR; CHOUDHURY; MICHENER, 2012). Segundo Krahe *et al.* (2020), apesar da falta de políticas adotadas pelos principais financiadores de pesquisa federais, um importante instrumento normativo para o estabelecimento de boas práticas de gestão de dados é o “*Code for the Responsible Conduct of Research*”, desenvolvido pelo Australia’s National Health and Medical Research Council (NHMRC), Australian Research Council (ARC) e Universities Australia, em 2007. Além de enfatizar boas práticas de gestão de dados de pesquisa e materiais primários, o código ressalta a importância de as instituições adotarem políticas institucionais que tratem da propriedade, armazenamento, retenção e acesso apropriado pela comunidade científica (CORTI *et al.*, 2019). Observa-se que diferentemente de outros países que geralmente vinculam a gestão de dados à abertura e ao compartilhamento de dados, a Austrália adota uma abordagem diferenciada, ao abordar o tratamento dos dados no âmbito das boas práticas científicas para conduta responsável na pesquisa.

No contexto europeu, os setes primeiros programas de financiamento à pesquisa e inovação da EU, careciam de uma política para os dados de pesquisa. Segundo Whyte (2012), a mudança começou em outubro 2010, quando Neelie Kroes, à época vice-presidente da European Commission (EC), anunciou o lançamento do relatório “*Riding the Wave: how*

³⁰ Treloar, Choudhury e Michener (2012) discutem o surgimento, princípios, objetivos do ANDS, um dos investimentos mais significativos da Austrália em infraestrutura de dados de pesquisa.

Europe can gain from the rising tide of scientific data” que abordava o cenário e desafios associados aos dados científicos. O relatório expressou a importância da gestão de dados nos programas de financiamento e recomendou que agências da EU solicitassem um PGD para todos os projetos financiados futuramente, a fim de obter uma visão estratégica do valor dos dados produzidos (SPICHTINGER; SIREN, 2018).

A partir de 2011, a EC realizou uma série de audiências a fim de compreender como poderia contribuir para circulação de conhecimento e acesso e preservação de informações científicas na era digital (EUROPEAN COMMISSION, 2012b). No ano seguinte, a EC publicou recomendações sobre o acesso e a preservação da informação científica, incentivando todos os Estados-membros ao desenvolvimento de capacidades para o acesso aberto aos resultados da investigação financiados com fundos públicos (EUROPEAN COMMISSION, 2012b). Após outras consultas, a Comissão implantou no oitavo Programa-Quadro, conhecido como Horizon 2020 (H2020), um projeto piloto de dados de pesquisa abertos, o *Open Research Data (ORD) Pilot*, com a expectativa que os dados fossem depositados em um repositório apropriado para acesso e reutilização dos dados gerados no âmbito do programa.

No Canadá, a gestão de dados se tornou uma prioridade para agências governamentais e instituições de pesquisa nos últimos anos. No âmbito das iniciativas financiadas pelo governo, em 2015, o Canadian Institutes of Health Research (CIHR), o Natural Sciences and Engineering Research Council of Canada (NSERC) e o Social Sciences and Humanities Research Council (SSHRC), coletivamente conhecidos como “Três-Agências”, apresentaram um esboço de declaração de princípios de gestão de dados com as expectativas gerais das agências relativas ao tratamento de dados de pesquisa e às responsabilidades dos pesquisadores, instituições e financiadores em atender a esses princípios. Atualmente, as três agências estão revisando uma nova política de gestão de dados de pesquisa alinhada com o estado de prontidão da comunidade de pesquisa canadense (CANADA, 2019, *on-line*). Além disso, a Canadian Association of Research Libraries (CARL), no âmbito da rede *Portage*³¹ lançada em 2015, tem adotado uma abordagem coordenada entre partes interessadas para a construção de uma cultura nacional de gestão de dados de pesquisa no país.

Na Alemanha, a premência em se abordar o tratamento destinado os dados de pesquisa financiada por recursos públicos, inicialmente, materializou-se na sétima recomendação “*Safeguarding and Storing of Primary Data*” enunciada no memorando “*Proposals for*

³¹ Fonte: <https://portagenetwork.ca/>.

Safeguarding Good Scientific Practice”, publicado pela German Research Foundation (*Deutsche Forschungsgemeinschaft* [DFG]) em 1998, recentemente substituído pelo código de conduta “*Guidelines for Safeguarding Good Research Practice*” (DEUTSCHE FORSCHUNGSGEMEINSCHAFT, 2019).

Em 2009, a DFG emitiu recomendações para o armazenamento seguro e disponibilidade de dados, solicitando aos pesquisadores que descrevessem como os dados seriam disponibilizados para futuras reutilizações (PRYOR, 2012). Em 2010, como parte de sua iniciativa intitulada “*Digital Information*”, a Alliance of German Science Organisations (2010), que reúne os principais institutos de pesquisa e financiadores alemães, anunciou um conjunto de princípios para o tratamento de dados de pesquisa referentes à segurança, acessibilidade, diferenças entre disciplinas científicas, reconhecimento científico, ensino e qualificação, uso de padrões e desenvolvimento de infraestrutura.

Especialmente no que diz respeito ao desenvolvimento de uma infraestrutura nacional, é oportuno citar a atual iniciativa alemã intitulada “*National Research Data Infrastructure*”, fruto de um acordo firmado entre o governo federal e os estados federados em novembro de 2018, que visa desenvolver um sistema nacional para a gestão de dados de pesquisa, incluindo a padronização dos dados de acordo com os princípios FAIR (ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT, 2020).

No Brasil, observa-se um alargamento da discussão nas instituições de pesquisa a partir do reconhecimento da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) de que a gestão de dados configura-se uma prática indispensável e preponderante na investigação científica (HENNING, 2019). Em 2017, a agência tornou o PGD um documento obrigatório para determinadas modalidades e chamadas de projetos de pesquisa submetidas à instituição (FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO, 2017).

No cenário brasileiro, ainda destaca-se a rede de apoio à gestão e abertura de dados de pesquisa em saúde, fruto de uma iniciativa internacional maior chamada GO FAIR que “busca o desenvolvimento de um ambiente global compartilhado voltado para a pesquisa e inovação baseadas em dados” (VEIGA; QUEIROZ, 2019, sem paginação). Conforme explicam Veiga e Queiroz (2019, sem paginação), a Rede de Implementação GO FAIR Brasil Saúde é “uma rede temática responsável pela elaboração de estratégias de implementação dos princípios FAIR no campo da saúde”, coordenada pelo ICICT da Fiocruz.

Diante do exposto, observa-se que a gestão de dados é uma questão de recorrente preocupação de governos, organizações financiadoras e acadêmicas e comunidade científica

em geral. Um ponto de partida para compreender a gestão de dados de pesquisa é examinar a noção complexa de dados, ou melhor, conforme sinaliza Borgman (2015), quando são dados.

4.2 DADOS DE PESQUISA NO CONTEXTO DA PESQUISA EM SAÚDE

O termo dados (*data*) surgiu na metade do século XVII derivado da palavra *datum* em latim, utilizada para se referir à ‘única peça de informação’ (SEMELER; PINTO, 2019). Para Borgman (2007, p. 119-120, tradução nossa³²), dados são “uma representação reinterpretável de informações de maneira formalizada, adequada para comunicação, interpretação, ou processamento”, que se materializam em “fatos, números, letras, e símbolos que descrevem um objeto, ideia, condição, situação ou outros fatores” (NATIONAL RESEARCH COUNCIL, 1999, p. 15). Segundo Semeler e Pinto (2019, p. 118), os dados podem ser “símbolos e representações de informação e conhecimento representáveis”, no entanto, não devem ser equiparados a informações ou conhecimentos, uma vez que “os dados não têm nenhum valor ou significado quando estão isolados” (COSTA, 2017a, p. 56).

A diversidade de dados de pesquisa entre domínios evidencia a complexidade em delimitá-los, uma vez que são provenientes de específicos domínios disciplinares, criados para diferentes propósitos e contextos (científicos ou não), por diferentes comunidades acadêmicas e científicas, por diferentes tecnologias e por meio de diferentes processos (SALES; SAYÃO, 2019a). Dados de pesquisa são provenientes de qualquer investigação científica sistemática envolvendo um processo de observação, experimento ou teste de hipóteses, que quando produzidos no contexto e interpretados produzirão novos conhecimentos (PRYOR, 2012).

Uma definição amplamente disseminada para o termo é a proposta pela OECD que os definem como “[...] registros factuais (resultados numéricos, registros textuais, imagens e sons) utilizados como fontes primárias para pesquisa científica e que são comumente aceitos na comunidade científica como necessários para validar resultados da pesquisa” (ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2007, p. 13, tradução nossa³³). No âmbito da pesquisa em saúde, o NIH entende que dados finais de pesquisa são o conjunto de dados computadorizados. Em um estudo clínico, os dados de pesquisa seriam o conjunto de dados computadorizados subjacentes às publicações, não os

³² No original: “[...] *A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing*” (BORGMAN, 2007, p. 119-120).

³³ No original: “*factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings*” (ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2007, p. 13).

relatórios de patologia ou outros documentos de fontes clínicas. A agência também considera dados brutos e variáveis derivadas como dados finais de pesquisa (NATIONAL INSTITUTES OF HEALTH, 2003, *on-line*).

É importante sinalizar que as definições concebidas pelo NIH (2003) e OECD (2007) são consoantes em não considerar como dados de pesquisa os cadernos de laboratório, conjuntos de dados parciais, análises preliminares, rascunhos de artigos científicos, planos para pesquisas futuras, relatórios de revisão por pares, comunicações com colegas ou objetos físicos, como cepas de bactérias e animais de teste, géis ou amostras de laboratório.

Segundo Shamoo e Resnik (2015), as amostras biológicas do organismo (sangue, tecido, células, etc.) linhas celulares, lâminas, géis, animais de laboratório são considerados materiais de pesquisa, enquanto os dados de pesquisa incluiriam as informações genômicas derivadas desses materiais (isto é, sequências de ácido desoxirribonucleico [DNA]). Portanto, os procedimentos para materiais de pesquisa são regidos por legislações diferentes daquelas tratadas para os dados de pesquisa (SHAMOO; RESNIK, 2015). No entanto, não há consenso na literatura sobre essas limitações, uma vez que muitas políticas de dados de pesquisa incluem esses materiais de pesquisa em suas definições de dados.

Sobre a questão, Jorge (2018) alerta sobre a necessidade de diferenciações entre os materiais e dados de pesquisa no âmbito das estratégias e planos institucionais, defendendo que ambos os assuntos não podem ser discutidos isoladamente, uma vez muitos estudos, inicialmente apoiam-se na análise dos materiais de pesquisa para em seguida gerar os dados para a pesquisa em saúde.

Na pesquisa em saúde, os dados de pesquisa refletem desdobramentos éticos, políticos e legais que precisam ser reconhecidos para preservar direitos fundamentais como privacidade, intimidade e honra dos sujeitos da pesquisa. Briney (2015) explica que dados que contenham informações de identificação pessoal estão sujeitos à legislação vigente em âmbito nacional que rege como os dados devem ser tratados. Essa legislação varia de país para país, logo, não existe uma estratégia comum que esteja em conformidade legislativa entre os países envolvidos em um estudo. Portanto, é fundamental que pesquisadores que se deparem com possíveis desacordos normativos consultem os especialistas sobre os requisitos específicos das leis que se aplicam os tipos de dados gerados na pesquisa (BRINEY, 2015).

No que diz respeito à legislação brasileira, é importante ressaltar a Lei Geral de Proteção de Dados Pessoais (LGPD) que embora não se aplique a dados pessoais acadêmicos, prevê que o tratamento de dados pessoais poderá ser realizado em estudos por órgão de pesquisa, garantindo, sempre que possível, a anonimização desses dados (BRASIL, 2018).

Nessa direção, verifica-se que determinados tipos de dados, sejam sensíveis ou confidenciais, podem apresentar altos riscos se forem disponibilizados abertamente. Além disso, existem dados que não foram criados para fins de investigação científica como os dados governamentais, de registros médicos e dados de pacientes no prontuário médico, mas quando usados para fins de análise e validação de resultados de pesquisa originais são dados com o potencial para pesquisa (JORGE, 2018; SAYÃO, SALES, 2019b).

Diante desse cenário, Sales e Sayão (2019b) propõe uma categorização para os dados a partir de oito características: quanto à **origem**, os dados de pesquisa, podem ser observacionais, experimentais e computacionais e os dados para pesquisa, são classificados como governamentais, registros e transacionais; quanto ao **grau de processamento**, classificam-se como dados brutos/primários/crus, dados intermediários/pré-processados/secundários e dados finais/processados/terciários; quanto à **abordagem**, classificam-se como qualitativos e quantitativos; quanto à **natureza**, podem ser número, multimídia, *software*, visualização, textual, artefato e processo; quanto ao **nível de sensibilidade**, classificam-se como dados não sensíveis/ostensivos, dados confidenciais e dados pessoais (anonimizados ou sensíveis); quanto à **materialidade**, são classificados como dados digitais e dados físicos; quanto à **perenidade**, são classificados como dados canônicos/referenciais e dados episódicos; quanto à **abertura**, em termos de grau de abertura, podem ser fechado, aberto, parcialmente aberto e embargado e em termos de tipo de licenças para reuso, como exemplo, a *creative commons*.

Nesse horizonte, Sayão e Sales (2015) afirmam que a categorização dos dados de pesquisa possibilita a identificação de quais estratégias de arquivamento e preservação deverão ser escolhidas e quais são suas especificidades durante tratamento dos dados. Os autores explicam que a heterogeneidade dos dados deve ser considerada durante a formulação de estratégias de gestão de amplo espectro que forneçam opções gerenciais e tecnológicas para o arquivamento persistente e para a curadoria digital. Na pesquisa em saúde, questões éticas e legais também precisam ser abordadas durante o gerenciamento de dados, visando maximizar seu o valor, enquanto protege os sujeitos da pesquisa.

4.3 GESTÃO, CURADORIA E PRESERVAÇÃO DE DADOS DE PESQUISA

Pesquisadores têm enfrentado desafios relacionados ao tratamento de dados de pesquisa, sobretudo, nas áreas relacionadas à gestão, curadoria e preservação de dados de pesquisa (WHYTE; TEDDS, 2011). A gestão de dados de pesquisa é uma questão complexa,

pois envolve várias atividades realizadas por diferentes atores. Segundo Cox, Pinfield e Smith (2014), a gestão de dados de pesquisa consiste no conjunto de atividades e processos associados ao ciclo de vida dos dados, envolvendo o projeto e criação de dados, armazenamento, segurança, preservação, recuperação, compartilhamento e reutilização, considerações éticas, legais e estruturas de governança. Tem o objetivo de “[...] garantir uma verificação confiável dos resultados e permitir pesquisas novas e inovadoras, baseadas nas informações existentes” (WHYTE; TEDDS, 2011 p. 1, tradução nossa³⁴).

A gestão de dados é considerada um processo ativo que agrega valor aos dados ao longo do seu ciclo de vida para além do previsto inicialmente por seus criadores, por meio de estratégias que permitam que tais recursos digitais permaneçam detectáveis, acessíveis e inteligíveis para pesquisas futuras. Para Sayão e Sales (2015, p. 5), a gestão de dados “tem um papel crucial como facilitador nos processos de compartilhamento dos dados, na garantia da sustentabilidade e acessibilidade dos dados em longo prazo”. Segundo Carlson (2014), a gestão ocorre nos estágios “ativos” do ciclo de vida dados, ou seja, no período em que os pesquisadores estão produzindo e fazendo uso dos dados para seus propósitos de pesquisa.

O planejamento é fase crucial no processo, sendo formalmente descrito por meio de ações e compromissos assumidos no PGD, um documento cada vez mais solicitado por financiadores da pesquisa para a concessão de bolsas. De acordo com Veiga *et al.* (2019, p. 29), o PGD “é uma ferramenta para apoiar na solução de problemas de gestão por ser um documento formal, dinâmico e vivo que descreve detalhadamente como os dados de pesquisa serão tratados durante todo o ciclo de vida da pesquisa e possivelmente após a sua conclusão”. Os melhores planos são periodicamente revisados conforme necessidades ou mudanças nos protocolos, políticas, tecnologias e equipe ao longo do ciclo da pesquisa (MICHENER, 2015).

Importa mencionar que há um amplo reconhecimento de que os planos podem ser de maior interação com máquinas e mais programáveis (PRÍNCIPE *et al.*, 2020). Conforme expressaram Veiga *et al.* (2019, p. 28), o modelo de PGD acionável por máquina “permite a participação de forma contínua e dinâmica de atores envolvidos em todo o processo de gestão”, como pesquisadores, revisores éticos, financiadores, bibliotecários de dados, entre outros. De acordo com os autores, os PGD acionáveis por máquinas possibilitam que partes do documento sejam geradas e compartilhadas automaticamente e que as informações contidas nele sejam mais facilmente recuperadas, compartilhadas e reutilizadas por outros serviços de informação e/ou pesquisadores.

³⁴ No original: “[...] ensure reliable verification of results, and permits new and innovative research built on existing information” (WHYTE; TEDDS, 2011 p. 1).

Para a execução de atividades voltadas para o gerenciamento de dados, agências e organizações de pesquisa estão adotando modelos de ciclo de vida de dados. Segundo Pennock (2007) e Tenopir *et al.* (2011), o ciclo de vida dos dados permite a verificação da proveniência dos dados digitais, sendo agora visto como complementar ao ciclo de vida da pesquisa. Os modelos de ciclo de vida fornecem uma estrutura para considerar as operações que precisam ser executadas em um registro de dados ao longo de sua vida útil, minimamente incluindo as etapas de criação, uso ativo, avaliação e seleção, transferência, armazenamento e preservação, acesso e reutilização (PENNOCK, 2007) e garantindo a manutenção da autenticidade, confiabilidade, integridade e usabilidade do dado (HIGGINS, 2008).

Vários modelos de ciclo de vida dos dados como o *Curation Lifecycle Model* do DCC e o *Data Lifecycle* do DataONE foram documentados na literatura analisada. As revisões realizadas por Ball (2012) e Cox e Tam (2018) sugerem que esses modelos possuem suas particularidades e objetivos específicos, visto que enquanto alguns são específicos para um campo ou instituição acadêmica, outros procuram generalizar para qualquer forma de pesquisa; enquanto alguns descrevem com mais detalhes as etapas de criação e reutilização de dados, outros são mais enfáticos nas etapas de curadoria de dados.

A abordagem do ciclo de vida dos dados vem sendo empregada na modelagem da curadoria digital como um mecanismo que auxilia na definição e ilustração dos processos complexos que ocorrem em distintos estágios dos dados na pesquisa, garantindo que todas as etapas necessárias sejam identificadas e planejadas, e necessárias ações sejam implementadas, na sequência correta (HIGGINS, 2008; WALLIS, 2012; RAY, 2014).

Conforme assinalam Sayão e Sales (2012, p. 184), a curadoria digital compreende a “gestão atuante e a preservação de recursos digitais durante todo o ciclo de vida de interesse do mundo acadêmico e científico, tendo como perspectiva o desafio temporal de atender a gerações atuais e futuras de usuários”, sendo os dados digitais um dos muitos tipos de recursos digitais (WALLIS, 2012). Orientada aos dados de pesquisa, Sayão e Sales (2012, p. 179) informam que a curadoria digital “envolve a gestão de dados de pesquisa desde o seu planejamento, assegurando a sua preservação por longo prazo, descoberta, interpretação e reuso”.

É oportuno esclarecer que, embora a gestão e curadoria sejam conceitos amplamente discutidos, a maneira como se sobrepõem e se relacionam ainda está em discussão. Na visão de Wallis (2012), a gestão de dados tende a implicar nas funções de armazenamento e documentação, enquanto a curadoria abrange um conjunto mais amplo de atividades de

suporte à reprodutibilidade e reuso de dados. No entanto, a autora reconhece que o termo gestão de dados é, atualmente, um termo mais significativo para a comunidade científica.

Koltay (2017) entende que a curadoria digital inclui a curadoria de dados e a preservação digital. No que diz respeito à curadoria de dados, o autor a considera como uma extensão das atividades de gerenciamento de dados, pois inclui serviços adicionais para preservar e agregar valor aos dados, permitindo a reutilização. No empreendimento científico, a agregação de valor é um fator significativo para a tão almejada obtenção de impacto (PRYOR, 2012). Estudos recentes demonstraram que pesquisadores que compartilham dados em arquivos públicos recebem mais citações (PIWOWAR; VISION, 2013).

De acordo com Carlson (2014), durante o processo de curadoria os dados não estão mais sendo desenvolvidos e, geralmente, são administrados por terceiros, como um gerenciador de repositório que desenvolve ações para aumentar o valor e a utilidade dos dados. A curadoria de dados inclui o arquivamento, definido por Lord e Macdonald (2003, p. 12, tradução nossa³⁵) como “uma atividade de curadoria que garante que os dados sejam adequadamente selecionados, armazenados e acessados e que sua integridade lógica e física seja mantida ao longo do tempo, incluindo segurança e autenticidade”. A curadoria de dados conecta o primeiro uso ao uso secundário dos dados, pois garante “que os resultados do projeto sejam adequados para arquivamento e que os ativos de pesquisa valiosos permaneçam adequados para reutilização” (WHYTE; TEDDS, 2011, sem paginação, tradução nossa³⁶).

Conforme escrevem Nascimento, Araújo e Arellano (2020, p. 530), a preservação digital visa “[...] assegurar a longevidade dos objetos digitais em seus acervos e garantir que permaneçam disponíveis, recuperáveis e compreensíveis pelo tempo que se fizer necessário”. Lord e Macdonald (2003) a consideram como uma atividade realizada dentro do arquivamento na qual itens específicos de dados são mantidos ao longo do tempo para que eles ainda possam ser acessados e compreendidos através de mudanças na tecnologia. Whyte e Tedds (2011) explicam que a preservação está em garantir que os dados de pesquisa permaneçam acessíveis por longo prazo para o uso secundário.

Portanto, observa-se que os dados de pesquisa para revelarem e transmitirem conhecimento por longo tempo exigem ações de gestão, curadoria e preservação constantes ao longo do ciclo de vida, que se iniciam na fase de planejamento do projeto de pesquisa, com a

³⁵ No original: “A curation activity which ensures that data is properly selected, stored, can be accessed and that its logical and physical integrity is maintained over time, including security and authenticity” (LORD; MACDONALD, 2003, p. 12).

³⁶ No original: “[...] that project results are fit to archive, and that valued research assets remain fit for reuse” (WHYTE; TEDDS, 2011, sem paginação).

formulação do PGD até depósito dos dados em um repositório adequado (SAYÃO; SALES, 2015; 2016). Os repositórios digitais demonstram-se instrumentos estratégicos para a gestão, compartilhamento e reuso dos dados, pois, segundo Sayão e Sales (2015), garantem a preservação ativa dos materiais digitais por longo prazo, permitindo que coleções de dados mantenham suas características de autenticidade, integridade e proveniência bem como sejam recuperadas, usadas, reusadas, compartilhadas e citadas por pesquisadores.

4.4 REPOSITÓRIOS DE DADOS, PADRÕES DE METADADOS E OS PRINCÍPIOS FAIR

Os repositórios digitais estão inseridos em um contexto de surgimento de ferramentas alternativas para a comunicação da ciência livre de barreiras de acesso. Segundo Weitzel (2006, p. 59), um repositório digital pode ser compreendido como “um arquivo digital que reúne uma coleção de documentos digitais”. Os repositórios digitais, sejam eles institucionais ou temáticos, visam promover, respectivamente, o acesso livre e irrestrito à literatura científica e acadêmica de uma instituição ou de um domínio do conhecimento.

Existe uma diversidade de conteúdos e formatos que podem ser armazenada nos repositórios, sendo os institucionais os mais abrangentes, incluindo tanto produção científica avaliada por pares como artigos científicos e outras tipologias que não foram submetidas ao processo formal de validação da ciência, como *preprints*, dados de pesquisa, entre outros arquivos multimídias (VEIGA, 2017).

Diante da amplitude de exigências compulsórias das agências de fomento e pressupostos da Ciência Aberta, as universidades e instituições de pesquisa começam a expandir os serviços oferecidos por seus repositórios institucionais – tradicionalmente voltados para publicações digitais – para o domínio heterogêneo e complexo dos dados, implementando sistemas de repositórios digitais para os dados de pesquisa, nas mais variadas configurações, plataformas tecnológicas e modelos de gestão (SAYÃO; SALES, 2016).

Os repositórios digitais de dados de pesquisa são uma extensão dos repositórios digitais, sendo “mantidos por conjuntos de ações que viabilizem o armazenamento de dados visando à otimização da coleta pelos pesquisadores, o que amplia as potencialidades de reuso destes dados” (MONTEIRO; SANT’ANA, 2018, p. 162).

De acordo com a revisão de Pampel *et al.* (2013), os repositórios de dados podem ser classificados como institucionais, disciplinares, multidisciplinares e orientados por projetos. Esses repositórios podem ter requisitos específicos como: assunto domínio/pesquisa; licença para acesso aos dados e seu reuso; formato de arquivo e estrutura de dados; metadados de

descrição; embargo; preços; ligação com periódicos; atribuição de identificação como o Identificador de Objeto Digital (*Digital Object Identifier* – DOI) (AVENTURIER; ALENCAR, 2016). Neste contexto, cita-se a iniciativa *Registry of Research Data Repositories* (re3data.org), um registro global de repositórios de dados de pesquisa que abrange repositórios de dados de diferentes áreas do conhecimento. Costa (2017b) afirma que cerca de 49% dos repositórios registrados no re3data.org é relacionado às ciências da vida.

Sayão e Sales (2016) observam que a implementação de repositórios de dados amplia a visibilidade dados, permitindo que sejam consultados e citados; permite a troca de experiência e o compartilhamento dos dados; oferece um ambiente tecnológico, gerencial e de padronização propício para a preservação de longo prazo; contribui para a formação da memória científica das instituições; oferece sistema de armazenamento seguro, esquemas de *backup* e segurança física; proporciona um ambiente apropriado para os processos de curadoria digital (avaliação, de adição de valor, reformatação, agregação e recriação de dados); aumenta o grau de reuso e reinterpretação dos dados; entre outros benefícios.

É importante salientar que os repositórios de dados são um dos muitos meios para ser compartilhar dados, outras formas são: publicação ou documento ampliado, que disponibiliza no próprio documento *links* para material suplementar, como um conjunto de dados; artigos de dados (*data papers*), publicados com revisão por pares em um periódico de artigos tradicionais que aceite artigos de dados; ou periódicos de dados (*data journals*), com revisão por pares, que só publicam artigos de dados (VEIGA, 2017). Além disso, Aventurier e Alencar (2016) ainda acrescentam que a depender da disciplina, os dados de pesquisa podem ser depositados em um ou mais repositórios que permitam o seu acesso.

Repositórios de dados têm significativa importância no contexto da colaboração científica internacional. Lee *et al.* (2009) apontam que os repositórios de dados permitem maior colaboração entre disciplinas e em maiores escalas, ao passo que proporcionam aos pesquisadores a produção de conhecimento livre de barreiras, por meio de informações abertamente acessíveis aos envolvidos em tal empreendimento científico. Repositórios bem estruturados permitem construir uma melhor compreensão dos problemas científicos e tecnológicos de importância social e identificação de soluções sustentáveis (LEE *et al.*, 2009).

Para a implementação e funcionamento dos repositórios de dados é indispensável o estabelecimento de padrões e ferramentas que permitam a descrição e divulgação dos dados de pesquisa (COSTA, 2017b). Nesse contexto, os metadados são fundamentais para a descrição padronizada dos dados, de modo a torná-los compreensíveis agora e no futuro. A ausência de metadados e documentação adequada que descreva os contextos e ferramentas

utilizadas para sua criação, armazenamento, adaptação e análise, tornam os dados de pesquisa sem valor e improváveis de serem descobertos, interpretados, gerenciados e efetivamente usados e reusados, conforme ressaltam Sayão e Sales (2016).

De forma geral, o termo “metadados” é utilizado para denotar “dados sobre dados” (BUCKLAND, 2006), no entanto, o mesmo possui uma definição mais ampla, sendo um “conjunto de informações essenciais, padronizadas e estruturadas que documentam os dados explicando sobre a proveniência, origem, propósito, autores, instituições envolvidas, termos de uso e detalhes técnicos e estruturais” (SAYÃO; SALES, 2015, p. 29).

Sayão e Sales (2015) atestam que grande parte dos repositórios incluem campos de metadados complementares para cada conjunto de dados depositado, de acordo com algum formato ou esquema que facilite a busca e a troca de metadados (interoperabilidade) entre repositórios. Entre os esquemas mais usados para descrição dos dados de pesquisa, os autores citam o *Dublin Core*, um padrão de metadados aplicável a várias disciplinas que fornece um esquema simples para a descrição de recursos da internet composto campos descritivos para autor, resumo, assunto, totalizando quinze elementos (SAYÃO, SALES, 2015). Vidotti *et al.* (2017, p. 228) afirmam que os padrões de metadados permitem “garantir que a descrição do recurso informacional seja recuperada e interoperável por diferentes sistemas de informação”.

Repositórios de dados, padrões de metadados, identificadores persistentes, políticas e PGD são os principais serviços e infraestruturas no ecossistema dos dados FAIR (EUROPEAN COMMISSION, 2018c) um acrônimo para *Findable, Accessible, Interoperable e Reusable* (localizáveis, acessíveis, interoperáveis e reusáveis) (WILKINSON *et al.*, 2016).

Em um *workshop* realizado na Holanda em 2014, especialistas, agências de fomento à pesquisa, editores científicos e outros atores reuniram-se para discutir a importância de se melhorar a infraestrutura de suporte à reutilização de dados acadêmicos (WILKINSON *et al.*, 2016). Segundo Wilkinson *et al.* (2016), esse encontro foi marcado pelo primeiro movimento para esboçar um conjunto de princípios e práticas orientadores visando facilitar a descoberta, acesso, integração, reutilização e citação de dados gerados no contexto da *e-Science*. Posteriormente, os princípios foram ajustados e publicados pelo grupo FORCE11³⁷.

Conforme informam Henning *et al.* (2019), cada letra do acrônimo remete a um conjunto de especificações que requerem conhecimentos técnicos e estruturais distintos. Os princípios FAIR são constituídos por 15 princípios que consideram que os dados e os metadados devam ser: *localizáveis*, sendo único e persistentemente identificável; *acessíveis*,

³⁷ Fonte: <https://www.force11.org/group/fairgroup/fairprinciples>.

na medida em que pode ser sempre obtido por máquinas e os seres humanos; *interoperáveis*, sendo acionáveis por máquinas, utilizando vocabulários compartilhados e/ou ontologias e sinteticamente analisados e semanticamente acessíveis à máquina e por fim, *reusáveis*, na medida em que são suficientemente bem descritos e podem ser automaticamente (ou com o esforço humano mínimo) vinculado ou integrado com outras fontes de dados, facilitando a citação adequada (FORCE11, [2014?]).

Henning (2019, sem paginação) ainda sinaliza que “tais princípios, embora criados dentro do contexto da ciência aberta, não tem a pretensão de impor padrões, muito menos de tratar de questões morais e éticas relativas à abertura”. Na verdade, esses princípios FAIR servem para orientar “boas práticas de gestão, compartilhamento, interoperabilidade voltados, quando possível, para o reuso dos dados” (HENNING, 2019, sem paginação).

De acordo com Hodson (2018), a implementação dos princípios FAIR requerer mudanças na prática e cultura de pesquisa de muitas comunidades, instituições de pesquisa e financiadores. Em 2018, a Comissão Europeia emitiu um relatório com quinze recomendações prioritárias para tornar os dados FAIR uma realidade, entre as questões apontadas está a necessidade de se desenvolver estruturas de interoperabilidade para o compartilhamento FAIR dentro de disciplinas e pesquisas interdisciplinares (EUROPEAN COMMISSION, 2018c).

A comissão defende que as comunidades de pesquisa precisam ser apoiadas para desenvolver estruturas de interoperabilidade que definam suas práticas para compartilhamento de dados, formatos de dados, padrões de metadados, ferramentas e infraestrutura (EUROPEAN COMMISSION, 2018c). No contexto da pesquisa interdisciplinar, essas estruturas de interoperabilidade devem ser articuladas em comum e adotar padrões globais quando relevante. Portanto, durante a construção de estruturas de interoperabilidade deve-se considerar a necessidade de apoiar pesquisas que trabalham além das fronteiras disciplinares tradicionais.

Apresentado o referencial teórico desta pesquisa, o próximo capítulo apresenta o percurso metodológico delineado para responder a questão de pesquisa e cumprir os objetivos inicialmente estabelecidos, descrevendo as escolhas metodológicas feitas, com vistas a dar maior transparência a todo processo desta pesquisa.

5 PERCURSO METODOLÓGICO

A abordagem metodológica aplicada nesta pesquisa parte do entendimento de Gil (2012, p. 8) que considera o método como o “caminho para se chegar a determinado fim. E método científico o conjunto de procedimentos intelectuais e técnicos adotados para se atingir o conhecimento”. Desse modo, este capítulo descreve as escolhas metodológicas realizadas para responder a questão de pesquisa e atingir os objetivos estipulados.

O objetivo geral da pesquisa foi investigar as orientações dos financiadores da pesquisa para gestão de dados produzidos em contexto de colaboração científica internacional no campo da saúde. Para efeito desta pesquisa, foi necessário identificar e analisar as expectativas e requisitos relacionados à gestão de dados de pesquisa consubstanciados no arcabouço normativo dos principais financiadores da pesquisa em saúde.

5.1 CARACTERIZAÇÃO DA PESQUISA

Este estudo caracteriza-se como exploratório-descritivo quanto aos seus objetivos. Exploratório porque pesquisas desta natureza são realizadas em áreas com pouco conhecimento acumulado e sistematizado, o que deste modo permite aproximação com o tema (VERGARA, 2011; GIL, 2012). Descritivo porque coleta, organiza e descreve sistematicamente os dados da pesquisa com a finalidade de expor detalhadamente características do objeto estudado (GIL, 2012). Nessa perspectiva, é importante ressaltar o caráter exploratório do estudo, uma vez que em mapeamento prévio da literatura, nacional e internacional, verificou-se a escassez de estudos com esta finalidade.

Do ponto de vista da forma de abordagem do problema, a escolha adotada foi a pesquisa qualitativa, pois o significado dos dados coletados é a preocupação central na abordagem de cunho qualitativo que tem como principal objetivo a percepção do fenômeno dentro do seu contexto (TRIVIÑOS, 1987). Isso implica que o material coletado seja predominantemente descritivo para que se possa analisar com profundidade o fenômeno estudado.

Já no que se refere aos meios de investigação, o presente estudo caracteriza-se como de caráter bibliográfico e documental. Segundo Lakatos e Marconi (2001, p. 183), a pesquisa bibliográfica “[...] abrange toda bibliografia já tornada pública em relação ao tema estudado, desde publicações avulsas, boletins, jornais, revistas, livros, pesquisas, monografias, teses, materiais cartográficos, etc. [...]”, proporcionando ao pesquisador contato direto com

contribuições de diversos autores sobre determinado assunto. Já a pesquisa documental distancia-se do caráter bibliográfico no que tange à natureza das fontes, pois “vale-se de materiais que não receberam ainda um tratamento analítico, ou que ainda podem ser reelaborados de acordo com os objetivos da pesquisa” (GIL, 2011, p. 51), como políticas, editais, relatórios, procedimentos, diretrizes, entre outros documentos dessa natureza.

5.2 DEFINIÇÃO DE CRITÉRIOS PARA COLETA E ANÁLISE DE DADOS

Para a operacionalização da coleta de dados e posterior análise, foi aplicada a pesquisa bibliográfica com vista estritamente à construção das bases teóricas da pesquisa, a partir de uma revisão de literatura do conhecimento, proporcionando a apropriação dos conceitos fundamentais para este estudo. Em um segundo momento do estudo, a pesquisa documental foi utilizada para a coleta de documentos que fornecem informações suficientes para realizar inferências a respeito de como os financiadores da pesquisa em saúde orientam a gestão de dados produzidos no âmbito de projetos ou programas de pesquisa de abrangência internacional. A seguir, as opções metodológicas são detalhadamente expostas.

5.1.1 Pesquisa bibliográfica

Para a pesquisa bibliográfica foram selecionadas as seguintes fontes de informações internacionais e nacionais: SCOPUS (via portal Capes), *Medical Literature Analysis and Retrieval System Online* (Medline) via PubMed, Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI) e a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD). É importante destacar que o Google Acadêmico (*Google Scholar*) também foi utilizado como um motor de busca complementar.

A base SCOPUS foi selecionada pelo caráter multidisciplinar que abrange documentos das mais variadas áreas do conhecimento científico como ciência, tecnologia, medicina, ciências sociais, artes e humanidades. A Medline/PubMED foi selecionada em virtude da sua cobertura temática na área médica e biomédica internacional. A BRAPCI foi selecionada por ser uma das fontes de informação mais relevantes para área da Ciência da Informação no Brasil, desse modo a partir da busca foi possível identificar o que está sendo abordado no país sobre o tema da pesquisa. A BDTD, mantida pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), foi escolhida por ser uma fonte de informação de acesso livre às teses e dissertações defendidas no âmbito de instituições brasileiras de ensino superior. O

Google Acadêmico foi utilizado como ferramenta auxiliar no processo da busca bibliográfica, pois fornece uma visão geral e aberta das publicações em diferentes fontes e tipos documentais.

Para a concretização da pesquisa bibliográfica, foi selecionada uma série de conceitos que julgamos representativos para a construção da revisão de literatura. Desta forma, a busca sistemática envolveu tanto termos controlados utilizados pelas bases de dados, bem como linguagem natural, na ausência de vocabulário controlado. O quadro 1 apresenta os blocos conceituais para a construção da pesquisa, as palavras chaves-chaves e sinônimos utilizados nas bases pesquisadas e a adaptação dos termos para o português e inglês.

Quadro 1 – Termos selecionados para a busca bibliográfica

| Idiomas | Bloco de conceitos | Palavras-chaves | Sinônimos |
|----------------|------------------------------------|------------------------------------|---|
| Português | Colaboração científica | Colaboração científica | Colaboração em pesquisa; pesquisa colaborativa; |
| Inglês | <i>Scientific collaboration</i> | <i>Scientific collaboration</i> | <i>Research collaboration;</i> <i>collaboration in research;</i> <i>collaborative research;</i> |
| Português | Colaboração internacional | Colaboração internacional | Pesquisa internacional |
| Inglês | <i>International collaboration</i> | <i>International collaboration</i> | <i>International research</i> |
| Português | Cooperação científica | Cooperação científica | Cooperação em pesquisa |
| Inglês | <i>Scientific cooperation</i> | <i>Scientific cooperation</i> | <i>Research cooperation;</i> <i>cooperation in research;</i> <i>Cooperative research;</i> |
| Português | Cooperação internacional | Cooperação internacional | Cooperação global; |
| Inglês | <i>International cooperation</i> | <i>International cooperation</i> | <i>Global cooperation;</i> |
| Português | Dados de pesquisa | Dados de pesquisa | Dados científicos; Dados digitais; |
| Inglês | <i>Research data</i> | <i>Research data</i> | <i>Scientific data; digital data</i> |
| Português | Gestão de dados de pesquisa | Gestão de dados de pesquisa | Gestão de dados científicos; Curadoria de dados; |
| Inglês | <i>Research data management</i> | <i>Research data management</i> | <i>Scientific data management;</i> <i>Data curation;</i> |
| Português | Pesquisa em saúde | Pesquisa em saúde | Pesquisa médica; pesquisa biomédica; |
| Inglês | <i>Health research</i> | <i>Health research</i> | <i>Medical research;</i> <i>Biomedical research;</i> |
| Português | Agência de fomento | Agência de fomento | Financiador de pesquisa; Financiamento em pesquisa; Fomento à pesquisa; |

| | | | |
|--------|-------------------------|-------------------------|--|
| Inglês | <i>Funding agencies</i> | <i>Funding agencies</i> | <i>Research Funders; Research funding; Research promotion;</i> |
|--------|-------------------------|-------------------------|--|

Fonte: Adaptado a partir de Estevão (2019).

Após a seleção dos termos mais representativos do tema da pesquisa para a construção das estratégias de busca, os procedimentos subsequentes foram buscas simples e combinadas de acordo com as especificidades de cada fonte de informação selecionada. Nas fontes de informação nacionais BRAPCI e BDTD (IBICT), procederam-se buscas simples e, quando possíveis, combinadas no campo que busca no título, palavra-chave e resumo, sem o uso de filtros. Na base de dados internacionais SCOPUS e Medline/PubMED foram realizadas buscas combinadas com os termos expostos no quadro 1 a partir do campo específico que busca no título, resumo, palavras-chave. Cabe esclarecer que não foram identificados descritores no Descritores em Ciências da Saúde (DeCS) e no Medical Subject Headings (MeSH), sendo assim, optou-se pelo uso de termos livres nas estratégias de busca realizadas na Medline/PubMED.

Efetuada a pesquisa bibliográfica, os dados coletados foram descarregados em formato CSV e exportados para o formato XLS na ferramenta *Excel*. A etapa seguinte consistiu na limpeza dos dados a partir de: 1) remoção de duplicatas; 2) análise do título, resumo e palavras-chaves. Em seguida, foi realizada a leitura integral dos documentos selecionados para a construção do referencial teórico desta pesquisa, que foi complementado com artigos, dissertações, teses, livros e relatórios recuperados em buscas exploratórias no Google acadêmico.

5.1.2 Pesquisa documental

A pesquisa documental foi utilizada para operacionalização da segunda etapa da pesquisa. O primeiro procedimento para a caracterização dos principais financiadores da pesquisa em saúde foi a seleção de uma fonte que fornecesse informações suficientes para tal, o que foi obtido na etapa da pesquisa bibliográfica quando foi identificado o artigo “*The 10 largest public and philanthropic funders of health research in the world: what they fund and how they distribute their funds*” de Viergever e Hendriks (2016)³⁸. Esse estudo foi então o ponto de partida para identificar as principais organizações públicas e filantrópicas de

³⁸ Identificou-se o referido artigo na pesquisa de Jorge (2018) que com base no estudo verificou iniciativas e práticas desenvolvidas por atores que participam e financiam pesquisas em saúde para o compartilhamento e abertura de dados nas rotinas científicas, mas também em situações de emergências em saúde pública.

financiamento à pesquisa em saúde no mundo bem como entender o que elas financiam e como distribuem seus recursos. Ainda que os dados coletados pelos autores sejam referentes ao período entre novembro de 2013 e agosto de 2014, esse estudo foi o que sistematicamente ofereceu informações para servir como fonte ao elencar os principais financiadores da pesquisa em saúde mundialmente. O quadro 2 a seguir fornece o *ranking* dos dez primeiros financiadores, de acordo com o total de gastos anuais em pesquisa em saúde relativos ao ano de 2013.

Quadro 2 – Principais financiadores da pesquisa em saúde

| Ranking | Financiador | Gastos | País/Sede |
|----------------|--|-----------------------|------------------|
| 1° | NIH | US\$ 26,081.3 bilhões | EUA |
| 2° | EC | US\$ 3717.7 bilhões | Bélgica |
| 3° | UK MRC | US\$ 1321.5 bilhões | Inglaterra |
| 4° | Institut National de la Santé et de la Recherche Médicale (Inserm) | US\$ 1041.2 bilhões | França |
| 5° | United States Department of Defense (US DoD) | US\$ 1017.7 bilhões | EUA |
| 6° | Wellcome Trust | US\$ 909.1 milhões | Inglaterra |
| 7° | CIHR | US\$ 883.6 milhões | Canadá |
| 8° | NHMRC | US\$ 777.6 milhões | Austrália |
| 9° | Howard Hughes Medical Institute (HHMI) | US\$ 752.0 milhões | EUA |
| 10° | DFG | US\$ 630.6 milhões | Alemanha |

Fonte: Adaptado a partir de Viergever e Hendriks (2016).

Considerando o tempo delimitado para a realização desta pesquisa, foram selecionados apenas os dois primeiros financiadores da pesquisa em saúde para análise, quais sejam: NIH e EC. Vale a pena destacar que não foi objetivo da pesquisa realizar uma análise comparativa entre os dois financiadores selecionados, pois embora as instituições selecionadas financiem a pesquisa em saúde, as mesmas apresentam características diferentes em seus respectivos sistemas de financiamento. Enquanto o NIH é uma agência norte-americana majoritariamente de fomento interno, a EC é uma instituição europeia que representa e defende os interesses da EU que, por sua vez, dispõe de programas de financiamento partilhados entre os Estados-membros que a compõe. Importa ainda ressaltar que durante a revisão de literatura foi possível perceber que tanto o NIH como a EC já haviam considerado a questão dos dados de pesquisa no âmbito de suas políticas de concessão. Desse modo, foi relevante para esta pesquisa identificar e analisar o(s) marco(s) normativo(s) das políticas para gestão de dados de pesquisa dentro dessas instituições.

Em seguida, para mapear e analisar os documentos normativos dos financiadores da pesquisa em saúde com políticas para gestão dos dados de pesquisas financiadas, o segundo procedimento consistiu em identificar nos portais eletrônicos (*sites*) de cada financiador o arcabouço normativo formalizado e instituído por seu corpo institucional. As estratégias empregadas para localizar esses documentos foram: buscas pelos termos ‘*data sharing*’, ‘*data management*’ ou ‘*data policies*’ e consultas na seção de políticas de cada financiador. Durante a busca foi possível verificar a diversidade de documentos normativos com políticas para a gestão de dados de pesquisa, incluindo guias, normas, declarações, políticas, diretrizes entre outros.

Para ainda atingir segundo objetivo específico da pesquisa, foi necessário delimitar quais elementos descritos nas políticas seriam analisados. Para tanto, o terceiro procedimento metodológico foi a realização de leituras técnicas de modelos de estrutura de avaliação de políticas de dados de pesquisa³⁹ presentes na literatura recuperada, com vistas a construir um modelo que fosse significativo para a análise do objeto de estudo que esteve centrado na gestão de dados de pesquisa em saúde, no contexto da colaboração científica internacional.

É oportuno esclarecer que os modelos analisados, a princípio, foram propostos para a análise de políticas institucionais, no entanto, foi possível extrair deles elementos-chaves para a análise de políticas de financiadores. Também cabe sinalizar que o modelo foi utilizado para avaliação das políticas no sentido mais amplo, ou seja, considerou-se por política o conjunto de recomendações, normas, diretrizes e regras emitidos pelos financiadores como condição para a concessão de recursos financeiros na pesquisa. Os elementos-chaves identificados foram agrupados na categoria grau de aplicação e cobertura de critérios, conforme ilustra o quadro 3 a seguir.

³⁹ Os modelos de avaliação de políticas de dados de pesquisa consultados foram os propostos por Digital Curation Centre ([2012?]), Horton e Digital Curation Centre (2016) e Monteiro (2019). O modelo final utilizado na pesquisa também foi complementado por elementos-chaves de políticas de dados de pesquisa identificados no estudo elaborado por Hodson e Molloy (2015), a pedido do Committee on Data of the International Science Council – CODATA.

Quadro 3 – Estrutura de avaliação de políticas de dados de pesquisa

| CATEGORIAS | | | | | | | | | | | | | | | | | |
|-------------------|------------|------------------|------------------------|--------|-------------------|---|-------------------------------|---------|----------------------------------|----------------------|---------------------------------|-------|--------------------------|-------------|--------------------------|-------|-------|
| Grau de aplicação | | | Cobertura de critérios | | | | | | | | | | | | | | |
| Recomendatório | Mandatório | Não identificado | Aspectos gerais | | | | Planejamento e infraestrutura | | | | | | Disponibilização e reuso | | | | |
| | | | Definição | Escopo | Responsabilidades | Monitoramento de conformidade e sanções | Plano de Gestão de Dados | Suporte | Curadoria/ Preservação por longo | Padrões de metadados | Repositórios e centros de dados | Custo | Limite de tempo | Propriedade | Compartilhamento Acesso/ | Reuso | Ética |
| ■ | ■ | ■ | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| ■ | ■ | ■ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ |
| ■ | ■ | ■ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

Legenda: ■ Grau de aplicação identificado

● Cobertura total

◐ Cobertura parcial

○ Nenhuma cobertura

Fonte: Adaptado a partir de Digital Curation Centre ([2012?]), Hodson e Molloy (2015), Horton e Digital Curation Centre (2016) e Monteiro (2019).

A primeira categoria de análise referiu-se ao grau de aplicação dos documentos analisados, que pôde ser classificado como “recomendatório” quando o documento dispõe de orientações e sugestões que se não cumpridas pelo pesquisador não impedem o recebimento de fomento à pesquisa; “mandatório” para documentos que apresentam exigências e obrigações que precisam ser atendidas para o financiamento e; “não identificado” quando o documento não apresenta o caráter recomendatório, nem o mandatório.

A segunda categoria de análise compreendeu a cobertura de critérios no que diz respeito aos “aspectos gerais”, “planejamento e infraestrutura” e “disponibilização e reuso”. A graduação da cobertura seguiu a escala proposta pelo modelo do Digital Curation Centre ([2012?]; 2016) e aplicado por Monteiro (2019) em que a escala de “cobertura total” foi utilizada para o critério que atendia a sua descrição completa; “cobertura parcial” para o critério que atendia parcialmente a sua descrição completa e; “nenhuma cobertura” quando o documento não fornecia informações referentes ao critério analisado.

Conforme descrito no quadro 4 a seguir, os elementos da subcategoria “aspectos gerais” referem-se aos compromissos, definições e aspirações dos principais *stakeholders* para a implantação de medidas que visam habilitar a política. A subcategoria “planejamento e infraestrutura” compreende os elementos essenciais para a implantação de medidas que forneçam as estratégias, ferramentas, recursos e infraestrutura para tornar os dados

preservados por longo prazo. A subcategoria “disponibilização e reuso” envolve os esforços necessários para tornar os dados localizáveis, acessíveis, interoperáveis e reusáveis pela comunidade científica e sociedade em geral.

Quadro 4 – Descrição dos elementos-chaves da categoria “cobertura de critérios” presentes na estrutura de avaliação de políticas de gestão de dados de pesquisa

| Crítérios | Elemento | Descrição |
|--------------------------------------|---|---|
| Aspectos gerais | Definição | Oferece definições dos termos principais da gestão de dados, ou seja, “dados”, “documentação”, “dados de pesquisa abertos”, “dados FAIR”, etc. |
| | Escopo | Especifica a quem se destina a política e quais dados de pesquisa devem ser contemplados |
| | Responsabilidades | Estabelece as responsabilidades dos <i>stakeholders</i> para garantir a preservação, disponibilidade e reutilização dos dados |
| | Monitoramento de conformidade e sanções | Indica como o cumprimento será monitorado, quais relatórios necessários e quais sanções podem ser impostas |
| Planejamento e infraestrutura | Plano de Gestão de Dados | O financiador recomenda ou exige um PGD como parte de todas as propostas de pesquisa |
| | Suporte | Define o papel que a instituição desempenha no suporte à gestão dos dados |
| | Curadoria/ Preservação de longo prazo | Fornece estipulações sobre manutenção de longo prazo e preservação de resultados de pesquisa |
| | Padrões de metadados | Fornece recomendações sobre os padrões de metadados que devem ser apropriados, ricos, de alta qualidade e estar em conformidade com as normas e boas práticas da comunidade |
| | Repositórios e centros de dados | Prevê o depósito de dados em um repositório para garantir o acesso e preservação ou <i>data center</i> para curadoria de dados não publicados |
| | Custo | Declaração sobre os custos da gestão dos dados |
| Disponibilização e reuso | Limite de tempo | Declaração sobre os prazos para tornar o conteúdo acessível e a duração mínima que os dados devem ser mantidos (preservados) |
| | Propriedade | Declaração sobre propriedade institucional dos dados da pesquisa |
| | Compartilhamento/ Acesso | Declaração sobre como os dados serão acessados e disponibilizados/compartilhados |
| | Reuso | Declaração sobre tipos de licenciamento e práticas de citação de dados |
| | Ética | Declaração sobre o uso ético/reutilização de dados pessoais |

Fonte: Adaptado a partir de Digital Curation Centre ([2012?]), Hodson e Molloy (2015), Horton e Digital Curation Centre (2016) e Monteiro (2019).

Na etapa de pré-análise, as informações dos documentos recuperados de cada financiador foram organizadas em planilhas no editor da Microsoft *Excel* (2010), com metadados referentes à instituição financiadora, título do documento, tipologia documental, data de publicação e endereço eletrônico do documento analisado. Cada planilha possui a estrutura de avaliação de políticas de dados de pesquisa descrita no quadro 3.

Para a análise dos dados, procedeu-se com uma leitura na íntegra dos documentos, com a finalidade de reunir informações que fornecessem respostas aos elementos contemplados dentro de cada categoria. Em seguida, foram extraídos dos documentos analisados os trechos significativos às categorias analisadas e posteriormente organizados nas respectivas planilhas. Sobre a análise das políticas para gestão de dados de pesquisa materializadas nos documentos normativos recuperados é importante mencionar que tais documentos foram analisados conjuntamente na avaliação final. Desse modo, se um critério era coberto totalmente em pelo menos um dos documentos analisados, a avaliação final do critério em questão corresponderia à “cobertura total”. Cabe ainda sinalizar que a análise dos documentos foi realizada de forma individualizada em virtude das particularidades de cada instituição, conforme mencionado anteriormente.

A partir dos critérios analisados, buscou-se identificar se entre os elementos contemplados na estrutura coexistiam especificações para a gestão de dados de pesquisa em saúde quando produzidos no âmbito da pesquisa colaborativa internacional. Embora não exclusivamente, tal análise deu ênfase ao critério “Responsabilidades”, visando identificar se os financiadores da pesquisa apresentam políticas relativas às responsabilidades dos pesquisadores partícipes de pesquisas colaborativas internacional no campo da saúde quanto ao gerenciamento dos dados produzidos.

O processo de gestão dos dados coletados nesta pesquisa está descrito no Plano de Gestão de Dados, elaborado a partir do modelo simplificado fornecido pela Fiocruz (apêndice A). O conjunto de dados gerados durante a etapa da pesquisa documental bem como o PGD da pesquisa foram depositados na comunidade da Fiocruz, disponível no repositório multidisciplinar de acesso aberto Zenodo e estão registrados sob o DOI 10.5281/zenodo.4469810.

Apresentado o percurso metodológico desta pesquisa, os próximos capítulos apresentam, respectivamente, os resultados obtidos e a análise dos materiais coletados.

6 APRESENTAÇÃO DOS RESULTADOS

Este capítulo apresenta os resultados obtidos nesta pesquisa, que teve o propósito de investigar se os principais financiadores da pesquisa em saúde dispõem de políticas para gestão de dados de pesquisa específicas para pesquisadores envolvidos em colaborações científicas internacionais. As duas seções do capítulo apresentam, respectivamente, uma contextualização das principais características institucionais dos dois financiadores da pesquisa em saúde definidos para o estudo, identificação dos principais marcos normativos das instituições para o campo de dados de pesquisa, seguida de uma breve descrição dos documentos normativos mapeados com políticas para gestão de dados de pesquisa.

6.1 NATIONAL INSTITUTES OF HEALTH (NIH)

O NIH, parte do Departamento de Saúde e Serviços Humanos dos Estados Unidos (*US Department of Health and Human Services – HHS*), é a maior agência norte-americana de pesquisa biomédica, com a missão de “buscar conhecimento fundamental sobre a natureza e o comportamento dos sistemas vivos e a aplicação desse conhecimento para melhorar a saúde, prolongar a vida e reduzir doenças e deficiências” (NATIONAL INSTITUTES OF HEALTH, 2019c, p. I-40). O NIH é formado por 21 institutos e 6 centros⁴⁰ e um escritório central (*Office of the Director – OD*).

Os institutos e centros (ICs) possuem missões distintas que se concentram em agendas particulares de pesquisa, geralmente orientadas a doenças ou sistemas corporais específicos. Cerca de 80% do orçamento do NIH é concedido pelos ICs que dispõem de orçamentos e estratégias de financiamento específicas para atender suas respectivas missões e prioridades de pesquisa. Os ICs financiam variados tipos de subvenções, acordos de cooperação e contratos que apoiam a pesquisa médica (NATIONAL INSTITUTES OF HEALTH, 2019a, *on-line*). Em geral, os subsídios do NIH são concedidos a organizações nacionais ou estrangeiras, públicas ou privadas, com ou sem fins lucrativos (NATIONAL INSTITUTES OF HEALTH, 2019c), embora a maior parte do financiamento seja destinada a organizações de pesquisa nos Estados Unidos (NATIONAL INSTITUTES OF HEALTH, 2021, *on-line*).

Muitos dos ICs possuem programas de colaborações internacionais vinculados em seus sítios eletrônicos, apoiando e financiando pesquisadores dentro e fora dos Estados

⁴⁰ Disponível em: <https://www.nih.gov/institutes-nih/list-nih-institutes-centers-offices>. Acesso em: 20 nov. 2020.

Unidos em programas internacionais (NATIONAL INSTITUTES OF HEALTH, 2018, *on-line*), como o *U.S.–Brazil Collaborative Biomedical Research Program*⁴¹, um programa de pesquisa colaborativa com financiamento conjunto entre institutos do NIH, Ministério da Saúde (MS) e o Ministério da Ciência, Tecnologia e Inovações (MCTIC) do Brasil.

O OD do NIH é o responsável por definir a política da agência e por planejar, coordenar e administrar os programas e demais atividades realizadas pelos ICs (NATIONAL INSTITUTES OF HEALTH, 2019b, *on-line*). O escritório é formado por escritórios específicos⁴² para lidar com questões científicas e administrativas, como o *Office of Extramural Research* (OER⁴³) e o *Office of Science Policy* (OSP⁴⁴) que exercem liderança no desenvolvimento de políticas e/ou normativas para o compartilhamento e gestão de dados.

O OER é um escritório de apoio à comunidade extramural de pesquisa do NIH, ou seja, pesquisadores externos que recebem recursos financeiros da agência. O escritório é responsável por fornecer recursos oficiais para todas as fases do financiamento de pesquisa e por garantir a integridade científica, responsabilidade pública e administração eficaz do portfólio de pesquisa externa do NIH (NATIONAL INSTITUTES OF HEALTH, 2020a, *on-line*). O órgão elabora instrumentos normativos como políticas, orientações, procedimentos e planos de apoio à comunidade receptora de recursos e aos ICs do NIH, de modo a guiar os pesquisadores no processo de obtenção de financiamento e compreensão de políticas e procedimentos federais (NATIONAL INSTITUTES OF HEALTH, 2020a, *on-line*). Dentre os documentos normativos emitidos pelo OER, destaca-se uma série de políticas de compartilhamento e orientações relacionadas a recursos de pesquisas financiadas pelo NIH⁴⁵.

Já o OSP é um escritório de assessoramento do OD voltado para o desenvolvimento de políticas sobre questões que incluem biossegurança, testes genéticos, compartilhamento de dados genômicos, proteção de indivíduos, organização e gestão do NIH e os resultados e valor da pesquisa financiada pela agência. O OSP conta com vários escritórios e programas para lidar com as questões científicas, éticas, jurídicas, sociais e de segurança que afetam a pesquisa biomédica, como a Divisão de Política de Compartilhamento de Dados Científicos⁴⁶, responsável por desenvolver políticas de pesquisa biomédica e recomendações no âmbito de

⁴¹ Disponível em: <https://grants.nih.gov/grants/guide/rfa-files/RFA-AI-18-054.html>. Acesso em: 15 nov. 2020.

⁴² Disponível em: <https://www.nih.gov/about-nih/what-we-do/nih-almanac/office-director-nih>. Acesso em: 15 nov. 2020.

⁴³ Disponível em: <https://grants.nih.gov/aboutoer/intro2oer.htm>. Acesso em: 17 nov. 2020.

⁴⁴ Disponível em: <https://osp.od.nih.gov/about-us/>. Acesso em: 16 nov. 2020.

⁴⁵ Disponível em: <https://grants.nih.gov/policy/sharing.htm>. Acesso em: 15 nov. 2020.

⁴⁶ Disponível em: <https://osp.od.nih.gov/scientific-data-sharing/>. Acesso em: 20 nov. 2020.

dois programas específicos: Genômica e Saúde⁴⁷ e Gestão de Dados de Pesquisa⁴⁸ (NATIONAL INSTITUTES OF HEALTH, [2020?], *on-line*).

O Programa Genômica e Saúde é responsável por analisar as questões científicas, éticas e sociais da pesquisa genética, especialmente as implicações da área genômica na saúde, emitindo recomendações e políticas sobre questões relevantes levantadas na pesquisa genômica e tecnologias emergentes, além de promover por meio da colaboração entre os IC e entre agências a implementação da medicina genômica.

Já o Programa de Gestão de Dados é responsável por fornecer recomendações e documentos deliberativos do NIH sobre questões associadas à gestão e compartilhamento de dados, como a ciência de dados, acesso público, ciência aberta e *big data*. Além disso, o programa endossa os sete princípios para promover o acesso a dados científicos e descobertas de pesquisa apoiadas pelo governo federal por meio da cooperação científica internacional, emitidos no relatório elaborado pela National Science and Technology Council (2016):

1. O progresso científico e a colaboração se beneficiam de um compromisso inicial e contínuo com o estabelecimento, descrição, curadoria, manutenção, validação, descoberta, acessibilidade e distribuição de dados científicos.
2. Os dados científicos devem ser disponibilizados abertamente na medida permitida por lei e sujeitos à privacidade, confidencialidade, segurança e outras restrições apropriadas (por exemplo, reconhecendo interesses proprietários, informações comerciais confidenciais e direitos de propriedade intelectual).
3. Dados científicos com suporte do governo devem estar disponíveis gratuitamente sempre que possível.
4. Os parceiros em atividades cooperativas de ciência e tecnologia internacionais devem estabelecer um plano de gestão de dados no início do projeto que considere o ciclo de vida completo dos dados científicos.
5. As agências federais devem encorajar a interoperabilidade técnica e jurídica para facilitar o compartilhamento internacional de dados científicos apoiados pelo governo, usando formatos compatíveis, publicamente disponíveis e de código aberto.
6. Os dados científicos (e publicações) apoiados pelo governo devem ser disponibilizados o mais cedo possível, com o tempo de divulgação e a duração de qualquer período de uso exclusivo explicitamente definido.
7. As agências federais devem trabalhar com parceiros internacionais de ciência e tecnologia para adotar políticas e padrões de dados que incentivem o compartilhamento aberto de dados para atividades colaborativas (NATIONAL SCIENCE AND TECHNOLOGY COUNCIL, 2016, p. 2).

Os princípios devem ser considerados por agências federais ao emitirem políticas e posicionamentos sobre compartilhamento e acesso aos dados. O governo dos EUA reconhece

⁴⁷ Disponível em: <https://osp.od.nih.gov/scientific-sharing/genomics-and-health/>. Acesso em: 20 nov. 2020.

⁴⁸ Disponível em: <https://osp.od.nih.gov/scientific-sharing/scientific-data-management/>. Acesso em: 20 nov. 2020.

que os dados de pesquisa abertamente acessíveis contribuem para o progresso científico e a colaboração internacional, sendo recursos catalisadores da inovação. Os princípios ratificam o posicionamento do governo em aumentar o acesso aberto aos dados gerados por agências federais ou por recursos públicos, visando promover a cooperação internacional em C&T para enfrentar os desafios globais (NATIONAL SCIENCE AND TECHNOLOGY COUNCIL, 2016).

O OSP também participa do NIH *Data Science Policy Council* (DSPC⁴⁹), conselho responsável por abordar os desafios e oportunidades do *big data* e ciência de dados na pesquisa biomédica, incluindo a administração responsável dos dados, questões bioéticas para os sujeitos da pesquisa e o público no que se refere à gestão, acesso e compartilhamento de dados, visando maximizar os possíveis benefícios públicos obtidos pela utilização de dados.

Dentro da estrutura do OD do NIH cabe ressaltar que há outros escritórios que embora não tenham a função de emitir políticas para a gestão de dados, são responsáveis por coordenar relevantes iniciativas e elaborar documentos para o campo dos dados de pesquisa no NIH, como o *Office of Data Science Strategy* (ODSS⁵⁰) que lidera a implementação do Plano Estratégico do NIH para Ciência de Dados (*Strategic Plan for Data Science*)⁵¹ e o *Office of Strategic Coordination* (OSC⁵²) que coordena programas relacionados aos dados, como o *Big Data to Knowledge* (BD2K)⁵³ e o *New Models of Data Stewardship*⁵⁴, iniciativas voltadas para desenvolver um ecossistema moderno de dados biomédicos.

6.1.1 Marcos normativos das políticas do NIH para gestão de dados de pesquisa

O compromisso do NIH em garantir que os resultados de pesquisas financiadas pela agência sejam compartilhados com a comunidade da pesquisa pode ser rastreado desde a década de 1990⁵⁵, com a emissão da política⁵⁶ para a disseminação de recursos de pesquisa desenvolvidos com fundos federais para serviços de saúde pública. Desde 1996, o NIH exige o compartilhamento de dados em várias áreas, como sequências de DNA, informações de

⁴⁹ Disponível em: <https://osp.od.nih.gov/scientific-sharing/nih-data-science-policy-council/>. Acesso em: 20 nov. 2020

⁵⁰ Disponível em: <https://datascience.nih.gov/about/odss>. Acesso em: 19 nov. 2020.

⁵¹ Disponível em: <https://datascience.nih.gov/strategicplan>. Acesso em: 19 nov. 2020.

⁵² Disponível em: <https://commonfund.nih.gov/>. Acesso em: 21 nov. 2020.

⁵³ Disponível em: <https://commonfund.nih.gov/bd2k>. Acesso em: 21 nov. 2020.

⁵⁴ Disponível em: <https://commonfund.nih.gov/data>. Acesso em: 21 nov. 2020.

⁵⁵ Segundo o National Research Council (2002), os dados de pesquisa foram incorporados nas discussões do governo a partir da emenda Shelby que solicitava ao *Office of Management and Budget* (OMB) alteração da Circular A-110, que estabelecia requisitos para concessões e acordos cooperativos do governo, exigindo que as agências federais garantissem que os dados produzidos com recursos públicos fossem disponibilizados por meio dos procedimentos estabelecidos na Lei de Liberdade de Informação (*Freedom of Information Act* – FOIA).

⁵⁶ Disponível em: <https://grants.nih.gov/grants/guide/notice-files/not96-184.html>. Acesso em: 23 nov. 2020.

mapeamento e coordenadas cristalográficas (NATIONAL INSTITUTES OF HEALTH, 1996). Nesse período, o PCD já era um requisito do NIH para alguns programas, como o projeto Genoma Humano.

No que diz respeito ao compartilhamento de dados gerados em outras áreas do domínio da saúde financiadas pelo NIH, a questão parece ter sido impulsionada a partir de março de 2002, quando o NIH solicitou comentários sobre seu projeto de declaração sobre compartilhamento de dados de pesquisa (NATIONAL INSTITUTES OF HEALTH, 2002). Os requisitos declarados no documento eram aplicados a pesquisadores externos e internos que buscavam financiamento em bolsas, acordos de cooperação e contratos. O documento também reconhecia que os direitos e privacidade dos participantes da pesquisa deveriam ser protegidos em todos os momentos. A declaração estava alinhada às políticas de muitas revistas científicas que publicavam as descobertas de pesquisas apoiadas pelo NIH.

Em decorrência do pedido, o NIH recebeu e analisou uma série de comentários de organizações científicas e do público em geral. Então, em 23 de fevereiro de 2003, o NIH divulgou a “*Final NIH statement on sharing research data*” endossando o compromisso em compartilhar os dados gerados por pesquisas financiadas, a fim de permitir a rápida tradução dos resultados da pesquisa em conhecimento, produtos e procedimentos para melhorar a saúde. A declaração também considerou as mudanças que haviam sido feitas naquele período para os padrões de privacidade de informações de saúde individualmente identificáveis da Regra de Privacidade do HHS (NATIONAL INSTITUTES OF HEALTH, 2003a). O que se seguiu à política geral do NIH para o compartilhamento de dados foi o surgimento gradativo de uma série de políticas com foco em temáticas específicas no âmbito dos institutos, centros, divisões e programas do NIH, totalizando atualmente 17 políticas específicas em vigor⁵⁷.

Importa também sinalizar que desde 2015 o NIH articula-se para expandir sua Política de Compartilhamento de Dados, compromisso firmado no documento “*NIH Public Access Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research*”, publicado em 2015 como resposta ao memorando do OSTP de 2013 (NATIONAL INSTITUTES OF HEALTH, 2015b). A partir de então, o NIH emitiu rascunhos, guias e informações suplementares para um novo projeto de política, seguido de solicitações de comentários públicos à comunidade em geral sobre a gestão de dados e estratégias e prioridades de compartilhamento de dados.

⁵⁷ Disponível em: https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_policies.html. Acesso em: 23 nov. 2020.

Em 29 de outubro de 2020, o NIH emitiu a nova Política do NIH para gestão e compartilhamento de dados, que exigirá que pesquisadores financiados pelo NIH apresentem um plano de gestão e compartilhamento de dados. A nova política endossa a importância de boas práticas de gestão de dados e ratifica a importância de maximizar o compartilhamento de dados de pesquisas financiadas ou conduzidas pelo NIH, considerando as limitações ou exceções justificadas para não fazê-lo (NATIONAL INSTITUTES OF HEALTH, 2020b).

6.1.2 Mapeamento de documentos normativos do NIH com políticas para gestão de dados de pesquisa

A política de compartilhamento do NIH impôs requisitos sobre os dados de pesquisas financiadas ou conduzidas pela agência como parte dos termos e condições para financiamento. Em decorrência, documentos orientadores foram emitidos pela agência para auxiliar os pesquisadores a atenderem à política, abordando questões como elementos-chaves para PCD, repositórios de dados, padrões de dados e compartilhamento de dados genômicos. O quadro 5 abaixo apresenta o arcabouço normativo recuperado para posterior análise.

Quadro 5 – Documentos normativos do NIH com políticas para gestão de dados de pesquisa

| Título do documento | Data de publicação |
|--|---------------------------|
| 1. <i>Final NIH Statement on Sharing Research Data</i> | Fevereiro de 2003 |
| 2. <i>NIH Data Sharing Policy and Implementation Guidance</i> | Março de 2003 |
| 3. <i>NIH Genome-Wide Association Studies (GWAS) Policy (Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS))</i> | Agosto de 2007 |
| 4. <i>Key Elements to Consider in Preparing a Data Sharing Plan Under NIH Extramural Support</i> | Dezembro de 2009 |
| 5. <i>Data Standards and Common Data Elements Resource Guide</i> | Setembro de 2011 |
| 6. <i>Data Repositories Resource Guide</i> | Setembro de 2012 |
| 7. <i>NIH Genomic Data Sharing Policy</i> | Agosto de 2014 |
| 8. <i>Guidance for Investigators in Developing Genomic Data Sharing Plans</i> | Julho de 2015 |
| 9. <i>NIH Grants Policy Statement (NIHGPS)</i> ⁵⁸ | Dezembro de 2019 |

Fonte: Elaboração própria.

O primeiro documento considerado para análise foi a “*Final NIH Statement on Sharing Research Data*”, publicada em 2003 como uma extensão da seção sobre compartilhamento de recursos de pesquisa da Política de Concessões do NIH. Em vigor até 25

⁵⁸ Publicada originalmente em outubro de 1998, a NIHGPS é revisado na virada de cada novo ano fiscal.

de janeiro de 2023, a política aplica-se aos candidatos que buscam recursos para pesquisa acima de US\$ 500 mil dólares em custos diretos em qualquer ano do período do projeto proposto por meio de doações, acordos de cooperação ou contratos (NATIONAL INSTITUTES OF HEALTH, 2003a). De acordo com a política, esses pesquisadores devem apresentar um PCD ou declarar porque o compartilhamento de dados não é possível. A política prevê que específicas chamadas de programas podem solicitar o PCD para solicitações com custos diretos inferiores a US\$ 500 mil dólares em um único ano (NATIONAL INSTITUTES OF HEALTH, 2003a).

O segundo documento recuperado foi o guia “*NIH Data Sharing Policy and Implementation Guidance*”, no qual é fornecida uma série de recomendações para a implementação de sua política de compartilhamento de dados. O guia declara que todos os dados gerados na pesquisa devem ser considerados para o compartilhamento, embora os dados exclusivos sejam particularmente importantes (NATIONAL INSTITUTES OF HEALTH, 2003b, *on-line*). O quadro 6 abaixo mostra algumas diferenciações conceituais para a definição de alguns tipos de dados de gerados nas pesquisas apoiadas pelo NIH.

Quadro 6 – Definições concebidas pelo NIH para alguns tipos de dados de pesquisa

| Termo | Definição |
|------------------------------|--|
| Dados de pesquisa finais | Registros factuais de materiais comumente aceitos na comunidade científica como necessários para validar os resultados da pesquisa. |
| Dados de pesquisa restritos | Conjuntos de dados que não podem ser distribuídos ao público em geral, devido a questões de confidencialidade do participante, licenciamento de terceiros ou acordos de uso ou considerações de segurança nacional. |
| Dados de pesquisa exclusivos | Dados que não podem ser replicados prontamente, como os dados gerados em pesquisas que são muito caras para replicar; em estudos de populações ou momentos únicos, como desastres naturais; em estudos de fenômenos raros, como doenças metabólicas raras. |

Fonte: Adaptado a partir de National Institutes of Health (2003b, *on-line*).

Sobre o compartilhamento de dados pessoais, o documento considera que os dados, antes de compartilhados, devam ser redigidos (redação de dados) para remover todos os identificadores diretos (por exemplo, nome, endereço, números de telefone) e indiretos (informações que podem levar à revelação dedutiva das identidades dos participantes) (NATIONAL INSTITUTES OF HEALTH, 2003b, *on-line*). O guia também aborda métodos para compartilhamento de dados que garantem o controle de acesso por terceiros, conforme detalhado no quadro 7 a seguir.

Quadro 7 – Métodos previstos pelo NIH para o compartilhamento de dados de pesquisa

| Método | Descrição |
|--|--|
| Sob os auspícios do investigador principal | Os investigadores podem enviar um CD com os dados ao solicitante ou postar os dados em seu site institucional ou pessoal. |
| Arquivo de dados | Os investigadores podem compartilhar os dados em uma instalação de arquivo de dados, um local onde dados legíveis por máquina são adquiridos, manipulados, documentados e finalmente distribuídos para a comunidade científica para análise posterior. |
| Enclave de dados | Em virtude de questões de confidencialidade, licenciamento, acordos de uso ou considerações de segurança nacional que proíbem a redistribuição dos dados, os investigadores podem compartilhar dados em um enclave de dados, um ambiente controlado e seguro no qual os pesquisadores qualificados podem realizar análises usando recursos de dados restritos. |
| Modo misto | Os investigadores podem desenvolver um modo misto para compartilhar dados que permita mais de uma versão do conjunto de dados e fornece diferentes níveis de acesso dependendo da versão. |

Fonte: Adaptado a partir de National Institutes of Health (2003b, *on-line*).

O NIH também considera o fornecimento de documentação para cada conjunto de dados compartilhado independentemente do método escolhido, visando garantir que outros possam compreender e usar o conjunto de dados de forma devida. Essa documentação deve fornecer informações a respeito da metodologia e os procedimentos usados para coletar os dados, detalhes sobre códigos, definições de variáveis, localizações de campos variáveis, frequências e semelhantes (NATIONAL INSTITUTES OF HEALTH, 2003b, *on-line*).

O terceiro documento é a “*NIH Genome-Wide Association Studies (GWAS) Policy*”, uma política relativa ao compartilhamento de dados de estudos de associação genômica ampla⁵⁹ gerados em pesquisas financiadas ou apoiadas pelo NIH. A política aplica-se a pesquisas de associação de todo o genoma, utilizando materiais genéticos e dados coletados prospectivamente e retrospectivamente. A política aborda os procedimentos de compartilhamento de dados GWAS, princípios de acesso a dados, propriedade intelectual e questões relacionadas à proteção dos participantes da pesquisa em todas as fases do estudo (NATIONAL INSTITUTES OF HEALTH, 2007). De modo a maximizar a disponibilidade de recursos que facilitem a pesquisa biomédica, a política propôs a criação de um repositório de

⁵⁹ Termo traduzido com base nos DeCS/MeSH. Um GWAS é “qualquer estudo de variação genética em todo o genoma humano que é projetado para identificar associações genéticas com características observáveis (como pressão sanguínea ou peso), ou a presença ou ausência de uma doença ou condição” (NATIONAL INSTITUTES OF HEALTH, 2007, *on-line*).

dados centralizado no NIH, conhecido como *The database of Genotypes and Phenotypes (dbGaP)* para arquivar e distribuir informações básicas sobre GWAS apoiado pelo NIH e conjuntos de dados genótipo-fenótipo disponíveis para estudos secundários. O repositório também aceita conjuntos de dados GWAS fornecidos por outras fontes (NATIONAL INSTITUTES OF HEALTH, 2007).

A política estabelece que todos os pesquisadores que recebem financiamento do NIH devem enviar para o repositório *dbGaP* dados e informações descritivas sobre os estudos, essas últimas em acesso aberto. Para preservar a privacidade dos participantes da pesquisa e a confidencialidade dos dados genômicos, a política prevê a redação de dados em nível individual (genotípicos e fenotípicos) antes de serem disponibilizados (NATIONAL INSTITUTES OF HEALTH, 2007). Nos termos desta política, o acesso aos dados GWAS é controlado. Os Comitês de Acesso aos Dados do NIH (*NIH Data Access Committees – DACs*) são os responsáveis por permitir o acesso aos dados apenas para utilizações em pesquisas que sejam consistentes com o consentimento do participante (NATIONAL INSTITUTES OF HEALTH, 2007).

O quarto documento recuperado para a análise foi o “*Key Elements to Consider in Preparing a Data Sharing Plan Under NIH Extramural Support*”, elaborado para auxiliar candidatos extramurais do NIH na preparação do PCD. De acordo com o documento, os candidatos devem descrever quais e quando dados serão compartilhados; quem terá acesso aos dados; onde serão localizados os dados a serem compartilhados e; como os investigadores irão localizar e acessar os dados (NATIONAL INSTITUTES OF HEALTH, 2009).

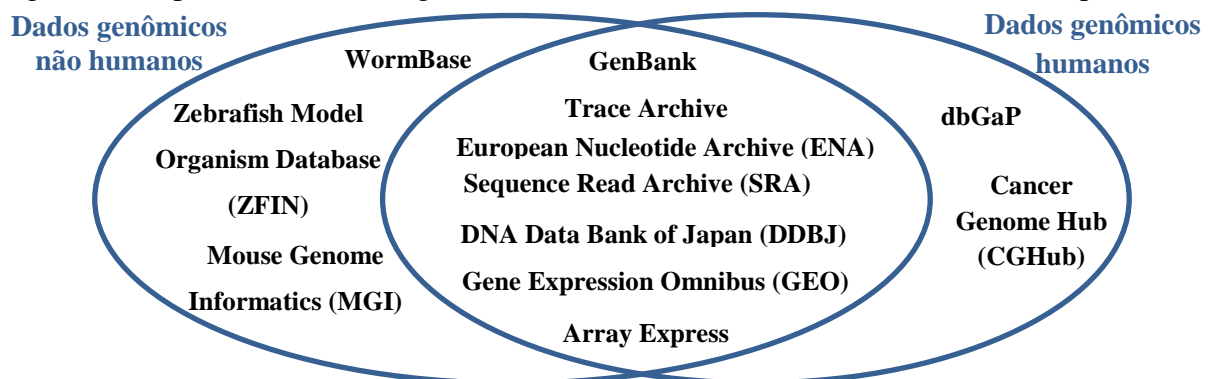
O quinto documento foi o “*Data Standards and Common Data Elements Resource Guide*”, desenvolvido para fornecer suporte à comunidade extramural do NIH na identificação de recursos úteis para padrões de dados e elementos de dados comuns utilizados por programas dos ICs do NIH. O uso de padrões de dados apropriados e elementos de dados comuns garante que os dados possam ser amplamente compartilhados entre os bancos de dados (interoperabilidade) e usados da maneira ideal na pesquisa (NATIONAL INSTITUTES OF HEALTH, 2011). Os padrões e elementos comuns recomendados foram: *Common Data Elements*, do National Institutes of Neurological Disorders and Stroke (NINDS); *CTEP Common Data Elements*, do National Cancer Institute (NCI); *Health Information Technology and Health Data Standards* do National Library of Medicine (NLM); *Enterprise Vocabulary Services (EVS)* do NCI; *Cancer Biomedical Informatics Grid (caBIG) Data Standards*, do NCI e; *National Database for Autism Research (NDAR) – Standards*, do Mental Health Data Archive.

O sexto documento “*Data Repositories Resource Guide*” fornece exemplos de repositórios de dados consistentes com as políticas de compartilhamento que podem ser usados para compartilhar dados gerados no âmbito dos programas financiados pela agência, sendo eles: *dbGaP*, desenvolvido para arquivar, curar e distribuir informações produzidas por estudos que investigam a interação de genótipo e fenótipo; NDAR, repositório de dados de pesquisa sobre o autismo; *The National Institute on Aging Genetics of Alzheimer's Disease Data Storage* (NIAGADS), repositório nacional de dados genotípicos que facilita o acesso de dados genotípicos a investigadores para o estudo da genética da doença de Alzheimer tardia (NATIONAL INSTITUTES OF HEALTH, 2012).

O sétimo documento é a “*Genomic Data Sharing (GDS) Policy*”, publicada como extensão da *GWAS Policy*, de 2007, considera tipos de dados genômicos adicionais. A política estabelece diretrizes que garantam o compartilhamento amplo e responsável de dados genômicos humanos ou não humanos em grande escala gerados em todas as pesquisas na área financiadas pelo NIH, independentemente do nível e mecanismo de financiamento (NATIONAL INSTITUTES OF HEALTH, 2014).

A política prevê que todos os candidatos a pesquisas na área genômica, independentemente da quantidade solicitada, forneçam um Plano de Compartilhamento de Dados Genômicos (PCDG) no momento da aplicação e, posteriormente, enviem um plano mais detalhado ao IC financiador antes da concessão. Para efeitos da política, dados genômicos em grande escala incluem os dados GWAS, matrizes de polimorfismos de nucleotídeos únicos⁶⁰ (*Single Nucleotide Polymorphisms* – SNP), dados de sequência genômica, transcriptômica, epigenômica e de expressão genética. A política também indica repositórios para dados genômicos humanos e não humanos, conforme ilustra figura 1.

Figura 1 – Repositório de dados genômicos humanos e não humanos recomendados pelo NIH



Fonte: Adaptado a partir de National Institutes of Health (2014).

⁶⁰ Termo traduzido com base nos DeCS/MeSH.

O oitavo documento é o “*Guidance for Investigators in Developing Genomic Data Sharing Plans*”, um guia que descreve as informações que deverão ser fornecidas em um PCDG, documento estipulado na política GDS. Os pesquisadores que buscam recursos financeiros em pesquisas que gerem dados genômicos humanos e não humanos em grande escala devem descrever: quais tipos de dados gerados; quais repositórios os dados serão submetidos e, para dados humanos, quais formas de acesso (irrestrito ou controlado); o cronograma de envio e liberação de dados; revisões do comitê de revisão institucional; uso apropriado dos dados; quais mecanismos alternativos para o compartilhamento de dados quando for solicitada a exceção à submissão de dados por não atender aos critérios de Certificação Institucional (NATIONAL INSTITUTES OF HEALTH, 2015a).

Por fim, o último documento selecionado para análise foi a “*NIH Grants Policy Statement*”, uma fonte da política da agência para os termos e condições de financiamento do NIH concedido a organizações nacionais e estrangeiras. O item 8.2 do documento aborda a disponibilidade de resultados de pesquisa, discutidos em consequência em políticas específicas sobre acesso público às publicações científicas, compartilhamento de recursos de pesquisa, propriedade intelectual e produtos de pesquisa provisórios. O documento aborda a Política de Acesso Público do NIH que exige aos beneficiários que tornem as publicações científicas revisados por pares acessíveis ao público pelo PubMed Central (PMC)⁶¹ (NATIONAL INSTITUTES OF HEALTH, 2019c).

No que diz respeito ao compartilhamento de recursos de pesquisa financiada pelo NIH, a política prevê que os investigadores os tornem disponíveis para fins de pesquisa a indivíduos qualificados dentro da comunidade científica. No entanto, o NIH considera os direitos dos destinatários e contratados de eleger e reter o título das invenções desenvolvidas com financiamento federal, conforme estabelecido na Lei Bayh-Dole (NATIONAL INSTITUTES OF HEALTH, 2019c). Sobre o compartilhamento de dados, a política ratifica que os direitos e privacidade dos participantes da pesquisa devem ser protegidos, dessa forma, os dados disponibilizados devem estar livres de identificadores que permitam identificar a identidade dos sujeitos da pesquisa (NATIONAL INSTITUTES OF HEALTH, 2019c).

Cabe destacar que a nova política de gestão e compartilhamento de dados não foi considerada para posterior análise, visto que a mesma apenas se tornará parte dos termos e condições de financiamento do NIH a partir de 25 de janeiro de 2023.

⁶¹ O PMC, mantido pelo NLM do NIH, é um arquivo digital gratuito de literatura de periódicos biomédicos e de ciências biológicas em texto completo e contém links para outras bases de dados de pesquisa, como GenBank e PubChem. Fonte: <https://publicaccess.nih.gov/faq.htm#753>.

6.2 EUROPEAN COMMISSION (EC)

Fundada em 1958, a Comissão Europeia, órgão executivo politicamente independente da EU, é responsável por elaborar as propostas legislativas europeias e por executar as decisões do Parlamento Europeu e do Conselho da EU. A estrutura organizacional da EC é composta por departamentos específicos (conhecidos como direções-gerais), departamentos de serviços e agências executivas, cada um responsável por domínios específicos.

Entre as agências executivas, responsáveis por gerenciar programas instituídos pela Comissão, destaca-se a Agência do Conselho Europeu de Pesquisa (*European Research Council Executive Agency – ERCEA*) destinada à administração dos programas de pesquisa da EU e por implementar a estratégia do Conselho Europeu de Pesquisa (*European Research Council – ERC*), um órgão público de financiamento da pesquisa científica e tecnológica europeia e um componente emblemático nos últimos dois Programas-Quadro de Investigação e Desenvolvimento Tecnológico da EU.

Diante da variedade de financiamento⁶² oferecidos pela EU para a pesquisa e a inovação, optou-se por mapear os documentos normativos no âmbito do Oitavo Programa-Quadro da EC (H2020), pois, conforme apontado na revisão de literatura foi no âmbito do programa que a preocupação acerca da necessidade de uma infraestrutura e políticas para gestão de dados de pesquisa foi ampliada. O H2020 é considerado o principal e maior programa de financiamento de pesquisa e inovação da EU com cerca de 80 bilhões de euros para recursos financeiros distribuídos ao longo de sete anos, compreendido entre 2014 e 2020.

Criado pelo Regulamento (EU) n° 1291/2013 do Parlamento Europeu e do Conselho da EU, o programa H2020 visa promover a cooperação internacional em matéria de ciência, tecnologia e inovação, de modo a contribuir para alcançar os objetivos da Estratégia Europa 2020 no sentido de reforçar a competitividade, contribuir para enfrentar os desafios sociais e apoiar as políticas externas da EU (EUROPEAN UNION, 2013b). Para tanto, o programa tem seu financiamento centralizado em áreas centrais como saúde, alterações demográficas, bem-estar, segurança alimentar, energia, ação climática, entre outras.

No H2020, a cooperação internacional em investigação e inovação é uma prioridade transversal, de acordo com a estratégia da EU (EUROPEAN COMMISSION, 2012a). Isso significa que parceiros internacionais (investigadores individuais, empresas, organizações de pesquisa, universidades, organizações não governamentais, etc.), independentemente do local

⁶² Disponível em: https://ec.europa.eu/info/research-and-innovation/funding/funding-opportunities/funding-programmes-and-open-calls_pt. Acesso em: 13 nov. 2020.

de estabelecimento ou residência, podem participar da maioria das oportunidades de financiamento do programa que envolve projetos colaborativos entre os Estados-membros e países associados⁶³, ações de apoio a investigadores individuais, parcerias público-privadas, entre outras (EUROPEAN COMMISSION, 2014). Os projetos colaborativos são aqueles realizados por consórcios de organizações que trabalham juntas em áreas específicas, com no mínimo três participantes de três diferentes Estados-Membros da EU ou países associados.

6.2.1 Marcos normativos das políticas do H2020 para gestão de dados de pesquisa

Desde sua criação em 2007, o ERC defende e promove o acesso aberto aos resultados de pesquisas financiadas por recursos públicos, incluindo a disponibilização de publicações científicas revisadas por pares e dados de pesquisa primários na internet de forma gratuita. Esse incentivo revestiu-se em uma obrigação formal⁶⁴ para subvenções concedidas pelo ERC ainda no âmbito do Sétimo Programa-Quadro realizado entre 2007 e 2013.

No H2020, a base jurídica para o acesso aberto é prevista no Regulamento (EU) n° 1291/2013. O regulamento estabelece que o acesso aberto aos dados de pesquisa resultantes do financiamento público deve ser promovido, respeitando as restrições acerca da privacidade, da segurança nacional e dos direitos de propriedade intelectual (EUROPEAN UNION, 2013b). Em resposta ao estipulado pelo legislador, durante o primeiro programa de trabalho (2014-2015) do H2020, a Comissão implementou o projeto-piloto para dados de pesquisa abertos (piloto ORD), em áreas selecionadas do programa com vistas a maximizar o acesso e a reutilização dos dados de pesquisa gerados pelos projetos financiados, levando em consideração o equilíbrio entre questões como abertura e proteção de informações científicas, comercialização e direitos de propriedade intelectual, privacidade, segurança, gestão e preservação de dados (EUROPEAN COMMISSION, 2018b).

Na versão revisada do programa de trabalho de 2017, o piloto ORD foi estendido a todas as áreas temáticas do H2020, a fim de tornar os dados de pesquisa abertos como a opção padrão (*'open data as the default'*). A participação no piloto tornou-se, em princípio, o padrão, embora mantenha opções de auto exclusão, realizada por meio de uma alteração no

⁶³ De acordo com os termos do Regulamento n° 1290/2013, país associado é um país terceiro que é parte em um acordo internacional com a União (EUROPEAN UNION, 2013a). Os países associados participam sob as mesmas condições que os Estados-membros da EU. Uma lista dos países associados está disponível em: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/3cpart/h2020-hi-list-ac_en.pdf

⁶⁴ O acesso aberto tornou-se uma obrigação formal por meio da Cláusula Especial n° 39 do documento *FP7 Grant Agreement – List of Special Clauses*. Disponível em: https://ec.europa.eu/research/participants/data/ref/fp7/92570/fp7-lga-clauses_en.pdf. Acesso em: 25 nov. 2020.

acordo de subvenção, em qualquer estágio do ciclo de vida do projeto, desde a fase de aplicação até após a assinatura do acordo de subvenção (EUROPEAN COMMISSION, 2017b). Desta forma, os beneficiários têm o direito de decidir não participar do piloto sem prejuízos na fase de avaliação de propostas, no entanto precisam declarar razões legítimas⁶⁵ para não fornecer acesso aberto aos dados de pesquisa.

O piloto ORD possui como pilares o desenvolvimento do PGD e a disponibilização dos dados de pesquisa em acesso aberto, quando possível (EUROPEAN COMMISSION, 2019). O PGD é um documento obrigatório para os projetos participantes do piloto ORD, no qual especificarão quais dados serão abertos, a menos que optem por sair do piloto, neste caso, a Comissão encoraja os pesquisadores a enviar um PGD voluntariamente, se for relevante para a pesquisa planejada (EUROPEAN COMMISSION, 2017c). O plano deverá detalhar quais dados serão gerados no projeto, se e como eles serão explorados ou disponibilizados para verificação e reutilização e, como eles serão curados e preservados, abordando aspectos relevantes para tornar os dados FAIR (EUROPEAN COMMISSION, 2017b).

A participação no piloto não obriga que todos os dados sejam abertos, ou seja, a participação é flexível, respeitando ao princípio *‘as open as possible, as closed as necessary’* (tão aberto quanto possível, tão fechado quanto necessário), “aberto” a fim de promover a reutilização e acelerar a pesquisa e “fechado” para salvaguardar a privacidade dos sujeitos, interesses comerciais e questões de confiabilidade e segurança.

6.2.2 Mapeamento de documentos normativos do H2020 com políticas para gestão de dados de pesquisa

Para a execução das políticas de dados de pesquisa abertos do H2020, a EC desenvolveu orientações específicas sobre gestão de dados. Conforme apontam Spichtinger e Siren (2018), a principal abordagem da política de dados da EU se refere ao acesso aberto, no entanto a gestão de dados foi incorporada no escopo da política, uma vez que a viabilização do acesso aos dados requer estratégias de gestão adequada, incluindo documentação,

⁶⁵ Algumas das razões são discriminadas no Anexo L do Programa de Trabalho de 2017 do H2020, quais sejam: a) Incompatibilidade com a obrigação do Horizon 2020 de proteger resultados que devem ser explorados comercial ou industrialmente; b) Incompatibilidade com a necessidade de confidencialidade em relação a questões de segurança; c) Incompatibilidade com regras de proteção de dados pessoais; d) Incompatibilidade com o objetivo principal do projeto; e) Se o projeto não irá gerar/coletar quaisquer dados de pesquisa; f) existem outras razões legítimas para não participar (EUROPEAN COMMISSION, 2017c).

descrição de metadados, armazenamento, curadoria e preservação dos dados, entre outras atividades.

Dessa forma, ainda que os documentos normativos recuperados que abordem os dados de pesquisa no âmbito do programa de financiamento europeu H2020 sejam majoritariamente direcionados às questões de compartilhamento e acesso aberto, tais documentos fornecem diretrizes para o gerenciamento dos dados de projetos participantes do piloto ORD. No quadro 8 a seguir são apresentados os documentos normativos recuperados para a análise.

Quadro 8 – Documentos normativos do H2020 com políticas para gestão de dados de pesquisa

| Título do documento | Data de publicação |
|---|---------------------------|
| <i>1. Multi-Beneficiary Model Grant Agreement: ERC Starting Grants, Consolidator Grants, Advanced Grants and Synergy Grants (H2020 ERC MGA — Multi)</i> | Outubro 2017 |
| <i>2. Multi-Beneficiary General Model Grant Agreement (H2020 General MGA – Multi)</i> | Outubro de 2017 |
| <i>3. H2020 Programme AGA – Annotated Model Grant Agreement</i> | Junho 2019 |
| <i>4. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020</i> | Março de 2017 |
| <i>5. Guidelines on FAIR Data Management in Horizon 2020</i> | Julho de 2016 |
| <i>6. Guidelines on Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020</i> | Abril de 2017 |
| <i>7. Open Research Data and Data Management Plans – Information for ERC grantees</i> | Julho de 2019 |
| <i>8. Horizon 2020 projects working on the 2019 coronavirus disease (COVID-19), the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and related topics: Guidelines for open access to publications, data and other research outputs</i> | Abril de 2020 |

Fonte: Elaboração própria (2020).

O Modelo de Acordo de Subvenção do ERC (*H2020 ERC MGA — Multi*) para projetos financiados pelo ERC no H2020 e o Modelo Geral de Acordo de Subvenção do H2020 (*H2020 General MGA – Multi*) estabelecem a base jurídica para o acesso aberto às publicações científicas e aos dados de pesquisa em seus artigos 29.2 e 29.3, respectivamente. O item 29.2 que trata do acesso aberto às publicações prevê que os dados necessários para validar os resultados de pesquisa devem ser depositados concomitantemente às suas publicações científicas revisadas por pares, preferencialmente em um repositório de dados.

Esse requisito não está relacionado à abertura dos dados, mas à gestão deles, visto que os repositórios de dados são um das estratégias para a gestão e preservação dos dados gerados pela pesquisa científica. (EUROPEAN RESEARCH COUNCIL, 2017c; EUROPEAN COMMISSION, 2017d).

A abertura é discutida no item 29.3 que apresenta os requisitos legais para o acesso aberto aos dados de pesquisa especificamente para projetos participantes do piloto ORD, com orientações ligeiramente diferentes entre os termos dos modelos de acordo H2020 do ERC e do H2020. Os documentos são consoantes em estabelecer que os beneficiários do ERC e da EC participantes do piloto ORD devem: a) realizar o depósito de dados digitais em um repositório de dados de pesquisa, adotando medidas que permitam, de forma rápida e gratuita a terceiros, o acesso, mineração e exploração de (i) dados e metadados necessários para validar os resultados apresentados em publicações científicas e (ii) outros dados e metadados associados conforme especificados no PGD; b) fornecer informações sobre as ferramentas e instrumentos disponíveis através do repositório escolhido, para que torne exequível a validação dos resultados da pesquisa e, quando possível, os mesmos deverão ser disponibilizados (EUROPEAN RESEARCH COUNCIL, 2017c; EUROPEAN COMMISSION, 2017d).

No entanto, o *H2020 General MGA – Multi* apresenta disposições específicas para o contexto da pesquisa em saúde. No item 29.3 o documento acrescenta que caso seja previamente estabelecido no programa de trabalho, os beneficiários envolvidos com ações de saúde que participam do piloto ORD devem depositar dados de pesquisa em repositórios de dados e fornecer em acesso aberto aos dados relevantes para atender uma emergência de saúde pública, quando solicitado pela Comissão e dentro do prazo definido. A exceção acontece quando a Comissão substitui a obrigação de acesso aberto por direitos de acesso especial para terceiros que precisam dos dados para atender à emergência de saúde pública, o que inclui o direito de acesso, mineração, exploração e reprodução dos dados gratuitamente (EUROPEAN COMMISSION, 2017d). Em situação de emergência em saúde pública, pesquisadores que não são participantes do piloto ORD podem ser solicitados a depositar dados de pesquisa digital em um repositório de dados ou conceder direitos de acesso específicos a terceiros que precisam dos dados para lidar com o cenário emergente (EUROPEAN COMMISSION, 2017d).

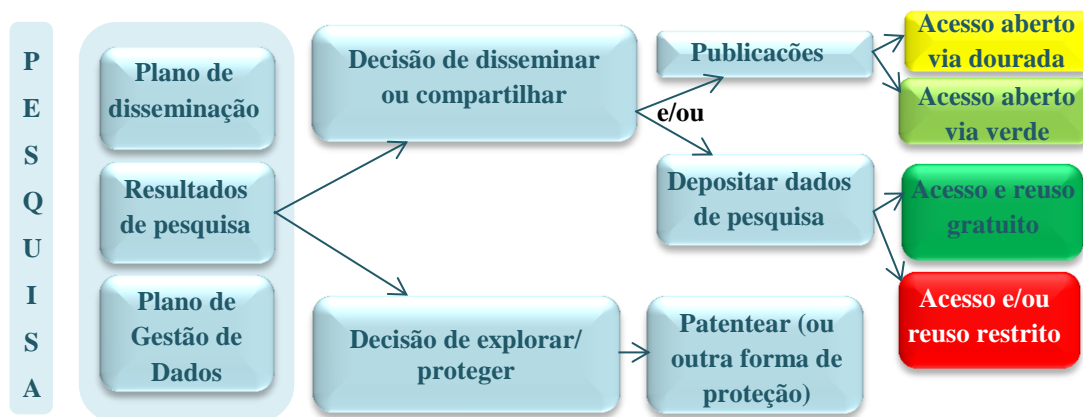
Os dois documentos assinalam que a participação do piloto não obriga que todos os dados sejam abertos. Apesar das exigências descritas acima, os beneficiários que participam do piloto ORD não devem fornecer acesso aberto a conjuntos de dados específicos se isso for

contra as obrigações de: i) proteger os resultados (incluindo pedidos de patente); ii) garantir a confidencialidade; iii) garantir a segurança e; iv) proteger dados pessoais. Os pesquisadores não têm de garantir acesso aberto a partes específicas dos dados de pesquisa se a realização do objetivo principal da pesquisa (para projetos H2020) ou se a realização de um dos objetivos da pesquisa (para projetos ERC financiados no âmbito do H2020) for prejudicada ao tornar os dados abertamente acessíveis (EUROPEAN RESEARCH COUNCIL, 2017c; EUROPEAN COMMISSION, 2017d). Neste caso, tanto o ERC como a EC prevê que os pesquisadores expliquem os motivos para não abrir os dados no PGD (EUROPEAN COMMISSION, 2019).

O terceiro documento recuperado foi “*H2020 Programme AGA*”, um guia do usuário de caráter informativo que visa explicar aos candidatos e aos beneficiários o modelo geral de contrato de subvenção do H2020. O documento fornece políticas para o compartilhamento de dados de pesquisa em contexto de emergência em saúde pública. No contexto de ações H2020 voltadas para emergências em saúde pública, o PGD é um documento obrigatório que deve ser apresentado como anexo à proposta ou, o mais tardar, antes da assinatura de concessão de financiamento (EUROPEAN COMMISSION, 2019).

O quarto documento foi o “*Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020*” que dispõe diretrizes sobre o acesso aberto a publicações científicas e dados a serem seguidas por pesquisadores com projetos financiados pelo H2020. Define acesso aberto aos dados de pesquisa como o direito de acessar e reutilizar dados de pesquisa digital (EUROPEAN COMMISSION, 2017a). O documento apresenta o fluxo para o acesso aberto a publicações científicas e aos dados no contexto mais amplo de divulgação e exploração, conforme mostra a figura 2 abaixo.

Figura 2 – Fluxos para o acesso aberto a publicações científicas e aos dados de pesquisa no contexto mais amplo de divulgação e exploração



Fonte: Adaptado a partir de European Commission (2017a, p. 4).

A ilustração mostra que os requisitos de acesso aberto não implicam na obrigação de publicar os resultados da pesquisa, sendo assim não impede que os pesquisadores explorem os resultados comercialmente, por exemplo, através da patente. No entanto, se os pesquisadores optarem pela publicação como o meio de disseminação ou compartilhamento, o depósito das publicações científicas em um repositório e o acesso aberto a essas publicações torna-se uma obrigação. Nesse contexto, a Comissão orienta fortemente que os pesquisadores depositem, ao mesmo tempo, os dados de pesquisa necessários para validar os resultados apresentados nas publicações científicas depositadas, visando garantir a adequada gestão de dados (EUROPEAN COMMISSION, 2017d; EUROPEAN COMMISSION, 2017a).

Já o acesso aberto aos dados de pesquisa não é uma obrigação no âmbito do programa H2020, no entanto está sendo ampliado pelo piloto ORD, conforme visto anteriormente. O piloto ORD aplica-se, principalmente: a) aos dados subjacentes (os dados necessários para validar os resultados apresentados em publicações científicas), incluindo os metadados associados (ou seja, metadados que descrevem os dados de pesquisa depositados) e; b) quaisquer outros dados (por exemplo, dados selecionados não diretamente atribuíveis a uma publicação, ou dados brutos), incluindo os metadados associados, conforme especificado e dentro dos prazos estabelecidos no PGD, ou seja, de acordo com o julgamento individual de cada projeto/beneficiário (EUROPEAN COMMISSION, 2017a).

O documento não aborda a gestão de dados, no entanto, orienta que participantes do piloto ORD considerem as diretrizes publicadas pela EC no documento “*Guidelines on FAIR Data Management in Horizon 2020*”. Dessa forma, o quinto documento considerado para análise fornece um conjunto de orientações para os beneficiários do H2020 para tornar os dados de pesquisa FAIR, a fim de garantir uma gestão efetiva dos dados (EUROPEAN COMMISSION, 2016). De acordo com European Commission (2016), os participantes do piloto ORD devem elaborar o PGD com descrições acerca do ciclo de vida da gestão de dados gerados por um projeto financiado. Com a finalidade de torná-los FAIR, os pesquisadores devem informar no documento: a) o tratamento dos dados de pesquisa durante e após o final do projeto; b) quais dados serão coletados, processados e/ou gerados; c) qual metodologia e padrões serão aplicados; d) se os dados serão compartilhados/tornados em acesso aberto e; e) como os dados serão curados e preservados (inclusive após o final do projeto).

O documento fornece um novo modelo de PGD (ANEXO A) inspirado no conceito FAIR, cuja utilização é recomendada, mas voluntária. O modelo é estruturado em um conjunto de perguntas que deverá ser respondido de acordo com o grau de detalhamento exigido pelo projeto, no entanto, não é necessário que todas as respostas sejam detalhadas na

primeira versão do PGD que tem de ser entregue no primeiro semestre de implementação do projeto (EUROPEAN COMMISSION, 2016). A Comissão considera o PGD como um documento vivo, deste modo, estabelece que o documento seja atualizado conforme o desenvolvimento do projeto e ocorrência de mudanças significativas. A atualização deve ser feita no mínimo no tempo da avaliação periódica do projeto (EUROPEAN COMMISSION, 2016).

Como suporte complementar ao desenvolvimento do PGD, a Comissão orienta que os pesquisadores utilizem os padrões de metadados desenvolvidos pelo *Metadata Standards Directory Working Group* (MSDWG), do *Research Data Alliance* (RDA)⁶⁶. O RDA fornece um diretório de padrões de metadados para dados de pesquisa que pode ser pesquisado em busca de padrões específicos de disciplinas e ferramentas associadas.

Para o depósito de publicações e dados, a Comissão orienta o uso do registro de repositório *Re3data*⁶⁷, recurso de informação para identificar o repositório de dados de pesquisa mais apropriado de acordo com a disciplina acadêmica, mantido pelo DataCite; o *Zenodo*⁶⁸, repositório aberto para a preservação e compartilhamento de publicações, dados e *softwares* que não fazem parte dos repositórios institucionais ou temáticos de comunidades de pesquisa, atualmente mantido pelo Open Access Infrastructure for Research in Europe (OpenAIRE)⁶⁹ e pelo European Organization for Nuclear Research (CERN) e; o *ScienceMatters*⁷⁰, uma plataforma de publicação científica que divulga observações únicas e validadas junto aos dados subjacentes às publicações. No que se refere à seleção de licença adequada para o reuso dos dados da pesquisa, a Comissão orienta o uso do assistente de licenças do serviço EUDAT B2SHARE⁷¹.

No entanto, as diretrizes dispostas nos documentos acima não são aplicadas em toda sua extensão a ações financiadas pelo ERC. Portanto, para a implementação do acesso aberto

⁶⁶ O RDA é uma iniciativa dirigida pela comunidade da pesquisa iniciada em 2013 pela EC, pela NSF e pelo National Institute of Standards and Technology (NIST) do governo americano e pelo Australian Government's Department of Innovation com o propósito de construir uma infraestrutura social e técnica para permitir o compartilhamento aberto e a reutilização de dados (RESEARCH DATA ALLIANCE, 2016).

⁶⁷ Disponível em: <https://www.re3data.org/>. Acesso em: 19 set. 2020.

⁶⁸ Disponível em: <https://zenodo.org/>. Acesso em: 19 set. 2020.

⁶⁹ O OpenAIRE é uma iniciativa financiada pela EU que visa apoiar a implementação das políticas de acesso aberto da EC. A infraestrutura técnica OpenAIRE atua como um agregador de metadados que descrevem objetos do ciclo de vida da pesquisa que são depositados em repositórios institucionais ou temáticos (por exemplo, ArXiv e o Europe PMC), em sistema de informação de editoras e periódicos de acesso aberto, em repositórios híbridos (por exemplo, Zenodo), em registros de entidades (por exemplo, OpenDOAR, *Re3data* e CORDA), entre outros tipos de fontes de dados (OPENAIRE, 2020, *on-line*).

⁷⁰ Disponível em: <https://www.sciencematters.io/>. Acesso em: 19 set. 2020.

⁷¹ O EUDAT B2SHARE é um serviço baseado na web para armazenamento e publicação de conjuntos de dados, destinado a cientistas europeus. Disponível em: <https://eudat.eu/services/userdoc/b2share-usage>. Acesso em: 19 set. 2020.

e do piloto ORD no âmbito do H2020, o ERC divulgou o documento “*Guidelines on Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020*”, no qual explicita que é dever dos beneficiários das subvenções do ERC concedidas no âmbito do H2020 a garantia de acesso aberto, gratuito e em linha para qualquer usuário a todas as publicações científicas revisadas por pares relacionadas aos seus resultados das pesquisas financiadas como também a adoção de melhores práticas para preservação dos arquivos de dados de pesquisa produzidos ao longo da investigação de modo a viabilizar o compartilhamento dos mesmos com outros pesquisadores quando não vinculados às restrições de direitos autorais, requisitos de confidencialidade ou cláusulas contratuais (EUROPEAN RESEARCH COUNCIL, 2017b).

O documento aborda a obrigatoriedade de um PGD para projetos ERC que participam do piloto ORD que detalhe questões para tornar os dados FAIR e outras questões para alocação de recursos e segurança de dados. Para tanto, a Agência Executiva do ERC emitiu diretrizes de implementação e um modelo de PGD (ANEXO B).

No que se refere ao repositório adequado para o depósito de publicações científicas e dados subjacentes às publicações (necessários para validar os resultados), o ERC recomenda a utilização de repositórios temáticos, quando existirem, como o *Europe PubMed Central*⁷² para publicações no domínio das ciências da vida, o *arXiv*⁷³ para publicações de Ciências Físicas e de Engenharia, o *OAPEN library*⁷⁴ para o depósito de monografias, capítulos de livros e outras publicações de texto longo e o *Zenodo* que pode ser usado na inexistência de repositórios institucionais ou temáticos da comunidade de pesquisa a qual o beneficiário faz parte (EUROPEAN RESEARCH COUNCIL, 2017b).

Para apoiar os pesquisadores com projetos financiados pelo ERC nos esforços em adotar boas práticas de gestão de dados de acordo com as particularidades de seus respectivos campos disciplinares, com base no trabalho do seu Grupo de Trabalho sobre Acesso Aberto, Gestão de Dados de Investigação e Ciência Aberta, o Conselho Científico do ERC publicou o documento “*Open Research Data and Data Management Plans – Information for ERC grantees*”. O sétimo documento considerado para a análise aborda requisitos para elaboração de PGD, depósito de dados, repositórios gerais para dados, preparação de metadados e dados, além de apresentar informações específicas para os dados de pesquisa abertos e deposição de dados nos domínios das Ciências da Vida, Ciências Sociais e Humanas e Ciências Físicas e

⁷² Disponível em: <http://europepmc.org/About>. Acesso em: 20 set. 2020.

⁷³ Disponível em: <https://arxiv.org/>. Acesso em: 20 set. 2020.

⁷⁴ Disponível em: <https://www.oapen.org/>. Acesso em: 20 set. 2020.

Engenharia. Diante da amplitude e variedade de práticas e padrões com relação à gestão, armazenamento e compartilhamento de dados entre as disciplinas, o ERC estabelece um conjunto de requisitos comuns entre os campos para a elaboração dos PGD de modo a atingir aos princípios FAIR, quais sejam, descrição do conjunto de dados, padrões de metadados, nome e identificador persistente para os conjuntos de dados, metodologia de curadoria, preservação e compartilhamento de dados (EUROPEAN RESEARCH COUNCIL, 2019).

No contexto da pesquisa em saúde, a Comissão emitiu um conjunto de diretrizes específicas para o acesso aberto às publicações, dados e outros resultados de pesquisas que trabalham com o tema da COVID-19, SARS-CoV-2 e tópicos relacionados, no âmbito dos projetos financiados pelo programa H2020, como propósito torná-los possivelmente usados no avanço da pesquisa e no enfrentamento à pandemia. Desse modo, o oitavo documento considerado para a análise “*Horizon 2020 projects working on the 2019 coronavirus disease (COVID-19), the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and related topics: Guidelines for open access to publications, data and other research outputs*” excede os requisitos de acesso aberto e obrigações legais consagradas no Acordo de Subvenção do H2020, a fim de abordar a atual situação de emergência em saúde pública.

O documento apresenta o cenário geral de esforços relacionados ao compartilhamento de dados no contexto da COVID-19, apoiadas pela *European Open Science Cloud (EOSC)*⁷⁵, como as iniciativas *EMBL-EBI COVID-19 Research Data*, *RDA Working Group on COVID-19*, *OpenAIRE’s Zenodo COVID-19 community*, *OpenAIRE’s COVID-19 Gateway*, *GOFAIR’s Virus Outbreak Data Network (VODAN)* e *BIP! Finder for COVID-19*.

O documento aborda também os princípios FAIR, diretrizes para acesso aberto a publicações, aos dados e outros produtos de pesquisa, PGD, infraestruturas de pesquisa europeia, recursos úteis como diretrizes do Conselho Europeu de Proteção de Dados (*European Data Protection Board – EDPB*), publicação e recursos de dados, padrões, repositórios e catálogos de dados, ferramentas para criação de PGD e outros serviços de suporte (EUROPEAN COMMISSION, 2020).

O conjunto de políticas para gestão de dados de pesquisa identificadas nos documentos apresentados nesta seção será analisado a seguir.

⁷⁵ A EOSC é uma iniciativa da EC que visa dar à União uma liderança global na gestão de dados de pesquisa e garantir que os cientistas europeus colham todos os benefícios da ciência baseada em dados (EUROPEAN COMMISSION, 2018a).

7 ANÁLISE DOS RESULTADOS

Nesta seção faz-se uma análise dos resultados obtidos no estudo a partir da avaliação das políticas para gestão de dados de pesquisa em saúde materializadas nos documentos normativos, a partir do modelo de avaliação que contempla as categorias grau de aplicação e cobertura de critérios. Em seguida, apresenta algumas considerações acerca dessas políticas para o contexto da colaboração científica internacional que foram observadas a partir da análise dos critérios estipulados.

7.1 ANÁLISE DAS POLÍTICAS PARA GESTÃO DE DADOS DE PESQUISA DO NIH

Foram analisados nove documentos normativos do NIH com políticas expressas para gestão de dados de pesquisa, sendo seis de caráter generalista e três de caráter mais específico na área genômica. A respeito dessa documentação analisada, importa salientar que grande parte da atenção da agência foi orientada ao aspecto do compartilhamento. A avaliação geral do conjunto das políticas é sintetizada no quadro 9 e discutida na sequência.

Quadro 9 – Avaliação geral das políticas para gestão de dados de pesquisa do NIH

| CATEGORIAS | | | | | | | | | | | | | | | | | |
|-------------------|------------|------------------|------------------------|--------|-------------------|---|-------------------------------|---------|---------------------------------------|----------------------|---------------------------------|--------------------------|-----------------|-------------|-------------------------|-------|-------|
| Grau de aplicação | | | Cobertura de critérios | | | | | | | | | | | | | | |
| Recomendatório | Mandatário | Não identificado | Aspectos gerais | | | | Planejamento e infraestrutura | | | | | Disponibilização e reuso | | | | | |
| | | | Definição | Escopo | Responsabilidades | Monitoramento de conformidade e sanções | Plano de Gestão de Dados | Suporte | Curadoria/Preservação por longo prazo | Padrões de metadados | Repositórios e centros de dados | Custo | Limite de tempo | Propriedade | Compartilhamento/Acesso | Reuso | Ética |
| | | | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |

Fonte: Elaboração própria a partir da análise dos documentos normativos.

No que diz respeito à análise da primeira categoria “grau de aplicação” verificou-se que os documentos apresentavam tanto políticas recomendatórias como mandatórias. O quadro 10 a seguir reúne políticas que eram consoantes entre documentos e seus respectivos graus de aplicação identificados.

Quadro 10 – Grau de aplicação das principais políticas para gestão de dados no NIH

| Políticas identificadas | Grau de aplicação |
|---|-------------------------------|
| Abordar o compartilhamento de dados no momento da inscrição | Mandatório |
| Elaborar PCD | Mandatório, salvo exceção |
| Abordar o compartilhamento de dados no documento de consentimento livre e esclarecido | Mandatório |
| Compartilhar os dados de pesquisa | Mandatório, salvo exceção |
| Redigir, desidentificar e anonimizar dados sensíveis antes do compartilhamento | Mandatório |
| Permitir acesso aos dados de pesquisa | Recomendatório, salvo exceção |

Fonte: Elaboração própria a partir da análise dos documentos normativos.

A política para abordar o compartilhamento de dados no momento da inscrição é mandatória para pesquisadores que concorrem a uma concessão ou enviam uma solicitação de financiamento ao NIH acima do valor de US\$500 mil dólares por ano, embora a decisão de compartilhá-los ou não caiba aos candidatos que deverão discriminar as razões para quando optam não fazê-lo. Nessa perspectiva, pondera-se que os projetos com investimentos menores, por via de regra, não precisam considerar o compartilhamento de dados, o que pode provocar certa invisibilidade a esses recursos de pesquisa e duplicação de esforços por terceiros para coleta de dados semelhantes. No entanto, cabe sinalizar que em documentos específicos da área genômica verificou-se que a política de considerar o compartilhamento de dados ainda no momento da inscrição é mandatória para todas as pesquisas que pretendem gerar dados, sem limitação de recursos financeiros solicitados.

Em seguida, observou-se que o PCD é um documento obrigatório apenas para os pesquisadores que consideram compartilhar os dados no momento da inscrição, ou seja, àqueles que declaram não ser possível o compartilhamento ficam desobrigados de atender à política de elaboração do documento. Tal política também demonstrou ser compulsória para pesquisas que geram dados genômicos humanos e não humanos.

A natureza sensível dos dados gerados nas pesquisas biomédicas destaca a importância do processo de consentimento do sujeito da pesquisa para uso dos dados tanto em estudos prospectivos como retrospectivos. Identificou-se como política mandatória no NIH que durante o processo de consentimento livre e esclarecido os pesquisadores informem aos participantes da pesquisa sobre o subsequente compartilhamento dos dados obtidos em repositórios de dados para fins de pesquisas futuras. Importa destacar que o NIH delega ao comitê de ética em pesquisa da instituição requerente a responsabilidade por determinar se um estudo é apropriado para ser submetido ao repositório de dados.

Observou-se que o compartilhamento de dados de pesquisa é uma política mandatória no NIH, que deve ser atendida de tal maneira que preserve a privacidade dos sujeitos e proteja a confidencialidade de seus dados. A exceção à política acontece quando os pesquisadores, ainda no momento da inscrição declaram não ser possível fazê-lo, conforme escrito anteriormente.

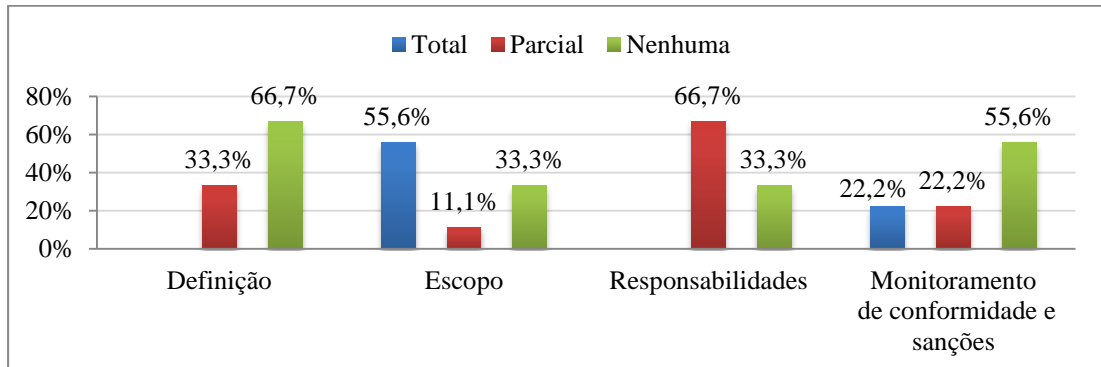
Nos documentos analisados, verificou-se que a política de redigir, “desidentificar” e “anonimizar” dados pessoais sensíveis para reduzir o risco de identificação dos sujeitos da pesquisa mostrou-se ser compulsória para pesquisas que se propõem à disponibilização de dados em ambientes recomendados pelo NIH. É importante salientar que a recomendação do uso de repositório de dados para o compartilhamento foi identificada em política para dados GWAS em 2007 e para outros tipos de dados, a indicação aparece em documentos de caráter recomendatório a partir de 2009. Nos documentos emitidos em 2003, que ainda vigoram nos termos e condições atuais para financiamento no NIH, os ambientes recomendados para compartilhamento de dados são os arquivos e enclaves de dados e *site* institucional ou pessoal, o que pode denotar certa descontextualização com as práticas atuais da ciência aberta e colaborativa.

A respeito da política de acesso aos dados de pesquisa, verificou-se que a mesma possui caráter recomendatório visando maximizar os benefícios do compartilhamento. No que diz respeito à questão do acesso, observa-se considerável ênfase dada à implementação de estratégias de proteção consistentes para salvaguardar a privacidade dos sujeitos da pesquisa por meio de níveis de acesso aos dados compartilhados em repositórios ou outros centros de dados, de modo a mitigar as chances de divulgação de informações que possam levar à revelação dedutiva das identidades dos participantes. Importa destacar que acesso público aos dados de pesquisa pode ser tornar mandatório para solicitações por meio da FOIA.

Dando prosseguimento, a segunda categoria analisada foi “cobertura de critérios”, na qual foi possível verificar o tipo de cobertura (total, parcial ou nenhuma) dada aos critérios contemplados nas seguintes subcategorias: a) Aspectos gerais: definição, escopo, responsabilidades, monitoramento de conformidades e sanções; b) Planejamento e Infraestrutura: PGD, suporte, curadoria/preservação de longo prazo, padrões de metadados, repositórios e centro de dados e custo; c) Disponibilização e reuso: limite de tempo, propriedade, compartilhamento/acesso, reuso e ética. Cabe lembrar que na avaliação final do conjunto de documentos analisados optou-se por considerar a cobertura mais completa (total ou parcial) identificada em cada critério.

A cobertura dos critérios contemplados na subcategoria “aspectos gerais” por documento é ilustrativamente demonstrada no gráfico 1.

Gráfico 1 – Porcentagem de cobertura da subcategoria “aspectos gerais” por critério (NIH)



Fonte: Elaboração própria a partir da análise dos documentos normativos.

Conforme demonstra a ilustração acima, os documentos normativos com políticas mandatórias e recomendatórias para gestão de dados de pesquisa do NIH oferecem cobertura total para os seguintes critérios:

- a) Escopo: os documentos mencionam que todos os dados gerados pela pesquisa financiada pela agência devem ser considerados para o compartilhamento quando possível, principalmente aqueles que não são facilmente replicados. Também foi possível identificar que as políticas se aplicam a pesquisas básicas, estudos clínicos e outros tipos de pesquisa que envolve seres humanos ou não, conforme a quantia de financiamento fornecido ou área da pesquisa.
- b) Monitoramento de conformidade e sanções: verificou-se que a conformidade com as políticas e regulamentos para os dados de pesquisa são monitoradas por meio dos programas dos ICs e o não cumprimento dos termos que regem o financiamento para pesquisa do NIH pode resultar na retenção de financiamento. Os documentos também abordam a necessidade de relatórios que descrevam os progressos no planejamento para o compartilhamento dos dados.

Dando prosseguimento à análise da subcategoria “aspectos gerais”, verificou-se que os documentos abordaram parcialmente os seguintes critérios:

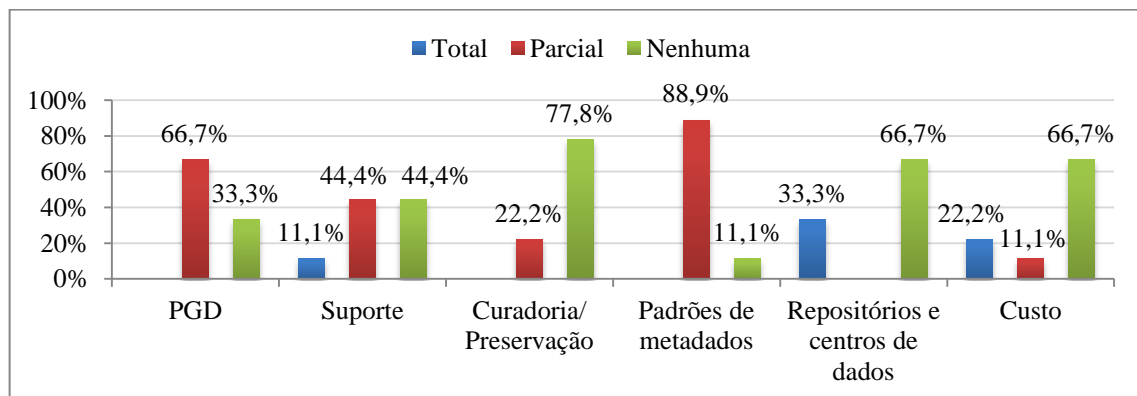
- a) Definição: os documentos analisados fornecem definições gerais sobre o que são “dados de pesquisa”, apresentando exemplos do que podem ser considerados ou não dados da pesquisa em estudos clínicos e pesquisas associadas ao genoma. No

entanto, durante a análise dos documentos não foi possível identificar considerações acerca dos dados de pesquisa abertos ou dados FAIR, conceitos centrais para promover a reprodutibilidade da pesquisa.

- b) Responsabilidades: os documentos não discriminam claramente em seção específica a distribuição de responsabilidades com a gestão de dados entre os envolvidos na pesquisa. No entanto, é possível observar que a maior parte das responsabilidades destinam-se aos pesquisadores e suas instituições que devem atender aos requisitos para o compartilhamento dos dados de acordo com as leis, regulamentos, regras e políticas aplicáveis no NIH. As responsabilidades do NIH concentram-se nas atividades de supervisão de implementação da política e fornecimento de serviços de dados.

Avançando na análise, o gráfico 2 mostra a porcentagem de cobertura dos critérios contemplados na subcategoria “planejamento e infraestrutura”.

Gráfico 2 – Porcentagem de cobertura da subcategoria “planejamento e infraestrutura” por critério (NIH)



Fonte: Elaboração própria a partir da análise dos documentos normativos.

Tendo como referência o gráfico acima, pode-se observar que os seguintes critérios foram abordados totalmente:

- a) Suporte: os documentos apontam para infraestruturas tecnológicas, padrões de dados, ontologias, entre outros recursos desenvolvidos dentro de programas dos ICs que compõem o NIH, visando garantir a interoperabilidade entre os bancos de dados desenvolvidos para o arquivamento e compartilhamento de dados.
- b) Repositórios e centros de dados: os documentos analisados evidenciam a preocupação do NIH em proporcionar um ambiente seguro e padronizado que

garantam a qualidade dos dados por longo prazo, permitindo o acesso e reuso dos dados em pesquisas futuras. Indicações de repositórios de dados são fornecidas, embora o NIH reconheça a diversidade desses ambientes de acordo com a disciplina e o propósito da pesquisa, deixando a critério do pesquisador a escolha do repositório mais apropriado.

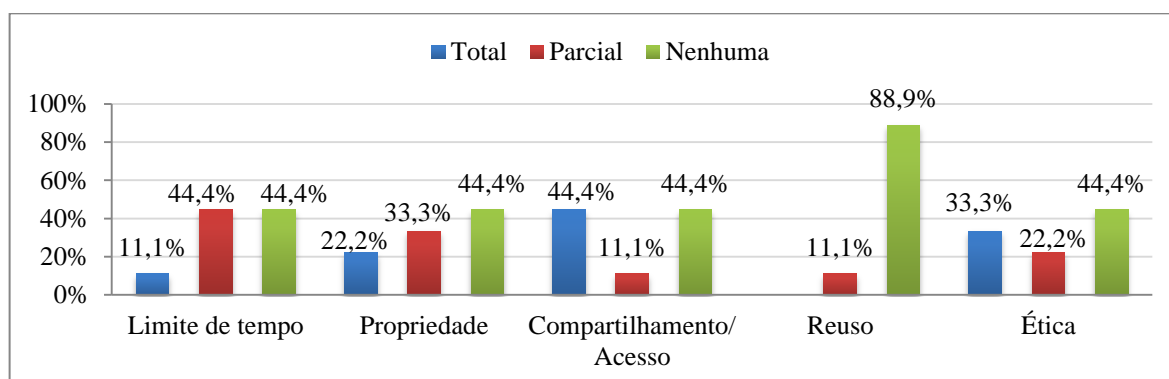
- c) Custo: solicitações para recursos financeiros para o tratamento e compartilhamento dos dados de pesquisa são elegíveis para pedidos de concessão no NIH.

No segundo momento da análise da subcategoria “planejamento e infraestrutura”, verificou-se que os documentos com políticas para a gestão de dados de pesquisa abordaram parcialmente os critérios a seguir:

- a) Plano de Gestão de Dados: O documento PGD não foi identificado como um requisito para financiamento em nenhum dos documentos analisados. Em seu lugar, é solicitado o PCD que enfatiza os aspectos da disponibilização e reutilização dos dados. Cabe considerar que apesar da expressiva ênfase no compartilhamento, percebe-se que o NIH também reconhece a necessidade de elaboração, durante o planejamento para o compartilhamento, da documentação dos dados (metodologia e os procedimentos usados para coletar os dados), uma prática determinante para a qualidade e preservação dos dados.
- b) Curadoria/Preservação por longo prazo: os documentos analisados carecem de políticas a respeito da curadoria e preservação dos dados de pesquisa por longo prazo. Tal ausência pode provocar o depósito de dados em ambientes descompromissados com a longevidade dos dados para reuso futuro. Por outro lado, os documentos com orientações para área de genômica abordam brevemente as vantagens de se dispor de um repositório central no NIH, pois permite o controle de padrões e qualidade de dados genotípicos e fenotípicos incluídos e garante o compromisso com a manutenção dos dados após a conclusão do estudo.
- c) Padrões de metadados: não foram identificadas políticas para o uso de padrões de metadados específicos para a descrição dos dados, no entanto, o NIH mostra-se preocupado com a interoperabilidade entre bancos de dados, fornecendo uma lista de padrões de dados que permitem a comunicação entre sistemas.

Dando sequência à análise das subcategorias, o gráfico 3 ilustra a porcentagem de cobertura dos critérios contemplados na subcategoria “Disponibilização e Reuso”.

Gráfico 3 – Porcentagem de cobertura da subcategoria “disponibilização e reuso” por critério (NIH)



Fonte: Elaboração própria a partir da análise dos documentos normativos.

Diante dos resultados retratados no gráfico acima, observa-se que os critérios abordados totalmente foram:

- a) Limite de tempo: há declarações para que o compartilhamento de dados não seja realizado após a aceitação para publicação das principais descobertas do conjunto de dados final, embora a disponibilidade anterior possa ser exigida aos dados genômicos. O NIH ainda prevê período de embargo de 30 a 60 dias para as instituições exercerem seus direitos de propriedade intelectual. O financiador em questão ainda espera que os dados sejam mantidos por um período de três anos após o encerramento de uma concessão ou acordo de contrato.
- b) Propriedade: os documentos informam que os beneficiários detêm os direitos sobre os dados resultantes de um projeto apoiado por uma subvenção, de tal modo que a titularidade de quaisquer invenções ou descobertas desenvolvidas ao abrigo dos respectivos acordos pode ser exercida conforme a Lei Bayh-Dole e as políticas da agência, exceto as sequências de DNA de ocorrência natural que não podem ser patenteadas nos Estados Unidos.
- c) Compartilhamento/Acesso: os documentos abordam múltiplas formas para o compartilhamento de dados, desde o envio de CD com os dados ao solicitante até o depósito em repositórios de dados. Os dados coletados ou gerados a partir de pesquisas envolvendo sujeitos humanos devem ser compartilhados por meio de estratégias eficazes para minimizar os riscos de divulgação não autorizada de identificadores pessoais. O compartilhamento não é condição para o acesso aos dados, que pode ser irrestrito ou controlado. Importa ressaltar que o documento de consentimento informado é um importante instrumento no qual a instituição

candidata determinará a adequação do depósito de dados em repositórios designados pelo NIH e do nível de acesso em que os dados serão disponibilizados.

- d) Ética: os documentos abordam o uso e reuso de dados pessoais, ratificando que a garantia dos direitos e privacidade dos indivíduos deve ser considerada e acordada no âmbito do comitê de ética em pesquisa das instituições candidatas. Para pesquisas que envolvam o compartilhamento de dados pessoais, deve-se pensar cuidadosamente o desenho do estudo, os documentos de consentimento informado e a estrutura do conjunto de dados antes do início do estudo. Como uma medida para salvaguarda as informações de saúde individualmente identificáveis, o NIH ainda orienta aos pesquisadores que obtenham um certificado de confidencialidade para se isentarem de liberação de dados sob uma das solicitações do FOIA e como uma medida adicional de proteção contra futura divulgação forçada para estudos que planejam coletar dados de associação ao genoma.

Finalizando a análise da última subcategoria “disponibilização e reuso”, verificou-se que os documentos normativos do NIH cobrem parcialmente o seguinte critério:

- a) Reuso: nos documentos analisados não foram identificadas indicações de licença de uso para o reuso dos dados de pesquisa. Quanto à citação de dados, o NIH considera que tal prática deve feita para reconhecer a fonte de dados na qual a pesquisa se baseia.

Diante do exame das políticas recomendatórias e mandatórias do NIH, pode-se afirmar que a agência demonstra forte compromisso com as questões éticas e legais inerentes a estudos que produzem dados a partir de seres humanos, chamando atenção dos pesquisadores para a garantia da integridade e privacidade dos sujeitos da pesquisa.

Por outro lado, as políticas não foram consistentes em descrever as responsabilidades dos pesquisadores, do financiador, dos curadores de dados, e outros atores envolvidos no processo da pesquisa, o que abre espaço para interpretações conflitantes e distribuição de funções e responsabilidades desequilibrada, sobretudo, entre os pesquisadores, os principais responsáveis pelos dados.

Por fim, chama-se a atenção para a insuficiência de abordagens para temas em voga na comunidade científica como os dados FAIR (principalmente padrões de metadados e licenças de uso), PGD e abertura de dados, o que pode estar relacionado à data de emissão de alguns documentos analisados que constam de mais de uma década sem atualização.

7.2 ANÁLISE DAS POLÍTICAS PARA GESTÃO DE DADOS DE PESQUISA NO ÂMBITO DO PROGRAMA HORIZON 2020

Foram analisados oito documentos normativos em vigor da EC no âmbito do programa de financiamento H2020, sendo quatro aplicáveis a todos os beneficiários do programa, três direcionados aos pesquisadores financiados pelo ERC em projetos do H2020 e um documento para beneficiários do programa H2020 com projetos de pesquisa associados à COVID-19. O quadro 11 abaixo sintetiza a avaliação do conjunto de políticas para gestão de dados de pesquisa identificado nos documentos normativos do programa.

Quadro 11 – Avaliação geral das políticas para gestão de dados de pesquisa do H2020

| CATEGORIAS | | | | | | | | | | | | | | | | | |
|-------------------|------------|------------------|------------------------|--------|-------------------|---|-------------------------------|---------|---------------------------------------|----------------------|---------------------------------|--------------------------|-----------------|-------------|-------------------------|-------|-------|
| Grau de aplicação | | | Cobertura de critérios | | | | | | | | | | | | | | |
| | | | Aspectos gerais | | | | Planejamento e infraestrutura | | | | | Disponibilização e reuso | | | | | |
| Recomendatório | Mandatário | Não identificado | Definição | Escopo | Responsabilidades | Monitoramento de conformidade e sanções | Plano de Gestão de Dados | Suporte | Curadoria/Preservação por longo prazo | Padrões de metadados | Repositórios e centros de dados | Custo | Limite de tempo | Propriedade | Compartilhamento/Acesso | Reuso | Ética |
| | | | | | | ● | ● | ○ | ● | ● | ● | ● | ● | ● | ● | ○ | ○ |

Fonte: Elaboração própria a partir da análise dos documentos normativos.

A primeira categoria analisada foi a “grau de aplicação” na qual foi possível verificar que as políticas para gestão de dados de pesquisa do programa de financiamento H2020 são tanto de caráter recomendatório como mandatório. Essas políticas cobrem todas as áreas temáticas do programa e seu grau de aplicação varia a depender da participação ou não do projeto no piloto ORD, conforme mencionado anteriormente.

No contexto da pesquisa em saúde, observou-se que os projetos voltados para ações de saúde participantes do piloto ORD e os projetos voltados para emergências em saúde pública devem atender a políticas compulsórias (embora existam exceções às regras), enquanto os projetos não participantes são fortemente orientados a atender as políticas para gestão dos dados que, em algumas circunstâncias, tornam-se obrigatórias.

Durante a análise foi possível verificar quatro principais políticas recorrentes relacionadas à gestão de dados de pesquisa endereçadas aos beneficiários do H2020 envolvidos em projetos de ações de saúde e os respectivos grau de aplicação das políticas, conforme mostra o quadro 12. A análise delimitou-se nas recomendações e exigências para a gestão de dados de pesquisa em saúde, embora as mesmas se apliquem a todas as áreas do programa H2020.

Quadro 12 – Grau de aplicação das principais políticas para gestão de dados de pesquisas envolvidas em ações de saúde no H2020

| Políticas identificadas | Ações de saúde | | |
|--|-----------------------------|---------------------------------|--|
| | Participantes no Piloto ORD | Não participantes no Piloto ORD | Voltadas para emergências de saúde pública |
| | Grau de aplicação | | |
| Elaborar o PGD | Mandatório | Recomendatório | Mandatório |
| Depositar dados subjacentes às publicações em repositório de dados | Mandatório | Recomendatório | Mandatório |
| Disponibilizar em acesso aberto os dados de pesquisa | Mandatório, salvo exceção | Mandatório por exceção | Mandatório, salvo exceção |
| Fornecer informações, através do repositório, sobre ferramentas e instrumentos de validação dos resultados | Mandatório | Não se aplica | Não identificado |

Fonte: Elaboração própria a partir da análise dos documentos normativos.

A política de elaboração do PGD é mandatória para projetos participantes do piloto ORD e projetos envolvidos em ações de resposta à emergência em saúde pública, embora seja recomendado que os projetos que optam por sair do piloto ORD enviem o PGD voluntariamente. Para Spichtinger e Siren (2018), a obrigatoriedade de elaborar um PGD para projetos participantes do piloto ORD pode proporcionar à Comissão e às suas agências maior monitoramento sobre a avaliação das práticas de compartilhamento de dados e evolução das competências de gestão de dados na pesquisa da EU.

A política de depósito de dados em repositórios é mandatória para projetos participantes do piloto ORD e projetos voltados para emergências em saúde pública. Para projetos de ações de saúde que não participam do piloto ORD, mas que optam pela publicação como o canal de disseminação de resultados no contexto amplo de divulgação e exploração da informação científica, tal política é fortemente recomendada, mas não obrigatória, exceto se a

CE reconhecer que os dados gerados na ação são relevantes para lidar com a emergência em saúde pública.

Nessa perspectiva, verificou-se que a política de depósito de dados de pesquisa em repositórios de dados é orientada para atingir dois propósitos diferentes. O primeiro objetivo visa garantir a gestão dos dados, o que torna o depósito mandatório para projetos ORD e projetos voltados para emergência em saúde pública e fortemente recomendado para projetos que não participam do piloto. O segundo objetivo visa facilitar o acesso aberto, o que torna a política mandatória para projetos ORD e voltados para emergência em saúde pública e pode se tornar mandatória para projetos não participantes do ORD se os dados forem relevantes para situações de emergência em saúde pública. No entanto, mesmo nessas situações o acesso aberto aos dados de pesquisa pode não ser concedido em alguns casos.

Dessa forma, a política de permitir o acesso aberto aos dados gerados em projetos de ações de saúde participantes do piloto ORD e voltados para emergências em saúde pública é mandatória, salvo exceções. Para projetos de ações de saúde que não participam do piloto ORD tal política pode se tornar mandatória por exceção, ou seja, caso o acesso aberto se faça necessário para lidar com uma emergência em saúde pública. Observa-se que mesmo o acesso aberto aos dados de pesquisa constituindo-se em uma política mandatória para projetos financiados pelo programa H2020 para atender as aspirações do cenário de emergência em saúde pública, os pesquisadores não são obrigados a fornecê-lo tendo em vista as razões descritas no quadro 13.

Quadro 13 – Razões para a não concessão de acesso aberto aos dados de pesquisas financiadas pelo programa europeu H2020

| Artigo | Razões | Ações de saúde | | |
|--------|---|-----------------------------|---------------------------------|------------------------------|
| | | Participantes no piloto ORD | Não participantes no piloto ORD | Emergências em saúde pública |
| 27 | Proteção dos resultados | x | x | x |
| 29.3 | Prejuízo ao objetivo principal da pesquisa (EC) ou um dos objetivos da pesquisa (ERC) | x | Não se aplica | Não identificado |
| 29.3 | Concessão de direitos de acesso especiais para terceiros | x | Não se aplica | x |
| 36 | Confidencialidade | x | x | x |
| 37 | Segurança dos dados | x | x | x |
| 39 | Proteção de dados pessoais | x | x | x |

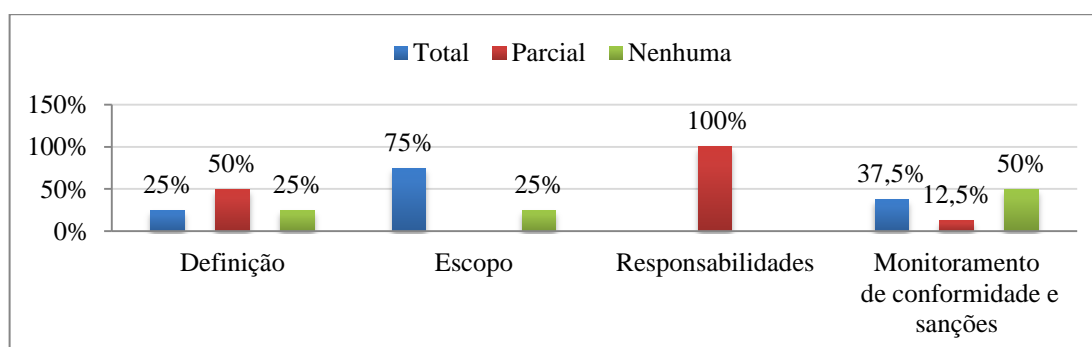
Fonte: Adaptado de European Commission (2019).

Por fim, identificou-se que a política de fornecer informações, através do repositório, sobre ferramentas e instrumentos de validação dos resultados é mandatória apenas para os projetos participantes do piloto ORD. Não foi possível identificar se a política era aplicada aos projetos voltados para emergência em saúde pública.

A partir da análise do grau de aplicação das políticas, foi possível observar que pesquisadores com projetos que não participam do piloto ORD ficam desobrigados de enviar um PGD com as informações de como os dados serão explorados, disponibilizados, curados e preservados para fins de verificação e reutilização. Além disso, esses pesquisadores, ao optarem pela exploração/proteção dos dados no contexto mais amplo de divulgação e exploração da informação científica, também ficam desobrigados de realizar o depósito dos dados de pesquisa em um repositório. Nesse contexto, pode-se inferir que a gestão dos dados produzidos pode vir a ser uma atividade negligenciada em virtude da ausência de um documento que formalize o compromisso do pesquisador com o tratamento desses recursos digitais e da desobrigação de depositar os dados em um ambiente tecnológico, gerencial e de padronização propício para a preservação de longo prazo (repositórios de dados).

A segunda categoria analisada foi “cobertura de critérios”. A porcentagem de critérios cobertos na primeira subcategoria “Aspectos gerais” é fornecida no gráfico 4 a seguir.

Gráfico 4 – Porcentagem de cobertura da subcategoria “aspectos gerais” por critério (EC)



Fonte: Elaboração própria a partir da análise dos documentos normativos.

Diante dos resultados retratados no gráfico acima, as políticas para gestão de dados de pesquisa do H2020 identificadas nos documentos normativos oferecem cobertura total para os seguintes critérios:

- a) Definição: os documentos apresentam definições acerca dos conceitos “dados de pesquisa”, “dados de pesquisa abertos”, “dados FAIR”, além de fornecerem exemplos de tipos de dados de pesquisa gerados pela pesquisa digital.

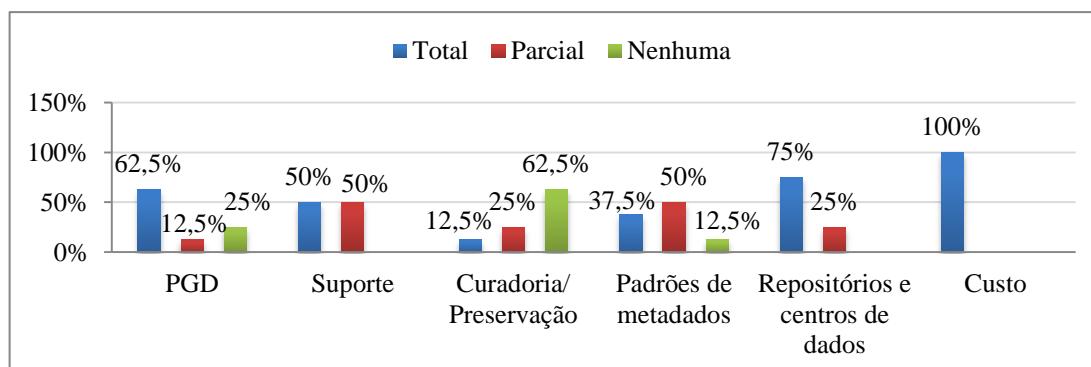
- b) Escopo: os documentos especificam que os tipos de dados cobertos pela política são os dados subjacentes à publicação científica e outros dados conforme especificado no PGD e a quem ela se destina, ou seja, aos beneficiários financiados ou cofinanciados no âmbito do H2020.
- c) Monitoramento de conformidade e sanções: os documentos declaram que os dados de pesquisa abertos serão monitorizados ao longo do H2020 e qualquer descumprimento às obrigações presentes no artigo 29 do modelo de acordo de subvenção que trata da disseminação dos resultados, acesso aberto e visibilidade do financiamento da EU pode levar a redução da subvenção.

Dando continuidade à subcategoria “aspectos gerais”, verificou-se que as políticas analisadas cobrem parcialmente o seguinte critério:

- a) Responsabilidades: os documentos definem as responsabilidades dos pesquisadores com o gerenciamento dos dados, envolvendo uma série de compromissos que deverão ser assumidos no PGD, embora não tenha seção específica que aborda esse critério. Os compromissos assumidos pela EC concentram-se no fornecimento de serviços de gestão de dados. Não foram identificadas as responsabilidades de outros *stakeholders* para garantir a preservação, disponibilidade e reutilização dos dados.

Avançando na análise das subcategorias, o gráfico 5 ilustra a porcentagem de cobertura dos critérios contemplados na subcategoria “Planejamento e infraestrutura”.

Gráfico 5 – Porcentagem de cobertura da subcategoria “Planejamento e infraestrutura” por critério (EC)



Fonte: Elaboração própria a partir da análise dos documentos normativos.

Conforme mostra o gráfico acima, a segunda subcategoria de análise “Planejamento e Infraestrutura” apresentou cobertura total para todos os critérios, conforme descrito abaixo:

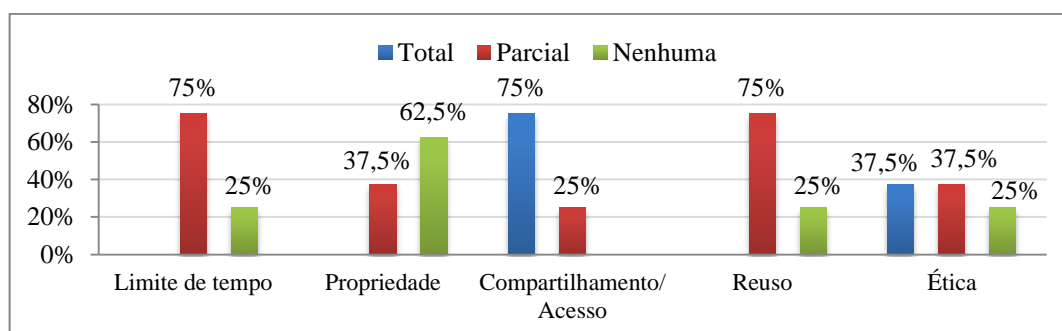
- a) Plano de Gestão de Dados: os documentos normativos analisados demonstram que tanto a EC quanto o ERC preveem a elaboração do PGD e estabelecem os prazos de envio do documento (diferentes para projetos ORD e projetos voltados para emergência em saúde pública) e o período de revisão do documento. Os documentos também oferecem as informações necessárias para a composição do PGD, visando tornar os dados de pesquisa gerados em projetos H2020 localizáveis, acessíveis, interoperáveis e reutilizáveis (FAIR).
- b) Suporte: os documentos analisados informam que a EC não oferece atualmente sua própria ferramenta *online* para PGD. Por outro lado, a Comissão e o ERC fornecem aos seus beneficiários modelos de PGD inspirados nos princípios FAIR, bem como permite o uso de ferramentas *online* para a criação do plano. A Comissão também fornece suporte por meio de serviços técnico e profissional específicos para o gerenciamento dos dados (incluindo um assistente de licença para dados de pesquisa) e recomenda sempre que for adequado e possível o uso de infraestruturas de pesquisa europeias para doenças emergentes, como a COVID-19.
- c) Curadoria/Preservação de longo prazo: os documentos estabelecem que os beneficiários verifiquem se o repositório escolhido para o depósito dos dados da pesquisa é sustentável ao longo do prazo, ou seja, considerem o armazenamento seguro dos dados; a descrição de forma padronizada com o uso de padrões de metadados; a localização, por meio do uso de um identificador persistente e; a acessibilidade e reutilização dos dados regidos por uma licença.
- d) Padrões de metadados: os documentos preveem que os beneficiários façam o uso de padrões de metadados específicos aceitos em sua comunidade de pesquisa, quando existentes. No entanto, o ERC reconhece que nem todas as disciplinas possuem padrões de metadados estabelecidos, desse modo, considera o *Dublin Core* como um padrão mínimo para a descrição de dados de pesquisa. Já a EC orienta que seus beneficiários busquem obter informações sobre padrões de metadados disciplinares no DCC e no *Research Data Alliance Metadata Directory*.
- e) Repositórios e centros de dados: os documentos analisados revelam que as políticas para repositórios e centros de dados não são prescritivas, desse modo, a decisão final sobre em qual ambiente depositar os dados gerados permanece com o pesquisador. O ERC orienta que seus beneficiários no programa H2020 façam o depósito de dados em repositórios institucionais, desde que tenham uma política de dados que garanta o suporte para armazenamento e compartilhamento de dados no

futuro. Nos documentos analisados do ERC e EC, o *Zenodo* é o principal repositório indicado para o depósito de dados gerados no âmbito do H2020, embora haja a indicação do *Dryad*, *Figshare*, entre outros. A EC oriente seus beneficiários a buscarem repositórios mais adequados aos seus campos disciplinares no *Re3data.org*, na *EUDAT Collaborative Data Infrastructure* (EUDAT CDI) e os certificados pelo *Core Trust Seal*. Para o domínio das ciências da vida, o ERC recomenda que os pesquisadores localizem repositórios de dados apropriados na lista do National Center for Biotechnology Information (NCBI)⁷⁶ do NIH.

- f) Custo: os documentos declaram que os beneficiários devem garantir recursos e planejamento orçamentário para gerenciamento de dados no PGD. Para tanto, a Comissão declara que os custos relacionados à implementação do piloto ORD (por exemplo, custos para fornecer acesso aberto, gestão de dados de pesquisa, curadoria e armazenamento de dados) são elegíveis se incorridos antes do final da ação e se todas as outras condições de elegibilidade forem cumpridas. Já o ERC no âmbito do H2020 não prevê aos beneficiários que decidem participar do piloto nenhum financiamento adicional para atividades de gestão de dados, enquanto para o acesso aberto os custos são elegíveis.

Dando sequência à análise das subcategorias, o gráfico 6 ilustra a porcentagem de cobertura dos critérios contemplados na subcategoria “Disponibilização e Reuso”.

Gráfico 6 – Porcentagem de cobertura da subcategoria “Disponibilização e Reuso” por critério (EC)



Fonte: Elaboração própria a partir da análise dos documentos normativos.

Quanto à cobertura de critérios da subcategoria “Disponibilização e reuso”, as políticas analisadas cobrem totalmente os seguintes critérios:

⁷⁶ Disponível em: <https://www.ncbi.nlm.nih.gov/guide/sitemap/>. Acesso em: 19 dez. 2020.

- a) Compartilhamento/Acesso: os documentos analisados mostram que as políticas para o compartilhamento e acesso não são prescritivas, mas sugerem que os pesquisadores financiados pelo ERC e pela EC sigam as melhores práticas, mantendo os dados FAIR para serem compartilhados com outros pesquisadores sempre que não houver restrições. Quanto ao acesso, o ERC e a EC reconhecem que nem todos os dados podem ser abertos, permitindo que os projetos financiados optem por sair do piloto ou aqueles que escolham permanecer possam definir no PGD quais os controles e limites de acesso. Importa destacar que as condições para o compartilhamento e acesso podem ser alteradas em situação de emergência em saúde pública.
- b) Ética: os documentos declaram que caso os projetos apresentem questões éticas ou legais que podem afetar o compartilhamento de dados deverão tratá-las no contexto da revisão de ética ao se candidatar a bolsa. Os beneficiários do ERC devem fornecer informações sobre as medidas tomadas para salvaguardar e proteger os dados sensíveis no PGD, embora também possa ser solicitados a enviarem um PGD de ética à Equipe de Ética da ERCEA. A EC considera que os dados pessoais devem ser processados de acordo com a legislação nacional e da EU aplicável, por exemplo, o Regulamento Geral de Proteção de Dados (*General Data Protection Regulation – GDPR*).

Dando prosseguimento à análise da subcategoria “Disponibilização e reuso”, as políticas avaliadas possuem cobertura parcial para os seguintes critérios:

- a) Limite de tempo: a abordagem para o limite de tempo para o depósito de dados é frequentemente descrita nos documentos como “o mais rápido possível”, embora as políticas considerem que embargos sejam permitidos e caso solicitados (visando garantir o tempo hábil para publicação ou solicitação de patente), a justificativa e o tempo aplicado deverá ser especificado no PGD. Em situações de emergência em saúde pública, a disponibilização e acesso aos dados são previstos no mais tardar, no prazo de 30 dias após a geração. No contexto da pandemia COVID-19, dados de pesquisa relevantes para lidar com a pandemia devem ser depositados e acessíveis imediatamente. O tempo de duração mínima de preservação dos dados não foi considerado nas políticas.
- b) Propriedade: os documentos analisados não versam sobre a propriedade que a instituição a qual beneficiário faz parte possui sobre os dados, embora prevejam

que, em alguns casos, os pesquisadores possam atrasar ou limitar o acesso aos dados da pesquisa para garantir a proteção da propriedade intelectual.

- c) Reuso: os documentos apontam para o aumento da reutilização de dados por meio de Licença de Atribuição Creative Commons (CC BY 4.0) aos dados depositados, ou por renúncia a todos os interesses associados a direitos autorais e proteção de banco de dados usando a ferramenta CC0 de domínio público. Dessa forma, os beneficiários devem declarar no PGD como os dados serão licenciados, quando serão depositados para a reutilização, por quanto tempo permanecerão reutilizáveis e justificativas para caso o reuso dos dados gerados não seja permitido. Quanto às práticas de citação de dados, não foram identificadas orientações específicas sobre o mecanismo para atribuir crédito aos conjuntos de dados nos documentos da EC. Já ERC identifica que a citação formal dos dados é um critério cada vez mais solicitado pelos editores, desse modo, orienta seus beneficiários no âmbito do H2020 à consulta de um material externo.

A análise das políticas do H2020 permitiu verificar que o programa está alinhado ao novo cenário científico de mudanças para todo o ciclo de vida da pesquisa, como novas preocupações acerca de como garantir que os dados sejam adequadamente gerenciados, compartilhados e preservados, tanto no decorrer como após a finalização do projeto da pesquisa.

De forma geral, as políticas do H2020 são abrangentes, apresentando recomendações genéricas para a gestão dos dados gerados em projetos de diferentes áreas financiadas. No entanto, também apresentam algumas orientações para dados de pesquisa em um nível mais disciplinar como no campo da saúde, com disposições acerca de questões éticas e legais relacionadas aos dados sensíveis e especificidades para o compartilhamento e acesso aos dados de pesquisa em situações de emergência em saúde pública que visam acelerar a disponibilização dos dados para a comunidade científica.

7.3 CONSIDERAÇÕES ACERCA DA GESTÃO DE DADOS DE PESQUISA EM SAÚDE NO CONTEXTO DA PESQUISA COLABORATIVA INTERNACIONAL

O cerne dessa pesquisa foi compreender se os financiadores da pesquisa em saúde dispõem de políticas específicas para a gestão de dados de pesquisas realizadas em colaborações entre pesquisadores de diferentes países. Nessa perspectiva, importa sinalizar

que os dois financiadores analisados apresentaram particularidades em seus respectivos sistemas de financiamento para colaborações internacionais, desse modo, cabe lembrar que não se objetivou elaborar um estudo comparado entre as duas agências analisadas, dadas as suas características únicas: o NIH, de caráter regional, que financia predominantemente projetos desenvolvidos nos EUA e a EC, de caráter multi-territorial, que fornece financiamento a projetos desenvolvidos por países europeus ou organizações internacionais. Nesse sentido, se não cabe falar em diferenças já que não são comparáveis, pela sua natureza, cabe sinalizar as características de cada uma delas na sua abordagem de financiamento no campo da saúde.

Na agência de fomento norte-americana NIH foi possível perceber que a colaboração internacional em pesquisa é realizada de formas diferentes entre os ICs. Sendo assim, alguns institutos consideram que a colaboração internacional em pesquisa acontece quando há uma concessão nacional com um componente estrangeiro, em que um pesquisador de uma instituição dos Estados Unidos trabalha em colaboração com um pesquisador empregado por uma organização de outro país. Outros ICs afirmam realizar a colaboração internacional por meio de programas⁷⁷ bilaterais com países selecionados em apoio a iniciativas globais de saúde ou apoiando pesquisadores dos Estados Unidos e dos países de baixa e média renda, conforme definido pelo Banco Mundial⁷⁸, em programas de treinamento ou de pesquisas⁷⁹ realizadas entre acordos cooperativos. É importante ressaltar que essas atuações identificadas não refletem, necessariamente, aos mecanismos de outros ICs que compõem a agência.

Durante a análise dos documentos normativos do NIH não foram identificadas políticas para a gestão de dados de pesquisas realizadas no âmbito da colaboração científica entre instituições e pesquisadores de diferentes países. No entanto, verificou-se que o critério “Responsabilidades”, analisado na subcategoria “Aspectos Gerais”, apresentou uma questão relevante para projetos de pesquisas realizados no âmbito de consórcio constituído pela organização beneficiária do financiamento do NIH e por uma ou mais organizações colaboradoras responsáveis por realizar partes das atividades da pesquisa financiada.

Nesses tipos de empreendimentos colaborativos, o beneficiário, destinatário direto e principal do financiamento do NIH e responsável perante a agência pelo desempenho do projeto, é responsável por estabelecer um acordo formal com cada participante do consórcio

⁷⁷ Disponível em: <https://www.nhlbi.nih.gov/about/translation-research-implementation-science/inequities-global-health>. Acesso em: 10 dez. 2020.

⁷⁸ Disponível em: <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519>. Acesso em: 10 dez. 2020.

⁷⁹ Disponível em: <https://www.fic.nih.gov/Funding/Pages/Collaborations.aspx>. Acesso em: 10 dez. 2020.

para negociar os arranjos necessários para atender aos requisitos científicos, administrativos, financeiros e de relatórios do subsídio. Entre outras disposições para garantir a conformidade com todos os regulamentos e políticas federais aplicáveis e facilitar o empreendimento colaborativo, o acordo de consórcio deverá incluir disposições em torno de questões com efeitos sobre a gestão dos dados, como:

- a) a propriedade e disponibilização dos dados produzidos no âmbito do acordo de consórcio, abordando se esses recursos estarão disponíveis gratuitamente para outros pesquisadores na comunidade científica ou serão fornecidos apenas a pesquisadores específicos.
- b) o compartilhamento de dados e invenções e a política de patentes do NIH, incluindo a exigência de comunicar as invenções ao beneficiário (coordenador) e que essa disposição seja aplicável a cada participante do consórcio e seus funcionários, a fim de assegurar que os direitos das partes do acordo de consórcio sejam protegidos e que o beneficiário possa cumprir suas responsabilidades para com o NIH.

Embora não tenha sido possível verificar políticas para gestão de dados de pesquisas realizadas em um cenário de colaboração internacional, a exigência para que atores envolvidos em projetos colaborativos interinstitucionais celebrem um acordo de consórcio abordando questões como propriedade, disponibilização e compartilhamento de dados denota certo reconhecimento do NIH quanto à necessidade de negociação e construção de acordos em torno de questões relacionadas à gestão dos dados gerados em pesquisas colaborativas.

No programa de financiamento europeu H2020, a colaboração em pesquisa é reconhecida como uma condição fundamental para o desenvolvimento das pesquisas financiadas. Os projetos colaborativos de pesquisa e inovação devem envolver minimamente três Estados-Membros da EU ou países associados, embora muitas áreas do H2020 consideram a colaboração internacional uma prioridade para atingir os propósitos da pesquisa, envolvendo pesquisadores de outras partes do mundo.

A análise realizada nos documentos do programa H2020 permitiu perceber que no arranjo dos projetos colaborativos de pesquisa e inovação do H2020 (consórcios de organizações), o coordenador⁸⁰ e os beneficiários⁸¹, oriundos de Estados-membros ou países

⁸⁰ O coordenador é o beneficiário que é ponto de contato central da Comissão/Agência e representa o consórcio, responsável por coordenar e gerir a subvenção (EUROPEAN COMMISSION, 2019).

⁸¹ Beneficiários são as entidades jurídicas que assinam o Acordo de Subvenção com a Comissão/Agência, ou seja, participam num projeto apoiado por uma subvenção da EU e são sujeitos à totalidade dos termos e condições do acordo (EUROPEAN COMMISSION, 2019).

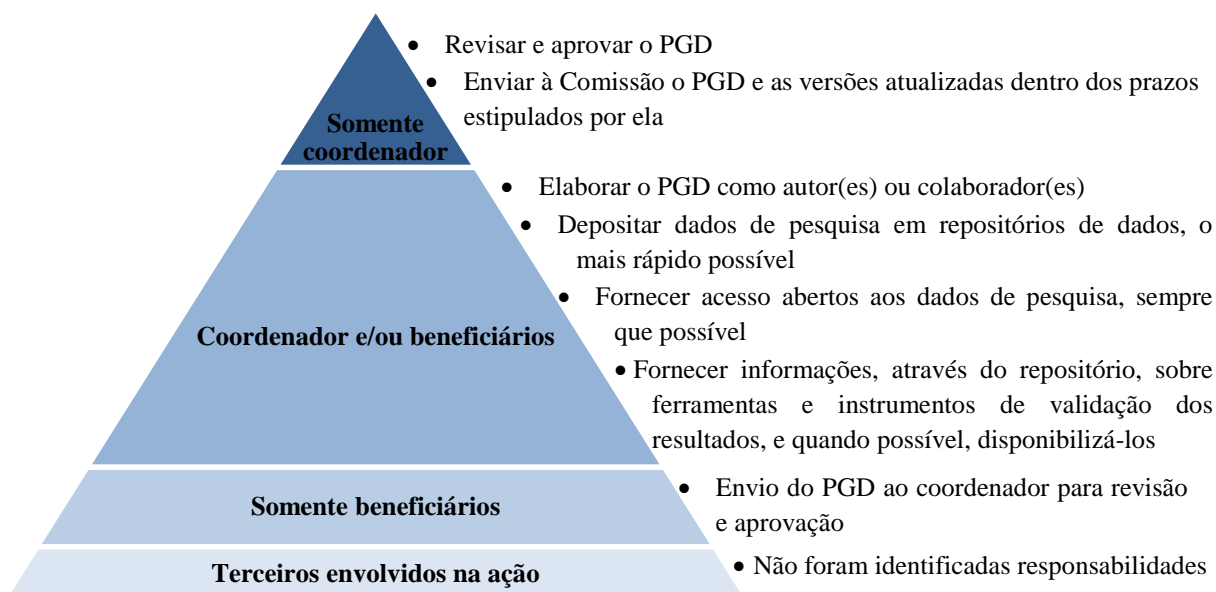
associados, não são os únicos participantes nesses empreendimentos que podem envolver também os chamados terceiros envolvidos na ação⁸² (incluindo parceiros internacionais, subcontratados, terceiros que dão contribuições em espécie, entre outros). Esses atores, embora não assinem o contrato de subvenção e, portanto não sejam beneficiários, ao participarem do projeto de pesquisa estão indiretamente vinculados à maioria das obrigações do contrato.

É importante pontuar que os beneficiários são os principais atores nos consórcios de pesquisa, portanto, devem, em conjunto, contribuir para a execução harmoniosa da subvenção. Para tanto, os beneficiários do H2020 também devem celebrar um acordo de consórcio que fixe os respectivos direitos e obrigações relativos à execução da pesquisa. A partir da avaliação das políticas do H2020, identificou-se que determinados critérios para gestão de dados de pesquisa apresentaram especificidades para o contexto de pesquisas colaborativas realizadas no âmbito dos consórcios nacionais e internacionais, conforme será analisado a seguir.

Quanto ao critério “Responsabilidades”, contemplado na subcategoria “Aspectos Gerais”, verificou-se que os compromissos que deverão ser assumidos com as atividades de gestão de dados estão concentrados nas figuras do coordenador e beneficiários. Importa assinalar que não foram identificadas declarações sobre os compromissos e as responsabilidades dos terceiros envolvidos na ação, uma vez que sob os termos do acordo de subvenção, cabe ao beneficiário garantir que as obrigações de seus terceiros envolvidos sejam cumpridas. A figura 1 a seguir apresenta a distribuição das principais responsabilidades identificadas durante análise dos documentos quanto às atividades relacionadas à gestão dos dados no âmbito dos consórcios.

⁸² Os terceiros envolvidos na ação são outras entidades que participam da ação, mas não assinam o Acordo de Subvenção, portanto, não são beneficiários (EUROPEAN COMMISSION, 2019). Não há vínculo contratual formal, ou seja, não estão formalmente vinculados aos termos e condições do acordo. No entanto, ao participarem do consórcio e por meio das obrigações dos beneficiários a que estão associados, passam a estar vinculados, indiretamente, à maioria das obrigações do acordo.

Figura 3 – Distribuição das principais responsabilidades identificadas para a gestão de dados de pesquisa no âmbito dos consórcios de pesquisa (EC)



Fonte: Elaboração própria a partir da análise dos documentos normativos.

No que diz respeito especificamente à responsabilidade de elaboração do PGD, infere-se que os beneficiários da ação podem elaborar o documento, ou seja, são autores do plano, no entanto, o coordenador é o responsável por revisá-lo e apresentá-lo à Comissão. Essa inferência baseia-se no dever dos beneficiários de apresentar ao coordenador do projeto quaisquer documentos ou informações exigidas pela Comissão, conforme declarado no artigo 41.2 do Modelo de Acordo de Subvenção do H2020.

Dando prosseguimento à análise, foram identificadas algumas recomendações para o critério analisado “Plano de Gestão de Dados”, referente à subcategoria “Planejamento e Infraestrutura”, conforme compiladas no quadro 14 a seguir.

Quadro 14 – Recomendações para a elaboração de PGD de projetos colaborativos no âmbito do H2020

| Recomendações | Descrição |
|--|--|
| Atualização | O PGD deve ser atualizado sempre que surgir mudanças nas políticas (por exemplo, novo potencial de inovação, decisão de solicitar uma patente) ou na composição do consórcio (por exemplo, entrada de novos membros do consórcio ou saída de membros antigos). |
| Definição de um cronograma de revisão | É recomendado que o consórcio defina no PGD um cronograma para revisão do documento. |

| | |
|---|--|
| Harmonização com o acordo de consórcio | A política para gestão de dados deve refletir o estado atual dos acordos de consórcio sobre gestão de dados. |
| Definição de responsabilidades | O PGD de um consórcio deverá especificar quem será o responsável pelo gerenciamento de dados no projeto. |

Fonte: Elaboração própria a partir da análise dos documentos normativos.

Quanto à subcategoria “Disponibilização e Reuso”, no critério “Propriedade” a Comissão declara que os interesses de proteção de dados, especialmente comerciais, deverão ser equilibrados entre os beneficiários participantes no projeto. A Comissão também prevê que os beneficiários tomem providências especiais em relação aos resultados produzidos por seus subcontratados, terceiros vinculados e parceiros internacionais, considerando que:

Se um terceiro gerar resultados, o beneficiário em causa deve obter todos os direitos necessários (transferência, licenças ou outros) do terceiro, para poder respeitar as suas obrigações como se esses resultados fossem gerados pelo próprio beneficiário. Na impossibilidade de obtenção dos direitos, o beneficiário deve abster-se de utilizar o terceiro para gerar os resultados (EUROPEAN COMMISSION, 2019, p. 237, tradução nossa⁸³).

Na perspectiva dos dados de pesquisa, infere-se que a reivindicação de direitos sobre os resultados gerados por terceiros envolvidos na ação possivelmente pode ser tornar um problema que necessite ser abordado nas instâncias de negociações entre os membros de um consórcio de pesquisa.

Por fim, no que diz respeito ao critério “Compartilhamento/Acesso”, a Comissão orienta que os beneficiários façam uso de protocolos harmonizados em colaboração com outros atores da pesquisa para facilitar o fornecimento do acesso aberto imediato a todos os dados de pesquisa relevantes para o enfrentamento de emergências de saúde pública. Ainda sobre o critério analisado, a Comissão possibilita que em consórcios com multiparticipantes, beneficiários específicos mantenham seus dados em acesso fechado, caso as justificativas estejam no acordo de consórcio firmado entre os membros e em conformidade com os motivos para a não concessão do acesso previstos no programa H2020. Para os consórcios de pesquisa nacionais e internacionais, O ERC também permite que seus beneficiários no âmbito do H2020 criem e utilizem bancos de dados visando promover a pesquisa na comunidade,

⁸³ No original: “*If a third party generates results, the beneficiary concerned must obtain all necessary rights (transfer, licences or other) from the third party, in order to be able to respect its obligations as if those results were generated by the beneficiary itself. If obtaining the rights is impossible, the beneficiary must refrain from using the third party to generate the results*” (EUROPEAN COMMISSION, 2019, p. 237).

embora o depósito adicional dos dados em um repositório público seja altamente recomendado para garantir curadoria, preservação e localização de longo prazo.

Mediante o exame das políticas para gestão de dados de pesquisa emitidas pelo NIH e EC foi possível constatar que esses financiadores antecipadamente reconhecem em seus documentos normativos que algumas questões relacionadas aos dados de pesquisa precisam ser negociadas entre os pesquisadores envolvidos em colaborações científicas (não necessariamente internacional). Portanto, dispõem no rol de suas normativas, orientações específicas como a definição de direitos e responsabilidades quanto à propriedade, compartilhamento e acesso aos dados, harmonização e atualização das expectativas descritas no PGD com a realidade das práticas vigentes de gestão dos dados no âmbito dos consórcios da pesquisa.

8 CONSIDERAÇÕES FINAIS

Este estudo investigou as orientações dos financiadores da pesquisa para gestão de dados produzidos em contexto de colaboração científica internacional no campo da saúde. As colaborações internacionais são uma realidade acadêmica, sobretudo na pesquisa em saúde, na qual facilitam a produção de conhecimento com o potencial de contribuir para a melhoria das condições de vida e saúde da população mundial. Considerou-se que as variações de entendimentos entre os atores envolvidos nesses contextos a respeito de questões relacionadas à gestão dos dados que não foram previamente abordadas e estabelecidas nos documentos normativos podem impactar diretamente a qualidade e integridade da pesquisa.

Desse modo, para atingir os propósitos desta pesquisa, em um primeiro momento foi necessária a identificação dos principais financiadores da pesquisa em saúde. A partir de uma revisão de literatura, foi possível identificar os dez maiores financiadores públicos e filantrópicos de pesquisa em saúde no mundo. A partir desse universo inicial, foram selecionados os dois primeiros do *ranking* mundial, ou seja, o NIH e a EC.

Algumas particularidades foram notadas nos dois financiadores no que diz respeito ao sistema de financiamento em pesquisa. Verificou-se que o NIH não expressa com clareza a natureza das pesquisas colaborativas internacionais que financia e demonstra priorização de financiamento para projetos de pesquisas desenvolvidas por pesquisadores e instituições norte-americanas. Por outro lado, a EC, no âmbito do H2020, explicita de forma mais contundente que o financiamento de projetos colaborativos com pesquisadores de Estados-membros e países associados é uma prioridade para alavancar a internacionalidade da pesquisa colaborativa.

Em seguida, a fim de constatar se as políticas para gestão de dados de pesquisa emitidas pelos financiadores selecionados apresentavam especificações para o contexto da pesquisa colaborativa internacional, realizou-se o mapeamento dos documentos que materializavam essas normativas. Nesse momento da pesquisa foi necessário delimitar um instrumento que possibilitasse a análise das políticas, de modo que fosse possível identificar as especificações em questão. A partir de uma revisão da literatura, foi identificada e adaptada para este estudo uma estrutura de avaliação de políticas de dados de pesquisa, que permitiu analisar o grau de aplicação das políticas e cobertura de critérios abordados.

O exame das políticas para a gestão de dados demonstrou que vários requisitos para essa atividade fazem parte das condições de concessão de financiamento nas pesquisas promovidas pelo NIH e EC. Os financiadores abordam de forma particular questões

relacionadas à gestão de dados, como a curadoria/preservação, compartilhamento, PGD, padrões de metadados, repositórios, acesso, abertura e questões éticas, por meio de recomendações e/ou obrigações que se não atendidas podem resultar na retenção de fundos. Os financiadores também consideraram em suas políticas disposições para os dados de pesquisa em saúde, considerando a necessidade de equilibrar as demandas de acesso e abertura dos dados e salvaguarda da privacidade dos sujeitos da pesquisa.

A respeito da cobertura fornecida para os critérios elencados na estrutura de avaliação, verificou-se que o NIH cobre totalmente 60% e a EC 73% dos critérios analisados. Esse resultado pode estar relacionado ao fato da atuação preponderante da EC no desenvolvimento de políticas recentes para dados de pesquisa que atendam aos pressupostos do movimento pelo acesso aberto às publicações científicas e aos dados de pesquisa bem como pela ciência aberta que traz para o cenário científico novas práticas mais colaborativas, abertas e transparentes para todo o processo da pesquisa.

No decorrer da análise, foram identificadas especificações para a gestão dos dados produzidos no contexto da pesquisa colaborativa, geralmente referida como consórcio de pesquisa. Importa sinalizar que as diferenças nos mecanismos de financiamento de pesquisa dificultaram a identificação da dimensão da ‘internacionalização’ das pesquisas de natureza colaborativa financiadas por essas instituições. Embora os financiadores tenham apresentado considerações acerca de questões envolvendo a gestão de dados produzidos em ambientes colaborativos, não foi claramente explicitado, sobretudo nos documentos do NIH, se essas pesquisas poderiam envolver pesquisadores de diferentes países. Portanto, conclui-se que essa lacuna de políticas específicas para a gestão de dados produzidos em colaborações internacionais podem colocar os pesquisadores em situações de conflitos normativos, uma vez que possivelmente colocarão ‘em jogo’ as normas de seus países e instituições.

Nos documentos normativos de ambos os financiadores foi possível identificar uma responsabilidade comum para participantes de projetos de pesquisas colaborativas, qual seja: a necessidade de celebração de um acordo de consórcio que, entre muitas questões científicas, administrativas e financeiras, estabeleça os respectivos direitos e obrigações com questões relacionadas à gestão de dados de pesquisa. Esse requisito demonstra que tanto o NIH como EC reconhecem que questões relativas à gestão de dados de pesquisa, como a propriedade, compartilhamento e acesso podem provocar interesses discordantes entre os participantes, desse modo, antecipam esse possível cenário de conflito, ratificando a importância da negociação de acordos antes do início do projeto.

Especificamente nos documentos da EC, outro aspecto que pôde ser observado foi a carência de informações a respeito da distribuição de funções e responsabilidades dos pesquisadores internacionais nos consórcios de pesquisa com as atividades de gestão de dados. Entretanto, importa sinalizar que a não identificação de responsabilidades não significa que esses pesquisadores não desempenham quaisquer atribuições na prática.

Ainda sobre a questão, cabe considerar que os pesquisadores envolvidos em projeto colaborativo, seja internacional ou não, em algum momento da pesquisa coletam, produzem e utilizam os dados da pesquisa. No entanto, as políticas do NIH e EC parecem centralizar grande parte das responsabilidades na figura do coordenador, o elo entre o financiador e o consórcio de pesquisa. Desse modo, pode-se concluir que o cenário da pesquisa é colaborativo, envolvendo atores heterogêneos como instituições de pesquisa (nacionais e internacionais), indústrias, entre outros, mas a responsabilização com as atividades de gestão dos dados é individual e centralizada na figura de um pesquisador (coordenador), em detrimento da responsabilização coletiva e distribuída entre todos os pesquisadores produtores de dados. Essa constatação merece ser investigada e confirmada em estudos futuros.

Portanto, uma possibilidade de pesquisa futura está em investigar a questão desta pesquisa diretamente com grupos de pesquisas formalmente institucionalizados de colaboração científica internacional no campo da saúde, a fim de verificar como os pesquisadores distribuem as responsabilidades e atribuições com a gestão dos dados e negociam os possíveis desafios que podem emergir sobre questões relativas aos dados da pesquisa em saúde que não foram previamente considerados nas políticas vigentes do financiador do projeto.

Cabe salientar que os resultados apresentados nesta pesquisa não caracterizam todas as políticas para gestão de dados das instituições de fomento à pesquisa em saúde. Uma limitação da pesquisa foi não analisar o cenário brasileiro, isso porque as escolhas metodológicas adotadas para identificar os principais financiadores da pesquisa em saúde restringiram o exame apenas do cenário norte-americano e europeu. Essa limitação suscita a necessidade de investigar a questão no âmbito das instituições de financiamento brasileiras que participam no financiamento de colaborações internacionais no campo da saúde.

Em contrapartida, vale destacar que quando instituições nacionais recebem financiamento do NIH e da EC para o desenvolvimento de suas pesquisas estão, automaticamente, responsabilizadas para seguir os critérios estabelecidos por esses financiadores. Portanto, ainda que não tenha sido investigado como os financiadores da pesquisa em saúde no Brasil orientam a gestão dos dados produzidos em contextos de

colaboração internacional, a pesquisa apresentou o cenário de políticas para a gestão dos dados que devem ser conhecidas e cumpridas por pesquisadores que recebem financiamento para projetos individuais ou estão envolvidos em projetos colaborativos internacionais financiados por essas organizações.

Uma possível contribuição desta pesquisa foi colaborar no entendimento acerca dos requisitos endereçados aos pesquisadores participantes de colaborações científicas internacionais no que diz respeito à gestão de dados de pesquisas a partir da perspectiva do financiador, de modo a proporcionar aos pesquisadores envolvidos nesse tipo de empreendimento científico maior compreensão acerca do seu papel no tratamento dos dados, com vistas a tornar a pesquisa mais credível, com maior integridade, mais eficiente, colaborativa e mais responsiva aos desafios da sociedade.

REFERÊNCIAS

- ALBAGLI, Sarita; APPEL, Andre Luiz; MACIEL, Maria Lucia. E-Science, ciência aberta e o regime de informação em ciência e tecnologia. **Tendências da Pesquisa Brasileira em Ciência da Informação**, João Pessoa, v. 9, n. 2, jan./jun. 2014. Disponível em: <https://brapci.inf.br/index.php/res/v/27541>. Acesso em: 15 abr. 2020.
- ALBAGLI, Sarita, CLINIO, Anne, RAYCHTOCK, Sabryna. Ciência Aberta: correntes interpretativas e tipos de ação. **Liinc em Revista, Rio de Janeiro**, v.10, n.2, p. 434-450, nov. 2014. DOI: <https://doi.org/10.18617/liinc.v10i2.749>. Disponível em: <http://revista.ibict.br/liinc/article/view/3593>. Acesso em: 22 out. 2020.
- ALBAGLI, Sarita. Ciência aberta em questão. *In*: ALBAGLI, Sarita; MACIEL, Maria Lucia; ABDO, Alexandre Hannud (org.). **Ciência aberta, questões abertas**. Brasília, DF: IBICT; Rio de Janeiro: UNIRIO, 2015. p. 9-25.
- ALLIANCE OF GERMAN SCIENCE ORGANISATIONS. **Principles for Handling Research Data**. 2 p., 2010. DOI: <https://doi.org/10.2312/ALLIANZOA.019>. Disponível em: https://gfzpublic.gfz-potsdam.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_2949914. Acesso em: 26 ago. 2020.
- ANDERSON, Melissa. International Research Collaborations: anticipating challenges instead of being surprised. *In*: EUROPA World of Learning. 61th ed. London, England: Routledge, 2011a. p. 14-18. Disponível em: https://slidelegend.com/international-research-collaborations-routledge-education-arena_59e116da1723ddc1d208d239.html. Acesso em: 16 dez. 2019.
- _____. Shifting Perspectives on Research Integrity. **Journal of Empirical Research on Human Research Ethics**, [S. l.], v. 13, n. 5, Nov. 2018. Disponível em: <https://journals.sagepub.com/doi/full/10.1177/1556264618813737>. Acesso em: 20 dez. 2019.
- ANDERSON, Melissa; STENECK, Nicholas. **International Research Collaborations: much to be gained, many ways to get in trouble**. New York: Routledge, 2011.
- ANDERSON, Melissa. What can be gained and what can go wrong in the context of different national research environments. *In*: ANDERSON, Melissa S.; STENECK, Nicholas H. **International Research Collaborations: much to be gained, many ways to get in trouble**. New York: Routledge, 2011b.
- AVENTURIER, Pascal; ALENCAR, Maria de Cléofas Faggion. Os desafios dos dados de pesquisas abertos. **RECIIS – Rev Eletron Comun Inf Inov Saúde**, Rio de Janeiro, v. 10, n. 3, p. 1-19, jul./set. 2016. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/1069/pdf1069>. Acesso em: 18 jul. 2020.
- BALL, Alex. **Review of data management lifecycle models**. Bath: University of Bath, 2012. Disponível em: <https://purehost.bath.ac.uk/ws/portalfiles/portal/206543/redm1rep120110ab10.pdf>. Acesso em: 5 jan. 2020.

BEAVER, D.; ROSEN, R. Studies in scientific collaboration: Part I, The professional origins of scientific Co-authorship. **Scientometrics**, Amsterdam, v. 1, n. 1, p. 65-84, Jan. 1978. Disponível em: <https://akademai.com/doi/abs/10.1007/BF02016840>. Acesso em: 15 dez. 2019.

BERLIN Declaration on Open Access to Knowledge in the Sciences and Humanities. Berlin, 2003. Disponível em: <https://openaccess.mpg.de/Berlin-Declaration>. Acesso em: 13 jun. 2020.

BHARDWAJ, Raj Kumar; BANKS, Paul (ed.). **Research data access and management in modern libraries**. Philadelphia: IGI Global, 2019.

BRAGA, Gilda Maria. Informação, Ciência e Política Científica: o pensamento de Derek de Solla Price. **Ci. Inf.**, Rio de Janeiro, v. 3, n. 2, p. 155-177, 1974. Disponível em: http://www.brapci.inf.br/_repositorio/2010/04/pdf_6e6a23fff6_0009903.pdf. Acesso em: 20 dez. 2019.

BOHNHORST, Mark A *et al.* Legal and regulatory considerations in international research collaborations. In: ANDERSON, Melissa S.; STENECK, Nicholas H. **International Research Collaborations**: much to be gained, many ways to get in trouble. New York: Routledge, 2011.

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

_____. **Scholarship in the Digital Age**: information, infrastructure, and the internet. London: The MIT Press 2007.

_____. The conundrum of sharing research data. **Journal of the American Society for Information Science and Technology**, Los Angeles, v. 63, n. 6, p. 1059–1078, Apr. 2012. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.22634>. Acesso em: 20 dez. 2019.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). **Diário Oficial da União**: seção 1, Brasília, DF, ano CLV, n. 157-A, 15 ago. 2018.

BRINEY, Kristin. **Data Management for Researchers**: organize, maintain and share your data for research success. Exeter: Pelagic Publishing, 2015.

BRINEY, Kristin; GOBEN, Abigail, ZILINSKI, Lisa. Do you have an institutional data policy? a review of the current landscape of library data services and institutional data policies. **Journal of Librarianship and Scholarly Communication**, [Oregon], v. 3, n. 2, p. 1-25, Fev. 2015. DOI: <http://doi.org/10.7710/2162-3309.1232>. Disponível em: <https://jlscc-pub.org/articles/abstract/10.7710/2162-3309.1232/>. Acesso em: 20 abr. 2020.

_____. Institutional, funder, and journal data policies. In: JOHNSTON, Lisa R (ed.). **Curating research data**: practical strategies for your digital repository. Chicago: Association of College and Research Libraries, 2017. cap. 2, p. 61-78.

BUCKLAND, Michael K. Descrição e pesquisa: metadados como infra-estrutura. **Brazilian Journal of Information Science: Research Trends**, [S. l.], v. 0, n. 0, p. 3–14, jul./dez. 2006. DOI: 10.5016/brajis.v0i0.26. Disponível em: <https://brapci.inf.br/index.php/res/v/14175>. Acesso em: 29 set. 2020.

BUDAPEST OPEN ACCESS INITIATIVE (BOAI). **Read the Budapest Open Access Initiative**, 2002. Disponível em: <https://www.budapestopenaccessinitiative.org/read>. Acesso em: 16 abr. 2020.

BUSS, Paulo Marchiori; FERREIRA, José Roberto. Ensaio crítico sobre a cooperação internacional em saúde. **RECIIS – R. Eletr. de Com. Inf. Inov. Saúde**, Rio de Janeiro, v.4, n.1, p.93-105, mar. 2010. DOI: 10.3395/reciis.v4i1.350pt. Disponível em: <https://www.arca.fiocruz.br/bitstream/icict/17386/2/11.pdf>. Acesso em: 10 dez. 2020.

BUSS, Paulo Marchiori; LEAL, Maria do Carmo. Saúde global e diplomacia da saúde. **Cad. Saúde Pública**, [S. l.], v.25, n. 12, p. 2540-2541, dez. 2009. DOI: <https://doi.org/10.1590/S0102-311X2009001200001>. Disponível em: https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2009001200001. Acesso em: 20 nov. 2020.

BUSS, Paulo Marchiori. Saúde global e diplomacia da saúde. **Cadernos de Saúde Pública**, v. 29, n. 1, jan. 2013. Disponível em: <http://www.scielo.br/pdf/csp/v29n1/02.pdf>. Acesso em: 13 nov. 2020.

BUSS, Paulo Marchiori; TOBAR, Sebastián. **Diplomacia em Saúde e Saúde Global: perspectivas latino-americanas**. Rio de Janeiro: Editora Fiocruz, 2017.

CANADA. Government of Canada. **Public Consultation Summary**. 2019. Disponível em: http://science.gc.ca/eic/site/063.nsf/eng/h_97905.html. Acesso em: 20 ago. 2020.

CARDOSO, Gustavo, JACOBETTY, Pedro; DUARTE, Alexandra. **Para uma ciência aberta**. Lisboa: Editora Mundos Sociais, 2012.

CARLSON, Jake. The use of life cycle models in developing and supporting data services. *In*: RAY, J. **Research Data Management: practical strategies for information professionals**. West Lafayette: Purdue University Press, 2014. p. 63-86.

CASSIMIRO, M. Integridade científica: experiências e desafios do Instituto Oswaldo Cruz para o desenvolvimento responsável da ciência. *In*: CASSIMIRO, Márcia de Cássia; DIÓS-BORGES, Marcelle Mourelle Perez (org.). **Integridade Científica, Saúde Pública, Bioética e Educação em Saúde no Instituto Oswaldo Cruz**. Porto Alegre: Editora Fi, 2017. 252 p.

CAVALCANTI, M. T.; SALES, L. F. Gestão de dados de pesquisa: um panorama da atuação da União Europeia. **Biblos: Revista do Instituto de Ciências Humanas e da Informação**, [S. l.], v. 31, n. 1, p. 73-98, jan./jun. 2017. DOI: <https://doi.org/10.14295/biblos.v31i1.5789>. Disponível em: <https://periodicos.furg.br/biblos/article/view/5789>. 10 jun. 2020.

CORTI, Louise *et al.* **Managing and sharing research data: a guide to good practice**. 2nd. Los Angeles: Saga, 2019.

COSTA, Maíra Murrieta. **Diretrizes para uma política de gestão de dados científicos no Brasil**. 2017a. 288 f. Tese (Doutorado em Ciência da Informação) – Faculdade em Ciência da Informação, Universidade de Brasília, Brasília, 2017a.

COSTA, Michelli. **Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika na perspectiva de pesquisadores**. 2017b. 269f. Tese (Doutorado em Ciência da Informação) – Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, 2017b.

COSTA, Michelli; LEITE, Fernando César Lima. Princípios e recomendações basilares para a comunicação dos dados de pesquisa. **Em Questão**, [S. l.], v. 23, n. 1, p. 87-112, jan./abr. 2017. Disponível em: <http://seer.ufrgs.br/index.php/EmQuestao/article/view/65623/38903>. Acesso em: 16 set. 2019.

COX, Andrew; VERBAAN, Eddy. **Exploring Research data Management**. London: Facet Publishing, 2018.

_____. How Academic Librarians, IT Staff, and Research Administrators Perceive and Relate to Research. **Library and Information Science Research**, [S. l.], v. 38, n. 4, p. 319-326, Nov. 2016. Disponível em: <http://shura.shu.ac.uk/8906/1/Verbaan%20how%20academic%20librarians.pdf>. Acesso em: 5 jan. 2020.

COX, Andrew *et al.* Developments in research data management in academic libraries: Towards an understanding of research data service maturity. **Journal of the Association for Information Science and Technology**, [S. l.], v. 68, n. 9, p. 2182-2200, Sep. 2017. DOI: <https://doi.org/10.1002/asi.23781>. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/asi.23781>. Acesso em: 15 jun. 2020.

COX, Andrew. Special issue: Research Data Management. **Program: electronic library and information systems**, [S. l.], v. 49, n. 4, Sep. 2015. DOI: <https://doi.org/10.1108/PROG-08-2015-0055>. Disponível em: https://www.emerald.com/insight/content/doi/10.1108/PROG-08-2015-0055/full/html?utm_source=TrendMD&utm_medium=cpc&utm_campaign=Program_TrendMD_0&WT.mc_id=Emerald_TrendMD_0. Acesso: 13 maio 2020.

COX, Andrew; TAM, Winnie. A critical analysis of lifecycle models of the research process and research data management. **Aslib Journal of Information Management**, [S. l.], v. 70, n. 2, p. 142–157, apr. 2018. DOI: <https://doi.org/10.1108/AJIM-11-2017-0251>. Disponível em: <http://eprints.whiterose.ac.uk/128251/>. Acesso em: 6 jan. 2020.

CUPANI, Alberto. A propósito do “ethos” da ciência. **Episteme**, Porto Alegre, v. 3, n. 6, p. 16-38, 1998.

CURDT, C.; HOFFMEISTER, D. Research data management services for a multidisciplinary, collaborative research project: Design and implementation of the TR32DB project database. **Program**, [S. l.], v. 49, n. 4, p. 494–512, Sep. 2015. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/PROG-02-2015-0016/full/html?fullSc=1>. Acesso em: 24 dez. 2019.

DEUTSCHE FORSCHUNGSGEMEINSCHAFT (DFG). **Guidelines for Safeguarding Good Research Practice**: code of conduct. 2019. Disponível em: https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp_en.pdf. Acesso em: 26 ago. 2020.

DIGITAL CURATION CENTRE. **Overview of funders' data policies**. [2012?] Disponível em: <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>. Acesso em: 30 ago. 2020.

DODSON, J. **Building partnerships of equals**: the role of funders in equitable and effective international development collaborations. London: UK Collaborative on Development Science, 2017.

DOUCETTE, Lise; FYFE, Bruce. Drowning in research data: addressing data management literacy of graduate students. *In*: MUELLER, Dawn M. (ed.). **Imagine, Innovate, Inspire**: the proceedings of the Acrl 2013 Conference. Indianapolis: Association of College and Research Libraries Conference, 2013. p. 165-171. Disponível em: http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/2013/papers/DoucetteFyfe_Drowning.pdf. Acesso em: 18 set. 2020.

ESTEVIÃO, Janete Saldanha. **Letramento informacional para o reuso de dados nas ciências sociais em ambientes virtuais de pesquisa**: proposta de requisitos e competências. 2019. Tese (Doutorado em Tecnologia e Sociedade) – Programa de Pós-graduação em Tecnologia e Sociedade, Universidade Tecnológica Federal do Paraná, 2019.

EUROPEAN COMMISSION. **Communication from the Commission to the European parliament, the Council, the European economic and social committee and the Committee of the regions**: Enhancing and focusing EU international cooperation in research and innovation: a strategic approach. Brussels, Sept. 2012a. Disponível em: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52012DC0497&from=EN>. Acesso em: 10 out. 2020.

_____. **Commission staff working document**: implementation Roadmap for the European Open Science Cloud. Brussels, Mar. 2018a. Disponível em: https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf#view=fit&pagemode=none. Acesso em: 15 out. 2020.

_____. **Guidelines on FAIR Data Management in Horizon 2020**. Version 3.0, Jul. 2016. Disponível em: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf. Acesso em: 11 out. 2020.

_____. **Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020**. Version 3.2, Mar. 2017a. Disponível em: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf. Acesso em: 13 out. 2020.

_____. **H2020 Programme**: AGA – Annotated Model Grant Agreement. Version 5.2, jun. 2019. Disponível em:

https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf. Acesso em: 20 out. 2020.

_____. **Horizon 2020 projects working on the 2019 coronavirus disease (COVID-19), the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and related topics:** Guidelines for open access to publications, data and other research outputs. Version 1.1, abr. 2020. Disponível em: https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-guidelines-oa-covid-19_en.pdf. Acesso em: 20 out. 2020.

_____. **Horizon 2020:** work programme 2016–2017. 2017b. Disponível em: https://ec.europa.eu/research/participants/data/ref/h2020/wp/2016_2017/main/h2020-wp1617-intro_en.pdf. Acesso em: 15 set. 2020.

_____. **Horizon 2020:** work programme 2016 – 2017. General Annexes. 2017c. Disponível em: https://ec.europa.eu/research/participants/data/ref/h2020/other/wp/2016-2017/annexes/h2020-wp1617-annex-ga_en.pdf. Acesso em: 19 set. 2020.

_____. **How to participate in Horizon 2020:** open to the world. 2014. Disponível em: <https://op.europa.eu/en/publication-detail/-/publication/7024b066-ef6c-4c46-8256-4c11ff0183aa/language-en/format-PDF/source-143421433>. Acesso em: 13 out. 2020.

_____. **Multi-Beneficiary General Model Grant Agreement (H2020 General MGA – Multi).** Version 5.0, Oct. 2017d. Disponível em: https://ec.europa.eu/research/participants/data/ref/h2020/mga/gga/h2020-mga-gga-multi_en.pdf. Acesso em: 20 out. 2020.

_____. **Online survey on scientific information in the digital age.** Luxembourg: Publications Office of the European Union, 2012b. Disponível em: https://ec.europa.eu/research/science-society/document_library/pdf_06/survey-on-scientific-information-digital-age_en.pdf. Acesso em: 06 jan. 2020.

_____. **Open Research Data (ORD):** the uptake in Horizon 2020. 2018b. Disponível em: <https://data.europa.eu/euodp/pt/data/dataset/open-research-data-the-uptake-of-the-pilot-in-the-first-calls-of-horizon-2020>. Acesso em: 11 out. 2020.

_____. **Turning FAIR into reality.** Luxembourg: Publications Office of the European Union, 2018c. Disponível em: https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf. Acesso em: 20 out. 2020.

EUROPEAN RESEARCH COUNCIL. **ERC Data Management Plan Template.** Version 1.0, Apr. 2017a. Disponível em: http://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-erc-tpl-oa-data-mgt-plan_en.odt. Acesso em: 20 set. 2020.

_____. **Guidelines on Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020.** Version 1.1. 21 Apr. 2017b. Disponível em: https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf. Acesso em: 20 set. 2020.

_____. **Multi-Beneficiary Model Grant Agreement:** ERC starting grants, consolidator grants and advanced grants. Version 5.0. 18 Oct. 2017c. Disponível em: http://ec.europa.eu/research/participants/data/ref/h2020/mga/erc/h2020-mgaerc-multi_en.pdf. Acesso em: 9 set. 2020.

_____. **Open Research Data and Data Management Plans Information for ERC grantees.** Version 3.1., Jul. 2019. Disponível em: https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf. Acesso em: 23 set. 2020.

EUROPEAN UNION. Regulamento (UE) n° 1290/2013 do Parlamento Europeu e do Conselho de 11 de dezembro de 2013, que estabelece as regras de participação e difusão relativas ao "Horizonte 2020 – Programa-Quadro de Investigação e Inovação (2014-2020)" e revoga o Regulamento (CE) n.o 1906/2006. **Jornal Oficial da União Europeia**, L 347, p. 81-103, dez. 2013a. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32013R1290&from=EN>. Acesso em: 14 out. 2020.

_____. Regulamento (UE) n° 1291/2013 do Parlamento Europeu e do Conselho de 11 de dezembro de 2013, que cria o Horizonte 2020 – Programa-Quadro de Investigação e Inovação (2014-2020) e que revoga a Decisão n.o 1982/2006/CE. **Jornal Oficial da União Europeia**, L 347, p. 104-173, dez. 2013b. Disponível em: https://ec.europa.eu/research/participants/data/ref/h2020/legal_basis/fp/h2020-eu-establact_pt.pdf. Acesso em: 15 out. 2020.

FECHER, Benedikt; FRIESIKE Sascha. Open Science: one term, five schools of thought. *In: BARTLING Sönke; FRIESIKE, Sascha (ed.). **Opening Science**: the evolving guide on how the internet is changing research, collaboration and scholarly publishing.* London: SpringerOpen, 2013. DOI: https://doi.org/10.1007/978-3-319-00026-8_14

FERREIRA, José Roberto *et al.* A experiência da Fundação Oswaldo Cruz em Relações Internacionais para o Desenvolvimento da Ciência e Tecnologia no Campo da Saúde. *In: BUSS, Paulo Marchiori; TOBAR, Sebastián. **Diplomacia em Saúde e Saúde Global**: perspectivas latino-americanas.* Rio de Janeiro: Editora Fiocruz, 2017.

FINKEL, M. *et al.* Managing collaborative research data for integrated, interdisciplinary environmental research. **Earth Science Informatics**, [S. l.], v. 1, p. 641–654, Jan. 2020. DOI: <https://doi.org/10.1007/s12145-020-00441-0>. Disponível em: <https://link.springer.com/article/10.1007/s12145-020-00441-0>. Acesso em: 23 dez. 2020.

FITZGERALD, A.; FITZGERALD, B.; PAPPALARDO, K. O futuro das políticas de dados. *In: HEY, T.; TANSLEY, S.; TOLLE, K.(org.). **O quarto paradigma**: descobertas científicas na era da eScience.* São Paulo: Oficina de textos, 2011. cap. 24, p. 210-216.

FONSECA E FONSECA, Bruna. **Colaboração como estratégia para instituições de ciência e tecnologia em saúde**: uma proposta de indicadores para análise organizacional. 2015. 239 f. Tese (Doutorado em Engenharia de Produção) – Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2015.

FORCE11. Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version b1.0 [2014?]. Disponível em: <https://www.force11.org/fairprinciples>. Acesso em: 14 out. 2020.

FRY, Jenny; SCHROEDER, Ralph; DEN BESTEN, Matthijs. Open science in e-science: contingency or policy? **Journal of Documentation**, [S. l.], v. 65, n. 1, p. 6-32, Ago. 2009. DOI: <https://doi.org/10.1108/00220410910926103>. Disponível em: https://repository.lboro.ac.uk/articles/Open_science_in_e-science_contingency_or_policy_/9416609/1. Acesso em: 10 jun. 2020.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. **Plano de gestão de dados**. São Paulo, 2017. Disponível em: <http://www.fapesp.br/gestaodedados>. Acesso em: 15 out. 2019.

GEORGHIOU, L. Global cooperation in research. **Research Policy**, Manchester, v. 27, n. 6, p. 611-626, Sep. 1998. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0048733398000547>. Acesso em: 12 dez. 2019.

GIFELS, Joe; VOLLMER, Sara H.; BIRD, Stephanie J. Editors' Overview: topics in the responsible management of research data. **Science and Engineering Ethics**, [S. l.], v. 16, p. 631-637, Oct. 2010. Disponível em: <https://link.springer.com/article/10.1007/s11948-010-9243-1>. Acesso em: 13 dez. 2019.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Editora Atlas S.A, 2012.

GOBEN Abigail; GRIFFIN, Tina. In aggregate: trends, needs, and opportunities from research data management surveys. **College & Research Libraries**, [S. l.], v. 80, n. 7, p. 903-924, Nov. 2019. DOI: <https://doi.org/10.5860/crl.80.7.903>. Disponível em: <https://crl.acrl.org/index.php/crl/article/view/23610>. Acesso em: 26 jun. 2020.

GONZÁLEZ DE GÓMEZ, M. N. Reflexões sobre ética da informação: panorama contemporâneo. In: GONZALEZ DE GOMEZ, M. N.; CIANCONI, R. de B. (org.). **Ética da Informação: Perspectivas e desafios**. Rio de Janeiro: Garamond, 2017. v. 1, cap.1, p. 19-44.

GRAY, Jim. Jim Gray on eScience: a transformed scientific method. Based on the transcript of a talk given by Jim Gray to the NRC-CSTB1 in Mountain View, CA, on January 11, 2007. In: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (org.). **The fourth paradigm: data-intensive scientific discovery**. Redmond, Washington: Microsoft Research, 2009.

GUIMARÃES, Maria Cristina Soares. Ciência aberta e livre acesso à informação científica: tão longe, tão perto. **RECIIS – Rev. Eletron. de Comun. Inf. Inov. Saúde**, Rio de Janeiro, v. 8, n. 2, p.139-152, jun. 2014. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/627>. Acesso em: 19 jun. 2020.

GUIMARÃES, Reinaldo. Pesquisa em saúde no Brasil: contexto e desafios. **Revista Saúde Pública**, [S. l.], v. 40, (N Esp), p. 3-10, abr. 2006. Disponível em: <https://www.scielo.br/pdf/rsp/v40nspe/30616.pdf>. Acesso em: 20 nov. 2020.

GUIMARÃES, Reinaldo; SOUZA Luís Eugênio Portela Fernandes de; SANTOS, Leonor Maria Pacheco. Ciência, Tecnologia e Pesquisa em Saúde. *In: GIOVANELLA, L., et al.* (org.). **Políticas e sistemas de saúde no Brasil** [online]. Rio de Janeiro: Editora FIOCRUZ, 2012. 2. ed. rev. e aum. ISBN 978-85-7541-349-4.

HEIDORN, Bryan P. Shedding light on the dark data in the long tail. **Library Trends**, [S. l.], v. 57, n. 2, p. 280-299, 2008. Disponível em: <https://arizona.pure.elsevier.com/en/publications/shedding-light-on-the-dark-data-in-the-long-tail-of-science>. Acesso em: 25 maio 2020.

HENNING, Patrícia Corrêa. Gestão de dados de pesquisa: uma demanda necessária para a geração de novos conhecimentos. **Revista de Pesquisa: Cuidado é Fundamental Online**, Rio de Janeiro, v. 11, n. 3, p. 1-2, abr. 2019. Disponível em: <http://www.seer.unirio.br/index.php/cuidadofundamental/article/view/8939>. Acesso em: 18 jun. 2019.

HENNING, Patrícia Corrêa. *et al.* Go fair e os princípios fair: o que representam para a expansão dos dados de pesquisa no âmbito da ciência aberta. **Em Questão**, v. 25, n. 2, p. 389-412, 2019. DOI: 10.19132/1808-5245252.389-412. Disponível em: <https://brapci.inf.br/index.php/res/v/113770>. Acesso em: 19 out. 2020.

HEY, Tony; TREFETHEN, Anne. The Data Deluge: an e-science perspective. *In: BERMAN, Fran; FOX, Geoffrey; HEY, Tony.* **Grid computing: making the global infrastructure a reality**. New York : J. Wiley, 2003. p. 809-824. DOI: <https://doi.org/10.1002/0470867167.ch36>. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/0470867167.ch36>. Acesso em: 20 jun. 2020.

HIGGINS, Sarah. The DCC Curation Lifecycle Model. **International Journal of Digital Curation**, Bath, v. 3, n. 1, p. 134-140, Dec. 2008. DOI: <https://doi.org/10.2218/ijdc.v3i1.48>. Disponível em: <http://www.ijdc.net/article/view/69>. Acesso em: 24 set. 2020.

HODSON, Simon; MOLLOY, Laura. **Current best practice for research data management policies**. 2015. DOI: <http://doi.org/10.5281/zenodo.27872>. Disponível em: <https://zenodo.org/record/27872#.YAGbmehKjIU>. Acesso em: 28 ago. 2020.

HODSON, Simon. Preface by Simon Hodson, Chair of the EC Expert Group on FAIR data. *In: EUROPEAN COMMISSION.* **Turning FAIR into reality**. Luxembourg: Publications Office of the European Union, 2018. Disponível em: https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf. Acesso em: 20 out. 2020.

HOLDREN, John P. Increasing Access to the Results of Federally Funded Scientific Research. **Memorandum for the Heads of Executive Departments and Agencies**, 22 fev. 2013. Disponível em: <https://obamawhitehouse.archives.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research>. Acesso em: 14 ago. 2020.

HORTON, L.; DIGITAL CURATION CENTRE. **Overview of UK Institution RDM Policies**, Digital Curation Centre, Version 6, Aug. 2016. Disponível em: <https://www.dcc.ac.uk/guidance/policy/institutional-data-policies>. Acesso em: 29 ago. 2020.

JANKOWSKI, Nicholas W. Exploring e-Science: an introduction. **Journal of Computer-Mediated Communication**, [S. l.], v. 12, n. 2, p. 549–562, Jan. 2007. DOI: <https://doi.org/10.1111/j.1083-6101.2007.00337.x>. Disponível em: <https://academic.oup.com/jcmc/article/12/2/549/4583044>. Acesso em: 15 jun. 2020.

JESUS, Elizabeth. **Molduras normativas e integridade da pesquisa em contexto de colaboração científica internacional**: uma perspectiva informacional. 2019. 303 f. Tese (Doutorado em Ciência da Informação) – Instituto de Arte e Comunicação Social, Universidade Federal Fluminense, Niterói, 2019.

JESUS, Elizabeth; GONZÁLEZ DE GÓMEZ, Maria Nélide. Integridade da pesquisa e molduras normativas em contextos de colaboração científica internacional. **Encontro Nacional de Pesquisa em Ciência da Informação**, n. XIX ENANCIB, 2018. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/103794>. Acesso em: 25 abr. 2020.

JONES, Sarah. Developments in research funder data policy. **International Journal of Digital Curation**, Bath, v. 7 n. 1, p. 114–125, Mar. 2012. DOI: <https://doi.org/10.2218/ijdc.v7i1.219>. Disponível em: <http://www.ijdc.net/article/view/209>. Acesso em: 3 jan. 2020.

JORGE, Vanessa. **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) – Escola de comunicação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2018.

KALLEBERG, Ragnvald. Scientific Misconduct, Plagiarism, and Institutional Control of Misconduct. In: WRIGHT, James D (org.). **International Encyclopedia of the Social & Behavioral Science**. 2nd ed. [S. l.]: Elsevier, 2015. p. 313-317.

KATZ, J. Sylvan; MARTIN, Ben R.; What is research collaboration? **Research Policy**, Brighton, [S. l.], v. 26, n. 1, p. 1-18, Mar. 1997. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0048733396009171>. Disponível em: 17 nov. 2019.

KOLTAY, Tibor. Research 2.0 and Research Data Services in academic and research libraries: priority issues. **Library Management**, [S. l.], v. 38, n.6/7, p. 345-353, Aug. 2017. DOI: <https://doi.org/10.1108/LM-11-2016-0082>. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/LM-11-2016-0082/full/html?skipTracking=true>. Acesso em: 10 out. 2020.

KRAHE, Michelle *et al.* A Research data management in practice: Results from a cross-sectional survey of health and medical researchers from an academic institution in Australia. **Health Information Management Journal**, [S. l.], v. 49, n. 2-3, p. 1-9, Mar. 2020. DOI: <https://doi.org/10.1177/1833358319831318>. Disponível em: <https://journals.sagepub.com/doi/10.1177/1833358319831318>. Acesso em: 19 jul. 2020.

LAKATOS, E. M.; MARCONI, M. A. **Fundamentos metodologia científica**. 4. ed. São Paulo: Atlas, 2001.

LATHAM, B. Research Data Management: defining roles, prioritizing services, and enumerating challenges. **The Journal of Academic Librarianship**, [S. l.], v. 43, n. 3, p. 263–265, May 2017. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0099133317301453>. Acesso em: 2 dez. 2019.

LEE, J. W. DataNet: an emerging cyberinfrastructure for sharing, reusing and preserving digital data for scientific discovery and learning. **AIChE Journal**, New York, v. 55, n. 11, p. 2757-2764, Nov. 2009. DOI: 10.1002/aic.12085. Disponível em: <https://deepblue.lib.umich.edu/handle/2027.42/64341>. Acesso em: 5 set. 2020.

LEWIS, M. Libraries and the management of research data. *In*: MCKNIGHT, S. (ed.). **Envisioning future academic library services: Initiatives, ideas and challenges**. London: Facet Publishing, 2010. p. 145-168.

LIMA, Ricardo de Arcanjo; VELHO, Lea Maria Leme Strini; FARIA, Leandro Innocentini Lopes de. Indicadores bibliométricos de cooperação científica internacional em bioprospecção. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 12, n. 1, p. 50-64, jan./abr. 2007. Disponível em: http://www.brapci.inf.br/_repositorio/2011/03/pdf_4500fe22f2_0015367.pdf. Acesso em: 10 dez. 2019.

LORD Philip; MACDONALD, Alison. **Data curation for e-Science in the UK: an audit to establish requirements for future curation and provision**. London: The JISC Committee for the Support of Research, 2003.

LUCAS, Adetokunbo O. International collaboration in health research. **Bulletin of the World Health Organization**, [S. l.], v. 83, n. 7, Jul. 2005. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2626293/>. Acesso em: 15 nov. 2020.

LUUKKONEN, T.; PERSSON, O.; SIVERTSEN, G. Understanding Patterns of International Scientific Collaboration. **Science, Technology, & Human Values**, [S. l.], v. 17, n. 1, p. 101-126, Jan. 1992. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/016224399201700106>. Acesso em: 19 dez. 2019.

LYON, L. **Dealing with data: roles, rights, responsibilities and relationships**. Bath: UKOLN, 2007. Disponível em: http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf. Acesso em: 15 dez. 2019.

MARTINSON, Brian C.; ANDERSON, Melissa S.; DE VRIES, Raymond. Scientists behaving badly. **Nature**, [S. l.], v. 435, n. 9, p. 737-738, June 2005. Disponível em: nature.com/articles/435737a. Acesso em: 4 dez. 2019.

MEADOWS, Arthur Jack. **A Comunicação científica**. Brasília, DF: Briquet de Lemos, 1999.

MERTON, Robert K. A ciência e a estrutura social democrática. *In*: _____. **Sociologia: teoria e estrutura**. São Paulo: Ed. Mestre Jou, 1968.

_____. Priorities in scientific discovery: a chapter in the sociology of science. **American Sociological Review**, [S. l.], v. 22, n. 6, p. 635-659, Dec. 1957. Disponível em: <https://www.jstor.org/stable/2089193?seq=1>. Acesso em: 20 nov. 2019.

MICHENER, William K. Ten simple rules for creating a good data management plan. **PLoS Comput Biol.**, v. 11, n. 10, Oct. 2015. DOI: 10.1371/journal.pcbi.1004525. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4619636/>. Acesso em: 13 out. 2020.

MITROFF, Ian. **The subjective side of Science**. Amsterdam, New York: Elsevier, 1974a.

_____. Norms and counter-norms in a select group of the Apollo moon scientists: a case study of the ambivalence of scientists. **American Sociological Review**, [S. l.], v. 39, n. 4, p. 579-595, Ago. 1974b. Disponível em: <https://www.jstor.org/stable/2094423?seq=1>. Acesso em: 15 nov. 2019.

MONTEIRO, Elizabete; SANT'ANA, Ricardo. Plano de gerenciamento de dados em repositórios de dados de universidades. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, [S. l.], v. 23, n. 53, p. 160-173, set./dez., 2018. DOI: <https://doi.org/10.5007/1518-2924.2018v23n53p160>. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v23n53p160>. Acesso em: 15 out. 2020.

MONTEIRO, Gabriela. **Mapeamento e análise das políticas institucionais de financiadores da pesquisa brasileira: cenário dos dados científicos abertos**. 2019. 146 f. Dissertação (Mestrado em Gestão da Informação) – Centro de Ciências Humanas e da Educação, Universidade do Estado de Santa Catarina, 2019.

MOREL, Regina Lucia de Moraes. **Ciência e Estado a Política Científica no Brasil**. São Paulo: T. A. Queiroz, 1979.

MUELLER, Suzana Pinheiro Machado. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ci. Inf.**, Brasília, v. 35, n. 2, p. 27-38, maio/ago. 2006. Disponível em: <http://revista.ibict.br/ciinf/article/view/1138>. Acesso em: 18 jun. 2020.

MULLINS, James L. The policy and institutional framework. In: RAY, J. **Research Data Management: Practical Strategies for Information Professionals**. West Lafayette: Purdue University Press, 2014. p. 25-44.

NASCIMENTO, Andrea; ARAÚJO, Luciana; ARELLANO, Miguel. Crise e oportunidades para a preservação digital da informação em saúde. **Reciis – Rev Eletron Comun Inf Inov Saúde**, Rio de Janeiro, v. 14, n.3, p. 520-3, jul./set. 2020. DOI: <https://doi.org/10.29397/reciis.v14i3.2223>. Disponível em: <https://www.reciiis.icict.fiocruz.br/index.php/reciis/article/view/2223>. Acesso em: 27 out. 2020.

NATIONAL ACADEMIES OF SCIENCES, ENGINEERING, AND MEDICINE. **Data Matters: ethics, data, and international research collaboration in a changing world: proceedings of a workshop**. Washington: The National Academies Press, 2018. DOI: <https://doi.org/10.17226/25214>.

_____. **Responsible science**: ensuring the integrity of the research process. Washington, DC: The National Academies Press, 1992. v.1.

NATIONAL INSTITUTES OF HEALTH (NIH). **About the Office of Extramural Research (OER) at NIH**. Dec. 2020a. Disponível em: <https://grants.nih.gov/aboutoer/intro2oer.htm>. Acesso em: 10 dez. 2020.

_____. **About us**. [2020?]. Disponível em: <https://osp.od.nih.gov/about-us/>. Acesso em: 15 nov. 2020.

_____. **Data repositories resource guide**. Set. 2012. Disponível em: https://grants.nih.gov/grants/data_repositories_samples.doc. Acesso em: 12 nov. 2020.

_____. **Data standards and common data elements resource guide**. Set. 2011. Disponível em: https://grants.nih.gov/grants/sharing_data_standards_guide.doc. Acesso em: 12 nov. 2020.

_____. **Final NIH Policy for data management and sharing**. Oct. 2020b. Disponível em: <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html>. Acesso em: 10 nov. 2020.

_____. **Final NIH Statement on Sharing Research Data**. Feb. 2003a. Disponível em: <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>. Acesso em: 7 nov. 2020.

_____. **Funding**. Mar. 2019a. Disponível em: <https://grants.nih.gov/funding/index.htm>. Acesso em: 2 nov. 2020.

_____. **Genomic Data Sharing Policy**. Aug. 2014. Disponível em: https://osp.od.nih.gov/wp-content/uploads/NIH_GDS_Policy.pdf. Acesso em: 20 nov. 2020.

_____. **Guidance for investigators in developing genomic data sharing plans**. Jul. 2015a. Disponível em: https://osp.od.nih.gov/wp-content/uploads/NIH_Guidance_Developing-GDS_Plans.pdf. Acesso em: 19 nov. 2020.

_____. **Information for Foreign Grants**. Jun. 2018. Disponível em: <https://grants.nih.gov/grants/foreign/index.htm>. Acesso em: 3 nov. 2020.

_____. **International Applications**. Jan. 2021. Disponível em: <https://www.niaid.nih.gov/grants-contracts/international-applications>. Acesso em: 15 jan. 2021.

_____. **Key elements to consider in preparing a data sharing plan under NIH extramural support**. Dec. 2009. Disponível em: https://grants.nih.gov/grants/sharing_key_elements_data_sharing_plan.pdf. Acesso em: 24 nov. 2020.

_____. **List of NIH Institutes, Centers, and Offices**. Aug. 2019b. Disponível em: <https://www.nih.gov/institutes-nih/list-nih-institutes-centers-offices>. Acesso em: 10 nov. 2020.

_____. **NIH Announces draft statement on sharing research data.** Mar. 2002. Disponível em: <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-02-035.html>. Acesso em: 5 nov. 2020.

_____. **NIH Data Sharing Policy and Implementation Guidance.** Mar. 2003b. Disponível em: https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm. Acesso em: 23 ago 2020.

_____. **[NIH Genome-Wide Association Studies (GWAS) Policy]**. Aug. 2007. Disponível em: <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html>. Acesso em: 11 nov. 2020.

_____. **NIH Grants Policy Statement.** [S. l.]: National Institutes of Health, 2019c. Disponível em: <https://grants.nih.gov/grants/policy/nihgps/nihgps.pdf>. Acesso em: 10 nov. 2020.

_____. **Plan for increasing access to scientific publications and digital scientific data from NIH funded scientific research.** Feb. 2015b. Disponível em: <https://grants.nih.gov/grants/nih-public-access-plan.pdf>. Acesso em: 15 jun. 2020.

_____. Public health service policy relating to distribution of unique research resources produced with phs funding. **NIH Guide**, v. 25, n. 23, Jul. 1996. Disponível em: <https://grants.nih.gov/grants/guide/notice-files/not96-184.html>. Acesso em: 4 nov. 2020.

NATIONAL RESEARCH COUNCIL (NRC). **A question of balance: private rights and the public interest in scientific and technical databases.** Washington: National Academy Press, 1999. Disponível em: <https://www.nap.edu/catalog/9692/a-question-of-balance-private-rights-and-the-public-interest>. Acesso em: 15 dez. 2019.

_____. **Access to research data in the 21st century: an ongoing dialogue among interested parties: report of a workshop.** Washington, D.C.: National Academies Press, 2002.

NATIONAL SCIENCE AND TECHNOLOGY COUNCIL. **Principles for promoting access to federal government-supported scientific data and research findings through international scientific cooperation.** [Washington, DC]: OSTP, 2016. Disponível em: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/NSTC/iwgodsp_principles_0.pdf. Acesso em: 18 nov. 2020.

NATIONAL SCIENCE FOUNDATION (NSF). **Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science**, 2015. Disponível em: https://www.nsf.gov/sbe/AC_Materials/SBE_Robust_and_Reliable_Research_Report.pdf. Acesso em: 03 jan. 2020.

NEYLON, Cameron. Building a culture of data sharing: policy design and implementation for research data management in development research. **Research Ideas and Outcomes**, [S. l.], v. 3, Oct. 2017. DOI: <https://doi.org/10.3897/rio.3.e21773>. Disponível em: <https://riojournal.com/article/21773/list/9/>. Acesso em: 3 jun. 2020.

OFFICE OF SCIENCE AND TECHNOLOGY POLICE. **Federal Research Misconduct Policy**. 2000. Disponível em: <https://ori.hhs.gov/federal-research-misconduct-policy>. Acesso em: 28 dez. 2019.

OLIVEIRA, Juliane *et al.* The global scientific research response to the public health emergency of Zika virus infection. **PLoS ONE**, [S. l.], v. 15, n. 3, p. 1-16, Mar. 2020. DOI: <https://doi.org/10.1371/journal.pone.0229790>. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0229790>. Acesso em: 12 jan. 2021.

OPENAIRE. **Aggregation and content provision workflows**. 2020. Disponível em: <https://www.openaire.eu/aggregation-and-content-provision-workflows#:~:text=What%20kind%20of%20data%20sources%20are%20in%20OpenAIRE%3F&text=Aggregator%20services%3A%20Information%20systems%20that,discovery%20of%20given%20research%20products>. Acesso em: 20 out. 2020.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). **OECD principles and guidelines for access to research data from public funding**. 2007. Disponível em: <https://www.oecd.org/sti/inno/38500813.pdf>. Acesso em: 20 jun. 2019.

_____. **Enhanced access to publicly funded data for science, technology and innovation**. Paris: OECD Publishing, 2020.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. **Relatório mundial da saúde 2013: pesquisa para a cobertura universal de saúde**. Geneva: Organização Mundial da Saúde. 2014. Disponível em: <https://apps.who.int/iris/handle/10665/85761>. Acesso em: 25 nov. 2020.

PAMPEL, Heinz; DALLMEIER-TIESSSEN, Sünje. Open Research Data: From Vision to Practice. In: BARTLING Sönke; FRIESIKE, Sascha (ed.). **Opening Science: the evolving guide on how the internet is changing research, collaboration and scholarly publishing**. London: SpringerOpen, 2013. DOI: https://doi.org/10.1007/978-3-319-00026-8_14

PAMPEL, Heinz *et al.* Making Research Data Repositories Visible: The re3data.org Registry. **PLoS ONE**, v. 8, n. 11, Nov. 2013. DOI: <https://doi.org/10.1371/journal.pone.0078080>. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0078080>. Acesso em: 15 out. 2020.

PARKER, Michael; KINGORI, Patricia. Good and bad research collaborations: researchers' views on science and ethics in global health research. **PLoS ONE**, [S. l.], v. 11, n. 10, p. 1-19, Oct. 2016. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0163579>. Acesso em: 3 dez. 2019.

PENNOCK, M. Digital Curation: a life-cycle approach to managing and preserving usable digital information. **Library & Archives**, Bath, n.1, Jan. 2007. Disponível em: https://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf. Acesso em: 7 jan. 2020.

PIMPLE, Kenneth D. Six domains of research ethics: a heuristic framework for the responsible conduct of research. **Science and Engineering Ethics**, Guildford, v. 8, n. 2, p.

191-205, Jun. 2002. Disponível em: <https://link.springer.com/content/pdf/10.1007/s11948-002-0018-1.pdf>. Acesso em: 28 dez. 2019.

PINFIELD, S; COX, A. M.; SMITH, J. Research data management and libraries: relationships, activities, drivers and influences. **PLoS ONE**, [S. l.], v. 9, n. 12, p. 1-28, Dec. 2014. DOI: <https://doi.org/10.1371/journal.pone.0114734>. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0114734#pone.0114734-Cox>. Acesso em: 04 jun. 2020.

PISANI, Elizabeth; ABOUZAHAR, Carla. Sharing health data: good intentions are not enough. **Bulletin of the World Health Organization**, Genebra, v. 88, n. 462, p. 462–466, 2010. DOI: 10.2471/BLT.09.074393. Disponível em: <https://www.who.int/bulletin/volumes/88/6/09-074393/en/>. Acesso em: 18 ago. 2020.

PISANI, Elizabeth *et al.* **Sharing health research data in low-resource settings**: supporting necessary infrastructure and building on good practices. A report submitted to the World Health Organization's Special Programme for Research and Training in Tropical Diseases (WHO/TDR) and the Wellcome Trust, 2018. Disponível em: https://wellcome.figshare.com/articles/journal_contribution/Sharing_health_research_data_in_low-resource_settings_Supporting_necessary_infrastructure_and_building_on_good_practices/6042047. Acesso em: 20 jun. 2020.

PIWOWAR, H. A.; VISION, T. J. Data reuse and the open data citation advantage. **Peer J**, San Diego, v. 1, e175, Oct. 2013. DOI: <https://doi.org/10.7717/peerj.175>. Disponível em: <https://peerj.com/articles/175/>. Acesso em: 20 set. 2020.

PRICE, Derek J. de Solla. **Little Science, Big Science – and Beyond**. New York: Columbia University Press, 1986.

PRÍNCIPE, Pedro *et al.* **Relatório técnico sobre ferramentas para a elaboração de planos de gestão de dados**. [Braga]: Universidade do Minho, 2020.

PRYOR, Graham. Why manage research data? *In*: PRYOR, Graham. (ed.) **Managing Research Data**. London: Facet Publishing, 2012.

_____. A patchwork of change. *In*: PRYOR, Graham; JONES, Sarah; WHYTE, Angus (ed.). **Delivering Research Data Management Services: Fundamentals of good practice**. London: Facet Publishing, 2014. cap. 1, p. 1-19.

RAY, J. **Research Data Management**: practical strategies for information professionals. West Lafayette: Purdue University Press, 2014.

READ, Kevin B. A two-tiered curriculum to improve data management practices for researchers. **PLoS ONE**, [S. l.], v. 14, n. 5, p. 1-14, may 2019. DOI: <https://doi.org/10.1371/journal.pone.0215509>. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0215509>. Acesso em: 02 nov. 2019.

RESEARCH DATA ALLIANCE. **About RDA**. 2016. Disponível em: <https://www.rd-alliance.org/about-rda>. Acesso em: 18 nov. 2020.

RIBEIRO, Helena. **Saúde Global: olhares do presente**. Rio de Janeiro: Editora Fiocruz, 2016.

ROYAL NETHERLANDS ACADEMY OF ARTS AND SCIENCES. **Responsible Research Data Management and the Prevention of Scientific Misconduct**. Amsterdam: KNAW, 2013.

ROYAL SOCIETY (THE). **Knowledge, networks and nations: global scientific collaboration in 21st century**. London: Elsevier, 2011. Disponível em: https://royalsociety.org/~media/Royal_Society_Content/policy/publications/2011/4294976134.pdf. Acesso em: 15 nov. 2019.

RUSSO, Marisa. Ética e integridade na ciência: da responsabilidade do cientista à responsabilidade coletiva. **Estud. av.**, São Paulo, v. 28, n. 80, p. 189-198, abr. 2014. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-40142014000100016&lng=en&nrm=iso. Acesso em: 02 jan. 2020.

SALES, Luana Farias; SAYÃO, Luís Fernando. A grande e a pequena ciência: análise das diferenças na gestão de dados de pesquisa. **Inf. & Soc.:Est.**, João Pessoa, v. 29, n. 3, p. 151-170, jul./set. 2019a. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/47615/27371>. Acesso em: 13 out. 2020.

_____. Uma proposta de taxonomia para dados de pesquisa. **Revista Conhecimento em Ação**, [S. l.], v. 4, n. 1, p. 31-48, jan/jun. 2019b. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/118944>. Acesso em: 19 jul. 2020.

SALLANS, Andrew; LAKE, Sherry. Data Management Assessment and Planning Tools. *In*: RAY, J. **Research Data Management: Practical Strategies for Information Professionals**. West Lafayette: Purdue University Press, 2014. p. 87-107

SANTOS, Paula X. *et al.* **Livro verde – ciência aberta e dados abertos: mapeamento e análise de políticas, infraestruturas e estratégias em perspectiva nacional e internacional**. Rio de Janeiro: Fiocruz, 2017.

SAYÃO, Luis Fernando; SALES, Luana Farias. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Inf. Inf.**, Londrina, v. 21, n. 2, p. 90-115, maio/ago., 2016. DOI: <http://dx.doi.org/10.5433/1981-8920.2016v21n2p90>. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939>. Acesso em: 16 set. 2020.

_____. A ciência invisível: revelando os dados da cauda longa da pesquisa. **Encontro Nacional de Pesquisa em Ciência da Informação**, n. XIX ENANCIB, 2018. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/103678>. Acesso em: 28 ago. 2020.

_____. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade: Estudos**, João Pessoa, v.22, n.3, p. 179-191, set./dez. 2012.

Disponível em: <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/12224>. Acesso em: 10 nov. 2019.

_____. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: Comissão de Energia Nuclear, 2015.

SCHMILLEN, Hanna. Library and information science education and escience: the current state of ALA accredited MLS/MLIS programs in preparing librarians and information professionals for escience needs. **Library and Information Science: student capstone projects**, [S. l.], Paper 1, 2015. Disponível em: https://digitalcommons.du.edu/lis_capstone/1/. Acesso em: maio 2016.

SCHÖPFEL, Joachim *et al.* Research data management in the French National Research Center (CNRS). **Data Technologies and Applications**, Bingley, v. 52, n. 2, Apr. 2018. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/DTA-01-2017-0005/full/html>. Acesso em: 20 nov. 2019.

SCHWARTZMAN, Simon. **Ciência e história da ciência**. [Rio de Janeiro]: FINEP, Grupo de Estudos sobre o Desenvolvimento da Ciência, 1976. Mimeografado.

SCIENCE EUROPE. **Practical guide to the international alignment of research data management**. Belgium: Science Europe, 2018.

SEMELER, Alexandre; PINTO, Adilson. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. **Ci.Inf.**, Brasília, DF, v.48, n.1, p.113-129, jan./abr. 2019. Disponível em: <http://revista.ibict.br/ciinf/article/view/4461>. Acesso em: 11 ago. 2020.

SHAMOO, A.; RESNIK, D. **Responsible Conduct of Research**. 2 ed. New York: Oxford University Press, 2015.

SHEARER, Kathleen. **Comprehensive brief on research data management policies**. [Government of Canada], 2015. Disponível em: <http://docplayer.net/17594465-Comprehensive-brief-on-research-data-management-policies.html>. Acesso em: 26 fev. 2020.

SILVA, Darly Henriques da. Cooperação internacional em ciência e tecnologia: oportunidades e riscos. **Revista Brasileira de Política Internacional**, Brasília, v. 50, n. 1, p. 5-28, jun. 2007. DOI: <https://doi.org/10.1590/S0034-73292007000100001>. Disponível em: https://www.scielo.br/scielo.php?pid=S0034-73292007000100001&script=sci_abstract&tlng=pt. Acesso em: 20 fev. 2020.

SONNENWALD, Diane. H. Scientific Collaboration. **Annual Review of Information Science and Techonology**, [S. l.], v. 41, n.1, p. 643-681, Oct. 2007. Disponível em: https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/aris.2007.1440410121?casa_token=TMicjw0-IHEAAAAA%3ATdE_Lvc1RFA0A7waJJzu3Xre9i8yIU-oMqRieQCCh4OQXZIZg2dP4hPaPcyu91Baq6moaY-TVEN. Acesso em: 3 dez. 2020.

SPICHTINGER, Daniel; SIREN, Jarkko. The development os research data management policies in Horizon 2020. *In*: KRUSE, Filip; THESTRUP, Jesper Boserup. **Research Data Management: a European Perspective**. Berlin: Walter de Gruyter, 2018.

STEHR, Nico. The Ethos of Science Revisited: Social and Cognitive Norms. **Sociological Inquiry**, [S. l.], v. 48, n. 3/4, p. 172–196, Jun. 1978. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-682X.1978.tb00825.x>. Acesso em: 27 nov. 2019.

STENECK, N. Fostering integrity in research: definitions, current knowledge and future directions. **Science and Engineering Ethics**, Guildford, v. 12, n. 1, p. 53–74, Mar. 2006. Disponível em: <https://link.springer.com/article/10.1007/PL00022268>. Acesso em: 29 nov. 2019.

_____. **ORI Introduction to the Responsible Conduct of Research**. Washington, DC: ORI, 2007.

STORER, Norman. A internacionalidade da ciência e a nacionalidade dos cientistas. In: BENDAVID, Joseph *et al.* **Sociologia da ciência**. Rio de Janeiro: Ed. FGV, 1975. p. 115-136.

TENOPIR, Carol *et al.* Data sharing by scientists: Practices and perceptions. **PLoS ONE**, [S. l.], v. 6, n. 6, p. 1-21, Jun. 2011. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101>. Acesso em: 19 dez. 2019.

TENOPIR, Carol *et al.* Research data management services in academic research libraries and perceptions of librarians. **Library & Information Science Research**, [S. l.], v. 36, n. 2, p. 84-90, Apr. 2014. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0740818814000255>. Acesso em: 26 dez. 2019.

TERRY, Robert F.; LITTLER, Katherine; OLLIARO, Piero L. Sharing health research data – the role of funders in improving the impact. **F1000Research**, [S. l.], v. 7, n. 1641, Dec. 2018. DOI: <https://doi.org/10.12688/f1000research.16523.2>. Acesso em: <https://f1000research.com/articles/7-1641>. Acesso em: 30 jun. 2020.

TRELOAR, A.; CHOUDHURY, G.S.; MICHENER, W. Contrasting national research data strategies: Australia and the USA. In: PRYOR, G. (ed.) **Managing Research Data**. London: Facet Publishing, 2012. cap. 9, p. 173-203.

TRIVIÑOS, Augusto N. S. **Introdução à pesquisa em ciências sociais: a pesquisa qualitativa em educação**. São Paulo: Atlas, 1987.

UK RESEARCH INTEGRITY OFFICE. **Code of practice for research: promoting good practice and preventing misconduct**. 2009. Disponível em: <https://ukrio.org/wp-content/uploads/UKRIO-Code-of-Practice-for-Research.pdf>. Acesso em: 26 dez. 2019.

VANNI, T. *et al.* International Scientific Collaboration in HIV and HPV: A Network Analysis. **PLoS ONE**, [S. l.] v. 9, n. 3, Mar. 2014. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0093376>. Acesso em: 18 dez. 2019.

VANZ, Samile. **As redes de colaboração científica no Brasil (2004-2006)**. 2009. 204 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Biblioteconomia e

Documentação, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, 2009.

VASCONCELOS, Sonia M. R. *et al.* The new geography of scientific collaborations: Changing patterns in the geography of science pose ethical challenges for collaborations between established and emerging scientific powers. **EMBO reports**, [S. l.], v. 13, n.5, p.404-407, May 2012. Disponível em: <https://www.embopress.org/doi/10.1038/embor.2012.51>. Acesso em: 15 dez. 2019.

VEIGA, Viviane. **A 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.

VEIGA, Viviane *et al.* Plano de gestão de dados FAIR: uma proposta para a Fiocruz. **Liinc em Revista**, Rio de Janeiro, v.15, n.2, p. 275-286, nov. 2019. Disponível em: <http://revista.ibict.br/liinc/article/view/5030/4349>. Acesso em: 13 jan. 2021.

VEIGA, Viviane; QUEIROZ, Claudete. Rede GO FAIR Brasil Saúde: uma rede de apoio à Gestão e Abertura de Dados de Pesquisa em Saúde no Brasil. *In: ENCONTRO IBÉRICO EDICIC*, 9., Barcelona, 2019. 1 p.

VELHO, Lea. Conceitos de Ciência e a Política Científica, Tecnológica e de Inovação. **Sociologias**, Porto Alegre, ano 13, n. 26, p. 128-153, jan./abr. 2011. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1517-45222011000100006. Acesso em: 20 nov. 2019.

_____. Redes regionais de cooperação em C&T e o Mercosul. **Parcerias Estratégicas**, n. 10, p. 58-74, mar. 2001. Disponível em: <http://www.bibliotecadigital.unicamp.br/document/?down=109529>. Acesso em: 20 mar. 2020.

VERGARA, Sylvia Constant. **Projetos e relatórios de pesquisa em administração**. 13. ed. São Paulo: Atlas, 2011.

VICENT-SAEZ, Ruben; MARTINEZ-FUENTES, Clara. Open Science now: a systematic literature review for an integrated definition. **Journal of Business Research**, [S. l.], v. 88, Jul. 2018. DOI: <https://doi.org/10.1016/j.jbusres.2017.12.043>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0148296317305441>. Acesso em: 23 out. 2020.

VIDOTTI, Silvana *et al.* Repositório de dados de pesquisa para grupo de pesquisa: um projeto piloto. **Informação & Tecnologia (ITEC)**, João Pessoa, v. 4, n. 2, p. 221-242, jul./dez. 2017. DOI: <https://doi.org/10.22478/ufpb.2358-3908.2017v4n2.40198>. Disponível em: <https://periodicos.ufpb.br/index.php/itec/article/view/40198>. Acesso em: 20 out. 2020.

VIERGEVER, Roderik F.; HENDRIKS, Thom C. C. The 10 largest public and philanthropic funders of health research in the world: what they fund and how they distribute their funds. **Health Research Policy and Systems**, [S. l.], v. 14, n.18, Feb. 2016. Disponível em:

<https://health-policy-systems.biomedcentral.com/articles/10.1186/s12961-015-0074-z>. Acesso em: 6 jan; 2020.

WAGNER, Carolina *et al.* **Science and Technology Collaboration: building capacity in developing countries?** California: RAND Science and Technology, 2001.

WALLIS, Jillian C. **The distribution of data management responsibility within scientific research groups.** 2012. 242 f. These (Doctorate of Philosophy in Information Studies) – University of California, Los Angeles, 2012.

WANG, Wei Min; GÖPFERT, Tobias; STARK, Rainer. Data Management in Collaborative Interdisciplinary Research Projects—Conclusions from the Digitalization of Research in Sustainable Manufacturing. **ISPRS Int. J. Geo-Inf.**, v. 5, n. 41, Jan. 2016. DOI: <https://doi.org/10.3390/ijgi5040041>. Disponível em: <https://www.mdpi.com/2220-9964/5/4/41>. Acesso em: 15 abr. 2020.

WEITZEL, Simone. O papel dos repositórios institucionais e temáticos na estrutura da produção científica. **Em Questão**, v. 12, n. 1, p. 51-71, jan./jun. 2006. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/11493>. Acesso em: 20 out. 2020.

WELLCOME TRUST. **Sharing research data to improve public health: full joint statement by funders of health research.** [2012]. Disponível em: <https://wellcome.org/what-we-do/our-work/sharing-research-data-improve-public-health-full-joint-statement-funders-health>. Acesso em: 20 nov. 2020.

_____. **Statement on data sharing in public health emergencies.** 2016. Disponível em: <https://wellcome.org/press-release/statement-data-sharing-public-health-emergencies>. Acesso em: 20 nov. 2020.

WILBANKS, J. I have seen the paradigm shift and it is us. *In*: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (org.). **The fourth paradigm: data-intensive scientific discovery.** Redmond, Washington: Microsoft Research, 2009.

WILKINSON, M. D. *et al.* The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, [S. l.], v. 3, n. 160018, Mar. 2016. DOI: 10.1038/sdata.2016.18. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/26978244/>. Acesso em: 20 out. 2020.

WILLIAM, M.; BAGWELL, J.; ZOZUS, M. Data management plans: the missing perspective. **Journal of Biomedical Informatics**, [S. l.], v. 71, p. 130-142, Jul. 2017. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1532046417300990>. Acesso em: 03 jan. 2020.

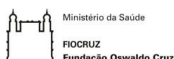
WHYTE, Angus. Emerging infrastructure and services for research data management and curation in the UK and Europe. *In*: PRYOR, G. (ed.) **Managing Research Data.** London: Facet Publishing, 2012. cap. 10, p. 205-234.

WHYTE, Angus; TEDDS, Jonathan. **Making the case for research data management.** DCC Briefing Papers. Edinburgh: Digital Curation Centre, 2011. Disponível em: <http://www.dcc.ac.uk/sites/default/files/documents/publications/Making%20the%20case.pdf>. Acesso em: 26 dez. 2019.

WORLD CONFERENCE ON RESEARCH INTEGRITY. **Montreal statement on research Integrity**. 2013. Disponível em: <https://wcrif.org/montreal-statement/file>. Acesso em: 18 nov. 2019.

ZIMAN, John Michael. **Conhecimento público**. Belo Horizonte: Ed. Itatiaia; São Paulo: Ed. da USP, 1979.

APÊNDICE A – Plano de Gestão de Dados

PLANO DE GESTÃO DE DADOS

Gestão de dados de pesquisa em saúde no contexto da colaboração científica internacional: um olhar a partir das políticas dos financiadores

Última atualização: 03/03/2021
- 12:47ID: 235

Dados do Pesquisador Principal

Nome do pesquisador principal:

Luisi Oliveira

E-mail do pesquisador principal:

luisi.costa43@gmail.com

Link do currículo Lattes:

<http://lattes.cnpq.br/2201050386747649>

Link do identificador persistente:

<https://orcid.org/0000-0002-1145-6754>

Afiliação:

Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICICT)

Etapa do Plano de Gestão de Dados

Durante a execução do projeto

Resumo do projeto

No contexto contemporâneo da ciência, os dados de pesquisa têm assumido um protagonismo central no processo de comunicação na ciência, dado o seu potencial de apoiar a transparência, reprodutibilidade na pesquisa e possível reuso para nova produção de conhecimento. No campo da saúde, há pelo menos uma década, financiadores internacionais de pesquisa começaram a exigir o compartilhamento de dados de maneira a maximizar os benefícios à saúde e à sociedade. Há um número crescente de recomendações e exigências sobre a gestão de dados nascidos de projetos financiados por agências de fomento à pesquisa, com o objetivo de tornar os dados compartilháveis e reutilizáveis por outros pesquisadores. Em contextos de colaboração científica internacional, que caracterizam a pesquisa do século XXI, a gestão de dados de pesquisa tornou-se um dos tópicos durante o processo de pesquisa que precisa ser reconhecido e negociado entre os envolvidos neste empreendimento científico para atender aos requisitos de diferentes sistemas de pesquisa a fim de garantir a qualidade e integridade da pesquisa. Nessa perspectiva, este projeto de pesquisa, de caráter exploratório-descritivo com abordagem qualitativa, teve como objetivo principal investigar as

orientações dos financiadores da pesquisa para gestão de dados produzidos em contexto de colaboração científica internacional no campo da saúde. A metodologia proposta consistiu na pesquisa documental para mapear os documentos normativos explícitos de agências e instituições de fomento à pesquisa em saúde que fornecem orientações sobre a gestão de dados de pesquisa, a fim de identificar as recomendações e exigências que estão sendo endereçadas aos pesquisadores envolvidos em projetos colaborativos internacionais. A análise dos documentos baseou-se em uma estrutura de avaliação de políticas de dados de pesquisa, adaptada a partir de modelos presentes na literatura, que contemplou as categorias grau de aplicação do documento e cobertura de critérios. A expectativa dessa dissertação foi de fornecer um panorama das recomendações e exigências relacionadas à gestão de dados de pesquisa estabelecidas pelos financiadores de pesquisa em saúde, contribuindo para um maior conhecimento a respeito das políticas e/ou normativas que precisam ser atendidas para garantir os compromissos e responsabilidades assumidos para o tratamento dos dados gerados no âmbito de pesquisas colaborativas internacionais.

Palavras-chave:

- Gestão de dados de pesquisa
- Acordos de Cooperação Científica e Tecnológica
- Dados de pesquisa
- Políticas de gestão de dados

O projeto tem financiamento?

Sim

- CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

Descrição dos dados coletados ou reúso de dados existentes

Seus dados serão coletados e/ou Produzidos:

Sim

Explique quais metodologias ou softwares serão utilizados se novos dados forem coletados ou produzidos:

Os dados da pesquisa foram coletados por meio de pesquisa documental. Os dados provenientes da pesquisa documental foram extraídos de documentos normativos emitidos pelos financiadores da pesquisa em saúde (National Institutes of Health e European Commission) com políticas para gestão dos dados de pesquisa. A pesquisa documental foi realizada nos sítios eletrônicos dos financiadores e as estratégias de busca empregadas para localizar esses documentos foram: buscas pelos termos 'data sharing', 'data management' ou 'data policies' e consultas na seção de políticas de cada financiador. Os documentos recuperados são de formato DOC e PDF, mas também foram identificadas políticas e normativas no próprio sítio eletrônico do financiador em linguagem HTML. Para cada financiador foi criado um arquivo em formato XLS, no software Excel (2010), com planilhas referentes à análise de cada documento recuperado, além de gráficos e quadros gerados a partir das análises realizadas. As planilhas referentes aos documentos analisados contêm metadados referentes ao nome do financiador, título do documento, tipologia documental, data de publicação e endereço eletrônico do documento analisado bem como a estrutura de avaliação de políticas de dados

de pesquisa, que contemplou as seguintes subcategorias: a) Grau de aplicação do documento: recomendatório, mandatório e não identificado; b) Cobertura de critérios: aspectos gerais; planejamento e infraestrutura e; disponibilização e reuso. Abaixo da estrutura de avaliação das políticas, foram dispostas as orientações específicas para a gestão de dados produzidos em contextos de colaboração científica internacional quando identificadas durante a análise dos documentos. Os arquivos gerados na pesquisa documental foram intitulados como "documentos_normativos_NIH_202101" e "documentos_normativos_EC_202101".

Seus dados serão reutilizados de outras fontes:

Não

Quais os tipos de dados que serão coletados, produzidos ou reutilizados?

- Textuais

Quais os formatos de dados que serão coletados, produzidos ou reutilizados?

- CSV - Comma-Separated Values
- DOC - Document Format
- PDF - Portable Document Format
- XLS - Excel Binary File Format

Qual o volume aproximado dos dados coletado, produzidos ou reutilizados:

Maior que 10 GB e menor que 100 GB

Documentação e Qualidade dos Dados

Indique os metadados específicos, se houver:

Os metadados usados para localizar os dados serão os utilizados pelo repositório de dados Zenodo, ambiente onde os dados serão depositados. Os metadados específicos utilizados nos arquivos de dados foram: nome do financiador; título do documento; tipologia documental; data de publicação; endereço eletrônico e; categorias de análise das políticas de dados de pesquisa, que foi dividida em duas subcategorias: 1) grau de aplicação do documento: a) recomendatório; b) mandatório; c) não identificado. 2) cobertura de critérios: a) aspectos gerais: i) definição; ii) escopo; iii) responsabilidades; iv) monitoramento de sanções e conformidades. b) planejamento e infraestrutura: i) plano de gestão de dados (PGD); ii) suporte; iii) curadoria/preservação por longo prazo; iv) padrões de metadados; v) repositórios e centros de dados; vi) custo. c) disponibilização e reuso: i) limite de tempo; ii) propriedade; iii) compartilhamento/acesso; iv) reuso; v) ética.

Armazenamento e Backup durante o processo de pesquisa

Onde os dados serão armazenados:

- Dispositivos de Armazenamento Externos como pendrive
- Laptops
- PCs
- Serviços de Nuvem Privado

Requisitos Legais e Éticos

A equipe tem ciência das normas relacionadas à avaliação ética na pesquisa:

Sim

A pesquisa envolve tratamento de dados em que se exija avaliação ética:

Não

A equipe está familiarizada com a legislação sobre o tratamento e proteção dos dados pessoais:

Sim

A pesquisa envolve tratamento de dados pessoais:

Não

Há previsão de período de embargo relacionado ao direito autoral:

Sim

A pesquisa envolve outras categorias de dados que exijam período de embargo para sua disponibilização:

Não

Informe a licença a ser aplicada à base de dados:

CC-BY - Creative Commons Attribution 4.0 International

Compartilhamento de Dados e Preservação a longo prazo

A pesquisa prevê compartilhamento de dados:

Sim

Como e quando os dados serão compartilhados:

O compartilhamento dos dados foi realizado no repositório Zenodo, durante o processo da pesquisa.

Existem possíveis motivações para submeter os dados à períodos de embargo no compartilhamento de dados:

Sim

Qual motivo para embargo:

- Outro

Outro motivo para embargo:

O conjunto de dados depositado no repositório Zenodo estará em acesso fechado até a aprovação da dissertação para validação da pesquisa e publicação de artigo científico para a comunicação dos resultados da pesquisa.

Qual o tempo estimado de embargo:

6 meses

ANEXO A – Componentes do PGD FAIR para projetos do Horizon 2020

| Componente PGD | Questões a serem tratadas |
|--|---|
| 1. Resumo de dados | <ul style="list-style-type: none"> • Indique o propósito da coleta / geração de dados; • Explique a relação com os objetivos do projeto; • Especifique os tipos e formatos de dados gerados / coletados; • Especifique se os dados existentes estão sendo reutilizados (se houver); • Especifique a origem dos dados; • Indique o tamanho esperado dos dados (se conhecido); • Descreva a utilidade de dados: para quem será útil; |
| 2. Dados FAIR 2.1. Tornar os dados localizáveis (incluindo disposições para metadados) | <ul style="list-style-type: none"> • Descreva a descoberta de dados (fornecimento de metadados); • Descreva a capacidade de identificação dos dados e consulte o mecanismo de identificação padrão. Você usa identificadores persistentes e exclusivos, como identificadores de objeto digital? • Esboço de convenções de nomenclatura usadas; • Descreva a abordagem da palavra-chave de pesquisa; • Descreva a abordagem para um controle de versão claro; • Especifique os padrões para a criação de metadados (se houver). Se não houver padrões em sua disciplina, descreva que tipo de metadados será criado e como; |
| 2.2 Tornando os dados acessíveis abertamente | <ul style="list-style-type: none"> • Especifique quais dados serão disponibilizados abertamente. Se alguns dados forem mantidos fechados, forneça uma justificativa para isso; • Especifique como os dados serão disponibilizados; • Especifique quais métodos ou ferramentas de <i>software</i> são necessários para acessar os dados. A documentação sobre o <i>software</i> é necessária para acessar os dados incluídos? É possível incluir o <i>software</i> relevante (por exemplo, em código-fonte aberto)? • Especifique onde os dados e metadados associados, documentação e código serão depositados; • Especifique como o acesso será fornecido caso haja alguma restrição; |

| | |
|---|--|
| 2.3. Tornando os dados interoperáveis | <ul style="list-style-type: none"> • Avalie a interoperabilidade de seus dados. Especifique quais vocabulários, padrões ou metodologias de dados e metadados você seguirá para facilitar a interoperabilidade; • Especifique se você usará vocabulário padrão para todos os tipos de dados presentes em seu conjunto de dados, para permitir a interoperabilidade interdisciplinar? Se não, você fornecerá mapeamento para ontologias mais comumente usadas? |
| 2.4. Aumente a reutilização de dados (por meio de licenças de esclarecimento) | <ul style="list-style-type: none"> • Especifique como os dados serão licenciados para permitir a maior reutilização possível; • Especifique quando os dados serão disponibilizados para reutilização. Se aplicável, especifique por que e por qual período um embargo de dados é necessário; • Especifique se os dados produzidos e/ou utilizados no projeto podem ser utilizados por terceiros, principalmente após o término do projeto? Se a reutilização de alguns dados for restrita, explique por quê; • Descrever processos de garantia de qualidade de dados; • Especifique por quanto tempo os dados permanecerão reutilizáveis; |
| 3. Alocação de recursos | <ul style="list-style-type: none"> • Faça uma estimativa dos custos para tornar seus dados justos; Descreva como você pretende cobrir esses custos; • Identifique claramente as responsabilidades pelo gerenciamento de dados em seu projeto; • Descreva os custos e o valor potencial da preservação a longo prazo; |
| 4. Segurança de dados | <ul style="list-style-type: none"> • Aborde a recuperação de dados, bem como armazenamento seguro e transferência de dados confidenciais; |
| 5. Aspectos éticos | <ul style="list-style-type: none"> • A ser abordado no contexto da revisão de ética, seção de ética do <i>Description of the Action</i> (DoA) e resultados de ética. Incluir referências e aspectos técnicos relacionados, se não forem cobertos pelo anterior; |
| 6. Outro | <ul style="list-style-type: none"> • Consulte outros procedimentos nacionais/ financiadores/ setoriais/ departamentais para gestão de dados que você está usando (se houver); |

**ANEXO B – Componentes do PGD para projetos financiados pelo ERC no âmbito do
Horizon 2020**

| Componentes PGD | Questões a serem tratadas |
|---|---|
| 1. Sumário | <ul style="list-style-type: none"> • Referência e nome do conjunto de dados; • Origem e tamanho esperado dos dados gerados/ coletados; • Tipos e formatos de dados; |
| 2. Tornando os dados localizáveis | <ul style="list-style-type: none"> • Descrição do conjunto de dados: metadados, identificadores persistentes e únicos, por exemplo, DOI; |
| 3. Tornando os dados abertamente acessíveis | <ul style="list-style-type: none"> • Quais dados serão disponibilizados abertamente e se alguns conjuntos de dados permanecerem fechados, os motivos para o não acesso; • Onde os dados e metadados associados, documentação e código são depositados (repositório?); • Como os dados podem ser acessados (são fornecidas ferramentas / métodos de <i>software</i> relevantes?); |
| 4. Tornando os dados interoperáveis | <ul style="list-style-type: none"> • Quais dados padrão ou específicos de campo e vocabulários e métodos de metadados serão usados; |
| 5. Aumentar a reutilização dos dados | <ul style="list-style-type: none"> • Quais dados permanecerão reutilizáveis e por quanto tempo, está previsto um embargo; • Como os dados são licenciados; • Procedimentos de garantia de qualidade dos dados; |
| 6. Alocações de recursos e segurança dos dados | <ul style="list-style-type: none"> • Custos estimados para tornar os dados do projeto em acesso aberto e valor potencial de preservação de dados em longo prazo; • Procedimentos para <i>backup</i> e recuperação de dados; • Transferência de dados confidenciais e armazenamento seguro em repositórios para preservação e curadoria de longo prazo; |

Fonte: European Research Council (2017a).