

## Compartilhamento de dados de pesquisa em saúde: iniciativas do National Institutes of Health (NIH)

Sharing health research data: National Institutes of Health (NIH) initiatives

Compartición de datos de investigación en salud: iniciativas de los National Institutes of Health (NIH)

Vanessa de Arruda Jorge<sup>1,a</sup>

[vanessaajorge@gmail.com](mailto:vanessaajorge@gmail.com) | <http://orcid.org/0000-0002-5298-9311>

Sarita Albagli<sup>2,b</sup>

[sarita.albagli@gmail.com](mailto:sarita.albagli@gmail.com) | <http://orcid.org/0000-0003-0030-8964>

<sup>1</sup>Instituto Nacional de Controle da Qualidade em Saúde. Rio de Janeiro, RJ, Brasil.

<sup>2</sup>Instituto Brasileiro de Informação em Ciência e Tecnologia. Rio de Janeiro, RJ, Brasil.

<sup>a</sup>Doutorado em Ciência da Informação pelo Instituto Brasileiro de Informação em Ciência e Tecnologia.

<sup>b</sup>Doutorado em Geografia pela Universidade Federal do Rio de Janeiro.

### Resumo

Este artigo apresenta um levantamento e uma análise das políticas e infraestruturas de compartilhamento de dados de pesquisa em saúde adotadas pelos institutos e centros que compõem o NIH – National Institutes of Health (Institutos Nacionais de Saúde), organismo norte-americano de pesquisa biomédica. A partir de pesquisa bibliográfica sobre definições e conceitos abordados neste estudo, o trabalho empírico consistiu na realização de buscas, nos sites dos institutos, centros e escritório central do NIH, de iniciativas de abertura e compartilhamento de dados de pesquisa. Localizaram-se no escritório central no NIH áreas responsáveis por essas ações, e foram identificadas políticas e repositórios de compartilhamento de dados cadastrados no diretório de busca da NLM – U.S. National Library of Medicine (Biblioteca Nacional de Medicina dos Estados Unidos). Como resultado, foram feitas análises sobre cada uma dessas iniciativas, considerando: as principais políticas de dados adotadas e seus principais objetivos, tipos e níveis de acesso e de publicação dos dados, formas de inserção e contribuição de dados, existência ou não de códigos de conduta, principais áreas de pesquisa envolvidas, tipos de repositórios (temáticos ou institucionais). As considerações finais fornecem subsídios para debates sobre diferentes tipos e abordagens de abertura e compartilhamento de dados de pesquisa científica, indicando questões pertinentes aos desdobramentos futuros da pesquisa.

**Palavras-chave:** Dados de pesquisa; Compartilhamento; Saúde; National Institutes of Health; Políticas; Repositórios de dados.

## Abstract

This paper presents a review and an analysis of health research data sharing policies and frameworks adopted by the institutes and centers of the National Institutes of Health (NIH), an American medical research agency. From a literature study of definitions and concepts approached in this study, the empirical research consisted in searches on the internet sites of the institutes, centers and the Office of the Director (central office at NIH) of initiatives to open access and sharing of research data. The responsible areas for these actions were located at the NIH Office of the Director and data sharing policies and repositories were identified in the National Library of Medicine (NLM) search directory. Then, an analysis of each one of these initiatives were carried out taking account: the main data policies adopted and their main objectives, kinds and levels of access and of published data, forms of insertion and contribution of data, whether or not codes of conduct are observed, main research areas involved, kinds of repositories (thematic or institutional). Final considerations provide inputs for discussions about different kinds and approaches to open access and scientific research data sharing, indicating suitable issues for future research developments.

**Keywords:** Research data; Sharing; Health; National Institutes of Health; Policies; Data repositories.

## Resumen

Este artículo presenta un levantamiento y el análisis de las políticas e infraestructuras de compartición de datos de investigación en salud adoptadas por los institutos y centros que componen el National Institutes of Health (NIH), agencia de investigación médica de los Estados Unidos. A partir de la investigación bibliográfica sobre definiciones y conceptos abordados en este estudio, el trabajo empírico consistió en la realización de búsquedas en los sitios de internet de los institutos, centros y oficina central de los NIH, de iniciativas de acceso abierto y compartición de datos de investigación. Se ubicaron en la oficina central de los NIH áreas responsables de esas acciones, y se identificaron políticas y repositorios de compartición de datos registrados en el directorio de búsqueda de la National Library of Medicine (NLM). Como resultado, se realizaron análisis sobre cada una de esas iniciativas, considerando: las principales políticas de datos adoptadas y sus principales objetivos, tipos y niveles de acceso y de publicación de los datos, formas de inserción y contribución de datos, existencia o no de normas de conductas, principales áreas de investigación involucradas, tipos de repositorios (temáticos o institucionales). Las consideraciones finales proporcionan contribuciones para debates sobre diferentes tipos y enfoques de acceso abierto y compartición de datos de investigación científica, indicando cuestiones pertinentes a los desdoblamientos futuros de la investigación.

**Palabras clave:** Datos de investigación; Compartición; Salud; National Institutes of Health; Políticas; Repositorios de datos.

---

### INFORMAÇÕES DO ARTIGO

#### Contribuição dos autores:

Concepção e desenho do estudo: Vanessa de Arruda Jorge e Sarita Albagli.  
Aquisição, análise ou interpretação dos dados: Vanessa de Arruda Jorge.  
Redação do manuscrito: Vanessa de Arruda Jorge e Sarita Albagli.  
Revisão crítica do conteúdo intelectual: Sarita Albagli.

**Fontes de financiamento:** CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico e Faperj - Fundação Carlos Chagas de Apoio à Pesquisa do Estado do Rio de Janeiro.

**Considerações éticas:** Não há.

**Agradecimentos/Contribuições adicionais:** Não há.

**Histórico do artigo:** Submetido: 26.fev.2018 | Aceito: 10.out.2018 | Publicado: 21.dez.2018.

**Apresentação anterior:** Não houve.

**Licença CC BY-NC atribuição não comercial.** Com essa licença é permitido acessar, baixar (download), copiar, imprimir, compartilhar, reutilizar e distribuir os artigos, desde que para uso não comercial e com a citação da fonte, conferindo os devidos créditos de autoria e menção à Reciis. Nesses casos, nenhuma permissão é necessária por parte dos autores ou dos editores.

## Introdução<sup>i</sup>

Entre os vários produtos e insumos da produção científica – publicações, patentes, relatórios, teses, dissertações, entre outros – há um crescente interesse nos dados de pesquisa. Estes podem ser definidos como os “registros factuais (resultados numéricos, documentos textuais, imagens e sons) usados como fonte primária para a pesquisa científica e que são comumente aceitos pelos pesquisadores como necessários para validar os resultados do trabalho científico”<sup>1</sup>. Os dados de pesquisa podem apresentar formatos, extensões e classificações próprios, além de tipologias e padrões documentais tradicionais já estabelecidos. Em alguns casos, os dados são criados para atender a uma finalidade específica (administrativa, por exemplo) e tornam-se dados de pesquisa científica de acordo com seu uso.

A abundância e a facilidade de acesso aos dados de pesquisa, potencializadas pelo desenvolvimento e a difusão de novas tecnologias da informação e comunicação, têm incentivado, por sua vez, políticas e práticas visando à sua abertura e compartilhamento, como parte do movimento mais amplo pela ciência aberta<sup>2</sup>. Neste contexto, ampliam-se as possibilidades de reuso de dados em novas pesquisas, de reprodutibilidade de experimentos, de verificação e validação da qualidade dos resultados da pesquisa científica, facilitando também a colaboração entre pesquisadores de diferentes partes do mundo.

Cabe, por outro lado, diferenciar os significados de abertura e de compartilhamento de dados. Um dado compartilhado nem sempre é um dado aberto. De acordo com a Open Knowledge International, os “dados são abertos quando qualquer pessoa pode livremente acessá-los, utilizá-los, modificá-los e compartilhá-los para qualquer finalidade, estando sujeito, no máximo, a exigências que visem preservar sua proveniência e sua abertura”<sup>3</sup>. Os dados abertos têm ainda “formato de arquivo não proprietário, cuja especificação esteja documentada publicamente e seja de livre conhecimento e implementação, livre de patentes ou qualquer outra restrição legal quanto à sua utilização”<sup>4</sup>. A licença de acesso e uso do dado deve ser aberta, ou seja, deve possibilitar “amplo acesso para que qualquer pessoa os utilize, os reutilize, e os redistribua, estando sujeito a, no máximo, a exigência de creditar a sua autoria e compartilhar pela mesma licença”<sup>5</sup>.

Já os dados compartilhados nem sempre seguem os princípios dos dados abertos. O compartilhamento de dados de pesquisa está geralmente pautado por acordos de acesso, que podem ser definidos como “o enquadramento processual estabelecido pelas instituições de pesquisa, agências e outros parceiros envolvidos, para determinar as condições de utilização de dados de pesquisa”<sup>1</sup>. Tais acordos normalmente restringem o compartilhamento entre partes determinadas e sob condições estabelecidas, sem uma preocupação com formatos não proprietários e que possam ser lidos por máquinas, nem com licenças que concedam amplo acesso de uso e redistribuição. Desta forma, a abertura e o compartilhamento de dados de pesquisa definem-se pelas condições de acesso e uso dos dados, suas condições de reutilização, reprodução e transparência.

Os dados variam conforme a área do conhecimento e o tipo de pesquisa. Sua gestão envolve desde questões técnicas e gerenciais, como planejamento, metadados, profissionais capacitados, infraestruturas disponíveis, normas e regulações até questões relacionadas com a cultura científica dominante, modo de funcionamento e disputas do campo científico<sup>6</sup>. Logo, analisar componentes que influenciam na governança do compartilhamento e abertura dos dados de pesquisa é importante para melhor entender o funcionamento dessas práticas e como elas se desenvolvem nas instituições científicas.

Neste trabalho, o foco recai nas políticas e repositórios de dados de pesquisa, devido à sua importância nas condições e termos em que se dão a abertura e o compartilhamento desses dados. As políticas de dados definem e regulam as condições gerais (ou, em alguns casos, específicas) de acesso aos dados produzidos ou coletados nas pesquisas e de seu uso, facilitando sua governança pelos diferentes atores do campo científico, como financiadores, pesquisadores e instituições de pesquisa. Já os repositórios digitais de dados de pesquisa constituem “infraestruturas de base de dados desenvolvidas para apoiar todo o ciclo da gestão de dados de pesquisa,

<sup>i</sup> Este trabalho faz parte de pesquisa de doutorado sobre o compartilhamento e a abertura de dados de pesquisa em saúde, desenvolvida no Programa de Pós-Graduação em Ciência da Informação/Ibict-UFRJ.

incluindo as ações mais dinâmicas e contundentes sobre os dados, que coletivamente são chamadas de curadoria de dados de pesquisa, que visam adicionar valor aos dados, avaliando, formatando, agregando e derivando novos dados”<sup>7</sup>.

A área da saúde, escolhida como foco deste trabalho, é uma área sensível e estratégica para sociedade. De acordo com a Declaração Universal dos Direitos Humanos, a saúde é um direito humano fundamental<sup>8</sup>, e, segundo a Constituição Federal do Brasil, trata-se de um direito de todos e um dever do Estado<sup>9</sup>. A pesquisa científica é peça fundamental para garantia deste direito, pois possibilita uma série de achados e inovações que contribuem para solução e prevenção de problemas de saúde, contribuindo para a melhoria da qualidade de vida das populações. A saúde é, por outro lado, um campo marcado por conflitos e disputas, por abranger diferentes interesses e pontos de vista, envolvendo desde a saúde pública, que procura representar o interesse social, até a indústria farmacêutica, que movimenta um mercado altamente lucrativo e tem foco em patentes. Ambas possuem forte atuação e grandes investimentos em pesquisa e desenvolvimento.

A instituição escolhida para análise deste trabalho foi o National Institutes of Health (NIH)<sup>ii</sup>, uma agência de pesquisa médica vinculada ao Departamento de Saúde dos Estados Unidos (U.S. Department of Health and Human Services), com sede em Bethesda, Maryland. O NIH iniciou suas atividades em 1798, quando foi criado o Marine Hospital Service<sup>iii</sup>. Tem como missão “buscar conhecimentos fundamentais sobre a natureza e o comportamento dos sistemas vivos e a aplicação destes conhecimentos para melhorar a saúde, prolongar a vida e combater doenças e deficiências”<sup>10,iv</sup>. Trata-se de um dos maiores centros de pesquisa biomédica do mundo, sendo uma das instituições pioneiras nas práticas de compartilhamento de dados de pesquisa em saúde.

## Procedimentos metodológicos

A partir de pesquisa bibliográfica sobre definições e conceitos mobilizados no texto, o trabalho empírico consistiu, em um primeiro momento, em pesquisa documental no site do NIH, buscando estruturas organizacionais que atuassem nas políticas de abertura e compartilhamento de dados de pesquisa. Em seguida, foi realizada pesquisa no diretório de busca da National Library of Medicine (NLM)<sup>v</sup>, vinculada ao NIH. Trata-se da maior biblioteca biomédica do mundo, cujas coleções digitais e físicas incluem livros, revistas, manuscritos, imagens e multimídia. A NLM coleta, organiza e disponibiliza informações científicas biomédicas, incluindo a PubMed/Medline e MedlinePlus, mantendo uma base de pesquisa integrada com dados de inúmeras instituições e de grupos de pesquisa em saúde nos Estados Unidos, incluindo dados genômicos (Projeto Genoma), químicos, toxicológicos (TOXNET) e ambientais, informação sobre drogas, dados de ensaios clínicos (ClinicalTrials.gov), padrões de dados de saúde, programas e informações de saúde para o consumidor.

A pesquisa no diretório da NLM foi realizada a partir do termo de busca ‘*data sharing*’ (compartilhamento de dados), tendo a análise concentrado-se nos assuntos de ‘políticas de compartilhamento de dados do NIH’ e ‘repositórios de compartilhamento de dados do NIH’. Duas listagens foram localizadas, uma sobre políticas<sup>vi</sup> e outra sobre repositórios<sup>vii</sup>. Os dados constantes nessas listagens foram analisados qualitativamente a partir de descritores que permitissem caracterizar, em linhas gerais, questões consideradas relevantes para análise das informações disponíveis na internet.

ii Disponível em: <https://www.nih.gov/>.

iii Disponível em: <https://www.nih.gov/about-nih/what-we-do/nih-almanac/chronology-events>.

iv “To seek fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to enhance health, lengthen life, and reduce illness and disability” (do original).

v Disponível em: [www.nlm.nih.gov/](http://www.nlm.nih.gov/).

vi Disponível em: [https://www.nlm.nih.gov/NIHbmic/nih\\_data\\_sharing\\_policies.html](https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_policies.html).

vii Disponível em: [https://www.nlm.nih.gov/NIHbmic/nih\\_data\\_sharing\\_repositories.html](https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html).

## Resultados e discussões

A estrutura organizacional do NIH é composta por um escritório central (NIH Office of the Director), 21 institutos e seis centros de referência<sup>viii</sup>. O escritório central atua na definição de políticas, no planejamento e coordenação de programas e de atividades do órgão. A esse escritório central estão vinculados diversos outros escritórios, dentre os quais cabe aqui citar dois pelo interesse que têm para os fins deste trabalho. Um deles é o Escritório de Política Científica (NIH Office of Science Policy), órgão de assessoramento que conta com uma Divisão de Política e Gestão de Compartilhamento de Dados Científicos (Scientific Data Sharing and Management Policy Division). Essa divisão é responsável pelo desenvolvimento de políticas de pesquisa biomédica, que englobam questões relacionadas ao compartilhamento e ao gerenciamento de dados de pesquisa científica visando promover a reprodutibilidade de seus resultados. O outro órgão é o Escritório de Pesquisa Extramuros (Office of Extramural Research), que viabiliza uma série de mecanismos visando garantir a integridade científica, a responsabilidade pública e a administração efetiva do portfólio de pesquisas externas do NIH. Esse escritório mantém uma série de normas de conformidade para recebimento de financiamentos e recursos para pesquisa do NIH, entre as quais uma especificamente orientada às políticas de compartilhamento de dados de pesquisa científica<sup>ix</sup>.

Os institutos e centros de referência do NIH<sup>x</sup> trabalham em temáticas de pesquisa específicas na área da saúde, em consonância com as normas e políticas desenvolvidas no escritório central. Muitas vezes focam em determinadas doenças ou sistemas corporais, e promovem pesquisas com instituições de outros países. O NIH financia ou participa de consórcios de pesquisas em saúde internacionais, inclusive no Brasil, por exemplo no *Parallel Funding Initiative for Collaborative Research Between Investigators in the USA and in the State of São Paulo, Brazil* (NOT-TW-16-001)<sup>xi</sup>, que envolve a Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp) e diversos institutos do NIH. Outro exemplo é o estudo *Zika in Infants and Pregnancy-ZIP*, que envolve a Fundação Oswaldo Cruz (Fiocruz) e NIH, a partir do National Institute of Allergy and Infectious Diseases (NIAID), Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) e National Institute of Environmental Health Sciences (NIEHS).

A partir da pluralidade de temáticas de pesquisas abarcadas por seus institutos e centros, o NIH desenvolveu uma política de compartilhamento de dados com caráter generalista. A Política de Compartilhamento de Dados do NIH<sup>xii</sup> (NIH Data Sharing Policy) foi publicada em 26 de fevereiro de 2003, como uma extensão da política de compartilhamento de recursos de pesquisa do órgão. Ela prevê que, para financiamento de pesquisas com valores superiores a U\$ 500 mil envolvendo apoio direto, é necessário o envio de plano de compartilhamento de dados com sua aplicação ou de documento justificando porque o compartilhamento de dados não é possível. Um ponto importante é o reconhecimento das regras de privacidade de informações de saúde individualmente identificáveis. A construção dessa política contou com comentários e contribuições de diversas organizações científicas e de mais de 150 indivíduos. O NIH disponibiliza uma série de documentos que estão associados a essa política<sup>xiii</sup> e que ajudam a melhor compreendê-la.

Com essa política de caráter generalista, o NIH incentiva a criação de outras políticas e planos de compartilhamento de dados específicos para seus institutos e centros. Neste contexto, verificou-se a existência de uma relação contendo 18 políticas de compartilhamento de dados do NIH (que inclui a

viii Disponível em: <https://www.nih.gov/institutes-nih/list-nih-institutes-centers-offices>.

ix Disponível em: <https://grants.nih.gov/policy/sharing.htm>.

x Disponível em: <https://www.nih.gov/institutes-nih/list-nih-institutes-centers-offices>.

xi Disponível em: <https://grants.nih.gov/grants/guide/notice-files/NOT-TW-16-001.html>.

xii Disponível em: <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>.

xiii Disponível em: [https://grants.nih.gov/grants/policy/data\\_sharing/](https://grants.nih.gov/grants/policy/data_sharing/).

NIH Data Sharing Policy), com os repositórios de dados em que elas estão em vigor. Essas políticas foram analisadas a partir dos seguintes descritores: nome do instituto, nome da política de compartilhamento de dados, ano, temática de destaque na política e repositórios associados (Quadro 1).

Quadro 1 — Relação de políticas de compartilhamento de dados disponibilizada no site do NIH

(Continua)

Instituto ou centro	Nome da política	Temática de destaque	Ano	Repositório associado
National Institutes of Health (NIH)	NIH Data Sharing Policy	Pesquisa em saúde	2003	Sem repositório específico
National Institutes of Health (NIH)	NIH Policy on Deposit of Atomic Coordinates into Structural Databases	Coordenadas atômicas de estruturas biomoleculares	1999	Protein Data Bank
National Institutes of Health (NIH)	NIH Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS)	Genômica		
National Institutes of Health (NIH)	Genomic Data Sharing Policy	Genômica	2014	dbGaP (para registro dos dados genômicos humanos) e repositórios de dados designados pelo NIH (para depósito de dados genômicos)
National Institutes of Health (NIH)	NDAR Grantees Data Sharing Policy	Autismo	2015	NDAR
National Human Genome Research Institute (NHGRI)	ENCODE Consortia Data Release, Data Use, and Publication Policies	Genômica	2009	ENCODE
National Human Genome Research Institute (NHGRI)	ENCODE Consortia Data Release, Data Use, and Publication Policies	Genômica	2009	ENCODE
National Institute of Allergy and Infectious Diseases (NIAID)	NIAID/DMID Data Sharing and Release Guidelines	Genômica	2017	GSCID, DMID Bioinformatics Resource Center, Trace Archive ou, quando apropriado, para o Short Read Archive, dbGaP, dbSNP, BEI Resources Repository
National Institute of Allergy and Infectious Diseases (NIAID)	Data Sharing Guiding Principles for the NIAID/DMID Systems Biology Program	Biologia sistêmica	2016 <sup>xiv</sup>	Repositórios associados ao Systems Biology Program (SBP)
National Institute of Allergy and Infectious Diseases (NIAID)	Human Immunology Project Consortium Data Sharing Plan	Imunologia	Não localizado	ImmPort

xiv Data da última atualização da página da internet.

(Conclusão)

Instituto ou centro	Nome da política	Temática de destaque	Ano	Repositório associado
National Institute on Aging (NIA)	Alzheimer's Disease Genetics Sharing Plan	Estudos genéticos da doença de Alzheimer	2008	NCRAD, NIAGADS, dbGaP
National Institute on Aging (NIA)	Alzheimer's Disease Neuroimaging Initiative (ADNI) Data Sharing and Publication Policy	Imagens e dados clínicos da doença de Alzheimer	2016	ADNI
National Institute on Drug Abuse (NIDA)	NIDA Data Sharing Policy	Genética	2018 <sup>xv</sup>	NIDA Genetics Consortium, NIDA Center for Genetic Studies Repository
National Institute on Drug Abuse (NIDA)	Clinical Trials Network Data Share Policy	Ensaio clínicos	Não localizado	CTN Data Share Website
National Heart, Lung, and Blood Institute (NHLBI)	NHLBI Policy for Data Sharing from Clinical Trials and Epidemiological Studies	Ensaio clínicos e estudos epidemiológicos	2014 <sup>xvi</sup>	BioLINCC
Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD)	Revised Resource Sharing Plan Instructions for Genetic Screens to Enhance Zebrafish Research and Enhancing Zebrafish Research with Research Tools and Techniques (PAR-08-138 and PAR-08-139)	Uso do peixe-zebra como modelo de vertebrados para pesquisas biomédicas e comportamentais	2008	Zebrafish International Resource Center (ZIRC)
National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)	The Environmental Determinants of Diabetes in the Young	Determinantes ambientais do diabetes no jovem	Desde 2003	NIDDK Repository
National Institute of Neurological Disorders and Stroke (NINDS)	Sharing Data via the Federal Interagency Traumatic Brain Injury Research (FITBIR) Informatics System	Pesquisa em traumatismo craniano	2017	FITBIR
Office of Rare Diseases Research (ORDR)	The Collaboration, Education, and Test Translation (CETT) Program's Guidelines for Data Collection and Sharing	Tradução de testes genéticos de doenças raras	Não localizado	dbGaP

Fonte: Adaptado de NIH Data Sharing Policies (2018)<sup>xvii</sup>.

As políticas listadas no quadro se “aplicam a práticas de compartilhamento de grandes grupos de pesquisadores e de dados”<sup>xviii</sup>, podendo existir requisitos e orientações específicas para pedidos individuais de acesso a dados, bem como para o compartilhamento de dados em determinados projetos, sendo evidente a predominância da área de genômica. A maioria está associada a um ou mais repositórios de dados, mostrando uma estreita relação entre a existência de políticas de compartilhamento e o funcionamento dos repositórios de dados.

Entre as políticas listadas, estão a NIH Data Sharing Policy, já comentada anteriormente, a NIH Policy on Deposit of Atomic Coordinates into Structural Databases (a política mais antiga citada na lista, do

xv Data da última atualização da página da internet.

xvi Data da última atualização da página da internet.

xvii Disponível em: [https://www.nlm.nih.gov/NIHbmic/nih\\_data\\_sharing\\_policies.html](https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_policies.html).

xviii “It includes policies at the NIH, IC, division, and program levels that apply to broad sets of investigators and data” (No original).

ano de 1999), a NIH Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS), a Genomic Data Sharing Policy e a NDAR Grantees Data Sharing Policy.

A NIH Policy on Deposit of Atomic Coordinates into Structural Databases prevê que as coordenadas atômicas de experiências de ressonância magnética nuclear e radiográfica de raios X que tiveram financiamento do NIH sejam depositadas em banco de dados apropriado no momento da submissão de um artigo de pesquisa. A NIH Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS) foi substituída pela Genomic Data Sharing Policy em 2014, que entrou em vigor em 2015. Essa última é mencionada como referência para outras políticas da área, sendo, portanto, uma política-chave nesse contexto. Ela propõe que os dados de pesquisa genômica derivados em larga medida de estudos financiados pelo NIH, envolvendo a espécie humana, organismos não humanos e modelos, sejam compartilhados por meio de um repositório de dados acessível ao público. Recomenda ainda que todos os estudos com dados genômicos humanos sejam registrados no repositório dbGaP e que os dados de organismos não humanos e modelos sejam depositados em um repositório de dados designado pelo NIH. Os dados não humanos podem ser submetidos a qualquer repositório de dados amplamente utilizado. A NDAR Grantees Data Sharing Policy dispõe que todos os dados resultantes das pesquisas sobre autismo financiadas pelo NIH envolvendo assuntos humanos devem ser submetidos ao repositório National Database for Autism Research (NDAR), juntamente com documentação de suporte apropriada para permitir o uso eficiente dos dados. Esta política tem forte relação também com o National Institute of Mental Health (NIMH), que mantém o NDAR.

Na área da genômica cita-se ainda a ENCODE Consortia Data Release, Data Use, and Publication Policies, do National Human Genome Research Institute (NHGRI). Essa política requer que os pesquisadores liberem dados primários, juntamente com uma interpretação inicial, para os bancos de dados públicos apropriados, assim que os dados forem verificados. Os membros dos consórcios também identificam os padrões de validação que serão aplicados nas análises subsequentes dos dados ou com experimentação adicional quando apropriado. Além disso, até que os dados sejam publicados em um periódico revisado por pares, o NHGRI pede aos usuários dos dados que os considerem inéditos e que sigam o protocolo científico padrão sobre o uso de dados não publicados. Todos os dados devem ser depositados em repositórios de dados, como o GenBank ou os Centros de Coordenação de Dados (DCCs) ENCODE / modENCODE, e esses dados de pré-publicação deverão estar disponíveis para todos.

O National Institute of Allergy and Infectious Diseases (NIAID), que lidera as pesquisas sobre doenças infecciosas, imunológicas e alérgicas, tem três políticas listadas, quais sejam: as NIAID/DMID Data Sharing and Release Guidelines, os Data Sharing Guiding Principles for the NIAID/DMID Systems Biology Program e o Human Immunology Project Consortium Data Sharing Plan. A primeira política citada estabelece princípios orientadores gerais e diretrizes específicas para planos de divulgação de dados do NIAID/DMID Omics Centers e ressalta que os dados humanos e não humanos sejam divulgados em uma linha de tempo que seja coerente com o duplo mandato do NIAID para apoiar pesquisa básica e clínica, bem como para responder a emergências de saúde pública. Também explica que as diretrizes e princípios estão alinhadas com princípios de dados atualmente recomendados, como os princípios FAIR – Findable, Accessible, Interoperable, Reusable (disponíveis, acessíveis, interoperáveis e reproduzíveis) para a liberação de dados. Esta orientação é relevante para discussão da abertura dos dados de pesquisa, considerando que a adoção dos FAIR tem a função de preparar os dados para o reuso, tendo seus princípios forte aproximação com os princípios dos dados abertos, anteriormente comentados.

A segunda política, Data Sharing Guiding Principles for the NIAID/DMID Systems Biology Program (SBP), estabelece princípios orientadores sobre o compartilhamento e a análise de dados de todo o NIAID, a partir da: 1) disponibilização de dados brutos aos pesquisadores do Instituto, incluindo dados brutos dos quais a análise final pode não estar completa; ou 2) realização de análises de amostras, sempre que possível, para maximizar o conteúdo de informações gerado por cada centro. Os requisitos do acordo de cooperação com a SBP explicam que

os dados de pesquisa, protocolos e modelos computacionais e estatísticos devem ser disponibilizados gratuitamente e publicamente à comunidade científica por meio dos *sites* dos centros ou de outras bases de dados públicas, de acordo com as políticas NIAID/DMID Data Sharing and Release Guidelines.

A terceira e última política listada, relacionada com o NIAID, é o Human Immunology Project Consortium (HIPC) Data Sharing Plan. Ela foi projetada para permitir a divulgação mais ampla de dados, além de proteger a privacidade dos participantes dos estudos. A política determina que os pesquisadores do HIPC devem depositar seus dados e metadados no repositório de dados ImmPort.

O National Institute on Aging (NIA) possui duas políticas listadas: a política Alzheimer's Disease Genetics Sharing Plan, que foi substituída em 2014 pelo Alzheimer's Disease Genomics Sharing Plan, que não consta na listagem disponível no quadro 1; e a Alzheimer's Disease Neuroimaging Initiative (ADNI) Data Sharing and Publication Policy, que trata da questão da anonimização dos dados sensíveis. Esta política recomenda o acesso completo e aberto às imagens do ADNI e dados clínicos anonimizados para indivíduos que se inscrevem e concordam com as condições do Acordo de uso de dados ADNI e que sejam submetidos à triagem limitada.

O National Institute on Drug Abuse (NIDA) também tem duas políticas listadas, sendo elas a NIDA Data Sharing Policy e a Clinical Trials Network Data Share Policy. A primeira política envolve os dados genéticos humanos produzidos em estudos financiados pelo NIDA, no NIDA Genetics Consortium (NGC). Prevê a submissão de um plano de compartilhamento de dados para participação no NGC. A segunda política explicita que os conjuntos de dados para os protocolos Clinical Trials Network (CTN) devem estar disponíveis no repositório NIDA Data Share (CTN data share) logo após a aceitação do artigo para publicação ou 18 meses, sendo este o período máximo de embargo.

Outra política que envolve ensaios clínicos é a NHLBI Policy for Data Sharing from Clinical Trials and Epidemiological Studies, do National Heart, Lung, and Blood Institute (NHLBI), que incentiva todos os candidatos a financiamento a elaborarem um plano de compartilhamento de dados ou a justificativa sobre o porquê do não compartilhamento dos dados (por exemplo, dados considerados sensíveis, protegendo os direitos de indivíduos e comunidades). De acordo com essa política, esses dados só devem ser disponibilizados em termos e condições consistentes, com o consentimento informado fornecido por participantes individuais, e conforme aprovado pelo Conselho de Revisão Institucional dos prêmios (IRB) e quaisquer leis e regulamentos locais, estaduais e federais. Os dados serão compartilhados através do repositório BioLINCC.

Outros institutos e centros têm políticas descritas no Quadro 1, como o Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), que, mais do que uma política de dados, trata-se de um requisito do edital de financiamento do NIH, intitulado Enhancing Zebrafish Research with Research Tools and Techniques (PAR-08-139). Ali determina-se que os pesquisadores devem incluir uma breve descrição de como os dados finais da pesquisa serão compartilhados ou explicar o porquê do não compartilhamento de dados, tendo como repositório de referência o Zebrafish International Resource Center (ZIRC).

O National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) desenvolve o estudo de coorte<sup>xix</sup> The Environmental Determinants of Diabetes in the Young (TEDDY), citado no Quadro 1. Os dados deste estudo – que incluem a fase de triagem, a fase de linha de base e conjuntos de dados de análise múltipla – estão disponíveis em repositórios de dados públicos. Os dados foram coletados no período de 2004-2010. Os pesquisadores que recebem recursos TEDDY devem concordar em reconhecer o repositório de dados do NIDDK.

---

xix É um tipo de desenho de estudo em saúde, em que se comparam experiências de grupos expostos e não expostos a um determinado fator de risco.

O National Institute of Neurological Disorders and Stroke (NINDS) mantém, em conjunto com o Department of Defense (DOD), o repositório Federal Interagency Traumatic Brain Injury Research (FITBIR), que tem a política Sharing Data via the Federal Interagency Traumatic Brain Injury Research (FITBIR) Informatics System. Ela prevê que os pesquisadores que submetem os dados ao repositório citado devem fornecer informações descritivas sobre seus estudos; fornecer dados genotípicos e fenotípicos codificados; e enviar uma requisição de submissão de dados para fornecer garantia de que todos os dados são submetidos ao Department of Defense (DOD) e ao National Institutes of Health (NIH), de acordo com as leis e regulamentos aplicáveis. Determina ainda que as identidades dos participantes da pesquisa não serão divulgadas no FITBIR.

A última política citada no Quadro 1 é a do Office of Rare Diseases Research (ORDR), ligado ao escritório central do NIH, denominada The Collaboration, Education, and Test Translation (CETT) Program's Guidelines for Data Collection and Sharing. Especifica que os dados clínicos anonimizados devem ser enviados e armazenados no repositório dbGaP. Para facilitar o maior acesso aos dados, as equipes colaborativas do CETT concordam em seguir os procedimentos de anonimização definidos na política do GWAS e desenvolver procedimentos e documentos educacionais/informativos.

Além da lista de políticas de compartilhamento de dados do NIH, foi encontrada, no diretório de pesquisa do NIH, uma outra lista contendo 73 repositórios de dados<sup>xx</sup>, verificando-se que 70 eram repositórios de compartilhamento de dados, e três eram *sites* de *software*. Trata-se de duas listas separadas, sendo que, na lista de políticas de dados, há uma coluna identificando os repositórios associados a cada uma dessas políticas. A partir dessas informações, foram feitas buscas complementares, buscando um maior detalhamento das informações inicialmente coletadas, de modo a realizar uma análise mais acurada, cujos resultados são apresentados no Quadro 2. As buscas foram orientadas pelos seguintes descritores: instituto ou centro, nome do repositório, tipo de repositório (institucional ou temático), temática, existência de código de conduta, descrição de quem pode inserir dados, descrição da disponibilidade de dados para qualquer interessado e descrição das políticas de acesso aos dados. Ressalta-se que o conteúdo dos *sites* em que foram realizadas essas buscas são dinâmicos e variáveis, ou seja, suas informações são atualizadas com alguma frequência.

Quadro 2 — Informações sobre repositórios de compartilhamento de dados do NIH

<b>Dados disponíveis para qualquer interessado?</b>	
Sim	52 74,3%
Mediante cadastro e aprovação	8 11,4%
Mediante cadastro	8 11,4%
Restrito a pessoas autorizadas	2 2,9%
	(Conclusão)
<b>Quem pode inserir dados?</b>	
Pesquisadores cadastrados	53 75,7%
Participantes do projeto	5 7,1%
Qualquer pesquisador da temática	4 5,7%
Não aceita submissão	1 1,4%

xx Pesquisa realizada em 1 de outubro de 2017.

Informação não localizada	6	8,6%
Em desenvolvimento	1	1,4%
<b>Normas para disponibilização dos dados</b>		
Sim	54	77,1%
Informação não localizada	15	21,4%
Em desenvolvimento	1	1,4%
<b>Há código de conduta?</b>		
Sim	29	41,4%
Não	3	4,3%
Informação não localizada	38	54,3%
<b>Tipo de repositório</b>		
Institucional	1	1,4%
Temático	69	98,6%
<b>Participação</b>		
Somente NIH	2	2,9%
NIH e outras instituições (nacionais e internacionais)	54	77,1%
Informação não localizada	14	20,0%

Fonte: As autoras (2017).

A partir desses resultados, destacam-se algumas análises e conclusões. Verifica-se, primeiramente que, na maioria dos repositórios (74,3%), os dados estão disponíveis para acesso por qualquer interessado. Cabe assinalar, por outro lado que, embora esse percentual seja alto, a análise e o uso dos dados podem ser feitos apenas por um público especializado, devido às diferentes linguagens e codificação utilizadas nas diversas áreas do conhecimento, como é o caso particularmente da saúde. Trata-se de sequências de números, símbolos, imagens e outros registros produzidos em ambientes controlados de pesquisa, que são compreensíveis, apenas ou principalmente, por pesquisadores que atuam nas respectivas áreas temáticas dos repositórios. Além disso, não se identificou um padrão de publicação dos dados que seja comum a todos os repositórios. Alguns dados podem ser baixados para o computador pessoal e outros mostram seus resultados apenas na própria página pesquisada. Outra questão a ser destacada é que esses dados, apesar de constarem em uma lista de repositórios de dados compartilhados, poderiam ser considerados compatíveis com a filosofia geral de dados abertos, já que podem ser acessados e utilizados por qualquer um sem prévio acordo formal entre partes. Porém não se pode afirmar que eles sejam de fato abertos, pois não foi possível verificar se os padrões e formatos adotados são passíveis de sua leitura por quaisquer máquinas.

Em segundo lugar, observa-se que a maioria desses repositórios aceita o depósito de dados não apenas de pesquisas financiadas pelo NIH, mas também de outras pesquisas que se relacionem com temáticas neles contempladas. Exige-se, para tanto, na maior parte (75,7% dos casos), que os pesquisadores efetuem seu cadastramento, sujeito à aprovação, para um possível depósito, provendo informações sobre os autores e suas pesquisas, bem como uma contextualização dos procedimentos de produção dos dados, tais como: procedência dos dados (incluindo seus metadados e a documentação apropriada que descreve o contexto), tipo de *software* e ambiente de produção, uso e armazenamento, garantindo um histórico, o que influencia diretamente a sua qualidade e a confiança de que poderão ser reutilizados. Em 5,7% não há necessidade de aprovação para submissão dos dados. O pesquisador que deseje submeter seus dados deve seguir as regras

de formatação e depósito dos dados, bem como se conectar através de sua conta de redes sociais, como por exemplo Google+, Facebook, ORCID etc. Por outro lado, 7,1% desses repositórios restringem a submissão de dados apenas a pesquisadores envolvidos em uma rede de pesquisa específica e/ou aqueles que participam de projeto ao qual o repositório está associado. Neles, não há outra opção de submissão de dados.

Já o descritor ‘normas para disponibilização dos dados’ possibilitou verificar se os repositórios tinham documentos ou declarações sobre o tipo de compartilhamento praticado, relacionando-os com as 18 políticas de compartilhamento de dados do NIH apresentadas no Quadro 1. Entre os repositórios associados às políticas listadas, 13 também aparecem na lista de repositórios de dados disponível no site do NIH, quais sejam: Protein Data Bank, BioLINCC, NIAGADS, dbGaP, dbSNP, BEI Resources Repository, ImmPort, Zebrafish International Resource Center (The Zebrafish Model Organism Database – ZFIN), NIDA Center for Genetic Studies Repository, CTN Data Share Website (NIDA Data Share), NIDDK Repository, NDAR e FITBIR. Considerando-se também os demais 57 repositórios, verificou-se que, no total, em 77,1% dos casos há políticas de dados declaradas. Cerca de 77,1% dos repositórios têm normas de disponibilização dos dados – ou seja, definem que dados são acessíveis, de que forma e quais as responsabilidades envolvidas – sendo que 74,3% disponibilizam os dados para qualquer interessado.

No descritor ‘códigos de conduta’, muitos repositórios não os possuem ou eles não são de fácil localização no seu site. O termo ‘código de conduta’ envolve o conjunto de normas comportamentais que guiam as práticas de uso dos dados, contribuindo para o respeito aos principais interessados. Os repositórios que apresentam códigos de conduta demonstram preocupação com questões éticas, como aquelas que envolvem a necessidade da correta citação de autoria dos dados ali depositados.

Outro aspecto observado refere-se à participação nacional e internacional nesses repositórios de dados. A maioria dos repositórios (77,1%) é mantida e/ou gerenciada por institutos ou centros do próprio NIH em associação com outras instituições (principalmente universidades) e tem participação de pesquisadores de diversas partes do mundo. Os grandes repositórios de dados possuem uma composição diferenciada, pois geralmente resultam de consórcios internacionais. Por exemplo, o GenBank é parte da International Nucleotide Sequence Database Collaboration, que compreende o ADN DataBank do Japão (DDBJ), o European Nucleotide Archive (ENA) e o GenBank no National Center for Biotechnology Information (NCBI). Outro exemplo é o Protein Data Bank (PDB), que é membro do wwPDB, um esforço colaborativo entre o PDBe (Reino Unido), o PDBj (Japão) e o BMRB (EUA) para garantir que o arquivo do PDB seja global e uniforme.

Destaca-se ainda a predominância de repositórios temáticos (98,6% dos casos). Como repositório temático considerou-se aqueles com “a preocupação de armazenar documentos com uma delimitação concisa de sua cobertura designada por um assunto, área do conhecimento ou temática específica”<sup>12</sup> e, como repositório institucional, a “reunião de repositórios temáticos sob a responsabilidade técnica e administrativa de uma instituição ou organismo”<sup>10</sup>. O Data and Specimen Hub (DASH) foi considerado um repositório institucional porque não há informações sobre a participação de outras instituições e o site desse repositório explica que ele é um recurso centralizado para pesquisadores armazenarem e acessarem dados anonimizados de estudos e pesquisas financiados pelo NICHD.

Por fim, observa-se que as áreas de genômica e genética são as que possuem mais projetos com disponibilização dos dados de pesquisa científica, incluindo repositórios de dados específicos sobre doenças de grande expressão na população, e outros como aqueles relacionados com o vírus HIV, a doença de Parkinson e o autismo. Não foram identificados repositórios de dados de pesquisa científica abertos em fármacos e imunobiológicos (foi encontrado apenas um repositório de estudos clínicos em Parkinson financiado por empresas produtoras de fármacos).

## Considerações finais

Realizar análises sobre abertura e compartilhamento de dados de pesquisa envolve a compreensão de uma diversidade de aspectos e questões, que estão dispersos em diferentes níveis de iniciativas e intervenções, especialmente em uma área como a da saúde, objeto de grande interesse social e econômico.

Neste estudo, o foco recaiu nas políticas e repositórios de dados de pesquisa adotadas por centros e institutos associados ao NIH para favorecer o compartilhamento de dados. A escolha justifica-se pela experiência do NIH, pioneira nesse campo há quase 15 anos, como agência governamental financiadora de pesquisas em saúde. Desta forma, o NIH e suas práticas podem servir de referência nos debates sobre diferentes tipos e abordagens de compartilhamento de dados de pesquisa científica, sem desconsiderar as diferenças existentes na cultura, nas políticas públicas e no funcionamento do sistema científico norte-americano em relação a realidades distintas, como a brasileira.

A análise das políticas de compartilhamento de dados do NIH mostraram que há situações em que diretrizes mais amplas e gerais podem ser utilizadas para sinalizar a necessidade do compartilhamento. E que, em outros casos, o detalhamento é imprescindível para a ação de compartilhar. Cada área tem uma especificidade de tipos de dados, padrões técnicos para descrição dos dados e até tempos distintos para disponibilização.

Observou-se que a opção pelo compartilhamento de dados com qualquer interessado é predominante nos repositórios mantidos pelo NIH. Verificou-se que a adoção de políticas e ações concretas de compartilhamento de dados já constitui requisito para o estabelecimento de cooperação científica entre o NIH e parceiros. Por outro lado, tais políticas não parecem alcançar áreas de maior interesse de mercado, como as de pesquisa em fármacos e imunobiológicos, embora estratégias de compartilhamento de dados venham sendo utilizadas no caso de pesquisas sobre doenças com potencial de lucratividade obtida com a produção de medicamentos, como o Alzheimer e o autismo. Isto pode indicar o reconhecimento, mesmo pela indústria, da comprovada importância do reuso de dados para aumentar a velocidade e a qualidade do avanço do conhecimento científico e, logo, do seu retorno para inovação e geração de novos produtos para o mercado.

A análise sobre as iniciativas do NIH também contribui para a reflexão sobre o papel e as formas de estruturação dos repositórios de dados, mostrando a possível tendência na direção de repositórios temáticos, bem como a preocupação com a proveniência dos dados compartilhados e com questões éticas sobre a conduta dos pesquisadores no reuso desses dados. A observação dos 70 repositórios de institutos e centros ligados ao NIH mostrou a complexidade de questões que envolvem seu funcionamento, como regras de uso, esquemas de classificação e arquivamento para preservação dos dados a longo prazo. Enfim, reforçou a necessidade de recursos financeiros, humanos e tecnológicos para manutenção dessas grandes infraestruturas e suas políticas.

O estudo aponta ainda para outras questões que merecem maior aprofundamento. Por exemplo, em que medida dados com potencial imediato de gerar patentes estariam dentro ou fora do escopo e abrangência das políticas de compartilhamento de dados de pesquisa do NIH. Ou, ainda, se as políticas de dados existentes têm sido suficientes para garantir o efetivo compartilhamento (e eventual abertura) desses dados. E, finalmente, quais os impactos concretos dessas políticas e infraestruturas para o reuso de dados em novas pesquisas na área da saúde.

## Referências

1. Organization for Economic Co-Operation and Development. Principles and guidelines for access to research data from public funding [Internet]. Paris; 2007 [cited 2016 Sept 13]. Available from: <http://www.oecd.org/sti/sci-tech/38500813.pdf>

2. Albagli S, Maciel ML, Abdo AH, organizador. Ciência aberta, questões abertas. Brasília: Ibiict; Rio de Janeiro: Unirio; 2015.
3. Open Data Handbook. O que são dados abertos? London: Open Knowledge International; [data unknown] [cited 2017 Nov 08]. Available from: [http://opendatahandbook.org/guide/pt\\_BR/what-is-open-data/](http://opendatahandbook.org/guide/pt_BR/what-is-open-data/)
4. Ministério do Planejamento, Orçamento e Gestão (BR). Secretária de Logística e Tecnologia da Informação. Instrução Normativa nº 4, de 12 de abril de 2012. Institui a Infraestrutura Nacional de Dados Abertos. DOU 2012 abr. 13;67-8.
5. Presidência da República (BR). Casa Civil. Subchefia de Assuntos Jurídicos. Decreto nº 8.777, de 11 de maio de 2016. Institui a Política Nacional de Dados Abertos do Poder Executivo Federal. DOU 2012 maio 12;21.
6. Bourdieu P. Usos sociais da ciência: por uma sociologia clínica do campo científico. São Paulo: Unesp; 2004.
7. Sayão LF, Sales LF. Algumas considerações sobre os repositórios digitais de dados de pesquisa. Inf [Internet]. 2016 maio/ago. [citado em 2018 nov. 06];21(2):90-115. doi: <http://dx.doi.org/10.5433/1981-8920.2016v21n2p90>
8. Organização das Nações Unidas. Declaração Universal dos Direitos Humanos. Assembleia Geral das Nações Unidas em Paris de 10 dezembro de 1948 [Internet]. Rio de Janeiro; 2009 [acesso em 2016 set. 13]. Disponível em: <https://nacoesunidas.org/wp-content/uploads/2018/10/DUDH.pdf>
9. Presidência da República (BR). Casa Civil. Subchefia de Assuntos Jurídicos. Constituição da República Federativa do Brasil de 1988. Brasília: Centro Gráfico; 1988.
10. National Institute of Health (US). What we do. Mission and goals [Internet]. Maryland; 2017 [cited 2018 Nov 06]. Available from: <https://www.nih.gov/about-nih/what-we-do/mission-goals>
11. National Library of Medicine (US). NIH data sharing policies [Internet]. Maryland; 2018 [cited 2018 Nov 06]. Available from: [https://www.nlm.nih.gov/NIHbmic/nih\\_data\\_sharing\\_policies.html](https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_policies.html)
12. Monteiro F, Bräscher M. Organização da informação em repositórios temáticos: o uso da modelagem conceitual [Internet]. Anais do 8º Encontro Nacional de Pesquisa em Ciência da Informação; 2007 out. 28-31; Salvador: UFBA; 2007 [citado em 2018 nov. 06]. Disponível em: <http://www.enancib.ppgci.ufba.br/artigos/GT2--261.pdf>