



24° ENANCIB
Encontro Nacional de Pesquisa em Ciência da Informação
Perspectivas Contemporâneas na Ciência da Informação
• Vitória - ES • Ancib • PPGCI/UFES



XXIV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XXIV ENANCIB

ISSN 2177-3688

GT 11 – Informação e Saúde

POLÍTICAS E DIRETRIZES NA GESTÃO DE DADOS GENÔMICOS NO ÂMBITO DA CIÊNCIA ABERTA

POLICIES AND GUIDELINES IN GENOMIC DATA MANAGEMENT IN THE CONTEXT OF OPEN SCIENCE

Fábio Bernardo da Silva – Fundação Oswaldo Cruz (FIOCRUZ)
Viviane Santos de Oliveira Veiga – Fundação Oswaldo Cruz (FIOCRUZ)

Modalidade: Resumo Expandido

Resumo: O advento dos sequenciadores de nova geração gera um volume significativo de dados genômicos, fundamentais para pesquisas biomédicas. Este trabalho tem como objetivo identificar as principais políticas e diretrizes na gestão de dados genômicos, no âmbito da ciência aberta. Metodologicamente, utilizou pesquisa bibliográfica e documental nas fontes *Fairsharing*, *Web of Science*, *PubMed* e Google Acadêmico. Os resultados obtidos proporcionam um panorama sobre as políticas e diretrizes de gestão de dados genômicos. Conclui que é imperativo a formulação de políticas, diretrizes e protocolos globais e abrangentes, que visem garantir a segurança dos dados compartilhados, a anonimização dos participantes da pesquisa e a adesão aos princípios da ciência aberta.

Palavras-chave: Gestão de dados genômicos; compartilhamento de dados, ciência aberta.

Abstract: The advent of next generation sequencers generates a significant volume of genomic data, essential for biomedical research. This work aims to identify the main policies and guidelines in the management of genomic data within the scope of open science. Methodologically, it used bibliographic research in the *Fairsharing*, *Web of Science*, *PubMed* and *Google Scholar*. The results obtained provide an overview of genomic data management policies and guidelines. It concludes that it is imperative to formulate global and comprehensive policies, guidelines and protocols, which aim to guarantee the security of shared data, the anonymization of research participants and adherence to the principles of open science.

Keywords: Genomic data management; data sharing, open science.

1 INTRODUÇÃO

Com o advento das novas tecnologias computacionais e de equipamentos conectados, em especial dos equipamentos de sequenciamento de DNA e RNA, conhecidos como *Next-Generation Sequencing* (NGS), que têm a capacidade de leitura rápida e eficiente de material genético em larga escala, em um tempo relativamente curto, gerando um grande volume de dados genômicos (Slatko; Gardner; Ausubel, 2018). Os dados genômicos são essenciais para entender a base genética de doenças, identificar variantes genéticas associadas a condições complexas e desenvolver novas abordagens terapêuticas. Sendo, por isso, fundamental a disponibilidade desses dados nos avanços em pesquisa genômica (Yakubu *et.al.*, 2018).

No entanto, a complexidade e a sensibilidade desses dados destacam a importância de políticas e diretrizes na gestão de dados genômicos, pois contêm informações pessoais e identificáveis sobre os indivíduos, o que levanta preocupações éticas e de privacidade do cidadão. O compartilhamento de dados genômicos em grande escala está comprometido em função de questões legais e éticas entre países, instituições e entre jurisdições, cujos fatores mais sensíveis são as questões com a privacidade e segurança dos dados, consentimento e a proteção dos direitos fundamentais dos seus titulares legais (Thorogood; Chokoshvili, 2023).

Nesse contexto, destaca-se a importância de políticas e diretrizes na gestão e compartilhamento de dados, abarcando os princípios FAIR - *Findable, Accessible, Interoperable, Reusable* (Wilkinson *et al.*, 2016), na gestão de dados genômicos e como as políticas e diretrizes podem promover a acessibilidade, interoperabilidade e reutilização desses dados no desenvolvimento de colaborações e pesquisas científicas. Como assegura Veiga (2019, p. 15), “Não basta compartilhar dados, eles precisam ser FAIR”. Os princípios FAIR abrangem não apenas dados, mas também metadados, pois são estes princípios que irão orientar a ampliação do acesso e possibilitar a correta interpretação e utilização dos dados pela comunidade científica.

Inúmeros projetos internacionais estabelecem objetivos de apresentar parâmetros para o compartilhamento ético, dentre eles destaca-se, como exemplo: *Human Variome Project* e *Global Alliance for Genomics and Health*, que apresentam responsabilidade e eficiência em dados genômicos e em outros relacionados à saúde humana. Devido à importância atribuída à disseminação e compartilhamento de dados de pesquisa no campo da genômica, é fundamental desenvolver protocolos e diretrizes que apoiem e facilitem aos

pesquisadores o tratamento adequado desses dados. A familiaridade com as regras e padrões que regem o campo científico nos estudos genômicos é crucial para a progressão desse campo e sua adesão aos princípios da ciência aberta.

2 OBJETIVO

Este artigo tem como objetivo geral analisar as principais políticas e diretrizes relacionadas à gestão de dados genômicos, bem como investigar o estado da arte dessas políticas no contexto da ciência aberta

3 MÉTODOS

O presente estudo se caracteriza como pesquisa exploratória, descritiva, de caráter qualitativo. Para alcançar o objetivo proposto, foi realizado o seguinte percurso metodológico: identificar as principais políticas e diretrizes na gestão de dados genômicos, no âmbito da ciência aberta, através de revisão bibliográfica e de levantamento documental.

As pesquisas documental e bibliográfica foram realizadas em quatro fontes de informação, sendo elas: *Fairsharing*, *Web of Science*, *PubMed* e Google Acadêmico. Foram utilizados termos livres e controlados. Os termos livres foram retirados da literatura da área. Os termos controlados foram identificados no Vocabulário Controlado da National Library of Medicine (Biblioteca Nacional de Medicina dos Estados Unidos), o Medical Subject Headings (**MeSH**).

O *Fairsharing* é uma fonte documental que visa viabilizar o uso dos princípios FAIR, promovendo padrões de dados, políticas e banco de dados, alinhados à estes princípios. Foi realizada em 16/06/2024, uma pesquisa com os seguintes termos: “policy” e “genomics”. Cinco documentos foram recuperados no Fairsharing.

Nas fontes *Web of Science*, *PubMed* e Google Acadêmico procedeu-se em 17/07/2024, utilizando-se operadores booleanos com os termos nos campos específicos assunto, título, resumo e palavras-chave, com o uso de filtro de data de publicação dos últimos dez anos. As pesquisas foram realizadas da seguinte forma: (1) Na base *Web of Science* aplicou-se a seguinte estratégia de busca: “genomic NEAR/2 data (Topic) and FAIR AND (polic* OR guideline* OR guidance) (All Fields)”; (2) Na base *PubMed* utilizou-se: “genomic NEAR/2 data (topic) or FAIR and (polic* OR

guideline* OR guidance) (All Fields)”; (3) Na base de dados Google Acadêmico utilizou-se a estratégia: “ ‘dados genômicos` FAIR (política OR diretriz OR orientação)”.

Nas bases Web of Science, PubMed e Google Acadêmico, 208 documentos foram inicialmente identificados. Após a remoção das duplicatas, restaram 166 artigos. Foram utilizados os seguintes critérios de inclusão e exclusão, após a leitura dos títulos e resumos dos documentos encontrados:

- Critérios de Inclusão: artigo científico ou normativa que abordasse políticas, diretrizes ou orientações para gestão de dados genômicos.

- Critérios de Exclusão: documentos publicados com mais de 10 anos.

Após a aplicação dos critérios de inclusão e exclusão foram selecionados os cinco documentos do Fairsharing e 16 artigos do levantamento bibliográfico. Após a leitura na íntegra dos 16 artigos, foram selecionados 10 artigos científicos, totalizando 15 itens entre artigos científicos e normativas no corpus da pesquisa.

4 RESULTADOS

Nesta seção serão apresentadas as principais políticas sobre gestão de dados genômicos identificadas no levantamento Bibliográfico e documental.

4.1 Estado da Arte das políticas e diretrizes para gestão de dados genômicos

Nesta subseção são apresentados os principais resultados obtidos dos textos selecionados e analisados, a partir do levantamento bibliográfico na Web Of Science, Pubmed e Google Acadêmico. Estes textos abordam conceitos de gestão de dados genômicos e a importância do grande volume de dados para novas perspectivas de pesquisa. Destacam a relevância da gestão de dados de pesquisa, mostrando sua importância tanto para publicações quanto para o financiamento por agências financiadoras, ressaltando a necessidade de criação e manutenção de políticas eficazes de gestão de dados.

4.1.1 Estudos Sobre a Gestão de Dados Genômicos

Como expõe Silva (2016, p. 74), “o fenômeno de dilúvio de dados pode ser ilustrado através do crescimento do GenBank, que é um banco de dados de sequências genéticas

mantido pelo NIH (EUA), e consiste em uma coleção anotada de todas as sequências de DNA publicamente disponíveis”. Atualmente, o Genbank dobra de tamanho a cada 18 meses. Esta tendência vem sendo confirmada em anos anteriores e deve ser mantida nos próximos anos (National Institutes of Health, 2024).

Outro ponto crucial é a transferência e armazenamento de dados. Quando se refere à transferência de grandes volumes de dados, possivelmente através de uma rede de longa distância, em muitos casos, transferi-los à nuvem computacional excede em muito o tempo necessário para processá-los (Issa *et al.*, 2013).

4.1.2 Gestão de Dados de Pesquisa

Como expõe Jorge (2018),

A gestão dos dados pode ser considerada uma disciplina, uma estratégia, uma ferramenta ou mesmo ter alguma outra denominação de acordo com as características daqueles que a adotam. Tem papel fundamental na regulação e operação estratégica do conhecimento científico produzido. Diversas instituições nacionais e internacionais, como universidades, financiadores e instituições de pesquisa, têm adotado políticas de gestão de dados de pesquisa, considerando-as essenciais para a qualidade, a integridade, a preservação e o acesso a esses dados (Jorge, 2018, p. 66).

Sob o ponto de vista de Cox, Pinfield e Smith (2014, p. 1-28),

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 (Cox, Pinfield e Smith, 2014, p. 1-28).

Tem como principal objetivo “[...] garantir uma verificação confiável dos resultados e permitir pesquisas novas e inovadoras, baseadas nas informações existentes” (Whyte; Tedds, 2011, p. 1, tradução própria)¹.

De outro modo, para Veiga (2017, p. 44),

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

os princípios da ciência aberta se baseiam no acesso aberto aos dados de pesquisa e às publicações científicas, principalmente às financiadas com recursos públicos; ferramentas e métodos de pesquisa abertos; processos de investigação colaborativos; a implementação de uma ciência cidadã; e a inovação aberta (Veiga, 2017, p. 44).

Na concepção de Veiga (2017, p. 24),

os dados de pesquisas necessitam ser preservados e compartilhados, porém apresentam especificidades que precisam de reflexão e análise. Dentre elas se destaca o fato de que os dados de pesquisa não possuem uma formalização estabelecida por serem de uma tipologia heterogênea em seu formato e em seu objeto. O seu compartilhamento envolve questões como ética e integridade na pesquisa, dados sensíveis, entre outros (Veiga, 2017, p. 248).

No contexto dos dados de pesquisa, os princípios FAIR (Findable, Accessible, Interoperable, and Reusable) foram recentemente estabelecidos com o objetivo de garantir o tratamento adequado dos dados para sua reutilização. Inicialmente, esses princípios foram definidos como um conjunto de orientações para tornar os dados localizáveis, acessíveis, interoperáveis e reutilizáveis, servindo como uma diretriz para aumentar a reutilização de dados (Wilkinson *et al.*, 2016, p. 1). No cenário de abertura e compartilhamento de dados, os princípios FAIR tornaram-se essenciais, dada a crescente complexidade e diversidade dos dados gerados no processo científico.

4.1.3 Políticas de gestão de dados

Na área biomédica, uma das principais questões é como proteger a enorme quantidade de dados existentes em meio digital, considerando principalmente a preservação dos direitos dos indivíduos. Em 2017, um grande avanço foi dado pela União Europeia, cujo Parlamento, após muitos anos de discussão, aprovou a Regulamentação Geral de Proteção de Dados (GDPR, na sigla em inglês), que entrou em vigor a partir de 2018. Seguindo estes passos, o Congresso Nacional aprovou, também em 2018, a chamada Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/18), que dispõe sobre a proteção de dados pessoais no Brasil e que entrou em vigor a partir de agosto de 2020 (Brasil, 2018).

Segundo a Academia Brasileira de Ciência (2020, p. 11),

toda política de gestão de dados deve considerar, no mínimo, três aspectos: a governança, a infraestrutura computacional (software e hardware) necessária para o armazenamento e disponibilização dos dados, e o treinamento e formação de pessoal que garantam a execução da política. O fator treinamento e formação é crucial para o sucesso da política e requer investimentos continuados. Neste contexto, abrange tanto os pesquisadores, quanto os profissionais que garantirão o funcionamento estável e duradouro dos repositórios.

Nesse contexto, a Research Data Alliance - RDA disponibiliza materiais que são recomendações-padrão para serem adotadas em todo o mundo. Uma tendência mundial importante é a de possibilitar a citação dos dados, independentemente de estarem ou não ligados a artigos científicos. Mas hoje existe a possibilidade de ser utilizado um identificador único para cada arquivo, reconhecido mundialmente, à semelhança do Digital Object Identifier - DOI para artigos científicos. De acordo com os princípios FAIR, é uma exigência que todos os dados tenham um identificador. As duas principais organizações mundiais de geração de DOI para dados são o *DataCite* e o *Handler Registry* (parte do *Handler.org*). Em ambos os casos, cada identificador gerado tem um custo, assim como os DOI de artigos em periódicos Academia Brasileira de Ciência (2020).

A leitura dos textos nos apontou para cinco aspectos relevantes em políticas de gestão e compartilhamento de dados, a saber: Aprovação ética, Compartilhamento de dados, Conformidade com a ciência aberta, Depósito de dados, Gestão de dados, Princípios FAIR, Propriedade intelectual livre, e Segurança dos dados. Estes aspectos serão utilizados para analisar as políticas identificadas no levantamento documental.

4.2 Identificação das e diretrizes relacionadas à gestão de dados genômicos

A pesquisa documental no Fairsharing retornou cinco políticas, das quais foram elencadas as seguintes diretrizes:

- (1) Open Access Science – Sanger Institute:** O instituto Sanger tem como objetivo o compartilhamento de dados, recursos, matérias e publicações. A política de compartilhamento de dados atende as conformidades de ciência aberta e os *softwares* desenvolvidos devem atender as diretrizes de *software* livre;
- (2) Wiley Advanced Genetics – Data Sharing/Repositories Policy:** A revista apoia suas publicações nas diretrizes internacionais para conduta ética em publicações

científicas (IUPAP), mantendo um elevado padrão ético das boas práticas científicas. A segurança das informações é implementada com técnicas administrativas que visam a integridade e a confidencialidade das informações contra o uso indevido;

- (3) **NIH² Genomic Data Sharing (GDS) Policy** : Responsável por disponibilizar a estrutura básica para o compartilhamento de dados genômicos gerados a partir de todas as pesquisas financiadas pelo Instituto. Usa a política GDS – Genomic Data Sharing, cujo objetivo é promover o compartilhamento de dados genômicos. Os dados genéticos devem ser desidentificados de acordo com a Lei de Responsabilidade e Portabilidade de Seguros de Saúde (HIPAA);
- (4) **Taylor and Francis All Life Jornal Data Sharing Policy** : A política de compartilhamento de dados obriga os autores a disponibilizar suas publicações e os dados de forma gratuita, os dados devem estar compatíveis com um repositório FAIR, com acesso e controles apropriados;
- (5) **2023 Final NIH Policy for Data Management and Sharing** : a política de compartilhamento e gestão incentiva práticas consistentes com os princípios FAIR, exige dos pesquisadores um plano de gestão e compartilhamento de dados.

O quadro 1, a seguir, apresenta as políticas na área de genômica em relação aos aspectos de gestão e compartilhamento de dados encontrados na literatura.

Quadro 1 – Políticas na área de genômica em relação aos aspectos de gestão e compartilhamento de dados

Aspectos	Políticas de Compartilhamento e Gestão de Dados				
	Open Acess Sciense-Sanger Institute	Wiley Advanced Genetics-Data Sharring/repositories Policy	NIH Genomic Data Sharing (GDS) Policy	Taylor and Francis All Life Jornal Data Sharing Policy	2023 Final NIH Policy for Data Management and Sharing
Aprovação ética	X	X	X	X	X
Conformidade com a ciência aberta	X	X	X	X	X
Depósito de dados	X	X	X	X	X
Gestão de dados	X	X	X	X	X
Princípios FAIR				X	X
Propriedade intelectual	X			X	
Segurança de dados	X		X	X	X

Fonte: Elaborado pelos autores (2024).

² National Institutes of Health (NIH).

Verifica-se que essas políticas apoiam o compartilhamento de dados, repositórios, gestão e segurança de dados. Apresentam questões éticas, conformidades com a ciência aberta e propriedade intelectual livre, porém é necessário fazer uma análise da articulação dessas políticas com os princípios FAIR de forma a garantir uma boa gestão de dados de pesquisa, que estejam em conformidade com os princípios de ciência aberta.

5 CONSIDERAÇÕES FINAIS

Realizar análises sobre a abertura e o 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.

Com esta pesquisa verificou-se que existe uma produção científica voltada para a reflexão sobre as políticas de gestão de dados no âmbito da pesquisa genômica. A literatura aponta que, apesar de existirem diretrizes importantes, como as promovidas por instituições como o NIH e o Sanger Institute, ainda há uma lacuna significativa na aplicação global de políticas que harmonizem o compartilhamento de dados genômicos com as normas de proteção de dados e privacidade.

Nesse contexto, identificou-se a existência de cinco políticas de gestão e compartilhamento de dados no campo da genômica. No entanto, foram observadas algumas fragilidades na formulação dessas políticas, principalmente, relacionado ao alinhamento dessas políticas com os princípios FAIR (Findable, Accessible, Interoperable e Reusable). Os princípios FAIR têm se mostrado um caminho promissor para a otimização da gestão de dados, facilitando o acesso e o reuso por pesquisadores em todo o mundo. São necessários estudos mais aprofundados para identificar as fragilidades e as necessidades de melhoria dessas políticas, incluindo a contribuição dos princípios FAIR na gestão de dados de pesquisa na área da genômica.

A gestão de dados genômicos representa um desafio no avanço da ciência biomédica, especialmente devido ao crescente volume de informações geradas por tecnologias de sequenciamento de nova geração. Essas informações são essenciais para pesquisas de grande impacto na área da saúde, permitindo a compreensão de doenças genéticas complexas e o

desenvolvimento de novas terapias. Contudo, o tratamento adequado desses dados exige políticas robustas que garantam sua reutilização de forma segura e ética, respeitando os princípios da ciência aberta.

REFERÊNCIAS

ACADEMIA BRASILEIRA DE CIÊNCIAS. **Abertura e gestão de dados:** desafios para a ciência brasileira. Ago. 2020. Disponível em: <https://www.abc.org.br/wp-content/uploads/2020/09/ABC-Abertura-e-Gest%C3%A3o-de-Dados-desafios-para-a-ci%C3%Aancia-brasileira.pdf>. Acesso em: 30 jun. 2024.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018.** Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, [2018]. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 30 jun. 2024.

ISSA, Shadi A. *et al.* Streaming support for data intensive cloud-based sequence analysis. **Biomed Research International**, New York, e791051, 2013.

JORGE, Vanessa de Arruda. **Abertura e compartilhamento de dados para pesquisa nas situações de emergência em saúde pública:** o caso do vírus Zika. Orientadora: Sarita Albagli. 2018. 263 f. Tese (Doutorado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Escola de Comunicação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2018. Disponível em: <https://www.arca.fiocruz.br/handle/icict/32604>. Acesso em: 30 jun. 2024.

NATIONAL INSTITUTES OF HEALTH (United States of America). **GenBank and WGS Statistics.** 2024. Disponível em: <http://www.ncbi.nlm.nih.gov/genbank/statistics>. Acesso em: 30 jun. 2024.

PINFIELD, S; COX, A. M.; SMITH, J. Research data management and libraries: relationships, activities, drivers and influences. **PLoS ONE**, San Francisco, 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: 4 jun. 2024.

SILVA, Fabricio Alves Barbosa da. Big data e nuvens computacionais: aplicações em saúde pública e genômica. **Journal of Health Informatics**, São Paulo, v. 8, n. 2, 2016. Disponível em: <https://jhi.sbis.org.br/index.php/jhi-sbis/article/view/336>. Acesso em: 28 jun. 2024.

SLATKO, Barton E.; GARDNER, Andrew F.; AUSUBEL, Frederick M. Overview of next-generation sequencing technologies. **Current protocols in molecular biology**, United States, v. 122, n. 1, p. e59, 2018.

THOROGOOD, Adrian; CHOKOSHVILI, Davit. Ethical and legal considerations in genomic data sharing: evolution of the discourse and the road ahead. *In*: VALDÉS, Erick; LECAROS, Juan Alberto (ed.). **Handbook of Bioethical Decisions**. Berlin: Springer Verlag, 2023. v. 1, p. 345-369.

VEIGA, Viviane Santos de Oliveira. Gestão de dados de pesquisa FAIR: dando um Jump em seus dados. *In*: ENCONTRO DA REDE SUDESTE DE REPOSITÓRIOS INSTITUCIONAIS, 1., 2019, Rio de Janeiro. **Anais eletrônicos [...]**. Rio de Janeiro: UFRJ, 2019.

VEIGA, Viviane Santos de Oliveira. **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. Orientadora: Cícera Henrique da Silva. 2017. 294 f. Tese (Doutorado em Ciências) – Instituto de Comunicação e Informação Científica, Fundação Oswaldo Cruz, Rio de Janeiro, 2017.

WHYTE, Angus; TEDDS, Jonathan. **Making the case for research data management**. Edinburgh: Digital Curation Centre, 2011.

WILKINSON, Mark D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, London, v. 3, art. 60018, Mar. 2016.

YAKUBU, Aminu *et al.* Model framework for governance of genomic research and biobanking in Africa – a content description. **AAS Open Research**, [S. l.], v. 1, 2018. pmid:30714023. 2nd version. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6354903/>. Acesso em: 30 jun. 2024.