



## **15ª Conferência Lusófona de Ciência Aberta (ConfOA) Acesso Aberto e Dados de Investigação Abertos: sistemas, políticas e práticas**

**Modalidade: Pecha Kucha**



### **CIÊNCIA ABERTA E OS DESAFIOS PARA O COMPARTILHAMENTO DE DADOS DE PESQUISA GENÔMICA.**

### ***OPEN SCIENCE AND THE CHALLENGER FOR SHARING GENOMIC RESEARCH DATA.***

**Fábio Bernardo da Silva**

Programa de Pós-Graduação em Informação e Comunicação em Saúde, PPGICS /  
ICICT/ Fundação Oswaldo Cruz (Fiocruz).

Rio de Janeiro, Rio de Janeiro, Brasil.

Orcid: [0009-0005-2922-6622](https://orcid.org/0009-0005-2922-6622)

[fabiober@ioc.fiocruz.br](mailto:fabiober@ioc.fiocruz.br)

**Viviane Santos de Oliveira Veiga**

Programa de Pós-Graduação em Informação e Comunicação em Saúde, PPGICS /  
ICICT/ Fundação Oswaldo Cruz (Fiocruz).

Rio de Janeiro, Rio de Janeiro, Brasil.

Orcid: [0000-0001-8318-7912](https://orcid.org/0000-0001-8318-7912)

[viviane.veiga@fiocruz.br](mailto:viviane.veiga@fiocruz.br)

#### **RESUMO:**

Este trabalho discute os desafios para o compartilhamento de dados de pesquisa genômica alinhados aos princípios FAIR. Para compreender melhor o campo, é proposto o Modelo de Fatores que Influenciam o Compartilhamento de Dados, que engloba a visão tecnológica e social do compartilhamento de dados, aliada aos princípios FAIR. Esse modelo apresenta os fatores de benefícios, custos e fatores contextuais na compreensão da percepção dos pesquisadores e o comportamento de compartilhamento de dados. Além disso, ao identificar as barreiras, fatores e problemas percebidos pelos pesquisadores é possível dirimi-los e minimizá-los. O modelo de fatores será utilizado junto aos pesquisadores que desenvolvem atividades de pesquisa na área de dados genômicos. Objetivando garantir que estes dados de pesquisa sejam reutilizáveis na ciência em prol da sociedade, apoiar a gestão de dados na pesquisa em genômica e levantar os requisitos para um modelo que busque garantir a gestão de dados FAIR na área de genômica, promovendo o compartilhamento de dados genômicos aberto e gratuito.

**Palavras-chave:** Gestão de dados de pesquisa; Princípios FAIR; Ciência Aberta.

## INTRODUÇÃO

### **Dados Genômicos e Ciência Aberta**

A disponibilidade de dados genômicos é fundamental para a pesquisa genômica (YAKUBU et. al., 2018). Dados genômicos são dados relacionados à estrutura e funções do genoma de um organismo, necessários ao seu desenvolvimento e incluem informações de sequência de moléculas nos genes de um organismo. Abarcando os dados de DNA, RNA, proteínas e suas funções da espécie humana, organismos não humanos e modelos. Dados de pesquisa genômicos são todos os dados gerados a partir de pesquisa científica na área.

Estudos mostram diversos desafios associados ao compartilhamento de dados, que incluem: a falta de padrões estabelecidos para os usuários de dados; a falta de crédito adequado para pesquisadores de países de baixa e média renda, que muitas vezes são apenas mencionados brevemente na seção de agradecimentos de publicações científicas; a possibilidade de perda de direitos de propriedade intelectual; o potencial uso indevido dos dados. Incluem ainda deficiências na governança da pesquisa genômica e dos biobancos; falta de quadros ético-legais abrangentes para orientar o compartilhamento de dados entre países de alta renda e baixa e média renda. Além disso, questões sobre o papel dos provedores de dados nas etapas da nova pesquisa, como, se devem revisar os resultados antes da publicação, colaborar na análise, aprovar previamente o plano de análise ou impor limitações à reutilização dos dados, continuam sendo áreas importantes de tensão e reflexão sobre o compartilhamento de dados (EKUSAI-SEBATTA et. al., 2023).

A produção e utilização de evidências científicas robustas sobre a segurança, a eficácia, a efetividade, a razão de custo-efetividade e o impacto orçamentário das tecnologias nos processos de avaliação; a garantia da participação social em todo o processo de gestão de tecnologias em saúde; a participação da comunidade e acesso aos dados disponibilizados abertamente, contribuem para a garantia do direito à informação, empoderando o cidadão e sua participação na construção de um sistema de saúde mais forte (VIEIRA; PIOLA; BENEVIDES, 2019).

Segundo Oliveira e Silva (2016, p. 9), na ciência aberta é primordial a cibe infraestrutura tecnológica que apoia o processo científico. Optou-se por desmembrar a dimensão morfológica e criar uma dimensão específica para tratar dos aspectos tecnológicos. A dimensão morfológica apresenta a estrutura da comunicação científica baseada no quarto paradigma, a qual pressupõe a cultura de compartilhamento.

De acordo com Sayão e Sales (2016, p. 96), a importância dos repositórios de dados de pesquisa é reconhecida pelos profissionais da informação, pelo fato de que havendo uma estrutura que garanta a reutilização desses dados, as possibilidades de avanços são mais rápidas na ciência, economizando em diversos recursos e ampliando novos temas de pesquisa.

Entretanto, assegura Wilkinson et Al (2017), “Não basta ter um Plano de Gestão de Dados: é preciso ser FAIR”. Saliencia, todavia, que não apenas os dados precisam ser FAIR, mas também os metadados precisam ser FAIR, pois são estes princípios que irão orientar a ampliação do acesso e possibilitar a correta interpretação e utilização dos dados pela comunidade científica.

## **DESENVOLVIMENTO**

### **Problema**

Segundo Henning et al (2019), no Brasil, em comparação com os países europeus e asiáticos, bem como com os Estados Unidos, a publicação de dados é relativamente baixa e deficitária. Este fato deve ser revisto pois o processo de pesquisa atual se fundamenta no compartilhamento de dados, interoperabilidade e reuso por pesquisadores em todo o mundo.

Muitas vezes, os dados são perdidos, ou ficam arquivados em computadores, ou são armazenados de forma inadequada, limitados a pequenos grupos de pesquisa. Grande parte destes volumes de dados, que são gerados a altos custos, são parcialmente analisados e em muitos casos, nem analisados são, deixando para serem vistos a posteriori e normalmente perdidos ou esquecidos ao longo do tempo. Muitos destes dados são gerados a altos custos, que em função do não aproveitamento dos mesmos, somam grandes quantidades de desperdícios de dinheiro público e privado (Henning et al, 2019).

A grande quantidade de dados que está sendo acumulada nos diferentes bancos de dados ao redor do mundo precisa, a partir das informações genômicas disponíveis, serem anotadas e interpretadas. Para este fim, é necessário que os diversos projetos interessados em trocar e integrar informações descrevam seus dados de forma padronizada, de modo a possibilitar com consistência a recuperação de informações (Campos, 2007). Ainda segundo Campos (2007, p. 117), “apesar dos conceitos emergentes de ciência aberta e compartilhamento de dados, no que tange ao volume de dados científicos em saúde, a adesão ao compartilhamento de dados, pelos pesquisadores brasileiros, está longe do que se conceitua como ideal”.

As questões norteadoras do presente projeto de pesquisa permeiam os seguintes questionamentos, que devem ser respondidos: a) Como garantir que os dados de pesquisa genômica sejam reutilizáveis na ciência em prol da sociedade? b) Como apoiar a gestão de dados na pesquisa em genômica alinhada aos princípios Fair? c) Quais são os principais requisitos para um modelo que busque garantir a gestão de dados FAIR na área de genômica?

## Abordagem

Esse projeto é uma pesquisa exploratória, de abordagem qualitativa e de natureza básica. A abordagem qualitativa se preocupa em explicar e se aprofundar em aspectos não quantificáveis das relações sociais (Gerhardt; Silveira, 2009).

A partir das características citadas acima, foi possível agregar estratégias que fossem apropriadas para auxiliar o desenvolvimento da pesquisa. Portanto, quanto aos procedimentos de pesquisa, serão adotados o tipo bibliográfico e a análise documental, que contribuirão para o entendimento pretendido sobre o assunto, ao final da pesquisa.

Os metadados FAIR são os elementos que transportam o conteúdo das informações que devem ser expressos por meio de técnicas de representação descritiva dos dados, condição essencial para o uso e reúso de dados de pesquisa (Dias et al., 2019).

Para criar soluções que apoiem o compartilhamento de dados de pesquisa genômica é necessário compreender os fatores contextuais e individuais, a partir da percepção dos pesquisadores da área de genômica. O Modelo de Fatores que influenciam no comportamento de compartilhamento de dados (MFDados), proposto por Veiga (2017), é um modelo que visa a elaboração de um roteiro de coleta de dados com os pesquisadores, para compreender as suas percepções e promover o compartilhamento de dados de pesquisa (Veiga, 2021). Portanto, acredita-se que o MFDados pode apoiar na detecção dos fatores, estímulos e barreiras no compartilhamento de dados genômicos, bem como compreender as percepções do pesquisador e o seu comportamento no compartilhamento de dados.

## CONSIDERAÇÕES FINAIS

Esse projeto de pesquisa de doutorado está em estágio inicial de desenvolvimento abarcando várias possibilidades discussões no âmbito da ciência aberta, biobancos e gestão de dados de pesquisa genômico.

O compartilhamento de dados de pesquisa é um dos pilares da ciência aberta. São vários os benefícios do compartilhamento de dados, mas os dados compartilhados só cumprirão seu papel no apoio aos avanços da pesquisa e à transparência científica se forem gerenciados adequadamente durante todo o ciclo de vida dos dados (KING, 2007, p. 173).

Neste contexto de compartilhamento de dados, o Modelo de Fatores de Dados – MFDados engloba a visão tecnológica e social do compartilhamento de dados, aliado aos princípios FAIR. Discorre sobre a compreensão da percepção dos pesquisadores e o comportamento de compartilhamento dos dados, bem como sobre a detecção das barreiras, fatores e problemas da área de genômica. Desta forma, o MF-Dados é uma ferramenta que pode apoiar a estruturação de uma proposta de modelo de gestão de dados para a área de pesquisa em genômica, desenvolvido juntamente com os stakeholders.

A gestão eficaz e o compartilhamento de dados genômicos pode colaborar com outras áreas de pesquisa, através da interoperabilidade, uso e reúso de dados, abertos e

compartilhados, contribuindo com novos conhecimentos em diversas áreas, dando origem a novas análises e investigações. Uma possível forma de utilização de dados genômicos abertos, na predição de casos clínicos regionalizados, depende da curadoria dos mesmos, e pode gerar uma base de dados de pesquisa, numa possível regionalização no direcionamento e no tratamento de saúde, no âmbito do SUS.

## REFERÊNCIAS

CAMPOS, Maria Luiza de Almeida. **Integração de Ontologias: o domínio da bioinformática**. RECIIS – Revista Eletrônica de Comunicação, Informação & Inovação em Saúde, Rio de Janeiro, v. 1, n. 1, p. 117-121, 2007. DOI: <https://doi.org/10.29397/reciis.v1i1.896>.

DIAS, Guilherme Ataíde *et al.* **Os princípios FAIR: viabilizando o reuso de dados científicos**. In: DIAS, Guilherme Ataíde; OLIVEIRA, Bernardina Maria Juvenal Freire de (org.). **Dados científicos: perspectivas e desafios**. João Pessoa: Editora UFPB, 2019.

EKUSAI-SEBATTA, Deborah; OCAN, Moisés; SINGH, Shenuka; KYADDONDO, David.; AKENA, Dickens.; NAKALEMBE, Loyce; et al. **Práticas de partilha de dados na investigação colaborativa do genoma humano em países de baixo e médio rendimento: um protocolo de revisão sistemática**. *PLoS ONE*, v. 18, n. 11, p. e0292996, 2023. DOI: <https://doi.org/10.1371/journal.pone.0292996>. Disponível em: <https://doi.org/10.1371/journal.pone.0292996>. Acesso em: 19 jun. 2024.

GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo (org.). **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009.

KING, Gary. **An introduction to the Dataverse network as an infrastructure for data sharing**. *Sociological Methods & Research*, v. 36, n. 2, p. 173–199, 2007. Disponível em: <https://gking.harvard.edu/sites/scholar.harvard.edu/files/gking/files/dvn.pdf>. Acesso em: 15 mai. 2024.

OLIVEIRA, Adriana Carla Silva de; SILVA, Edilene Maria da. Ciência aberta: dimensões para um novo fazer científico. **Informação & Informação**, Londrina, v. 21, n. 2, p. 5-39, 2016. DOI: <https://doi.org/10.5433/1981-8920.2016v21n2p5>. Disponível em: <https://www.uel.br/revistas/uel/index.php/informacao/article/view/27666/20113>. Acesso em: 12 fev. 2023.

SAYÃO, Luis Fernando; SALES, Luana Farias. **Subsídios para a construção de um modelo de avaliação de sistemas de gestão de dados de pesquisa**. *Ponto de Acesso*, Salvador, v. 12, n. 3, p. 80-108, dez. 2016.

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**. Rio de Janeiro: Fiocruz, 2017. Disponível em: <https://www.arca.fiocruz.br/handle/icict/26842>. Acesso em: 10 set. 2023.

VEIGA, Viviane Santos de Oliveira; SILVA, Cícera Henrique da; BORGES, Maria Manuel. Modelo de fatores que influenciam no comportamento de compartilhamento de dados de pesquisa (MFDados). *In*: BORGES, Maria Manuel; CASADO, Elias Sanz (org.). **Sob a lente da Ciência Aberta: olhares de Portugal, Espanha e Brasil**. Coimbra: Imprensa da Universidade de Coimbra. 2021. p. 153-187.

VIEIRA, Fabiola Sulpino; PIOLA, Sergio Francisco; BENEVIDES, Rodrigo Pucci de Sá e. **Vinculação orçamentária do gasto em saúde no Brasil: resultados e argumentos a seu favor**. Rio de Janeiro: Ipea, 2019.

WILKINSON, Mark D. *et al.* **The FAIR Guiding Principles for scientific data management and stewardship**. *Scientific Data*, v. 3, article n. 600 18, mar. 2016.

WILKINSON, M. D. *et al.* **Interoperability and FAIRness through a novel combination of Web technologies**. *PeerJ Computer Science*, v. 3, p. e110, 2017. DOI: <https://doi.org/10.7717/peerj-cs.110>. Disponível em: <https://doi.org/10.7717/peerj-cs.110>. Acesso em: 19 ago. 2024.

YAKUBU, A. *et al.* **Quadro modelo para a governação da investigação genômica e dos biobancos em África – uma descrição do conteúdo**. *Pesquisa Aberta da AAS*, v. 1, 2018. PMID:30714023.