

**Antecedentes...**

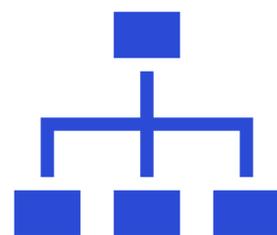
**Piloto  
Repositório  
Institucional de  
Dados de Pesquisa  
da Fiocruz  
CINCO/VPEIC**

**Rio de Janeiro, 30 de julho de 2020**

# Contexto Fiocruz



Política de gestão,  
compartilhamento e  
abertura de dados para  
pesquisa da Fiocruz



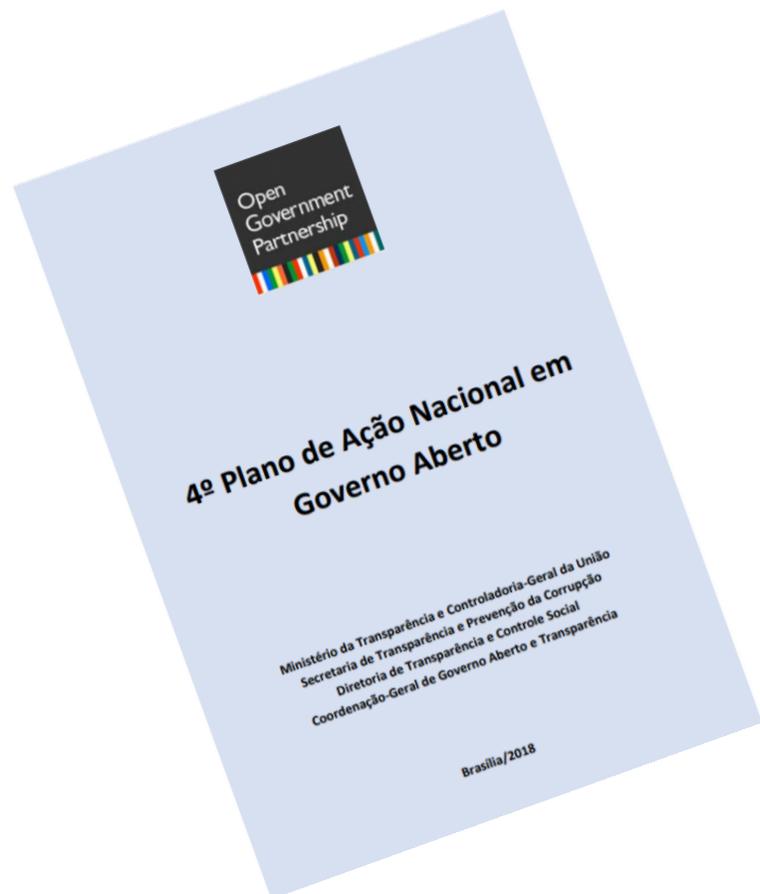
Estratégias e ações  
para estruturar  
processos científicos  
mais colaborativos e  
transparentes



Infraestrutura  
tecnológica para  
depósito e publicação  
dos dados

# Contexto Nacional

Compromisso 3 que tem como propósito “estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil”



MINISTÉRIO DA  
CIÊNCIA, TECNOLOGIA,  
INOVAÇÕES E COMUNICAÇÕES

PÁTRIA AMADA  
BRASIL  
GOVERNO FEDERAL

Embrapa

Ministério da Saúde  
FIOCRUZ  
Fundação Oswaldo Cruz

RNP

ibict 65  
Instituto Brasileiro de Informação  
em Ciência e Tecnologia

CAPES

CNPq

UnB

CNEN

SciELO

JARDIM BOTÂNICO  
DO RIO DE JANEIRO

ANCIB  
Associação Nacional de Pesquisa e  
Pós-Graduação em Ciência da Informação

LABFGV  
LABORATÓRIO DE POLÍTICAS PÚBLICAS  
FUNDAÇÃO GETÚLIO VARGAS - RIO

ABEC  
BRASIL  
Associação Brasileira de Editores Científicos

# Contexto Internacional

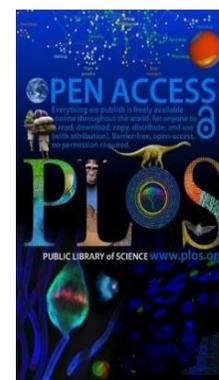
## Políticas Públicas

Holanda, Portugal, França, Austrália, EUA e Reino Unido

## Revistas científicas

Exigência de dados, preprints, open peer review, data papers

## Financiadores (exigência de Plano de Gestão de Dados)



The NEW ENGLAND JOURNAL of MEDICINE



BILL & MELINDA GATES foundation



# Repositório de dados de pesquisas (RDP)

**re3data.org**  
REGISTRY OF RESEARCH DATA REPOSITORIES

+ 2000  
repositórios  
registrados

GBIF (Global Biodiversity Information Facility)  
Acesso livre e aberto a dados de biodiversidade

Occurrences: 1,421,349,621  
Datasets: 53,557  
Publishing institutions: 1,619  
Peer-reviewed papers using data: 4,521

JOHNS HOPKINS UNIVERSITY OF MEDICINE CORONAVIRUS RESOURCE CENTER

COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins U...

Total Confirmed: 11,565,541

Confirmed Cases by Country/Region/Sovereignty:

- 2,922,000 US
- 1,623,284 Brazil
- 697,413 India
- 686,777 Russia
- 305,703 Peru
- 298,557 Chile
- 287,290 United Kingdom

World Map, U.S. Map, Critical Trends

**VectorBase**  
Bioinformatics Resource for Invertebrate Vectors of Human Pathogens

**IMMPORT Shared Data**  
Your site for searching and downloading shared data

**GEINFO**  
INFRAESTRUTURA DE DADOS ESPACIAIS DA EMBRAPA

Armazenamento, organização, curadoria e disponibilização de dados espaciais para aprimorar a pesquisa, desenvolvimento e inovação da agropecuária brasileira.

CAMADAS, MAPAS, DOCUMENTOS

**PubMed**  
labs

**IRD** Influenza Research Database

# Exemplo - Depósito dos dados de pesquisa

Research | Open Access | Published: 05 May 2020

## The neovaginal microbiome of transgender women post-gender reassignment surgery

Kenzie D. Birse, Kateryna Kratzer, Christina Farr Zuend, Sarah Mutch, Laura Noël-Romas, Alana Lamont, Max Abou, Emilia Jalil, Valdiléa Veloso, Beatriz Grinsztejn, Ruth Khalili Friedman, Kristina Brolden, Fridberg Bradley, Vanessa Poliquin, Fan Li, Carolyn Yanavich, Adam Burgener & Grace Aldrovandi

*Microbiome* 8, Article number: 61 (2020) | Cite this article  
1570 Accesses | 11 Altmetric | Metrics

### Abstract

#### Background

Gender reassignment surgery is a procedure some transgender women (TW) undergo for gender-affirming purposes. This often includes the construction of a neovagina using existing penile and scrotal tissue and/or a sigmoid colon graft. There are limited data regarding the composition and function of the neovaginal microbiome representing a major gap in knowledge in neovaginal health.

#### Results

Metaproteomics was performed on secretions collected from the neovaginas ( $n = 5$ ) and rectums ( $n = 7$ ) of TW surgically reassigned via penile inversion/scrotal graft with ( $n = 1$ ) or without ( $n = 4$ ) a sigmoid colon graft extension and compared with secretions from cis vaginas ( $n = 32$ ). We identified 541 unique bacterial proteins from 38 taxa. The most abundant taxa in the neovaginas were *Porphyromonas* (30.2%), *Peptostreptococcus* (9.2%), *Prevotella* (9.0%), *Mobiluncus* (8.0%), and *Jonquetella* (7.2%), while cis vaginas were primarily *Lactobacillus* and *Gardnerella*. Rectal samples were mainly composed of *Prevotella* and *Roseburia*. Neovaginas (median Shannon's  $H$  index = 1.33) had higher alpha diversity compared to cis vaginas (Shannon's  $H = 0.35$ ) ( $p = 7.2E-3$ , Mann-Whitney  $U$  test) and were more similar to the non-*Lactobacillus* dominant/polymicrobial cis vaginas based on beta diversity (perMANOVA,  $p = 0.001$ ,  $r^2 = 0.342$ ). In comparison to cis vaginas, toll-like receptor response, amino acid, and short-chain fatty acid metabolic pathways were increased ( $p < 0.01$ ), while keratinization and cornification proteins were decreased ( $p < 0.001$ ) in the neovaginal proteome.

Download PDF

Sections: Figures, References

Abstract, Introduction, Results, Discussion, Conclusions, Methods, Availability of data and materials, Abbreviations

Advertisement: Prueba gratuita de la base de datos de Springer Protocols, ya disponible en la Biblioteca Virtual del SSPA

## Availability of data and materials

16S rRNA gene sequence files and metadata for the TW and CW samples used in this study have been deposited in Figshare ([https://figshare.com/articles/TW\\_neovaginal\\_rectal\\_buccal\\_seq/11690382](https://figshare.com/articles/TW_neovaginal_rectal_buccal_seq/11690382); [https://figshare.com/articles/CW\\_16S/11710227](https://figshare.com/articles/CW_16S/11710227)). Data sets including unrarefied OTU tables from 16S rRNA gene sequence data, metadata, protein spectral count data, and R scripts used in this study are available in GitHub ([https://github.com/kmbirse/Birse\\_etal\\_Neovaginal-Microbiome](https://github.com/kmbirse/Birse_etal_Neovaginal-Microbiome)).

Supplemental Table 1 | Supplemental Table 2 | Supplemental Table 3 | Supplemental Table 4 | Supplemental Table 5 | Supplemental Table 6 | Supplemental Table 7 | Supplemental Table 8 | Supplemental Table 9

40168\_2020\_804\_MOESM7\_ESM.xlsx (171.5 kB)

Cite Download (171.5 kB) Share Embed + Collect (you need to log in first)

### Additional file 7 of The neovaginal microbiome of transgender women post-gender reassignment surgery

Dataset posted on 07.05.2020, 06:07 by Kenzie D. Birse, Kateryna Kratzer, Christina Farr Zuend, Sarah Mutch, Laura Noël-Romas, Alana Lamont, Max Abou, Emilia Jalil, Valdiléa Veloso, Beatriz Grinsztejn, Ruth Khalili Friedman, Kristina Brolden, Fridberg Bradley, Vanessa Poliquin, Fan Li, Carolyn Yanavich, Adam Burgener, Grace Aldrovandi

Additional file 7: Supplemental Table 1. Percent coverage and mean normalized bacterial spectral counts of KEGG ko level functions from neovaginal and cis vaginal compartments. Supplemental Table 2. Proteins differentially abundant between neovaginas and cis vaginas. Supplemental Table 3. Enriched host pathways positively associated with the neovaginal compartment compared to the cis vaginal compartment. Supplemental Table 4. Enriched host pathways negatively associated with the neovaginal compartment compared to the cis vaginal compartment. Supplemental Table 5. Human protein correlates of bacterial diversity as measured by Shannon's H index. Supplemental Table 6. Enriched host pathways positively associated with bacterial diversity as measured by Shannon's H index. Supplemental Table 7. Enriched host pathways negatively associated with bacterial diversity as measured by Shannon's H index. Supplemental Table 8. Immune cell protein set signatures that overlap with signatures enriched in the neovagina. Supplemental Table 9. Taxa included in the curated database and each taxa's protein detection from initial TREMBL bacteria database search.

FUNDING  
Canadian Institutes of Health Research Fiocruz PMA National Institute of Allergy and Infectious Diseases

TIMELINE  
Submission date: 19.09.2019 | Acceptance date: 10.02.2020 | Publication date: 05.05.2020

KEYWORDS(S)  
Transgender women, Neovaginal, Gender reassignment surgery, Microbiome, Metaproteomics

LICENCE  
CC BY + CC0

EXPORT  
RefWorks, BibTeX, Ref. manager

Log in | Sign up

	A	B	C	D	E
1	Supplemental Table 1 Percent coverage and mean normalized bacterial spectral counts of KEGG ko level functions from neovaginal and cis vaginal compartments				
2					
3	ko number	Function	% coverage cis vagina	% coverage neovagina	Mean bacterial spectral count cis vagina
4	ko00750	Vitamin B6 metabolism	0	0.6	0.01
5	ko00340	Histidine metabolism	0	0.4	0.01
6	ko00072	Synthesis and degradation of ketone bodies	0	0.2	0.01
7	ko00280	Alanine, leucine, and isoleucine degradation	0	0.2	0.01
8	ko00310	Lysine degradation	0	0.2	0.01
9	ko00362	Benzotriazole degradation	0	0.2	0.01
10	ko00380	Tryptophan metabolism	0	0.2	0.01
11	ko00900	Terpenoid backbone biosynthesis	0	0.2	0.01
12	ko01212	Fatty acid metabolism	0	0.2	0.01
13	ko00300	Lysine biosynthesis	0	0.4	0.01
14	ko01210	2-Oxocarboxylic acid metabolism	0	0.4	0.01
15	ko00561	Glycerolipid metabolism	0	0.2	0.01
16	ko00760	Nicotinate and nicotinamide metabolism	0	0.2	0.01
17	ko03320	PPAR signaling pathway	0	0.2	0.01
18	ko00460	Cyanamino acid metabolism	0.03125	0.2	20.4203875
19	ko00670	One carbon pool by folate	0.09375	0.2	78.2745971875

Supplemental Table 1 | Supplemental Table 2 | Supplemental Table 3 | Supplemental Table 4 | Supplemental Table 5 | Supplemental Table 6 | Supplemental Table 7 | Supplemental Table 8 | Supplemental Table 9

40168\_2020\_804\_MOESM7\_ESM.xlsx (171.5 kB)

MD5: d15d7ba0989627d4c6e05828169c |

Descrição dos dados de pesquisa

Dados de pesquisa e documentação

# Dados de pesquisa da Fiocruz em acesso aberto pelo mundo



**Datasets: 13.253**



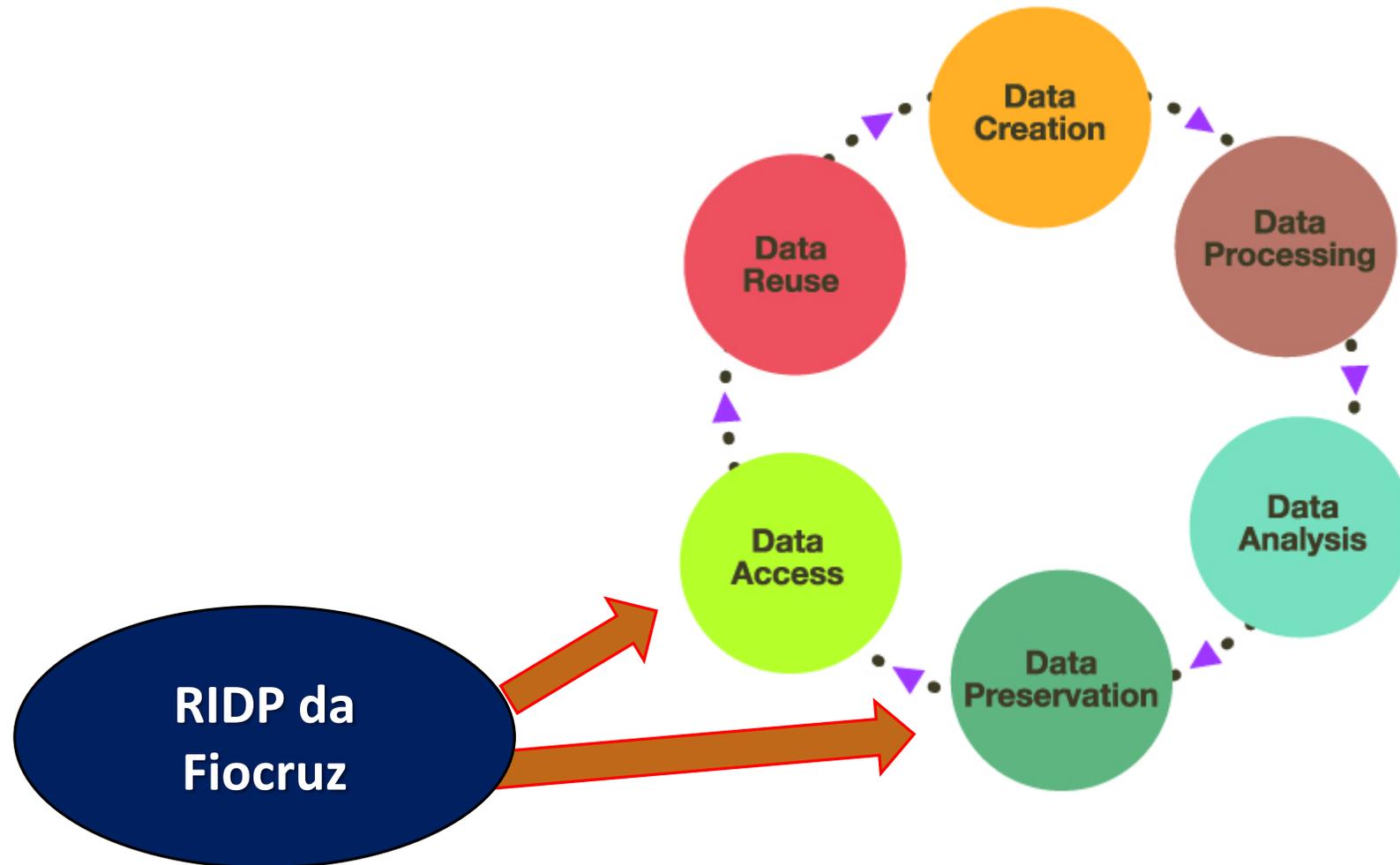
**Datasets: 4.686**



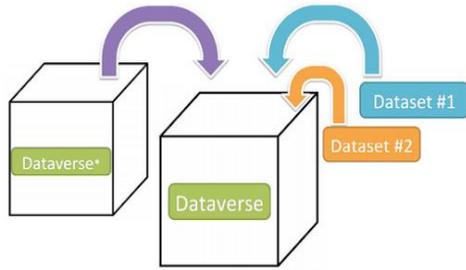
**Datasets: 5.211**

# Repositório Institucional de Dados de Pesquisas (RIDP)

## Ciclo de vida dos dados de pesquisa



## Ferramenta Escolhida

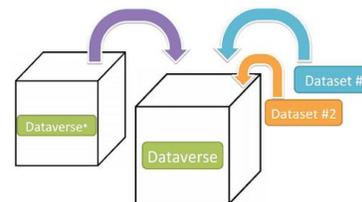


Lançado em **2006** pelo Harvard's Institute for Quantitative Social Science: **15 anos de Dataverse**

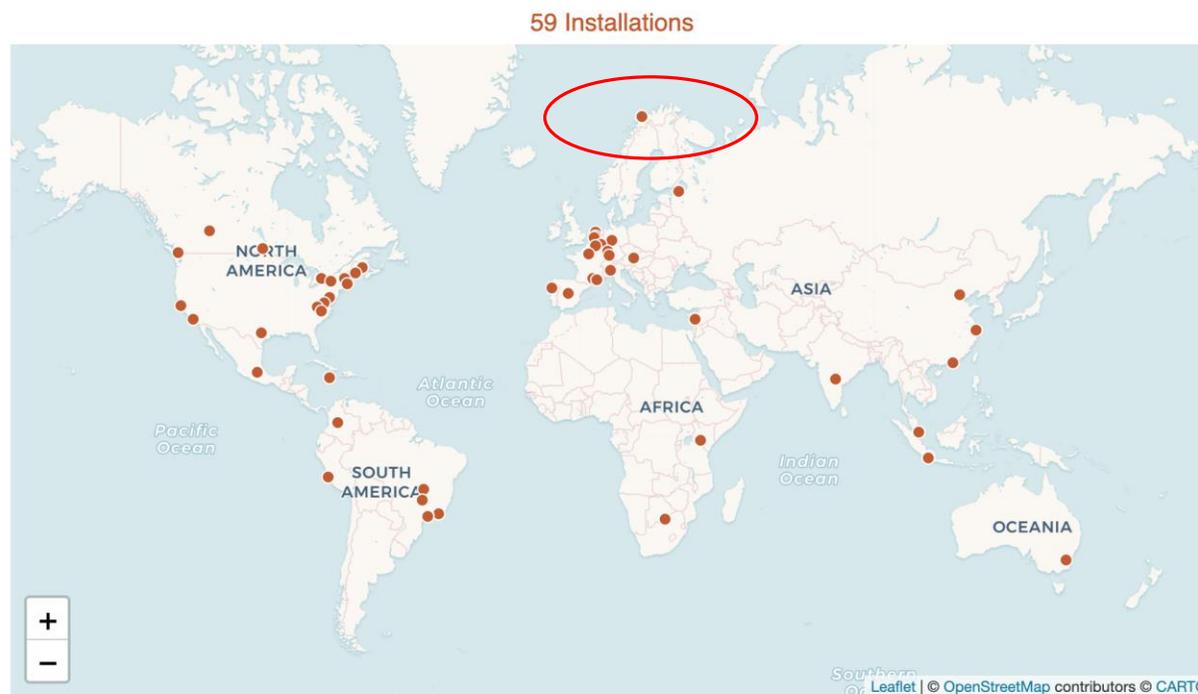
O **Dataverse** é um **Repositório de Dados** de código aberto para **compartilhar, preservar, citar, explorar e analisar dados de pesquisa**. **Pesquisadores, autores de dados, editores, distribuidores de dados e instituições afiliadas** recebem o **crédito apropriado** por meio de uma **citação de dados** com um **identificador persistente** (p. ex.: DOI ou Handle). Cada **conjunto de dados** contém **metadados descritivos** e **arquivos de dados** (incluindo documentação e código que acompanham os dados) (DATAVERSE PROJECT, 2018, tradução nossa).

Harvard Dataverse é o **maior** site de repositórios com:

**Dataverse** 



- **4.000 dataverses**  
(coleções de conjuntos de dados)
- **99.145 conjuntos de dados**
- **475.304 arquivos**
- **16 milhões de downloads**





**8.000 arquivos** disponibilizados no Dataverse se constituem em **documentos** relacionados aos **98.000 “conjuntos de dados”** de pesquisa

**Documentação**

**Dados de Pesquisa**

**Códigos**

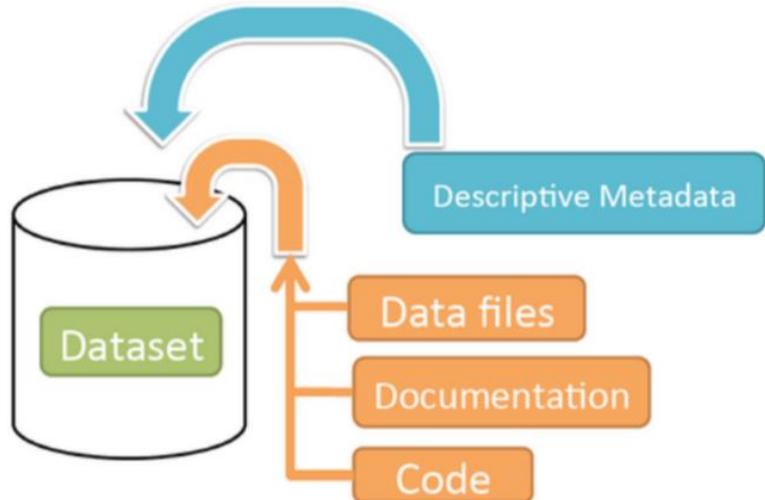
**HARVARD**  
Dataverse

Search - About User Guide Support Sign Up Log In

 Virus Epidemiology and Control (VEC) Dataverse (Kemri Wellcome Trust Research Programme, Kilifi, Kenya) Population dynamics of viral pathogens informing intervention strategies

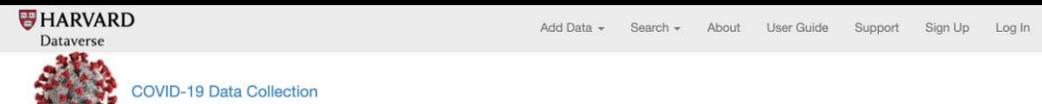
Harvard Dataverse > KWTRP Research Data Repository > Virus Epidemiology and Control (VEC) Dataverse > **Replication Data for: Whole genome sequencing and phylogenetic analysis of Human metapneumovirus strains from Kenya and Zambia**

<input type="checkbox"/>	 <b>EKamau_HMPV_WGS_Readme.txt</b> Plain Text - 4.5 KB - Aug 5, 2019 - 0 Downloads MD5: 94e1f85ded6a0a8b4e99f460ba7de65f Dataset readme file <b>Documentation</b>	<input type="button" value="Download"/>
<input type="checkbox"/>	 <b>Identity_graph_HMPVA_Ggene.csv</b> Comma Separated Values - 3.2 KB - Aug 5, 2019 - 0 Downloads MD5: 85b9d82a093f56f425a618da56dbba64 <b>Data</b>	
<input type="checkbox"/>	 <b>Identity_graph_HMPVA_SHgene.csv</b> Comma Separated Values - 2.4 KB - Aug 5, 2019 - 0 Downloads MD5: 5eec8e812e0c9cdd1a81e7d31a7cf551 <b>Data</b>	
<input type="checkbox"/>	 <b>Identity_graph_HMPVB_Ggene.csv</b> Comma Separated Values - 3.6 KB - Aug 5, 2019 - 0 Downloads MD5: 991131141a43d62276cd3083fd78a7d9 <b>Data</b>	
<input type="checkbox"/>	 <b>Identity_graph_HMPVB_SHgene.csv</b> Comma Separated Values - 2.6 KB - Aug 5, 2019 - 0 Downloads MD5: c8a3d807c5e88443678bcf3b68291802 <b>Data</b>	<input type="button" value="Download"/>
<input type="checkbox"/>	 <b>script_2Jul2019.R</b> R Syntax - 3.0 KB - Aug 5, 2019 - 0 Downloads MD5: 64531365d4f6caaeaf95549d170fdccd Replication code in R <b>Code</b> @mercecrosas	<input type="button" value="Download"/>



17

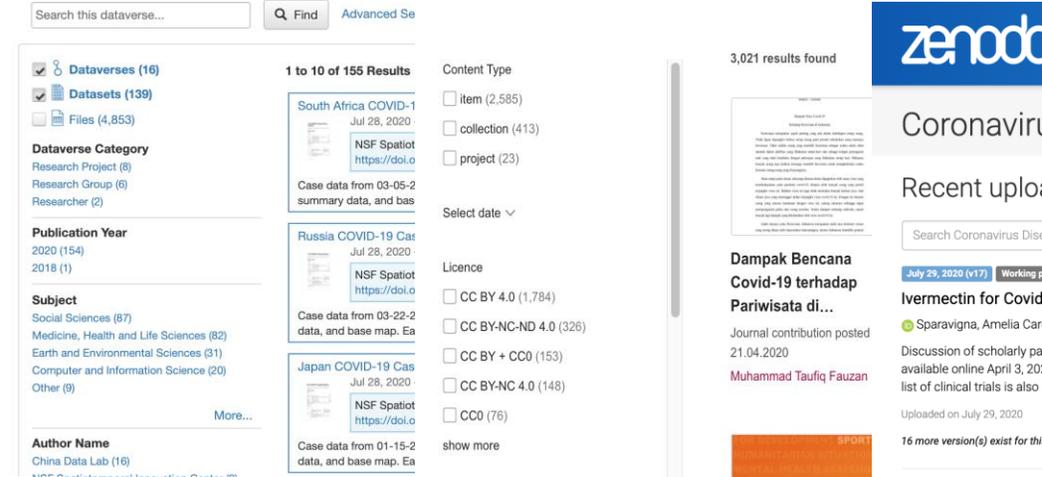
# Agenda de Pesquisa: Pandemia Covid-19



Harvard Dataverse COVID-19 Data Collection



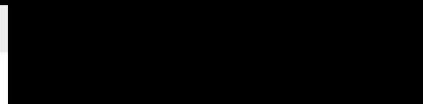
figshare covid-19



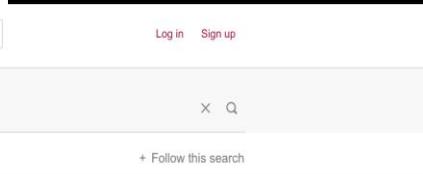
zenodo Coronavirus Disease Research Community - COVID-19



COVID-19 Data Sharing / BR  
Banco de dados compartilhados FAPESP



Dataverses (16)  
Datasets (139)  
Files (4,853)



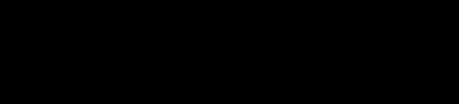
Content Type  
item (2,585)  
collection (413)  
project (23)



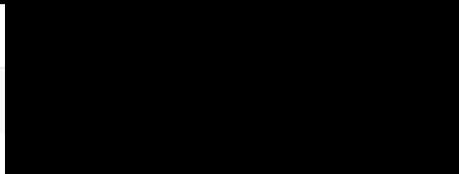
Recent uploads  
Sparavigna, Amelia Carolina;  
Ivermectin for Covid-19



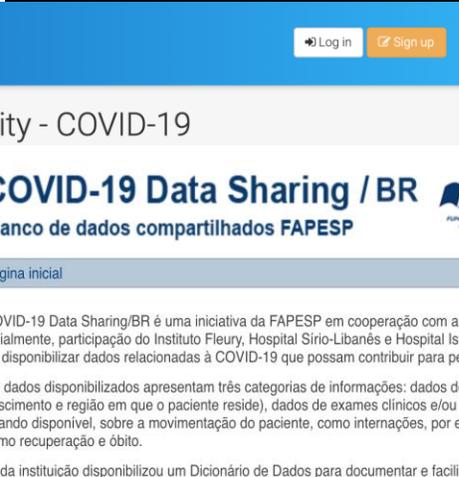
Buscar no Repositório  
Navegar  
Discover



South Africa COVID-19  
Jul 28, 2020  
NSF Spatiot  
https://doi.org/10.7927/H729-9K94



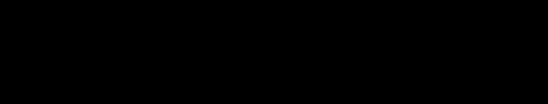
need help?



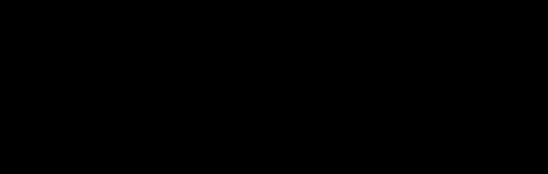
3,021 results found  
Dampak Bencana Covid-19 terhadap Pariwisata di...  
Journal contribution posted 21.04.2020  
Muhammad Taufiq Fauzan



COVID-19 Data Sharing / BR  
Banco de dados compartilhados FAPESP



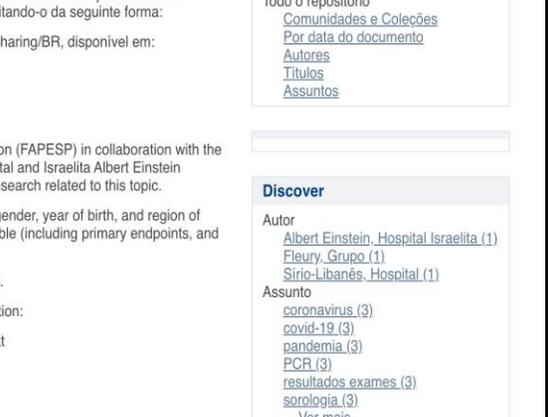
Russia COVID-19 Cases  
Jul 28, 2020  
NSF Spatiot  
https://doi.org/10.7927/H729-9K94



show more



July 29, 2020 (v17) Working paper Open Access  
Ivermectin for Covid-19  
Sparavigna, Amelia Carolina;  
Discussion of scholarly papers published on Ivermectin, starting from the article of Caly et al. in available online April 3, 2020, and of some news that we can find about the use of this drug for a list of clinical trials is also given. The proposed di



COVID-19 Data Sharing / BR  
Banco de dados compartilhados FAPESP

# Composição de Grupo de Trabalho - Fase 1



## **CINCO / VPEIC**

**Coordenação do GTCA no desenvolvimento das estratégias para gestão, compartilhamento e abertura de dados para pesquisa da Fiocruz;  
Representação da Fiocruz no Compromisso 3 do 4º Plano Nacional de Governo Aberto.**



## **ICICT**

**Experiência na coordenação da Rede de Bibliotecas e do Repositório Arca**



## **COGETIC**

**Prover serviços de infraestrutura tecnológica e segurança da informação para sistemas e rede Fiocruz**



## **COC**

**Experiência na coordenação do Sistema de Gestão de Documentos Arquivísticos e do Preservo.**



## **IBICT e RNP**

**Responsáveis por conduzir as ações que envolvem a criação de infraestrutura tecnológica federada de repositórios de dados de pesquisa no Brasil**

# Linha do tempo – Fase 1



**5.421 *datasets* da Fiocruz disponíveis em repositórios internacionais**

Pesquisa em 10/02/2020 no Portal OpenAIRE

**Convite em julho de 2019**

**Início em agosto 2019**

**Atualização do planejamento em outubro de 2019**

**Entrega em Abril de 2020**



**Decisão de usar dados disponíveis em repositórios de dados multidisciplinares**



Repositório Institucional da Fiocruz



**zenodo**



**figshare**



**239**

**Datasets**

# Testes realizados - Fase 1



**Cadastro e  
Importação  
dos dados**



**Links de  
datasets**



**Metadados  
padrão**



**Tamanho  
dos  
arquivos**



**Perfis**



**Publicação,  
embargo e  
acesso**



**Remoção e  
restrição de  
arquivos**



**Segurança**



**Autenticação institucional  
via Shibboleth**



**Interoperabilidade  
com Archivematica**



**Coleta de dados  
pelo OASISBR**



## Entregas realizadas - Fase 1

*Fluxos de depósito e publicação*

*Estratégias segurança do repositório*

*Estratégias para preservação dos dados*

*Aquisição do DOI via CoNciência (CNPq)*

*Relatório Final*



# Desafios para o Repositório de Dados da Fiocruz

## Depósito

Tipos de dados

Padrões de metadados

Vocabulários temáticos

Interoperabilidade

Integração com plataformas

Licenças, cessão e termos de uso

## Publicação

Aplicação do DOI

Aberto

Compartilhado

Embargado

Fechado

## Preservação digital

Prazos de guarda para os dados

Integração com Archivemática para os dados com longos prazos de guarda

## Certificação

Confiabilidade e qualidade

Normas internacionais

## Acesso

Indicadores

Métricas

Reúso

## Fase 2 - RIDP



Rio de Janeiro, 30 de julho de 2020

# Composição de Grupo de Trabalho - Fase 2



**CINCO/VPEIC - Coordenação, organização das agendas e assessoria jurídica**



**•ICICT - Plano de gestão dos dados e Curadoria**



**•COC - Identificação dos tipos de dados de pesquisa**



**•COGETIC - Infraestrutura tecnológica e segurança da informação**



**•GESTEC - Aspectos jurídicos relacionados à inovação e desenvolvimento tecnológico**



**•VPPCB - Acompanhamento da fase 2 como parte interessada**



**•VPGDI - Acompanhamento da fase 2 como parte interessada**



**•IBICT, RNP e CNPQ - Convidados externos**

## Objetivos- Fase 2



**Elaborar proposta para organização e regras para criação das comunidades e coleções temáticas do Repositório Institucional de Dados de Pesquisa da FIOCRUZ (RIDP)**



**Identificar tipos de dados de pesquisa e diversidade de metadados que atendam as necessidades de representação dos dados de pesquisa da FIOCRUZ depositados no Dataverse**



**Realizar testes de publicação dos dados com DOI**



**Disponibilizar os dados de pesquisa coletados para a comunidade Fiocruz**

## Objetivos- Fase 2



**Instituir licenças e termos de consentimento e uso do repositório (auto declaração para depósito dos dados);**



**Elaborar uma proposta de indicadores para acompanhamento gerencial em painéis (*dashboards*)**



**Estabelecer estudo para definição de prazos de guarda dos dados, visando a preservação a longo prazo dos dados com valor de reuso e histórico para pesquisa**



**Avaliar as funcionalidades do RDPI, de caráter experimental, aplicado aos pesquisadores e corrigir eventuais falhas antes de sua implantação definitiva**



**Estabelecer governança do RIDP**

# Cronograma

Macro atividade	Semana 1	Semana 2	Semana 3	Semana 4
Início do piloto 1 (apresentação da pesquisa)				
Identificação dos tipos de dados				
Aplicação do plano de gestão de dados				
Levantamento das questões jurídicas				
Depósito e curadoria dos dados				
Avaliação do piloto				



*Obrigada!*

**Equipe CINCO/VPEIC**