

Piloto Repositório Institucional de Dados de Pesquisa da Fiocruz

Fase 1

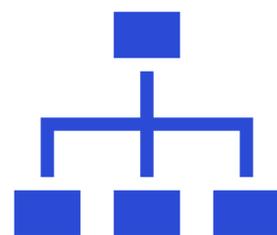
CINCO/VPEIC

Rio de Janeiro, 08 de julho de 2020

Contexto Fiocruz



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



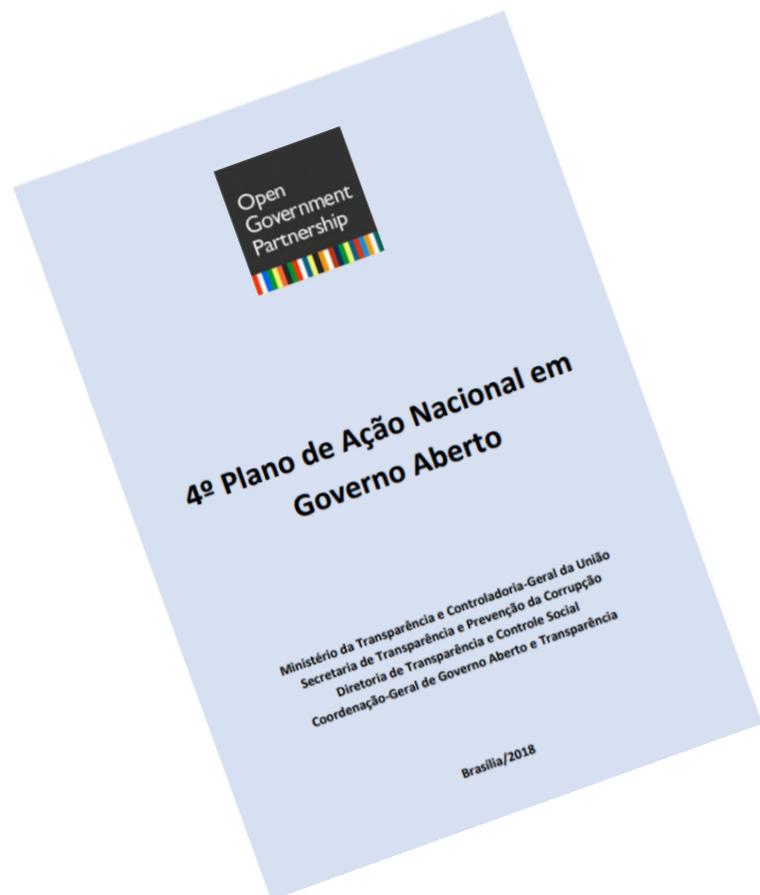
Estratégias e ações
para estruturar
processos científicos
mais colaborativos e
transparentes



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

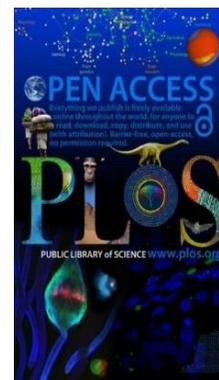
Revistas científicas

Exigência de dados e data papers

Financiadores



BILL & MELINDA
GATES *foundation*



The NEW ENGLAND
JOURNAL of MEDICINE

SCIENTIFIC DATA

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](#), [Emília Jalil](#), [Valdiléa Veloso](#), [Beatriz Grinsztejn](#), [Ruth Khalili Friedman](#), [Krstina Brolden](#), [Frideborg 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.

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/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).

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, Emília Jalil, Valdiléa Veloso, Beatriz Grinsztejn, Ruth Khalili Friedman, Krstina Brolden, Frideborg 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.

19 views **0** downloads **0** citations

READ THE PEER-REVIEWED PUBLICATION
The neovaginal microbiome of transgender women post-gender reassignment surgery

SPRINGER NATURE

CATEGORIES
 • Biochemistry
 • Medicine
 • Microbiology
 • Cell Biology
 • Biotechnology
 • Ecology
 • Biological Sciences not elsewhere classified
 • Cancer
 • Hematology

KEYWORD(S)
 Transgender women Neovaginal
 Gender reassignment surgery
 Microbiome Metaproteomics

LICENCE
 CC BY + CC0

EXPORT
 RefWorks
 BibTeX
 Ref. manager

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

figshare Browse Search on figshare... 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	Valine, leucine and isoleucine degradation	0	0.2	0.01
8	ko00310	Lysine degradation	0	0.2	0.01
9	ko00362	Benzoate 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

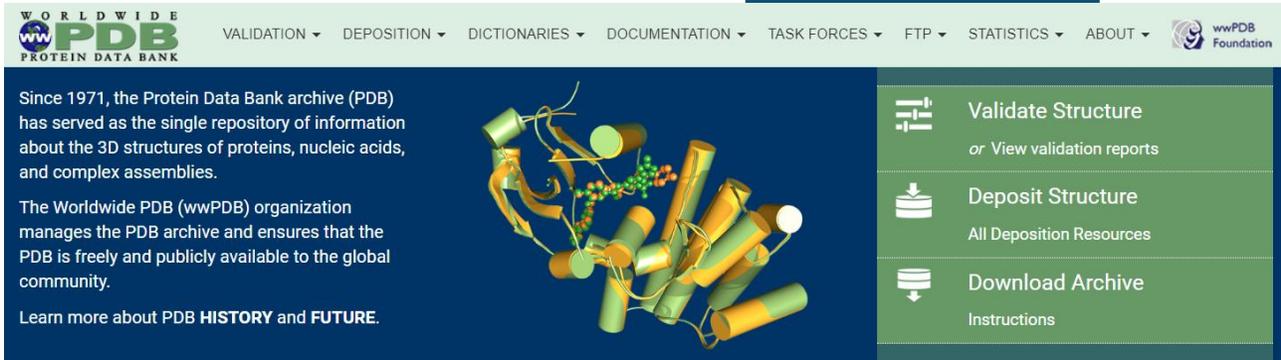
40168_2020_804_MOESM7_ESM.xlsx (171.5 kB)

MDS: d15d7ba0989627d6acfe05828169c | [↩](#)

Repositório de dados de pesquisas (RDP)

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

+ 2000
repositórios
registrados



WORLDWIDE PDB PROTEIN DATA BANK

VALIDATION ▾ DEPOSITION ▾ DICTIONARIES ▾ DOCUMENTATION ▾ TASK FORCES ▾ FTP ▾ STATISTICS ▾ ABOUT ▾

Since 1971, the Protein Data Bank archive (PDB) has served as the single repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies.

The Worldwide PDB (wwPDB) organization manages the PDB archive and ensures that the PDB is freely and publicly available to the global community.

Learn more about PDB **HISTORY** and **FUTURE**.

- Validate Structure
or View validation reports
- Deposit Structure
All Deposition Resources
- Download Archive
Instructions



VectorBase
Bioinformatics Resource for Invertebrate Vectors of Human Pathogens



IMMPORT
Shared Data
Your site for searching and downloading shared data



IRD Influenza Research Database

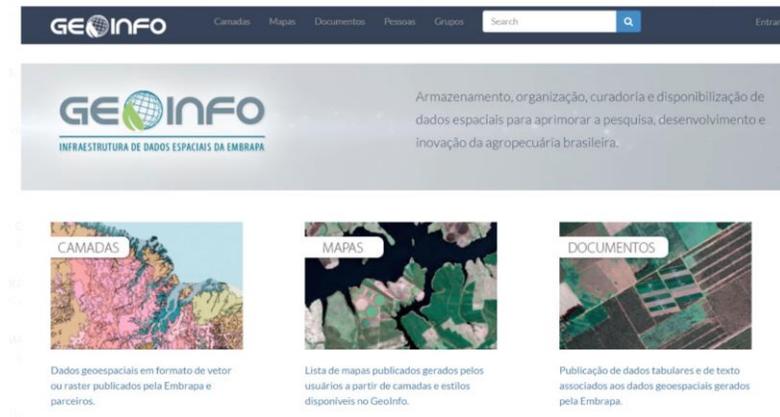


GBIF Global Biodiversity Information Facility

Acesso livre e aberto a dados de biodiversidade

OCCURRENCES SPECIES DATASETS PUBLISHERS RESOURCES

1,421,349,621 Occurrence records
53,557 Datasets
1,619 Publishing institutions
4,328 Peer-reviewed papers using data



GEoINFO

Camadas Mapas Documentos Pessoas Grupos

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

INFRAESTRUTURA DE DADOS ESPACIAIS DA EMBRAPA

- CAMADAS: Dados geoespaciais em formato de vetor ou raster publicados pela Embrapa e parceiros.
- MAPAS: Lista de mapas publicados gerados pelos usuários a partir de camadas e estilos disponíveis no Geoinfo.
- DOCUMENTOS: Publicação de dados tabulares e de texto associados aos dados geoespaciais gerados pela Embrapa.

Dados de pesquisa da Fiocruz em acesso aberto pelo mundo

Busca simples: "FIOCRUZ"



Total datasets: 5211



Total datasets: 226



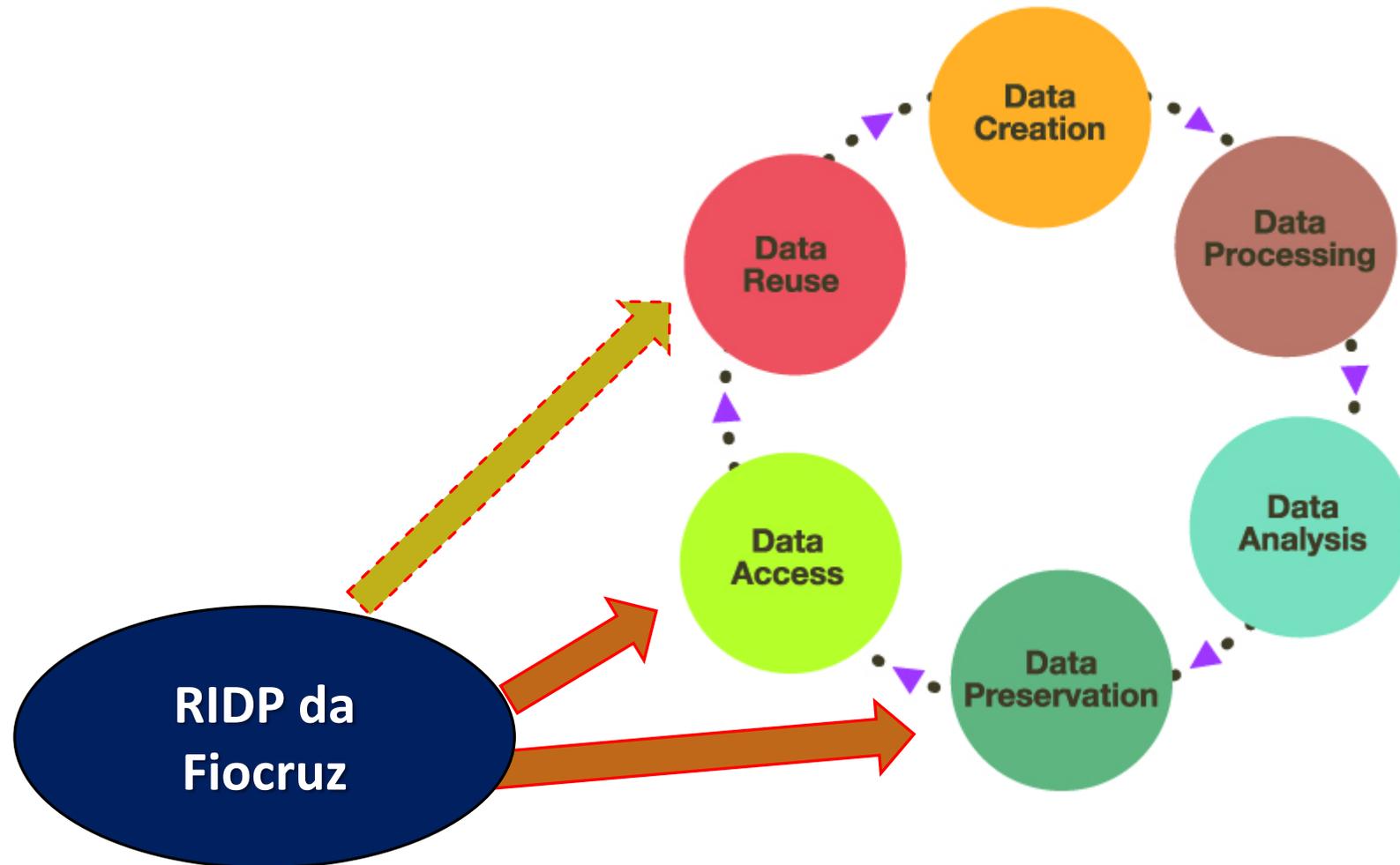
Total datasets: 152



Total de dados associados: 2876

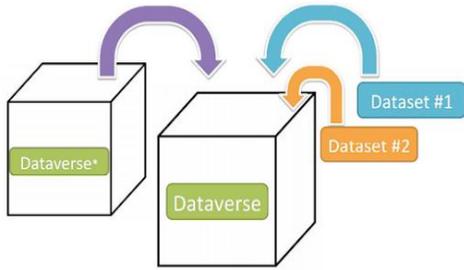
Repositório Institucional de Dados de Pesquisas (RIDP)

Ciclo de vida dos dados de pesquisa



Piloto fase 1 - Ferramenta Escolhida

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



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

- **3.888 dataverses**
(coleções de conjuntos de dados)
- **98K conjuntos de dados**
- **560.000 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

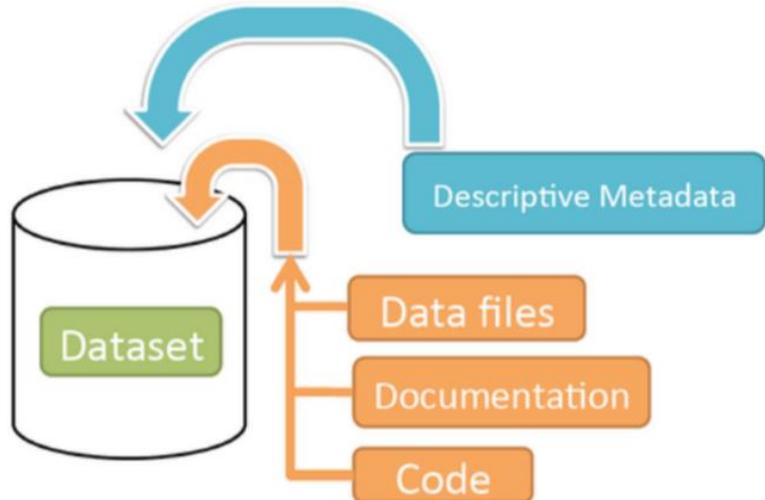
HARVARD
Dataverse

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



Dataset

Descriptive Metadata

Data files

Documentation

Code

17

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



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



Fase 2 – Resultados esperados

Pesquisadores – demandas e necessidades sobre depósito e publicação

Depósitos alinhados aos padrões temáticos (tipos de dados, metadados e vocabulários)

Definição de requisitos para entrada no repositório

Disponibilização de termos de uso e licenças

Governança do Repositório



Desafios para o Repositório de Dados da Fiocruz

Depósito

Tipologia 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

Integridade, autenticidade e boas práticas de qualidade

Normas internacionais como Core Trust Seal, ISO

Acesso

Indicadores

Metricas

Reúso



Obrigada!

Equipe CINCO/VPEIC