

# Ciência Aberta e o papel dos Repositórios

**Maria de Fátima M. Martins Corrêa**

*Vice-Presidência de Educação, Informação e Comunicação (VPEIC/FIOCRUZ)*



22 de Julho de 2021

# Acesso Aberto e Ciência Aberta: Tem diferença?

O **acesso aberto (open access)** é o acesso imediato, gratuito, sem requisitos de registro, a materiais digitais educacionais, acadêmicos, científicos etc., principalmente os artigos de revistas científicas revisadas por pares.



<https://www.opensocietyfoundations.org/explainers/what-open-access>



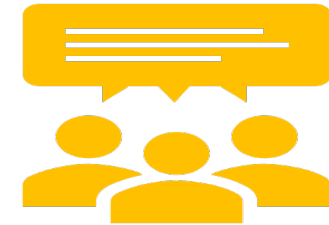
A **ciência aberta** vai além do **acesso aberto** à literatura científica e dados de pesquisa, incluindo a **abertura de todo o processo de pesquisa...**



# Ciência Aberta: O que é?



“É uma mudança de paradigma na forma de fazer ciência”



“Aberta, colaborativa e feita com e para a sociedade” (ANLADA & ABADAL, 2018)

# Acesso Aberto e Ciência Aberta: Tem diferença?

## 1. Acesso livre e gratuito:

- publicações científicas
- dados de pesquisa descritos e/ou analisados



## 1. Acesso livre e gratuito:

- publicações científicas
- dados de pesquisa (FAIR)

## 2. Métricas abertas

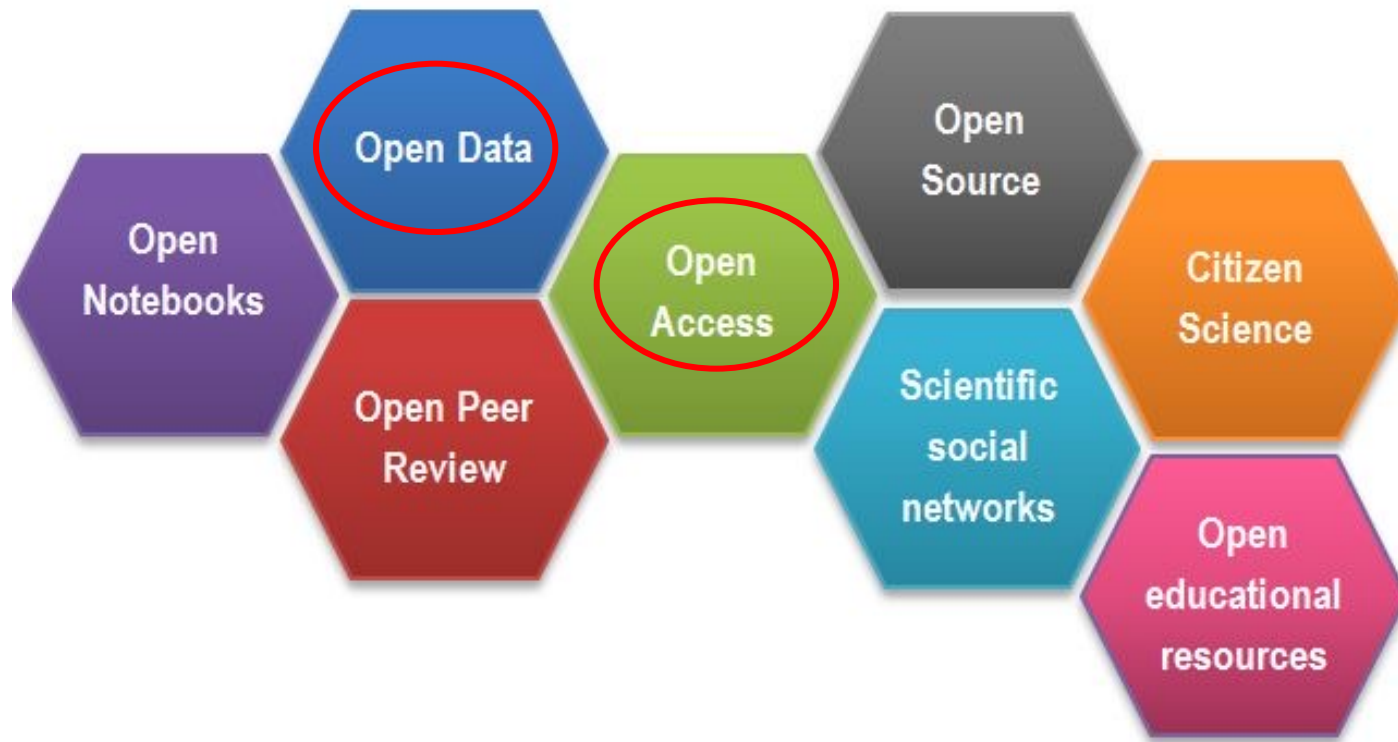
## 3. Software livre

## 4. Recursos educacionais abertos

## 5. Revisão por pares aberta

## 6. Ciência cidadã

# Ciência Aberta: Para que?



“**Ciência aberta** é, por sua vez, um termo **guarda-chuva**, que envolve **múltiplos níveis** e **escopos de abertura**” (ABLAGI; APPEL; MACIEL, 2014)

## Ciência para todos



Transparência



Qualidade



Colaboração



Acesso



Reprodutibilidade



Participação social

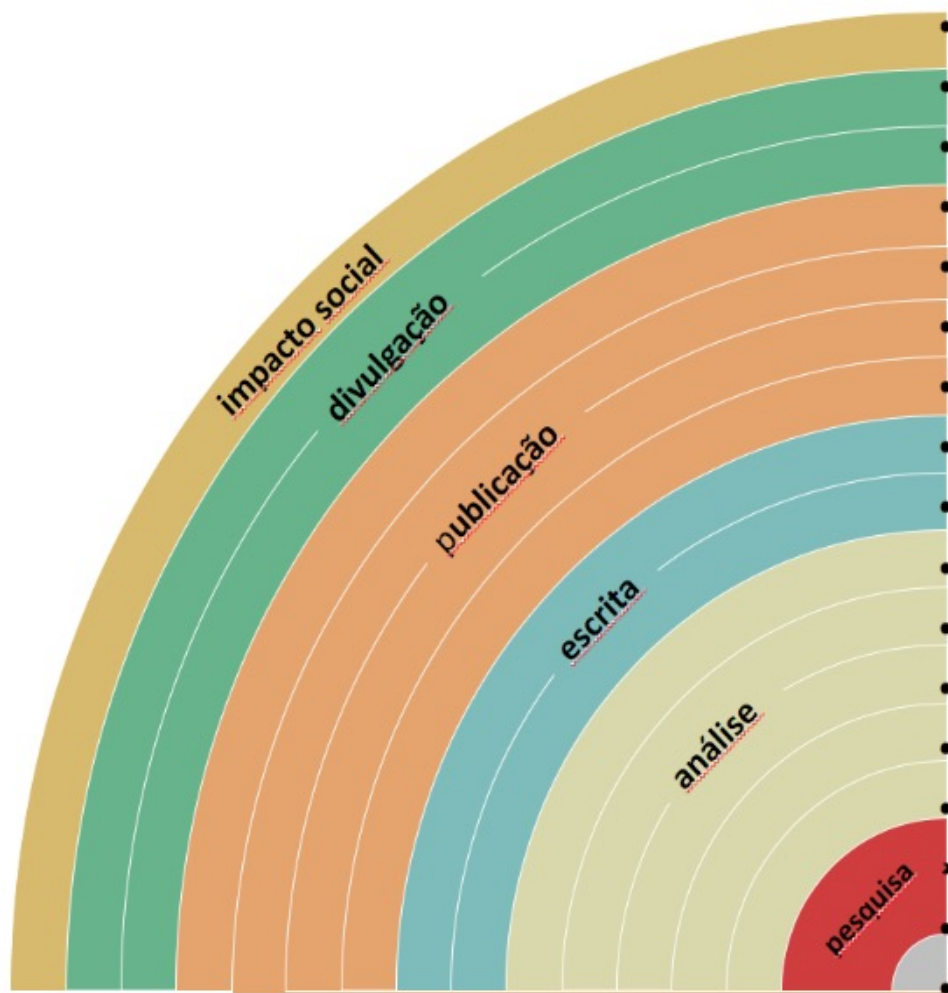


Reutilização em diversos contextos



Visibilidade

# O pesquisador necessita realizar sua pesquisa de forma mais aberta: Como?



- adicionar uma avaliação alternativa, p. ex.: [Altmetrics](#)
- comunicar através das mídias sociais, p. ex.: [Twitter](#)
- compartilhar posters e publicações, p. ex.: [FigShare](#)
- usar licenças abertas, p. ex.: [CC0](#) ou [CC-BY](#)
- publicar em acesso aberto, via “verde” ou “dourada” [DOAJ](#)
- usar a revisão aberta de pares, p. ex.: [PubPeer](#)
- compartilhar os preprints, p. ex.: [OSF](#), [arXiv](#) ou [bioRxiv](#) [bioRxiv](#)
- usar formatos abertos, p. ex.: [Jupyter](#) ou [CoCalc](#)
- abrir o arquivo em XML, p. ex.: [Overleaf](#) ou [Authorea](#)
- compartilhar protocolos e fluxogramas, p. ex.: [Protocols.io](#)
- compartilhar cadernos de laboratório, p. ex.: [OpenNotebookScience](#)
- compartilhar códigos, p. ex.: [GitHub](#) com GNU/MIT license
- compartilhar dados, p. ex.: [Dryad](#), [Zenodo](#) or [Dataverse](#) [zenodo](#) [Dataverse Network](#)
- pre-registrar, p. ex.: [OSF](#) ou [AsPredicted](#)
- comentar abertamente, p. ex.: [Hypothes.is](#)
- usar bibliotecas de referências compartilhadas, p. ex.: [Zotero](#)
- compartilhar ideias e objetivos, p. ex.: [RIO](#)

# Contribuições: Quais são?



- ✓ **Produz novos recursos de informação**, viabilizada pela **retroalimentação** da atividade científica;
- ✓ Altera a **dinâmica** da rede pois o pesquisador pode valer-se de novos **insumos de informação**, ou seja, os **conjuntos de disponibilizados pesquisadores**;
- ✓ O pesquisador também precisará **disponibilizar** os **dados** que produziu, evitando a **perda de dados**;
- ✓ O **acesso** a estes recursos **democratiza** a **produção científica** e **beneficia** toda a sociedade.

# Dados de Pesquisa: O que são?



**Dados** que são **usados** como **fontes primárias** para validar atividades e/ou resultados de **pesquisas, questões técnicas** ou **científicas** ou **artísticas**, e que são usados como **evidências** no **processo de pesquisa**.

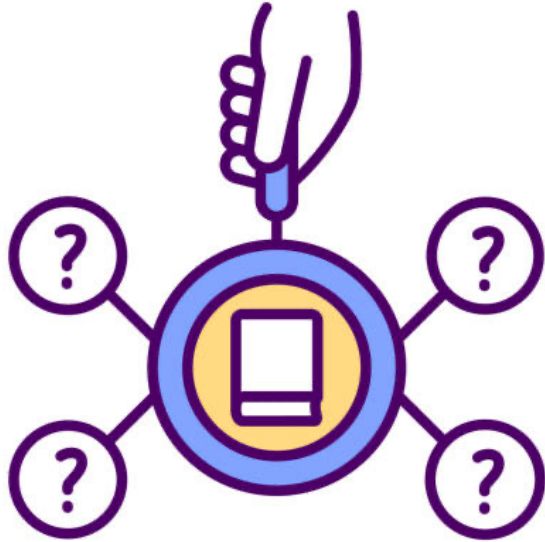
Os dados podem ser:

- ✓ **experimentais**
- ✓ **observacionais**
- ✓ **operacionais**
- ✓ **setor público**
- ✓ **monitoramento**
- ✓ **processados ou reutilizados**

Glosario (CASRAI)

<https://casrai.org/term/research-data>





Os **dados** incluem: **cadernos de laboratório**, **cadernos de anotações de campo**, **dados de pesquisa primária** (incluindo dados em papel ou computadorizados), **questionários**, **fitas de áudio**, **vídeos**, **desenvolvimento de protótipos**, **fotografias**, **filmes**, **resultado de testes**, etc.

As informações sobre a **origem dos dados** também podem incluir:

- ✓ **Como**;
- ✓ **Quando**,
- ✓ **Onde foi coletado e**
- ✓ **Com que** (por exemplo, **instrumentos**).

O **código de software** usado para **gerar**, **comentar** ou **analisar** os dados também podem ser considerados como **dados**.



Grupo de Trabajo de “Depósito y Gestión de datos en Acceso Abierto” del proyecto RECOLECTA. La conservación y reutilización de los datos científicos en España. Informe del grupo de trabajo de buenas prácticas [en línea] Madrid: Fundación Española para la Ciencia y la Tecnología, FECYT (2012) [Consulta 14/01/2013]. Disponible en [WWW.FECYT.ES](http://WWW.FECYT.ES)

# Dados Abertos: O que são?

São dados de pesquisa que:

- Estão disponíveis gratuitamente na internet;
- Permite ser baixado, copiado, analisado, usado, reusado, etc.
- Não tem barreiras econômicas, legais ou técnicas (além de ter acesso à internet)

Sparc Europe (2017). What is Open Data? [Online]  
<https://sparceurope.org/what-we-do/open-data/what-is-open-data/>

**WORLD VIEW** *A personal take on events*



**Open your minds and share your results**  
*An open approach is the best way to maximize the benefits of research for both scientists and the public, says Geoffrey Boulton.*

There is a compelling case for having open access to scientific papers, to enhance the efficacy and reach of scientific communication. But important though this is, the open-access debate has drawn attention away from a deeper issue that is at the heart of the scientific process: that of 'open data'. In an attempt to focus much-needed attention on this subject, I chaired a group that produced *Science as an Open Enterprise*, a policy report from the Royal Society published last week.

Open enquiry has been at the heart of science since the first scientific journals were printed in the seventeenth century. Publication of scientific theories — and the supporting experimental and observational data — permits others to identify errors, to reject or refine theories and to reuse data. Science's capacity for self-correction comes from this openness to scrutiny and challenge.

Modern techniques to gather, store and manipulate data make this more difficult. In the 1980s, I published a paper that presented seven hard-won data points showing the relationship between stress and velocity beneath a glacier. Two years ago, I was involved in an analogous experiment on the Antarctic ice sheet that created more than a billion times more data points. No journal could publish these data, so for them to be accessible, the only option was to deposit the information in a recognized repository, complete with metadata (data about data), and to signpost it in published papers, preferably through live links in the papers' electronic versions.

In the Royal Society report, we argue that this procedure must become the norm, required by journals and accepted by the scientific community as mandatory. As scientists, we have some way to go to achieve this. A recent study of the 50 highest-impact journals in biomedicine showed that only 22 required specific raw data to be made available as a condition of publication. Only 40% of papers fully adhered to the policy and only 9% had deposited the full raw data online (A. A. Alsheikh-Ali *et al.* *PLoS ONE* 6, e24357; 2011).

We also need to be open towards fellow citizens. The massive impact of science on our collective and individual lives has decreased the willingness of many to accept the pronouncements of scientists unless they can verify the strength of the underlying evidence for themselves. The underlying some of the claims of climate science — was in part a motivation for the Royal Society's current report. It is vital that climate scientists accede to requests from members of the public for data not seen to hide behind closed laboratory doors, but engages seriously with the public.

There is, of course, a problem in making data sets open to non-specialists. They are rarely in the form of an Excel spreadsheet, an illusion

under which many politicians labour in their laudable but problematic calls for open data. True openness requires data to be not only accessible, but also intelligible, assessable (who produced the data, what are their qualifications, do they have conflicts of interest?) and reusable. Everyone will benefit from a more open approach. The digital and communications revolutions bring opportunities for research that demand openness and a willingness to share data. These include the assembly of massive data sets from diverse sources, and linking them to allow data integration, dynamic updating and the manipulation of data within electronic publications. Such data-led science offers ways to explore massive data sets for patterns and relationships. Yet this, too, presents a problem. Too often, we scientists seek patterns in data that reflect our preconceived ideas. And when we do publish the data, we too frequently publish only those that support these ideas. This cherry-picking is bad practice and should stop.

For example, there is strong evidence that the partial reporting of the results of clinical trials, skewed towards those with positive outcomes, obscures relationships between cause and effect. We should publish all the data, and we should explore them not just for preconceived relationships, but also for unexpected ones. Without rigorous use and manipulation of data, science merely creates myths. At the same time, communications technologies are displacing the printed page from its dominant role as the medium of scientific communication. This is already exploiting the collective intelligence of the scientific community and shifting the social dynamic of research towards collaboration. This shift has not been mandated by research councils, governments or national academies, but is the consequence of scientists finding more productive and creative ways to do science. Pathfinder disciplines include bioinformatics, astronomy, mathematics, nanotechnology and social and health statistics. Likewise, to extend the reach and depth of these approaches does not need top-down orchestration. It merely requires some constraints to be removed and some enabling changes to be made.

What about costs? Data curation should be viewed as a necessary cost of research. Creative data generation should be a source of primary esteem and a criterion for promotion. We need a revolution in the role of the science library, with data scientists supporting the management of data strategies for both institutions and researchers. We need strategic funding to develop software tools to automate and simplify the creation and exploitation of data sets. And above all, we need scientists to accept that publicly funded research is a public resource. ■

**SCIENCE'S CAPACITY FOR SELF-CORRECTION COMES FROM ITS OPENNESS TO SCRUTINY AND CHALLENGE.**

**Geoffrey Boulton** is regius professor of geology emeritus at the University of Edinburgh, UK.  
e-mail: g.boulton@ed.ac.uk

© 2012 Macmillan Publishers Limited. All rights reserved. 28 JUNE 2012 | NATURE | 11

Boulton G (2012). *Open your minds and share your results.* Nature 486: 441. doi:10.1038/486441a

# Compartilhar os dados: Quais as razões?

## Pesquisadores /instituições

- 📁 Maior visibilidade e impacto => citações dos dados
- 📁 Qualidade / melhores práticas
- 📁 Possibilidade de participação em novas redes de colaborações

OPEN

## Revistas Científicas

- 📁 Transparência
- 📁 Ética e integridade científica
- 📁 Facilita a reprodutibilidade

OPEN

OPEN

## Agências de Fomento

- 📁 Aumenta a visibilidade e o reuso das pesquisas financiadas

OPEN

PGD

## Políticas Públicas

- 📁 Promove políticas públicas baseadas em evidências /dados
- 📁 Promoção dos direitos humanos e da democracia

OPEN

## Sociedade

- 📁 Transferência mais rápida de conhecimento
- 📁 Promove o avanço da ciência
- 📁 Participação social

OPEN

OPEN

Incentivo por compartilhar/disponibilizar

OPEN

Exigência de compartilhar/disponibilizar

PGD

Exigência de Plano de gestão de dados

# Meus dados: Posso compartilhar?



- Exigências agências de fomento, editores científicos e colaboradores;
- Riscos de compartilhar ou não os dados;
- Comprometer o objetivo do projeto – direitos de propriedade industrial;
- Proteção de dados pessoais;
- Confidencialidade em relação à segurança;
- Uso previsto dos dados;
- Métodos de compartilhamento de dados....

# Repositório de Dados: O que é?



São grandes infraestruturas de banco de dados configuradas para coletar, gerenciar, arquivar e acessar conjuntos de dados para análise, compartilhamento e reúso.



Um banco de dados especializado em metadados de arquivos de dados / conjuntos de dados / bancos de dados.



Difere do repositório de publicação principalmente em sua capacidade de:

- > Armazenar metadados em diferentes níveis e/ou hierarquias;
- > Armazenar e receber arquivos de dados em vários formatos para preservação de longo prazo.



# Repositório de Dados: Por que?



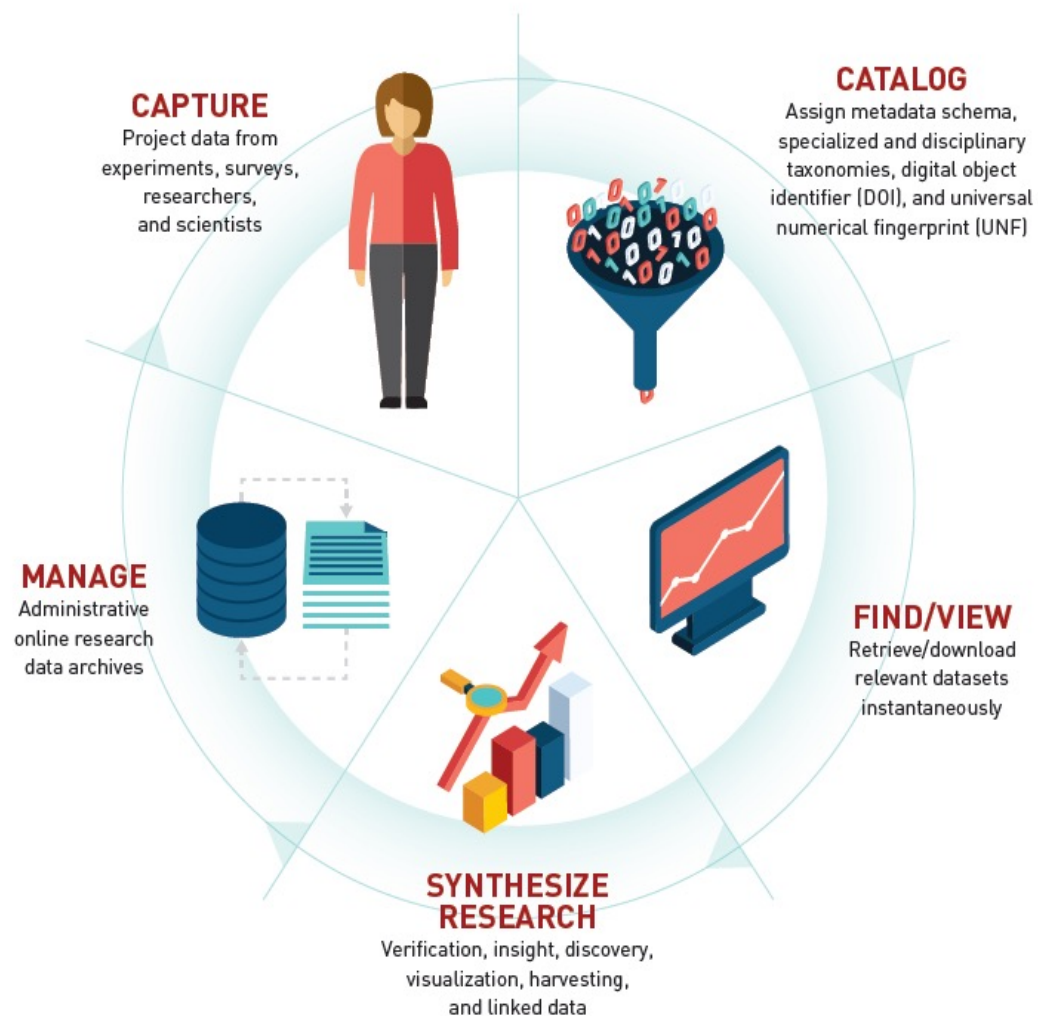
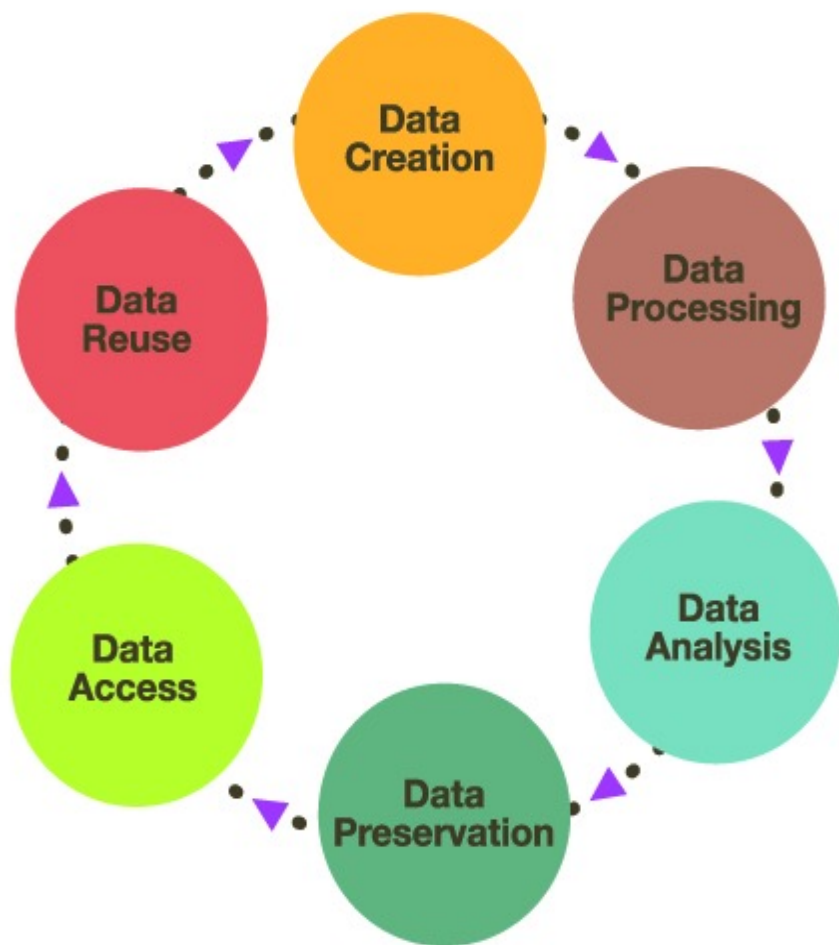
- Identificadores persistentes únicos e citáveis;
- Controle de acesso;
- Termos e licenças de uso;
- Preservação de dados - migrar para novos formatos ou emular formatos antigos;
- Diretrizes do repositório para depósito;
- Backup profissional;
- Padrões de repositório garantem comprometimento e qualidade (metadados, curadoria, etc.)

**Um repositório poderá fornecer estes serviços e muito mais!!!**

**Implementação e  
gestão de um  
repositório de dados:  
Quais são os pontos  
estratégicos?**



# Dados de pesquisa e Repositório de dados: E o ciclo de vida?





# Repositórios de Dados de Pesquisa: O que significa “Acesso”?



ACESSO	LIVRE	RESTRITO	FECHADO
<b>Repositório</b>	Aberto	Aberto ou Restrito	Fechado
<b>Dados</b>	Aberto (embargado, restrito ou fechado)	Restrito (embargado ou fechado)	Fechado
<b>Upload de dados</b>	Aberto ou restrito	Fechado, aberto ou restrito	Fechado

# Quais são os tipos de Repositórios de Dados?

## Multidisciplinares



## Especializados/Temáticos



## Institucionais



# Repositório de dados: Como localizar?

re3data.org  
REGISTRY OF RESEARCH DATA REPOSITORIES

+ 3000  
repositórios  
registrados

re3data.org

Search

Browse ▾

Suggest

Resources ▾

Contact



Filter

Subjects ⊕

Content Types ⊕

Countries ⊕

AID systems ⊕

API ⊕

Certificates ⊕

Data access ⊕

Data access restrictions ⊕

Database access ⊕

Database access restrictions ⊕

Database licenses ⊕

Data licenses ⊕

Data upload ⊕

Data upload restrictions ⊕

Enhanced publication ⊕

Institution responsibility type ⊕

Institution type ⊕

Keywords ⊕

Metadata standards ⊕

PID systems ⊕

Provider types ⊕

Quality management ⊕

Repository languages ⊕

Software ⊕

Syndications ⊕

Repository types ⊕

Versioning ⊕

Search...

Search

Toggle short help

← Previous 1 2 3 4 5 6 7 ... 109 Next →

Sort by ▾

Found 2709 result(s)

## PubChem



Subject(s)

Basic Biological and Medical Research Chemistry Biology Life Sciences Natural Sciences

Content type(s)

Databases Images Structured graphics Scientific and statistical data formats Raw data Plain text Structured text

Country

United States

Pubchem contains 3 databases. 1. PubChem BioAssay: The PubChem BioAssay Database contains bioactivity screens of chemical substances described in PubChem Substance. It provides searchable descriptions of each bioassay, including descriptions of the conditions and readouts specific to that screening procedure. 2. PubChem Compound: The PubChem Compound Database contains validated chemical depiction information provided to describe substances in PubChem Substance. Structures stored within PubChem Compounds are pre-clustered and cross-referenced by identity and similarity groups. 3. PubChem Substance: The PubChem Substance Database contains descriptions of samples, from a variety of sources, and links to biological screening results that are available in PubChem BioAssay. If the chemical contents of a sample are known, the description includes links to PubChem Compound.

## World Data System



International Council for Science World Data System

Subject(s)

Humanities and Social Sciences Life Sciences Natural Sciences Engineering Sciences

Content type(s)

Standard office documents Images Scientific and statistical data formats Raw data Plain text Archived data Structured text

Country

Japan International

The Prototype Data Portal allows to retrieve Data from World Data System (WDS) members. WDS ensures the long-term stewardship and provision of quality-assessed data and data services to the international science community and other stakeholders

# Exemplo - Depósito dos dados de pesquisa

## PLOS NEGLECTED TROPICAL DISEASES

OPEN ACCESS PEER-REVIEWED  
RESEARCH ARTICLE

Artigo

### Factors associated with quality of life in patients with Chagas disease: SaMi-Trop project

Nayara Dornela Quintino, Ester Cerdeira Sabino, José Luiz Padilha da Silva, Antonio Luiz Pinho Ribeiro, Ariela Mota Ferreira, Gabriela Lemes Davi, Claudia Di Lorenzo Oliveira, Clareci Silva Cardoso

Published: May 27, 2020 • <https://doi.org/10.1371/journal.pntd.0008144>

Article	Authors	Metrics	Comments	Media Coverage
---------	---------	---------	----------	----------------

#### Abstract

- Author summary
- Introduction
- Methods
- Results
- Discussion
- Acknowledgments
- References

Reader Comments (0)  
Media Coverage (0)  
Figures

#### Abstract

Chagas disease (CD) is a neglected tropical disease and its main consequence is chronic chagasic cardiomyopathy (CCC). The clinical manifestations of CD can cause functional limitation and impact on the perception of quality of life (QoL). We evaluated the QoL profile of patients with CD living in remote regions, and their association with socio-demographic, behavioral, and clinical characteristics. A cross-sectional study, using the baseline of a cohort study (SaMi-Trop) was conducted in the state of Minas Gerais, Brazil, an endemic region for CD. The cohort is composed of 1959 patients with CD. The patients were interviewed, and blood collection and electrocardiogram (ECG) were performed. The WHOQOL-BREF scale was used to assess QoL. Descriptive statistics were performed and, for the QoL scale, measures of central tendency and dispersion were calculated. Inflated beta regression models constructed using the gamlss framework were used, and fitted using R software. A significance level of 5% for all analyses was adopted. QoL was evaluated in 625 participants. The Environment (57.66) and Social Relationships (73.17) domains presented the lowest and highest QoL score respectively. There was no statistically significant difference between QoL domains in the presence of CCC. In the multivariate analysis, the factors associated with a lower QoL were in the Physical domain age increase (OR: 0.95, CI: 0.91–0.99) and using an angiotensin converting enzyme inhibitor (OR 0.89; CI: 0.80–0.99); in the Social Relations domain a history of acute myocardial infarction (OR: 0.75; CI: 0.61–0.92); and in the Environment domain age increase (OR 0.94, CI 0.91–0.97). The factor associated with higher QoL was the use of angiotensin receptor blockers (OR: 1.15; CI 1.04–1.26). The findings of this study reinforce the importance of QoL indicators for planning care and clinical management of patients from remote regions with CD.

**Copyright:** © 2020 Quintino et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability:** Data are available in <https://figshare.com/s/6e7bdf135cd5b256d48b>.

**Funding:** The SAMI-TROP cohort study is supported by The National Institutes of Health-NIH (grant number P50 AI098461-02). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.



Data View Variable View

paper\_qol.sav (55.25 kB)

MD5: 04eaea8b2b487b37567785a0ad514b66

Download (55.25 kB)

This item is shared privately

Descrição dos dados de pesquisa

### SaMi-Trop Quality of Life Study

Dataset modified on 24.04.2020, 18:06

SaMi-Trop is a prospective cohort study with two-year follow-up, including one baseline visit.

FUNDING

The SAMI-TROP cohort study is supported by The National Institutes of Health-NIH, P50 AI098461-02, <https://www.niaid.nih.gov/>.

CATEGORIES

- Biomarkers
- Diseases
- Health Care

KEYWORD(S)

Chagas Disease Quality of life

LICENCE



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Table: Data List																			
2	study_id	Sexo	idade	alfabetizado	Estado_Civil_Final	Renda_Corrig	num_residentes	habito_cigarro	atividade_fisica	diabetes	iam	hipertensao	Alcool	Numero_Medicamentos_Cat	Digoxina1	IECA	Diureticos	ARA	Betabloqueador	Amiodarona
3	2	0	60	0	0	672	7	1	0	0	1	1	0	2	0	0	1	1	0	0
4	3	1	42	1	1	672	2	.	0	0	0	0	.	.	0	0	0	0	0	0
5	8	1	66	1	1	550	2	0	0	0	0	0	0	1	0	0	1	1	0	0
6	11	0	51	1	0	500	1	2	0	0	0	1	1	1	1	0	1	0	0	0
7	12	1	58	1	3	670	1	0	1	0	0	0	0	1	0	1	1	0	0	0
8	15	1	67	0	3	670	2	0	0	0	0	0	0	2	0	1	1	0	0	0
9	16	0	36	0	0	500	7	0	0	0	0	0	0	2	0	1	1	0	0	0
10	17	1	75	0	3	1300	3	0	1	0	0	1	0	2	0	1	1	0	1	1
11	18	0	53	1	3	70	1	0	0	0	0	1	0	2	0	1	1	0	1	0
12	21	0	73	0	1	1300	2	1	1	1	0	1	0	1	0	1	0	0	0	0
13	23	0	0	1	1	500	3	1	1	0	0	1	0	0	0	0	0	0	0	0
14	26	0	66	1	1	1660	2	0	1	0	0	1	0	0	0	1	1	0	0	0
15	27	1	63	0	1	2000	3	0	0	1	0	1	0	3	0	0	1	1	0	0
16	31	1	51	1	3	1000	3	0	1	1	0	1	0	2	0	1	1	0	1	0
17	29	1	83	0	3	1300	2	0	0	0	0	1	0	1	0	0	1	1	0	1

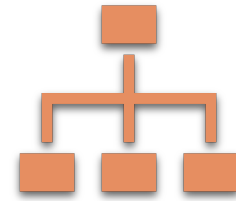
URL onde os dados foram depositados

Dados de pesquisa

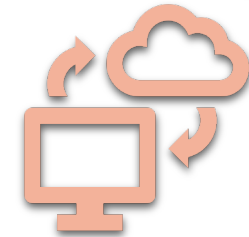
# Contexto Fiocruz



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



Manifesto de Acesso Aberto  
a Dados da Pesquisa  
Brasileira para Ciência  
Cidadã (IBICT)



Fiocruz assume  
Coordenação  
do  
Marco 4 na OGP



Implementação do  
Arca Dados



2011

2014

2016

2018

2019

2020

2021

Acesso Aberto

Compromisso 3:  
Estabelecer mecanismos de  
governança de dados científicos  
para o avanço da Ciência  
Aberta no Brasil

Piloto Fase 1  
Repositório de Dados  
Recursos Educacionais  
Abertos

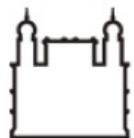
Piloto Fase 2  
Repositório de Dados  
Política de gestão,  
compartilhamento e  
abertura de dados  
da Fiocruz

Parceria para Governo Aberto  
OGP (Open Government  
Partnership  
Iniciativa Multilateral Internacional



07/12/2020

SEI/FIOCRUZ - 0437852 - Portaria da Presidência



Ministério da Saúde

FIOCRUZ  
Fundação Oswaldo Cruz

Boletim de Serviço Eletrônico em 07/12/2020

#### Portaria da Presidência

PORTARIA Nº 5760, de 07 de dezembro de 2020

A Presidente da Fundação Oswaldo Cruz, no uso das atribuições que lhe são conferidas pelo Decreto de 03 de janeiro de 2017, da Presidência da República e pelo Decreto nº 8.932, de 14 de dezembro de 2016 - Estatuto da Fiocruz,

#### RESOLVE:

##### 1.0 - PROPÓSITO

Instituir a "Política de gestão, compartilhamento e abertura de dados para pesquisa", aprovada pelo Conselho Deliberativo da Fiocruz em reunião de 24 de setembro de 2020.





<https://www.arca.fiocruz.br/handle/icict/46408>




# Estratégias Iniciais da Fiocruz

## *Dados para pesquisa*



### Política institucional

-  Termo de Referência
-  Consulta interna




### Capacitação

-  Curso EAD
-  Oficinas
-  Disciplina transversal

### Ferramentas e infraestrutura

-  Plano de Gestão de Dados
-  Repositório de dados

### Projetos pilotos

-  Identificação de tipos de dados
-  Aspectos éticos e jurídicos
-  Fluxos



# The Dataverse Project

- ⌘ Uma **plataforma de código aberto** para **publicar, citar e armazenar os arquivos de dados de pesquisa**
- ⌘ Desenvolvido no **Harvard's Institute for Quantitative Social Ciência (IQSS)** desde 2006 e com subsídios, em colaboração com instituições ao redor do mundo.
- ⌘ Construído para oferecer **suporte** a diversos **tipos de dados, usuários e fluxos de trabalho**

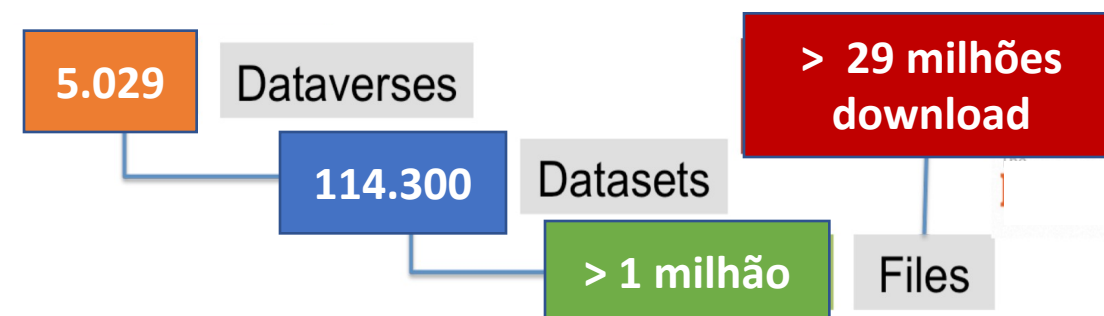
<https://dataverse.harvard.edu/dataverse/harvard>



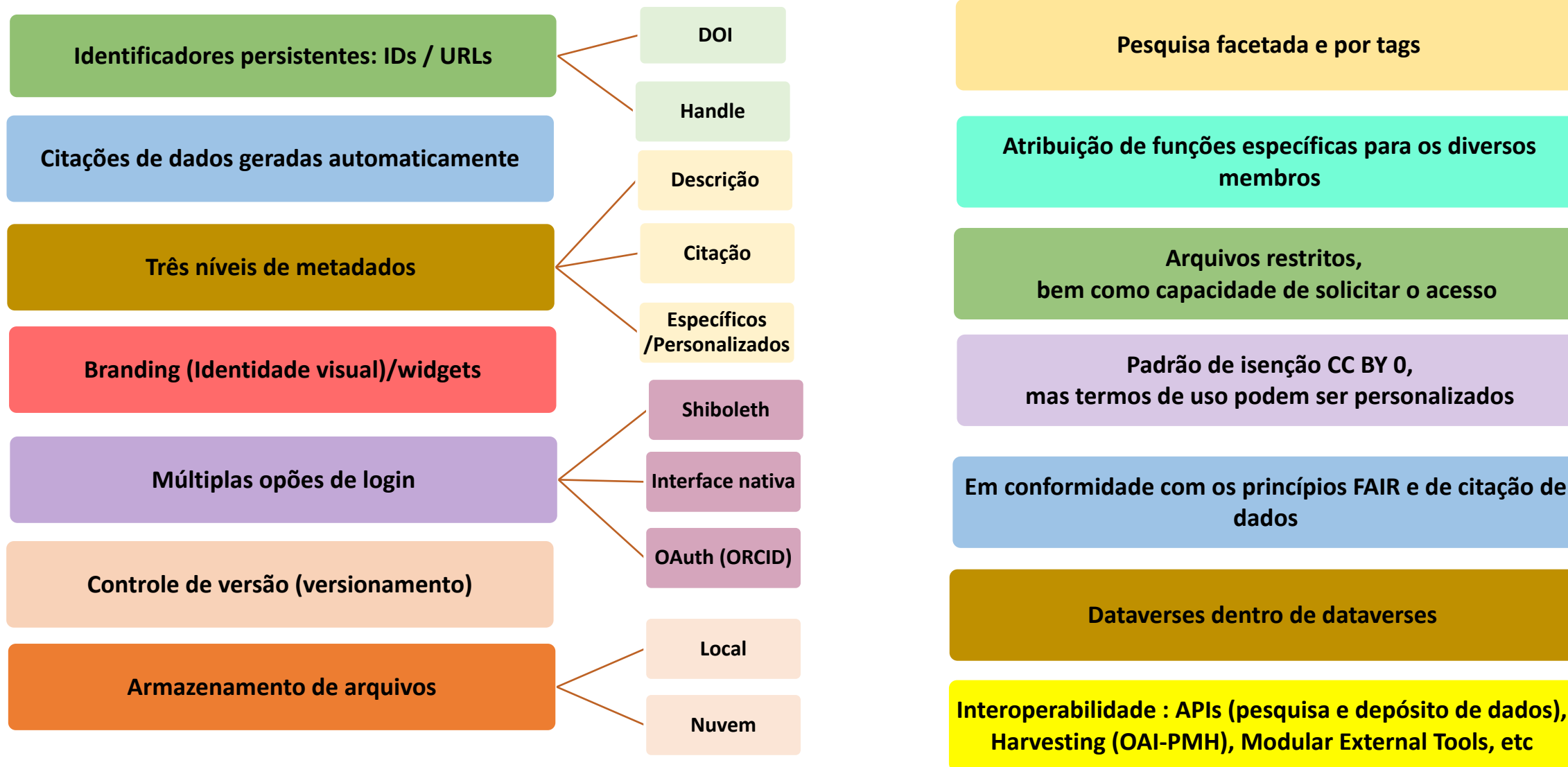
## Quem usa o Dataverse?

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

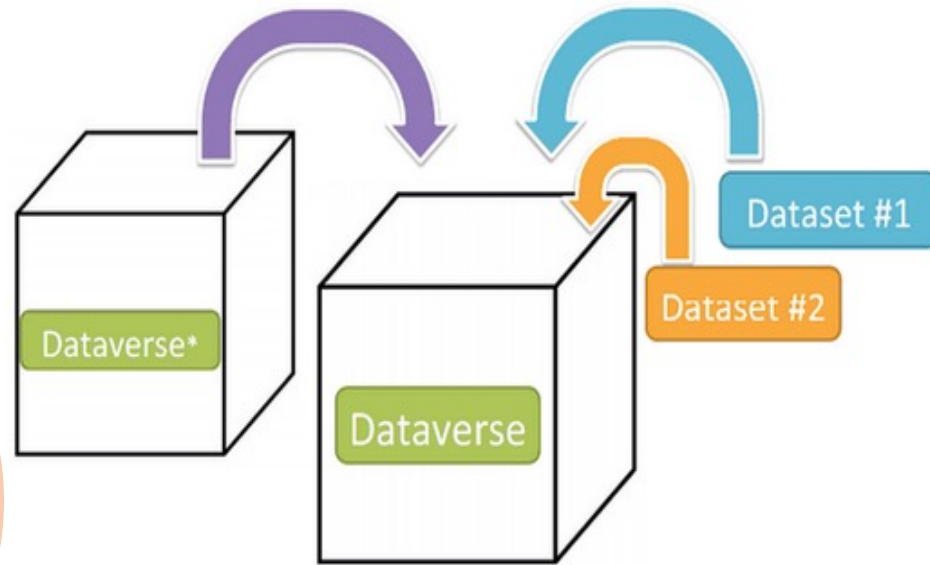
- ⌘ **69 instalações** no mundo



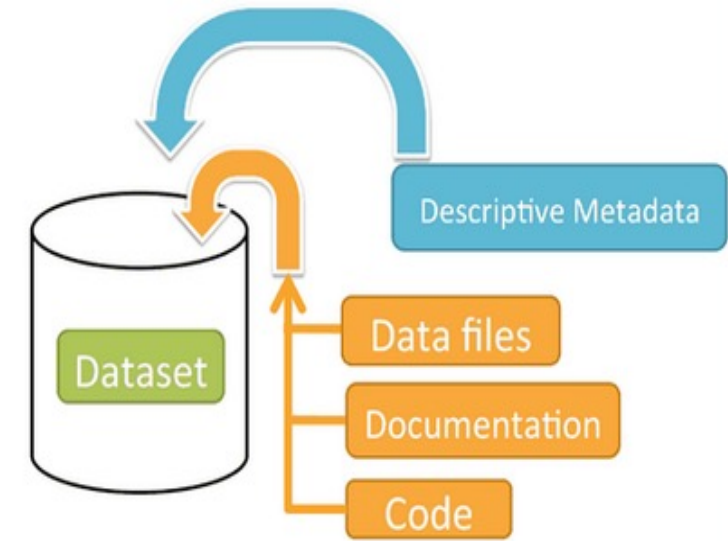
# Dataverse: Quais as vantagens?



O que é um dataverse e um dataset?



Container for your **Datasets** and/or **Dataverses\***



Container for your data, documentation, and code.

\* Dataverses can now contain other Dataverses (this replaces Collections & Subnetworks)

Um **Dataverse** é um “container” de **conjuntos de dados (dataset)** e/ou outros **Dataverses** (conhecido como Sub-Dataverse). O **conjunto de dados** é efetivamente um “container” para armazenar todos os **dados de pesquisa, código, documentação de apoio e metadados** que descrevem os dados.

# DATAVERSE: Destaques importantes

Dataverse de Teste para Validação (FUNDAÇÃO OSWALDO CRUZ)

FIOCRUZ > Dataverse de Teste para Validação >

## (TESTE DO DATAVERSE) Modelos matemáticos aplicados a Sars-CoV-2: Plataforma interativa para avaliação de cenários e modelos matemáticos aplicados a Sars-CoV-2

Version 1.1

Citação



Oliveira, Juliane Fonseca de, 2021, "(TESTE DO DATAVERSE) Modelos matemáticos aplicados a Sars-CoV-2: Plataforma interativa para avaliação de cenários e modelos matemáticos aplicados a Sars-CoV-2", <https://doi.org/10.80458/OWDUC9>, FIOCRUZ, V1

Cite Dataset ▾

Learn about [Data Citation Standards](#).

Identificador persistente

Versionamento

Access Dataset ▾

Contact Owner

Share

Dataset Metrics ?

Files

Metadata

Terms

Versions

1 Download ?

Search this dataset...

Find

Filter by

File Type: All ▾

Access: All ▾

File Tag: All ▾

1 to 2 of 2 Files



N1.2.pdf

Adobe PDF - 287.2 KB - Jan 28, 2021 - 1 Download  
MD5: e8afb310a86ce1c3dc2745677d154a94



N1.pdf

Adobe PDF - 287.2 KB - Jan 28, 2021 - 0 Downloads  
MD5: e8afb310a86ce1c3dc2745677d154a94

Tag PDF Documentação

Controle do tipo de acesso

Files Metadata Terms Versions

Todas as informações do dataset

Dataset	Summary	Contributors	Published
1.1	<a href="#">View Details</a>	Hataanderson Luiz Cabral dos Santos	Jan 28, 2021
1.0	This is the first published version.	Hataanderson Luiz Cabral dos Santos	Jan 28, 2021



Um parte dos **arquivos** disponibilizados no Dataverse se constituem em **documentos** relacionados aos **“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"/>

Dataset

Descriptive Metadata

Data files

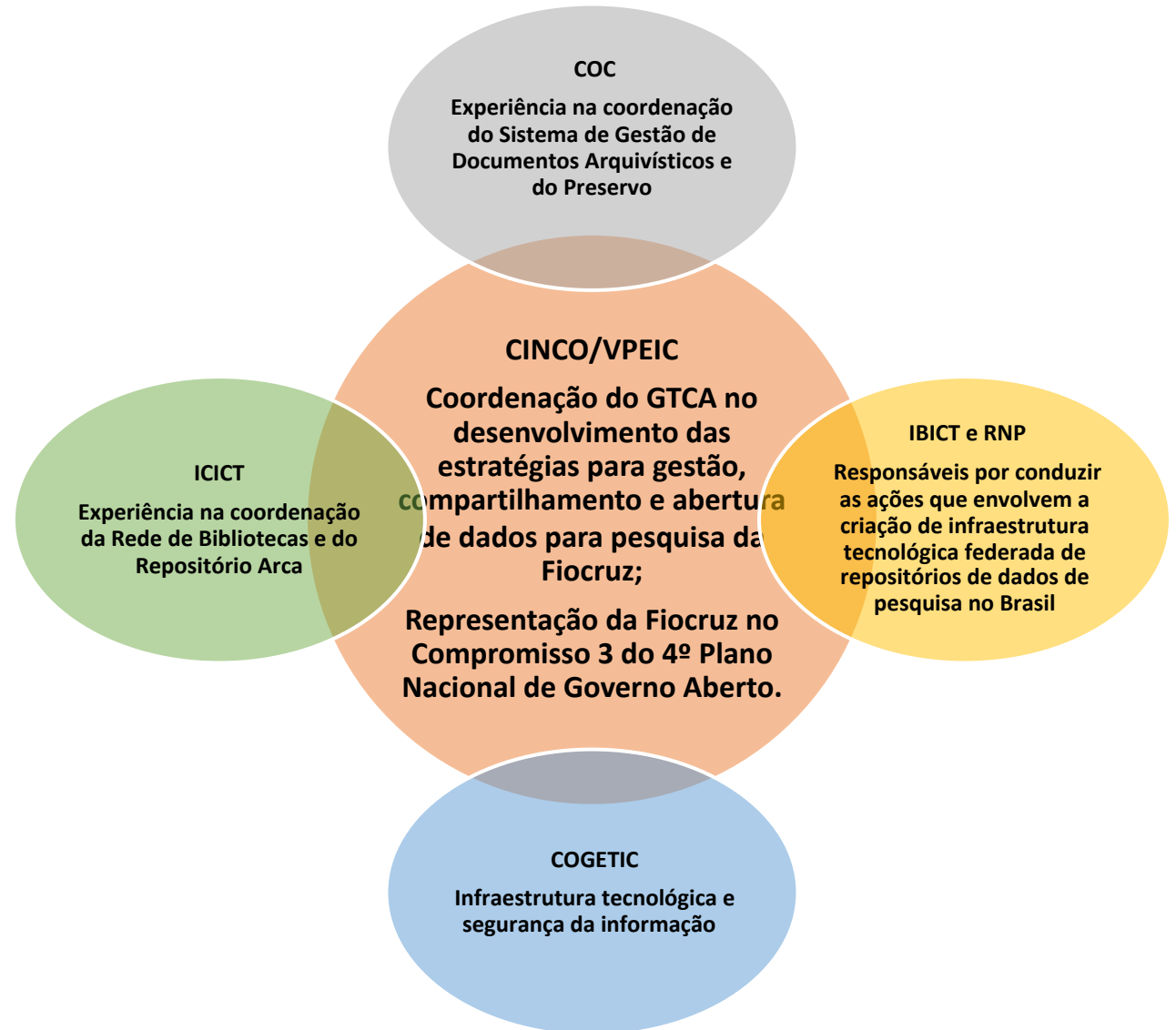
Documentation

Code

# Composição de Grupo de Trabalho



## Fase 1 – Arca Dados



# Testes realizados - Fase 1



**Cadastro e  
Importação  
dos dados**



**Links de  
datasets**



**Metadados  
padrão**



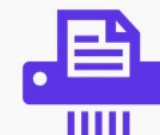
**Tamanho  
dos  
arquivos**



**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)*

---

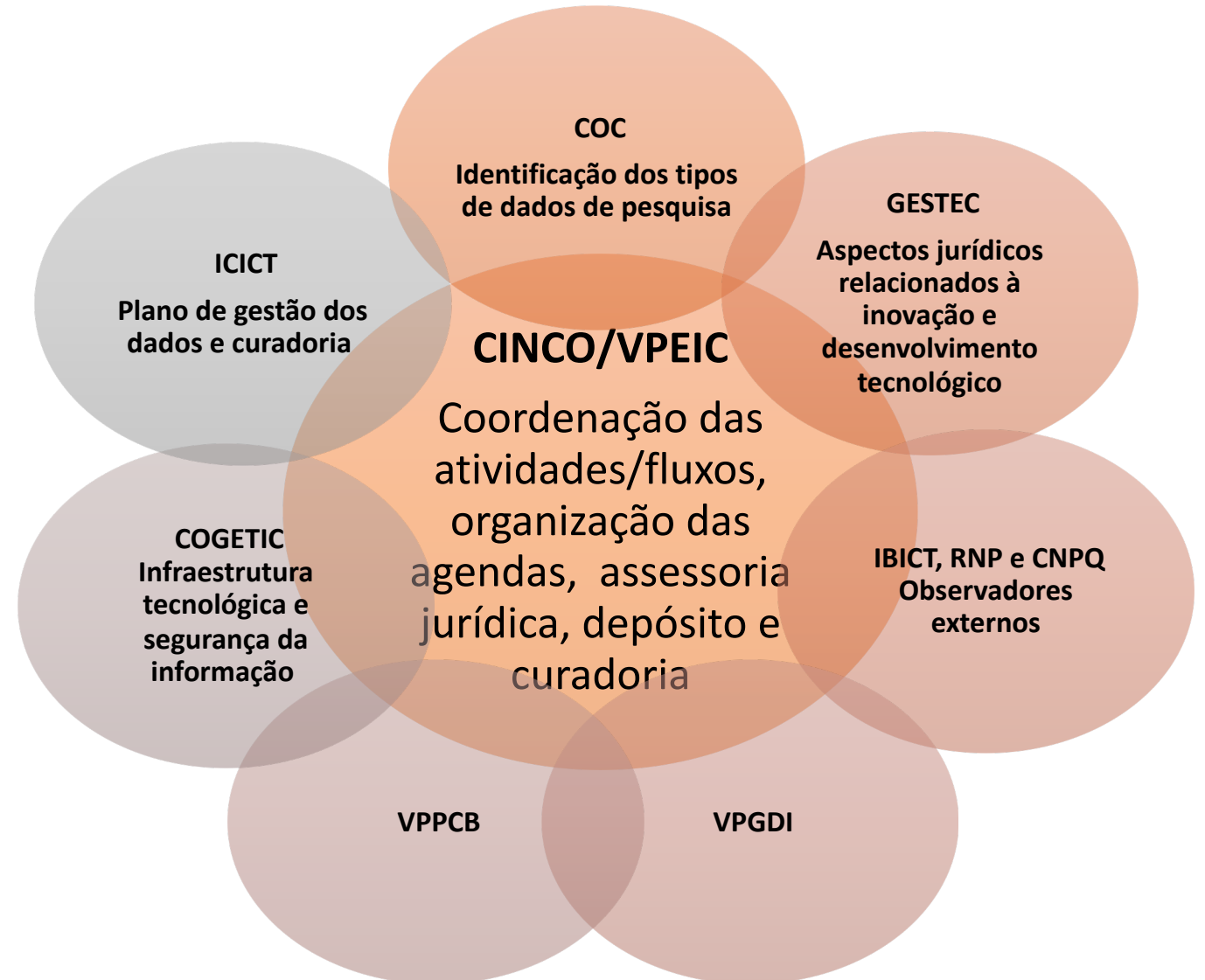
Relatório Final



# Funções Grupos de Trabalho



**Fase 2 – Arca Dados e processos necessários para depósito e publicação de dados**



Preenchimento do Plano de Gestão de Dados

Identificação e análise dos tipos de dados de pesquisa

Levantamento da conformidade em aspectos éticos e jurídicos

Avaliação se dados tem potencial de patentes

Infraestrutura tecnológica e segurança da informação

Depósito com curadoria de metadados

Inclusão de Termos de uso e Licenças

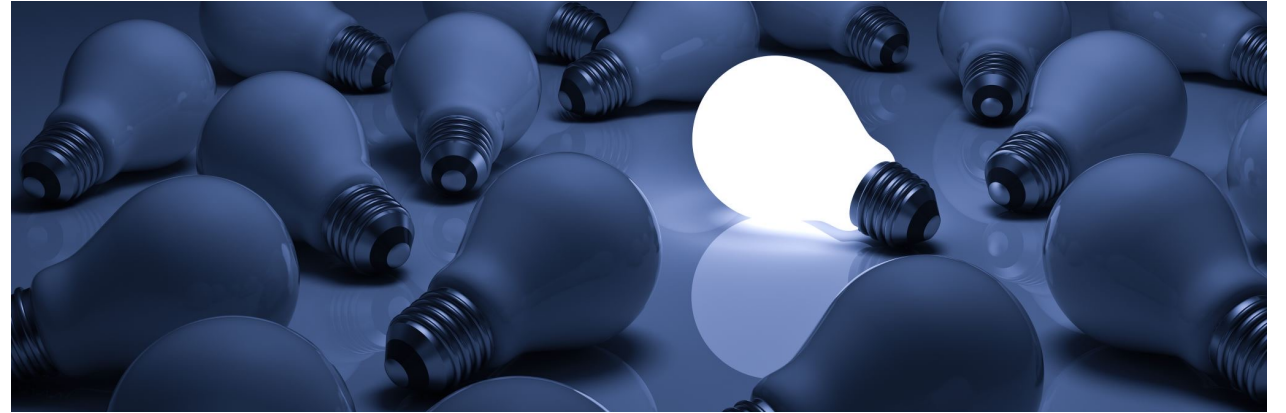
Publicação dos dados

Plano de Preservação Digital para guarda permanente em RDC

**Etapas do Piloto – Fase 2**



# Temáticas dos Pilotos



**Saúde Mental – “Morar em Liberdade”- Luta Antimanicomial**

**Vigilância da Doença de Chagas em Minas Gerais – Programa de controle da Doença de Chagas**

**Rede de Pesquisa Clínica e Aplicada em Chikungunya (Replick) - Consórcio de estudos clínicos aplicados**

**Monitoramento de casos de síndrome respiratória aguda grave (SRAG) notificados no SIVEP-Gripe - Infogripe**



# ARCA DADOS



**FIOCRUZ**

Oswaldo Cruz Foundation:  
an institution in the service of life

Fundação Oswaldo Cruz

Fiocruz

<https://dadosdepesquisa.fiocruz.br/>

 Metrics

663 Downloads

 Contact  Share

 Edit ▾

Criada em 25 de maio de 1900 - com o nome de Instituto Soroterápico Federal -, a Fundação Oswaldo Cruz nasceu com a missão de combater os grandes problemas da saúde pública brasileira. Hoje, a instituição, vinculada ao Ministério da Saúde, tem como missão "promover a saúde e o desenvolvimento social, gerar e difundir conhecimento científico e tecnológico, ser um agente da cidadania".



Fiocruz Brasília



Fiocruz Minas



Instituto Nacional de Infectologia  
Evandro Chagas (INI)



Instituto Oswaldo Cruz (IOC)



Search this dataverse...

 Find

[Advanced Search](#)

 Add Data ▾

 **Dataverses (36)**

 **Datasets (675)**

 **Files (192,844)**

#### Dataverse Category

Organization or Institution (18)

Laboratory (6)

Department (5)

Research Project (4)

Research Group (2)

**Metadata Source**

1 to 10 of 711 Results

 Sort ▾

[Pharmacokinetics of chloroquine and primaquine in health volunteers: evaluation of chloroquine or primaquine blood levels in three randomized, single-dose, crossover studies](#) 



Jun 24, 2021 - Estudos da Farmacocinética em Humanos

Daher, André; Pinto, Douglas Pereira; Fonseca, Laís Bastos; Esteves, Alessandra Lanzillotta; Medeiros, Juliana J. Soares; Mendonça, Jorge Souza, 2021, "Pharmacokinetics of chloroquine and primaquine in health volunteers: evaluation of chloroquine or primaquine blood levels in three randomized, single-dose, crossover studies", <https://doi.org/10.35078/F8RSR9>, Fundação Oswaldo Cruz, V1, UNF:6:i9vBIIfvLPgLNPLpViMQ== [fileUNF]

Vivax is a neglected disease. There is an irrefutable need for better treatments with higher acceptability and efficacy. The treatment efficacy is influenced by many factors including bioavailability. Hence a straightforward strategy to improve vivax treatment efficacy is the dep...

# Indicadores de depósito de patentes da Fundação Oswaldo Cruz (Fiocruz) no contexto da ciência, tecnologia e inovação no período de 1988 a 2020

Version 2.0



Grando, Rafaela Lora; Silva, Waldeyr Mendes Cordeiro; Pereira-Silva, Marcus Vínicius (Coord.), 2021, "Indicadores de depósito de patentes da Fundação Oswaldo Cruz (Fiocruz) no contexto da ciência, tecnologia e inovação no período de 1988 a 2020", <https://doi.org/10.35078/YN2UTF>, Fundação Oswaldo Cruz, V2

Cite Dataset ▾

[Learn about Data Citation Standards.](#)

Access Dataset ▾

Contact Owner

Share

Dataset Metrics ⓘ

2 Downloads ⓘ

## Description ⓘ

A pesquisa teve como objetivo o estudo métrico do depósito de patentes depositadas pela Fundação Oswaldo Cruz (Fiocruz) no contexto da ciência, tecnologia e inovação até fevereiro de 2021. (2021-05-25)

## Subject ⓘ

Medicine, Health and Life Sciences; Social Sciences

## Keyword ⓘ

Indicadores de Patentes, Patents Indicators, Inovação Tecnológica, Technological Innovation, Propriedade Intelectual, Intellectual Property, Fiocruz, Brasil, Brazil

## Related Publication ⓘ

Acessar o dashboard do Observatório em Ciência, Tecnologia e Inovação em Saúde da Fiocruz

Files

Metadata

Terms

Versions

Search this dataset...

Find

Filter by

File Type: All

Access: All ▾

File Tag: All ▾

Sort ▾

1 to 2 of 2 Files

Download



Metodologia\_Patentes.docx

MS Word - 80.5 KB - Jun 7, 2021 - 1 Download

MD5: ebad744cef64fb25331693d3dcf568af

Metodologia do estudo métrico aplicado para identificar o depósito de patentes da Fundação Oswaldo Cruz (Fiocruz) no contexto da ciência, tecnologia e inovação até fevereiro de 2020

Indicadores de Patentes Indicadores de CT&I Inovação Tecnológica Fiocruz Documentation



Planilha\_Patentes\_Fiocruz\_Orbit\_fev2021.csv

Comma Separated Values - 270.8 KB - Jun 7, 2021 - 1 Download

MD5: 80a92d21de8ad7f07afe74db58c567eb

Planilha com os indicadores de depósito de patentes da Fundação Oswaldo Cruz (Fiocruz) no contexto da ciência, tecnologia e inovação até fevereiro de 2020

Data Indicadores de Patentes Indicadores de CT&I Inovação Tecnológica Fiocruz



Citação dos dados com o Identificador Persistente (DOI)



Conjunto de arquivos



Metadados



Licenças, cessão e termos de uso



Versionamento



Tags



EXEMPLO DATASET



# Desafios

## 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



Fechado



## Preservação digital

Prazos de guarda para os dados

Integração com Archivemática para os dados com longos prazos de guarda



## Certificação

Core Trust Seal

## Acesso

Citação



Indicadores

Métricas



# Onde estamos?

Finalização das pendências da Fase 2 com produção de documentos de apoio

Discussão da governança e fluxos do Arca Dados para lançamento na Fiocruz

D2.2 - Apoiar, de forma prioritária, os projetos de pesquisa sobre os quais recaem exigências dos órgãos financiadores sobre gestão de dados para pesquisa, e a solicitação de dados por revistas científicas.

Promovendo o trabalho de sensibilização...

# CAPACITAÇÃO



- **Programa de Formação Modular sobre Ciência Aberta (curso EAD)\***

Promoção: VPEIC, Escola Corporativa, Campus Virtual e Universidade do Minho.

\*Cerca de 20244 alunos inscritos até Junho/2021

<https://cutt.ly/OnYmB9m>

- **Disciplina transversal aos Programas de Pós-Graduação**

Oferecida a partir de outubro 2021



- **Oficinas**

A partir do lançamento das ferramentas

**FIOCRUZ CAMPUS VIRTUAL**

### Séries do Programa

Série 1 (20h)	Série 2 (20h)	Série 3 (20h)	Série 4 (20h)
<b>Fundamentos da Ciência Aberta</b>	<b>Marcos Legais</b>	<b>Pesquisa Aberta</b>	<b>Educação Aberta</b>
Curso 1 (10h)   7 aulas O que é Ciência Aberta?	Curso 1 (10h)   4 aulas Propriedade Intelectual aplicada à Ciência Aberta	Curso 1 (10h)   4 aulas Acesso Aberto	Curso 1 (10h)   4 aulas Panorama da Educação Aberta
Curso 2 (10h)   7 aulas Panorama Histórico da Ciência Aberta	Curso 2 (10h)   5 aulas Dados Abertos Direito de Acesso à Informação e Proteção de Dados Pessoais	Curso 2 (10h)   5 aulas Dados Abertos	Curso 2 (10h)   4 aulas Recursos Educacionais Abertos
Saiba mais   Iniciar curso	Saiba mais   Iniciar curso	Saiba mais   Iniciar curso	Saiba mais   Iniciar curso

**Ministério da Saúde**  
FIOCRUZ  
Fundação Oswaldo Cruz  
Coordenação-Geral de Gestão de Pessoas

**FIOCRUZ 120 ANOS**  
PATRIMÔNIO DA SOCIEDADE BRASILEIRA

**ESCOLA CORPORATIVA FIOCRUZ**

**Universidade do Minho**

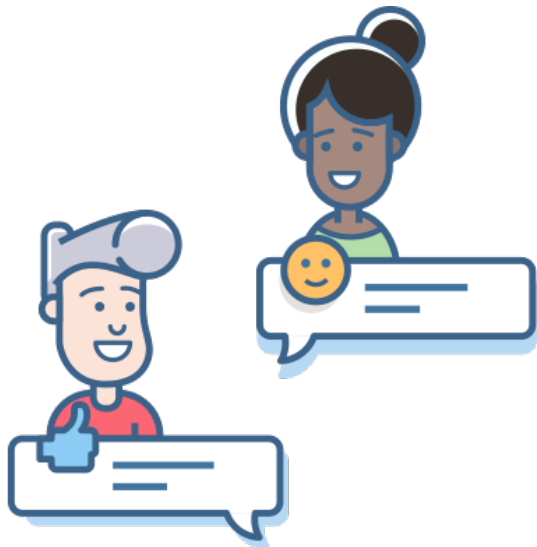


## Compartilhar dados:

- ✓ Facilita a reprodutibilidade e avaliação da pesquisa;
- ✓ Promove a colaboração;
- ✓ Permite o avanço da ciência;
- ✓ É justo e sustentável.

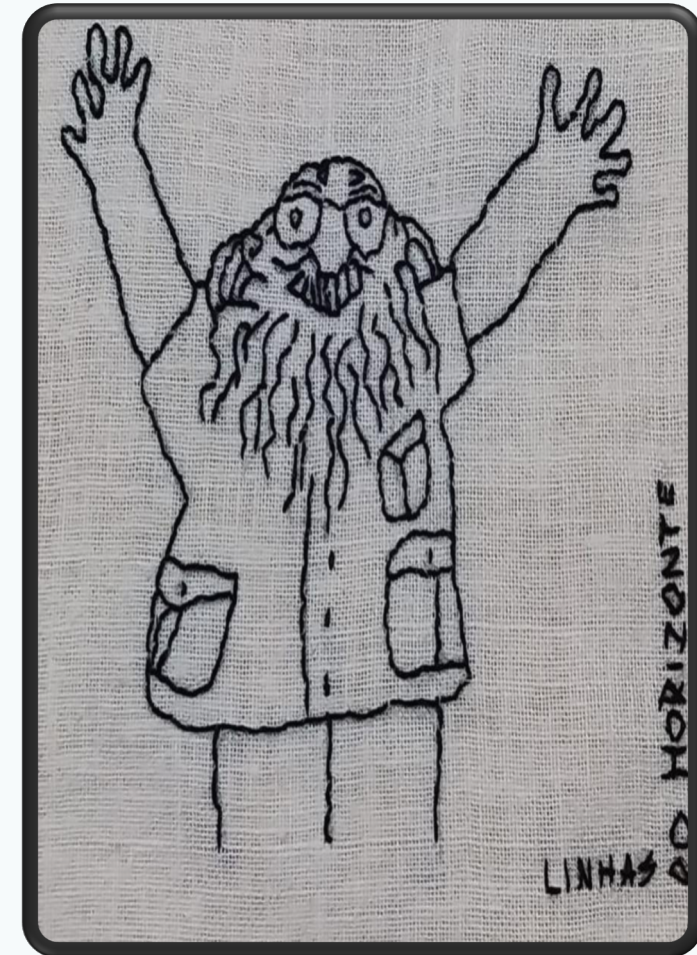


Embora...



- ✓ Os pesquisadores reconhecem os benefícios dos dados abertos, mas as práticas ainda são limitadas;
- ✓ Requer infraestruturas e apoio técnico;
- ✓ Não existe um caminho uniforme, as práticas variam segundo as áreas de conhecimento, disciplina, áreas geográficas, etc.

*“Todos nós sabemos alguma coisa.  
Todos nós ignoramos alguma coisa.  
Por isso, aprendemos sempre.”*  
Paulo Freire



[maria.correa@fiocruz.br](mailto:maria.correa@fiocruz.br)