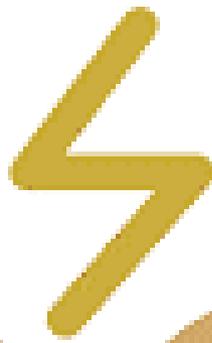




**I ENCONTRO**

Rede Sudeste de Repositórios Institucionais

**28 a 30 MAIO 2019**



**10**

**MANDAMENTOS  
PARA UM BOM  
REPOSITÓRIO**

Luana Sales – PPGCI-  
IBICT

luanasales@ibict.br

Luís Sayão – CNEN-CIN

lsayao@cnen.gov.br

# Agenda

Para que repositórios de dados de pesquisa?

O que são repositórios de dados de pesquisa?

O Desafio dos Repositórios

O que é um bom repositório?

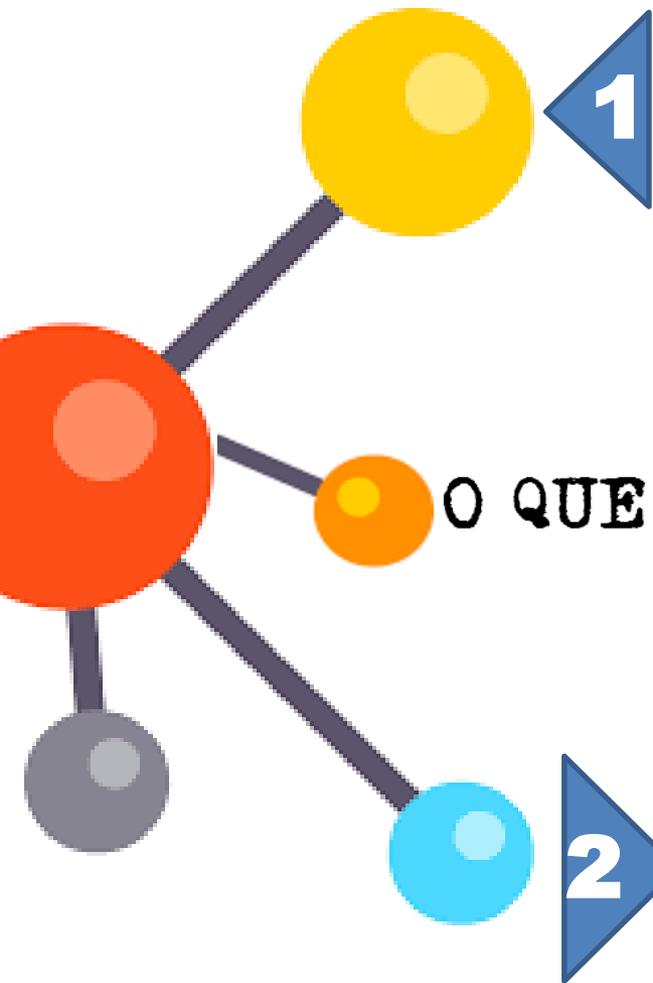
As cinco leis dos dados de pesquisa

Os 10 mandamentos do bom repositório

# Para que REPOSITÓRIO DIGITAL DE DADOS DE PESQUISA



- A primeira exigência para a curadoria é o dado estar disponível em um repositório



## DEPOSITAR & COMPARTILHAR

infraestruturas que assegurem o máximo de **confiabilidade, estabilidade e acessibilidade** e que facilitem o trabalho de **arquivamento, compartilhamento** e **reconhecimento de autoria** para os seus dados

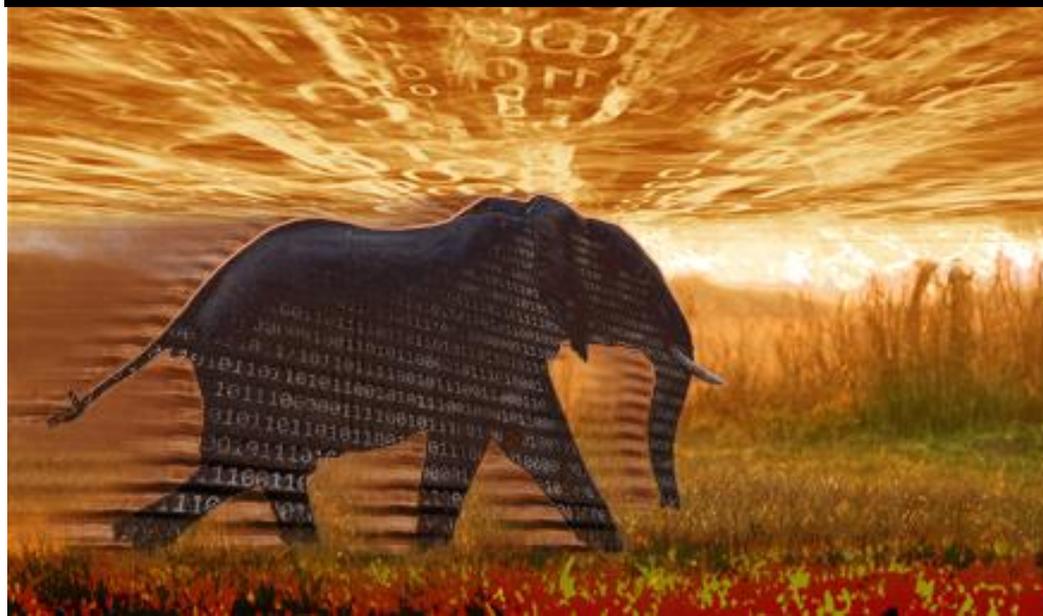
## O QUE PRECISAM OS PESQUISADORES?

### DESCOBRIR E ACESSAR

precisam **encontrar coleções de dados** de pesquisa, saber como acessá-las e sob que condições podem reutilizar esses dados e assim dar prosseguimento às suas pesquisas **confiando na autenticidade e proveniência dos dados** coletados ou gerados por outros pesquisadores.

# AFINAL, O QUE SÃO REPOSITÓRIOS DE DADOS DE PESQUISA

“ BANCOS DE DADOS CIENTÍFICOS DIGITAIS QUE GARANTEM O ACESSO A RESULTADOS DE PESQUISA AGORA E NO FUTURO, TENDO COMO PERSPECTIVA PRIMORDIAL O ACESSO ABERTO AO QUE É CONSIDERADO TAMBÉM UM PATRIMÔNIO DIGITAL DA HUMANIDADE (SURF FOUNDATION, 2013)



## ○ REPOSITÓRIO INSTITUCIONAL DE DADOS DE PESQUISA

Extensão dos modelos de dados dos RI's

Links entre os recursos

Multidisciplinares

**CARPE DIEM (IEN/CNEN); Edimburgh Data Share**

## ○ REPOSITÓRIO TEMÁTICO DE DADOS DE PESQUISA

Domínios ou tipos de dados específicos

**Bio Models Data Base (modelos computacionais na área de biologia)**

**GenBank**

## ○ REPOSITÓRIO MULTIDISCIPLINAR DE DADOS DE PESQUISA

Dados gerados/coletados por várias áreas

Gestão mais complexa: heterogêneos, licenças, formatos...

**Dryad, Figshare, Zenodo e o Dataverse**

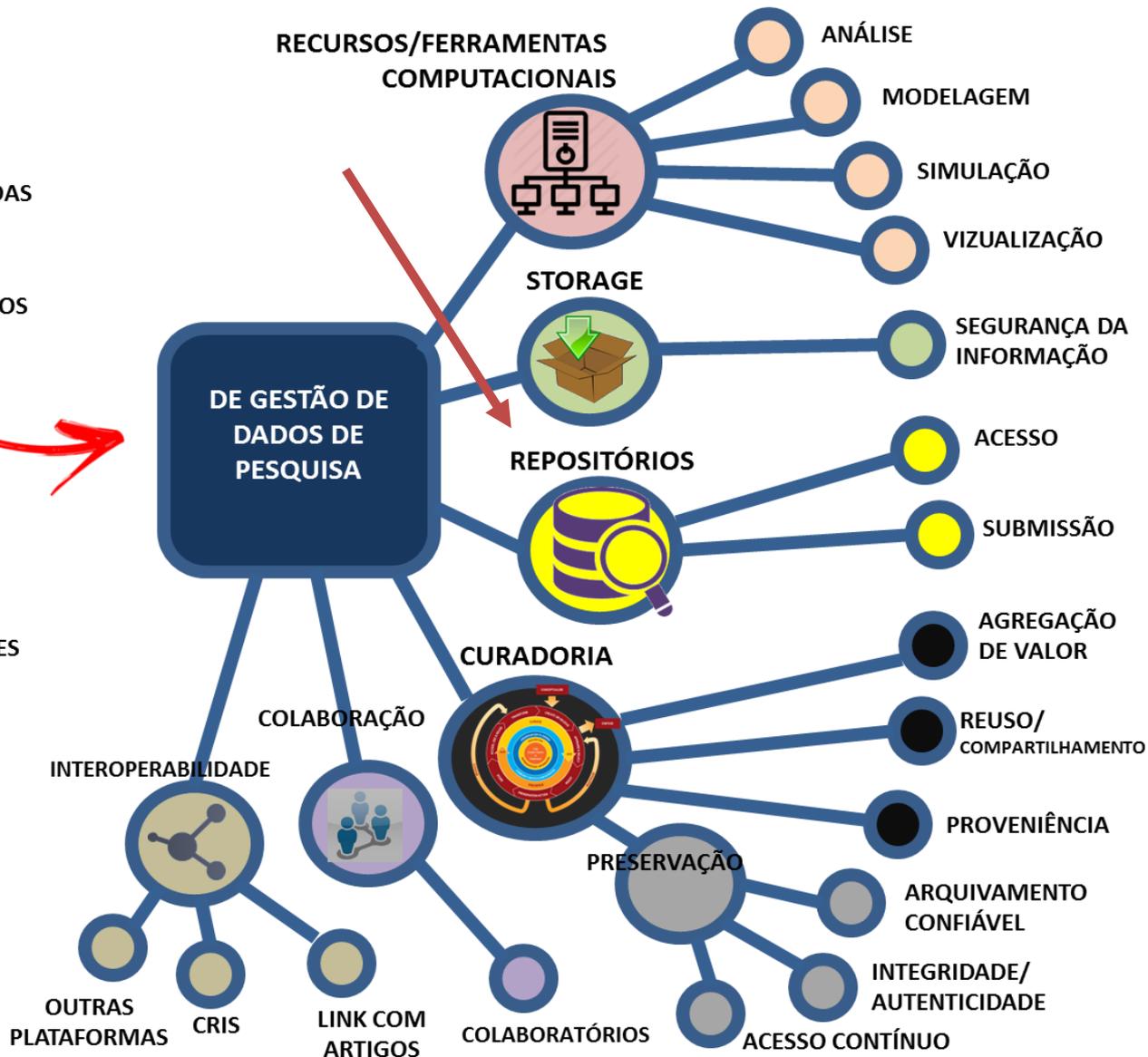
## ○ REPOSITÓRIO DE DADOS DE PESQUISA DE UM PROJETO ESPECÍFICO

Coleções de dados resultados de projetos de pesquisa ou resolução de problemas específicos.

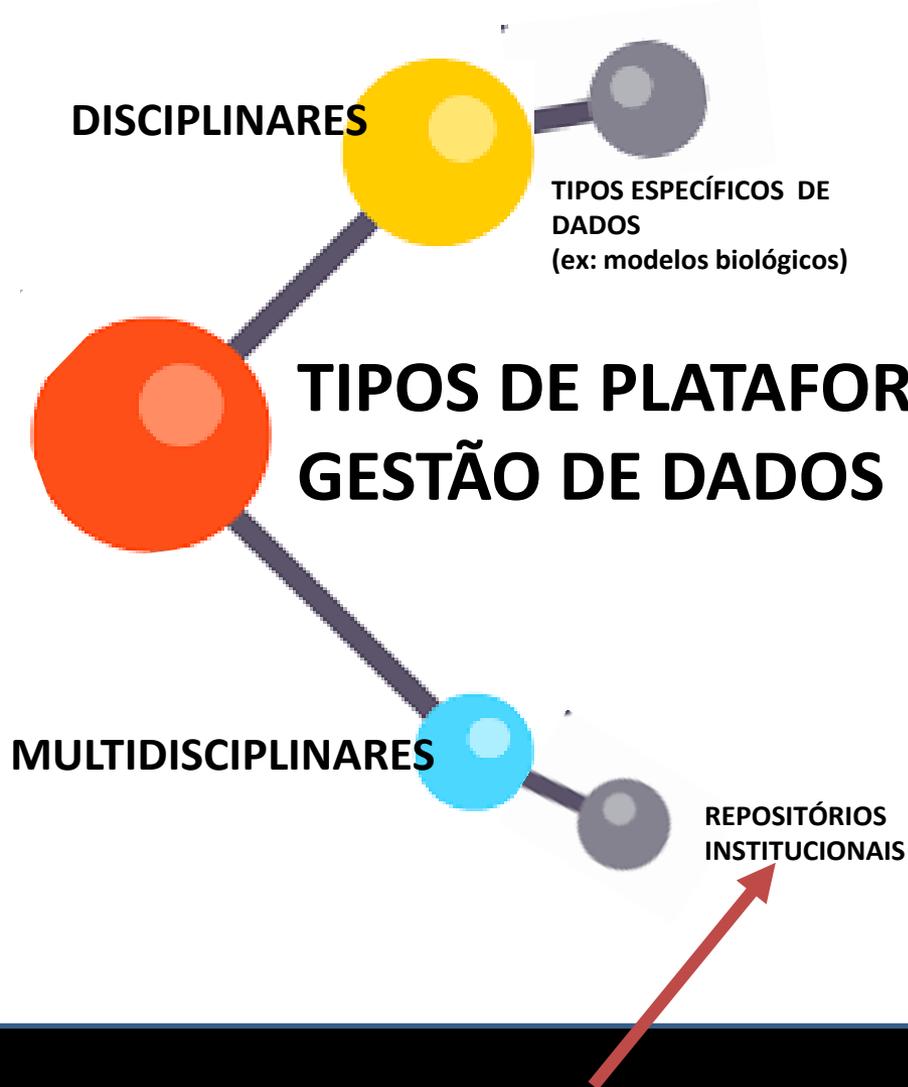
**Scientific Drilling Database**

# CIBERINFRAESTRUTURA DE DADOS DE PESQUISA

## FONTES DE DADOS



## POLÍTICA DE DADOS DE PESQUISA



As **PLATAFORMAS DISCIPLINARES** se voltam para domínios específicos ou para tipos particulares de dados. Em geral possuem modelos de dados adequados à representação das coleções de dados e oferecem uma **CARTEIRA DE SERVIÇOS** mais orientadas, como curadoria e visualização.

Essas plataformas estão abertas para publicar qualquer tipo de dados, e são especialmente desenvolvidas para dar apoio a publicação de *datasets* produzidas no âmbito da ciência chamada de **“CAUDA LONGA”** – domínios científicos nos quais um grande número de relativamente pequenos laboratórios ou de pesquisadores individuais produzem a maioria dos resultados científicos.

# BioModels Database



Advanced

- [BioModels Home](#)
- [Models](#)
- [Submit](#)
- [Support](#)
- [About BioModels](#)
- [Contact us](#)

BioModels Database is a repository of computational models of biological processes. Models described from literature are manually curated and enriched with cross-references. All models are provided in the Public Domain. More information about BioModels Database can be found in the [FAQ](#).

## Models published in the literature

*Browse*

*Alternative access*



[Manually curated](#)  
(630 models)



[Non curated](#)  
(983 models)



[Gene  
Ontology  
classification](#)



[Gene  
Ontology tree](#)



[Advanced  
search](#)

## Models automatically generated from pathway resources (Path2Models)

*Browse*

*Alternative access*



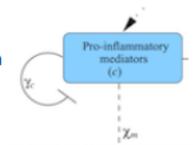
[Metabolic](#) (112,898 models)  
[Non-metabolic](#) (27,531 models)



## Model of the month

May, 2017

A mathematical model on inflammation resolution: prediction suggests that an effective treatment protocol would be to target macrophage phagocytosis alongside neutrophil apoptosis.



[Access this model of the month.](#)

## News

16 January 2017 [Metabolic network and logical models](#)

Archives of [metabolic network](#) and [logical models](#) are available for download.

9 January 2017 [Models on neurodegeneration: review article](#)

Published in CPT:PSP [Mechanistic models on neurodegenerative disease processes](#)

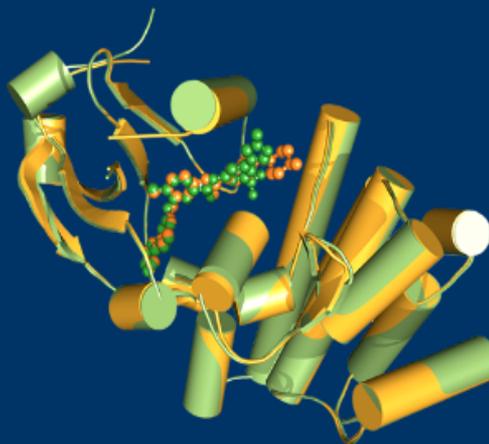
9 January 2017 [New Models in BioModels](#)

BioModels now provides 1603 literature-based models.

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**

### wwPDB Members

wwPDB data centers serve as deposition, annotation, and distribution sites of the PDB archive. Each site offers tools for searching, visualizing, and analyzing PDB data.

#### BMRB

- › **Biological Magnetic Resonance Data Bank**



Collects NMR data from any experiment and captures assigned chemical shifts, coupling constants, and peak lists for a variety of macromolecules; contains derived annotations such as hydrogen exchange rates, pKa values, and relaxation parameters.

#### PDBe

- › **Protein Data Bank in Europe**



Rich information about all PDB entries, multiple search and browse facilities, advanced services including PDBePISA, PDBeFold and PDBeMotif, advanced visualization, and validation of NMR and EM structures

### wwPDB Resources

#### Data Dictionaries

- › **Macromolecular Dictionary (PDBx/mmCIF)**
- › **Small Molecule Dictionary (CCD)**
- › **Peptide-like antibiotic and inhibitor molecules (BIRD)**

#### Annotation

- › **Procedures and policies**
- › **Improvements for consistency and accuracy**

#### Community Input: Task Forces and Working Groups

- › **Validation Task Forces (X-ray, NMR, 3DEM)**
- › **Small Angle Scattering Task Force**
- › **PDB/mmCIF Working Group**
- › **Hybrid/Integrative Methods Task Force**
- › **Ligand Validation Workshop**

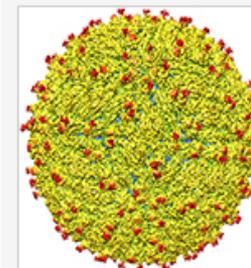
#### PDB Data Growth & Usage Statistics

### News & Announcements

**March 31, 2016**

- › **Zika Virus Structure Released**

The 3D structure of the Zika virus, determined by cryo-electron microscopy, has been released. Public availability of the Zika virus atomic coordinates to medical researchers worldwide will accelerate new antiviral drug and vaccine development.



*Photo Credit: Purdue University*

**Read more**

- 1 DEPOSIT DATA**  

- 2 GET PERMANENT IDENTIFIER**  

- 3 WATCH YOUR CITATIONS GROW!**  

- 4 RELAX, YOUR DATA ARE DISCOVERABLE AND SECURE**  


<http://dx.doi.org/10.5061/dryad.20>

[Submit data now](#)

[How and why?](#)

## Search for data

[Advanced search](#)

## Browse for data

[Recently published](#) [Popular](#)

### Recently published data

Méndez-Janovitz M, Macías García C (2017) Data from: Do male fish prefer them big and colourful? Non-random male courtship effort in a viviparous fish with negligible paternal investment. *Behavioral Ecology and Sociobiology*  
<https://doi.org/10.5061/dryad.9s380>

Dominoni DM, Åkesson S, Klaassen R, Spoelstra K, Bulla M (2017) Data from: Methods in field chronobiology. *Philosophical Transactions of the Royal Society B*  
<https://doi.org/10.5061/dryad.511h1>

Becker DJ, Streicker DG, Altizer S (2017) Data from: Using host species traits to

## Latest from @datadryad

### Tweets by @datadryad

 Dryad Retweeted 

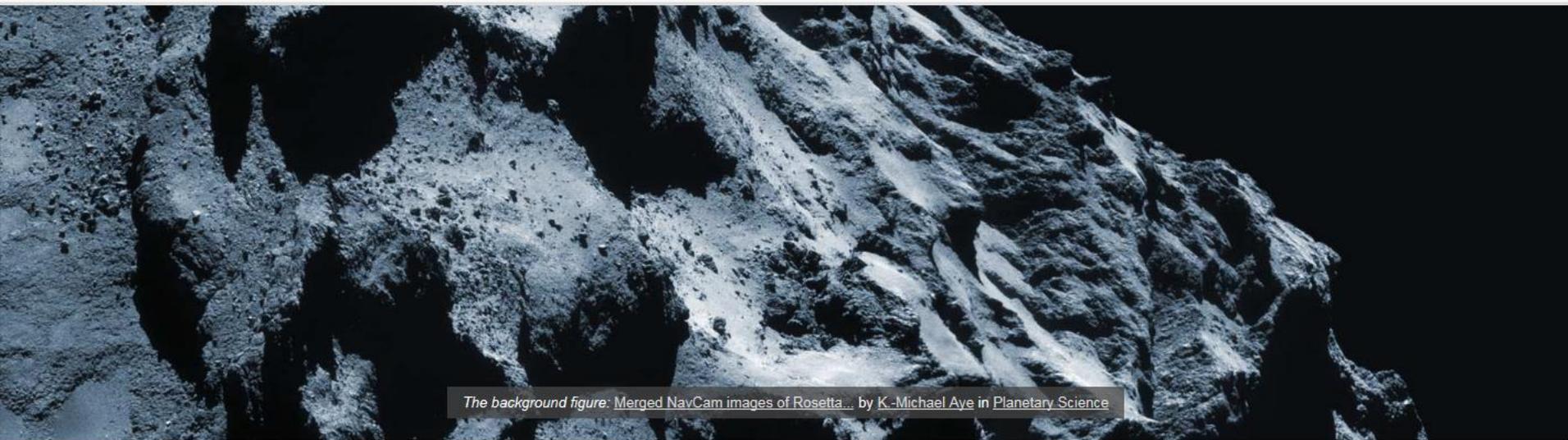
 **UC Curation Center**  
@UC3CDL

OrgID Working Group issues Request for Information. Interested in developing open identifier registry? Submit Nov15  
[uc3.cdlib.org/2017/10/10/org...](http://uc3.cdlib.org/2017/10/10/org...)

  6h

 Dryad Retweeted 

 **Simon Hodson**



The background figure: Merged NavCam images of Rosetta... by K-Michael Aye in Planetary Science

## Discover research from figshare

### FEATURED CATEGORIES

[Agricultural and Veterinary Sciences](#)[Astronomy, Astrophysics, Space Science](#)[Biological Sciences](#)[Built Environment and Design](#)[Chemistry](#)[Commerce, Management, Tourism and Services](#)[Earth and Environmental Sciences](#)[Engineering](#)[Health Sciences](#)[Humanities](#)[Information And Computing Sciences](#)[Language, Communication and Culture](#)[Mathematics](#)[Meta Science](#)[Physics](#)[Psychology](#)[Social Science](#)[Studies in Creative Arts and Writing](#)[Studies in Human Society](#)



Share, archive, and get credit for your data. Find and cite data across all research fields.

 Find

[Advanced Search](#)

 Add Data

 **Dataverses (2,348)**

 **Datasets (74,726)**

 **Files (353,618)**

**Dataverse Category**

- [Research Project \(716\)](#)
- [Researcher \(673\)](#)
- [Organization or Institution \(195\)](#)
- [Journal \(116\)](#)
- [Research Group \(63\)](#)

[More...](#)

**Metadata Source**

- [Harvested \(51,643\)](#)
- [Harvard Dataverse \(25,431\)](#)

**Publication Date**

- [2015 \(15,557\)](#)
- [2011 \(9,626\)](#)
- [2012 \(8,151\)](#)
- [2016 \(4,658\)](#)
- [2009 \(4,020\)](#)

[More...](#)

**Subject**

- [Social Sciences \(33,161\)](#)
- [Medicine, Health and Life Sciences \(1,709\)](#)

1 to 10 of 77,074 Results

 Sort ▾

Replication Data for: Differential Registration Bias in Voter File Data: A Sensitivity Analysis Approach 

Oct 11, 2017 - American Journal of Political Science (AJPS) Dataverse



Brendan Nyhan; Christopher Skovron; Rocio Titiunik, 2016, "Replication Data for: Differential Registration Bias in Voter File Data: A Sensitivity Analysis Approach", doi:10.7910/DVN/LCDBRU, Harvard Dataverse, V2, UNF:6:dfpHOaXg54q+BviR8l0bDA==

The widespread availability of voter files has improved the study of political participation in American politics, but the lack of good data on non-registrants creates difficult inferential issues. Most notably, observational studies that examine turnout rates among registrants o...

Replication Data for Institutions, Norms and Accountability: A Corruption Experiment with Northern and Southern Italians 

Oct 11, 2017 - Journal of Experimental Political Science



Zhang, Nan, 2017, "Replication Data for Institutions, Norms and Accountability: A Corruption Experiment with Northern and Southern Italians", doi:10.7910/DVN/XCSQ9C, Harvard Dataverse, V1, UNF:6:QYDNL1Y9qm/rzgiHk0SA==

Anti-corruption research has highlighted the potential for grassroots monitoring to improve governance outcomes, but the conditions under which citizens are willing to report bribery remain under-studied. Are individuals from some societies socialized into a "culture of corruptio...

Replication Data for: "Adoption of Distributed Solar across India" 

Oct 10, 2017 - Chao-yo Cheng Dataverse



Michaël Aklin; Cheng, Chao-yo; Johannes Urpelainen, 2017, "Replication Data for: "Adoption of Distributed Solar across India"", doi:10.7910/DVN/GITJQ4, Harvard Dataverse, V1, UNF:6:NRwNn3JA+Gg4Ct6TcS6ORw==

This package includes all datasets and codes (in R) to replicate all analyses and figures in "Adoption of Distributed Solar across India," a forthcoming article on Energy for Sustainable Development. To cite the dataset: Aklin, Michaël; Cheng, Chao-yo; Urpelainen, Johannes. Repli...

Chao-yo Cheng Dataverse (UCLA) 



Oct 10, 2017

## Recent uploads

August 30, 2017 (v2) Working paper Open Access

View

### Introducing Parsl: A Python Parallel Scripting Library

Babuji, Yadu; Brizius, Alison; Chard, Kyle; Foster, Ian; Katz, Daniel S.; Wilde, Michael; Wozniak, Justin

Researchers frequently rely on large-scale and domain-specific workflows to conduct their science. These workflows may integrate a variety of independent software functions and external applications. However, developing and executing such workflows can be difficult, requiring complex...

Uploaded on September 15, 2017

1 more version(s) exist for this record

September 6, 2017 (v4) Software Open Access

View

### geodynamics/pylith: PyLith v2.2.1

Brad Aagaard; Charles Williams; Matthew Knepley

Added new examples. examples/3d/subduction: New suite of examples for a 3-D subduction zone. This intermediate level suite of examples illustrates a wide range of PyLith features for quasi-static simulations. examples/2d/subduction: Added quasi-static spontaneous rupture earthquake cycle...

Uploaded on September 6, 2017

3 more version(s) exist for this record

August 31, 2017 (v1) Figure Open Access

View

### Fig. 1 in Vanmanenia orcicampus, a new species of loach from the Plain of Jars, Laos (Teleostei: Gastromyzontidae)

Maurice Kottelat

Fig. 1. Vanmanenia orcicampus, MHNG 2767.094, holotype, 47.5 mm SL; Laos: Nam Ngum on Plain of Jars.

Uploaded on September 6, 2017

Zenodo now supports DOI versioning!



Read more about it, in our newest blog post.

### Using GitHub?



Just Log in with your GitHub account and click here to start preserving your repositories.

### Zenodo in a nutshell

- **Research. Shared.** — all research outputs from across all fields of research are welcome! Sciences and Humanities, really!
- **Citeable. Discoverable.** — uploads get a Digital Object Identifier (DOI) to make them easily and uniquely citeable.
- **Communities** — create and curate your own community for a workshop, project, department, journal, into which you can accept or reject uploads. Your own complete digital repository!
- **Funding** — identify grants, integrated in reporting lines for research funded by the European Commission via OpenAIRE.
- **Flexible licensing** — because not everything is under Creative Commons.
- **Safe** — your research output is stored safely for the future in the same cloud infrastructure



# PLATAFORMAS PARA GESTÃO DE DADOS DE PESQUISA: Qual a melhor?



**INVENIO** é um framework para **bibliotecas digitais, repositórios digitais, arquivos multimídia ou repositório de dados de pesquisa** disponíveis na web. <http://invenio-software.org/>

Seguem exemplos de utilizadores:

**Zenodo**, um serviço do CERN que permite o compartilhamento de dados de pesquisa, utiliza o Invenio – <https://zenodo.org/>

A **biblioteca digital das Nações Unidas** também utiliza este framework em seu repositório – <https://digitallibrary.un.org/>



**CKAN** (Comprehensive Knowledge Archive Network) é um sistema de depósito e gerenciamento de dados que oferece ferramentas para publicação, compartilhamento, descoberta e uso de dados. É um sistema **voltado a governos nacionais e regionais, companhias e organização que desejam disponibilizar seus dados** (IBICT). O **CKAN** é uma [aplicação web](#) de catalogação de dados desenvolvido pela [Open Knowledge Foundation](#). O CKAN mantém uma descrição dos dados nele armazenados, bem como outras informações úteis, como em que formatos eles estão disponíveis, quem é o seu dono e se eles estão livremente disponíveis, e com quais assuntos ele se relaciona. Tal como ocorre num [wiki](#), outros usuários podem melhorar ou acrescentar informações, enquanto o CKAN guarda um histórico versionado (Wikipedia).



**DATAVERSE** is an open source web application to share, preserve, cite, explore, and analyze research data. It facilitates making data available to others, and allows you to replicate others' work more easily. Researchers, data authors, publishers, data distributors, and affiliated institutions all receive academic credit and web visibility. <http://dataverse.org>

Dataverse supports DataCite DOIs. You can configure your installation following this guide: <http://guides.dataverse.org/en/latest/installation/config.html?highlight=datacite>



O **DSpace** é um software dedicado à implementação de repositórios e que vem sendo amplamente adotado no Brasil. **DSpace** <http://www.dspace.org>

DSpace is an open source software platform that enables organisations to: capture and describe digital material using a submission workflow module, or a variety of programmatic ingest options distribute an organisation's digital assets over the web through a search and retrieval system preserve digital assets over the long term.

# Desafios na Gestão de Repositórios

O Problema  
dos  
Repositórios





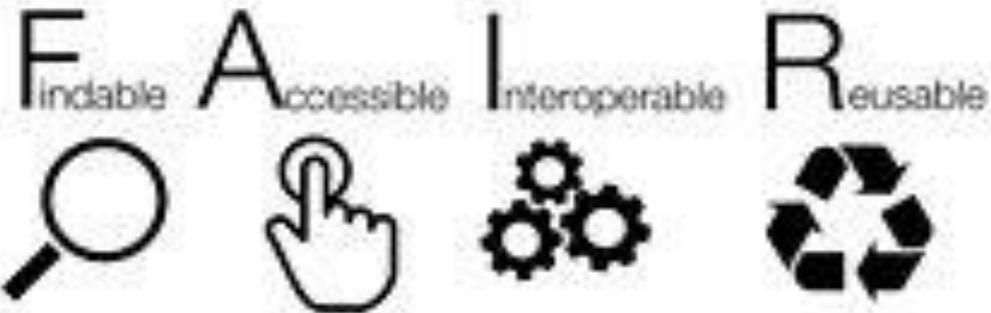
# DESAFIOS DOS REPOSITÓRIOS

- Subutilizados por diversos motivos
  - SÃO INVISÍVEIS
  - SÃO MUITO GENERICOS
  - NÃO TEM CONEXÃO COM A COMUNIDADE
  - NÃO POSSUI SERVIÇOS
  - SÃO COMPULSORIOS
  - PROBLEMAS TECNOLÓGICOS
  - DUPLICAÇÃO DE ESFORÇOS
  - NÃO DA FEEDBACK AO USUÁRIO

# COMO DEFINIR UM BOM REPOSITÓRIO?

UM BOM REPOSITÓRIO É  
AQUELE QUE:

- É localizado
- É acessado
- É interoperável
- Seus dados são reusáveis



# +50%

DOS ACHADOS  
NÃO FORAM  
PUBLICADOS (FERGUNSON ET AL, 2014)



O COMPARTILHAMENTO PODE REVELAR VALORES IMPORTANTES OCULTOS NESSES DADOS

# Dados de pesquisa precisam estar visíveis

Dados de pesquisa se tornam visíveis a partir do momento que são gerenciados, preservados e disseminados por meio de serviços criados sobre plataformas





# LOCALIZADO



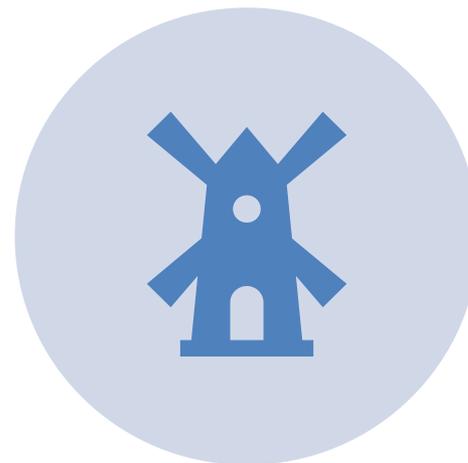
- O repositório precisa ser visível



# REPOSITÓRIO ACESSADO



**TEM QUE SER CONFIÁVEL**



**TEM QUE TER SERVIÇOS**

# Como adicionar confiança aos repositórios digitais



信  
頼

As instituições de patrimônio **já são confiáveis**. Elas são encarregadas dos materiais e dos objetos que documentam nossa herança cultural. Elas são instituições reconhecidas historicamente como confiáveis para armazenar e preservar esses valiosos materiais e para promover o acesso a eles.

Porém, como a informação digital é menos tangível que outros materiais e muito mais mutável, **confiança e credibilidade podem ser mais difícil de provar**.

A certificação é um elemento essencial para confiabilidade dos repositórios digitais  
**PADRÕES, BOAS PRÁTICAS, GESTÃO, TECNOLOGIA, ORGANIZAÇÃO**



As organizações estão crescentemente interessadas em avaliar suas **infraestruturas de preservação digital** tomando como referência arcabouços padronizados

**AUDITORIA E CERTIFICAÇÃO**

**AUTOAVALIAÇÃO**

**A IMPORTÂNCIA DOS PROCESSOS DE AVALIAÇÃO RESIDE NO FATO DELES PROMOVEREM A CONFIANÇA NA USABILIDADE, SUSTENTABILIDADE E PERSISTÊNCIA POR LONGO PRAZO DOS DADOS DISPONÍVEIS PARA COMPARTILHAMENTO.**

# CERTIFICAÇÃO & AUDITORIA

A IMPORTÂNCIA DOS PROCESSOS DE AVALIAÇÃO RESIDE NO FATO DELES PROMOVEREM A CONFIANÇA NA USABILIDADE, SUSTENTABILIDADE E PERSISTÊNCIA POR LONGO PRAZO DOS DADOS DISPONÍVEIS PARA COMPARTILHAMENTO.



# O DATA SEAL APPROVAL (DSA) É UM PROCESSO DE AUTOAVALIAÇÃO PARA ARQUIVOS DIGITAIS QUE OBJETIVA ESPECIFICAMENTE OS ARQUIVOS QUE DETÉM DADOS



OS DADOS PODEM SER **ENCONTRADOS** NA INTERNET

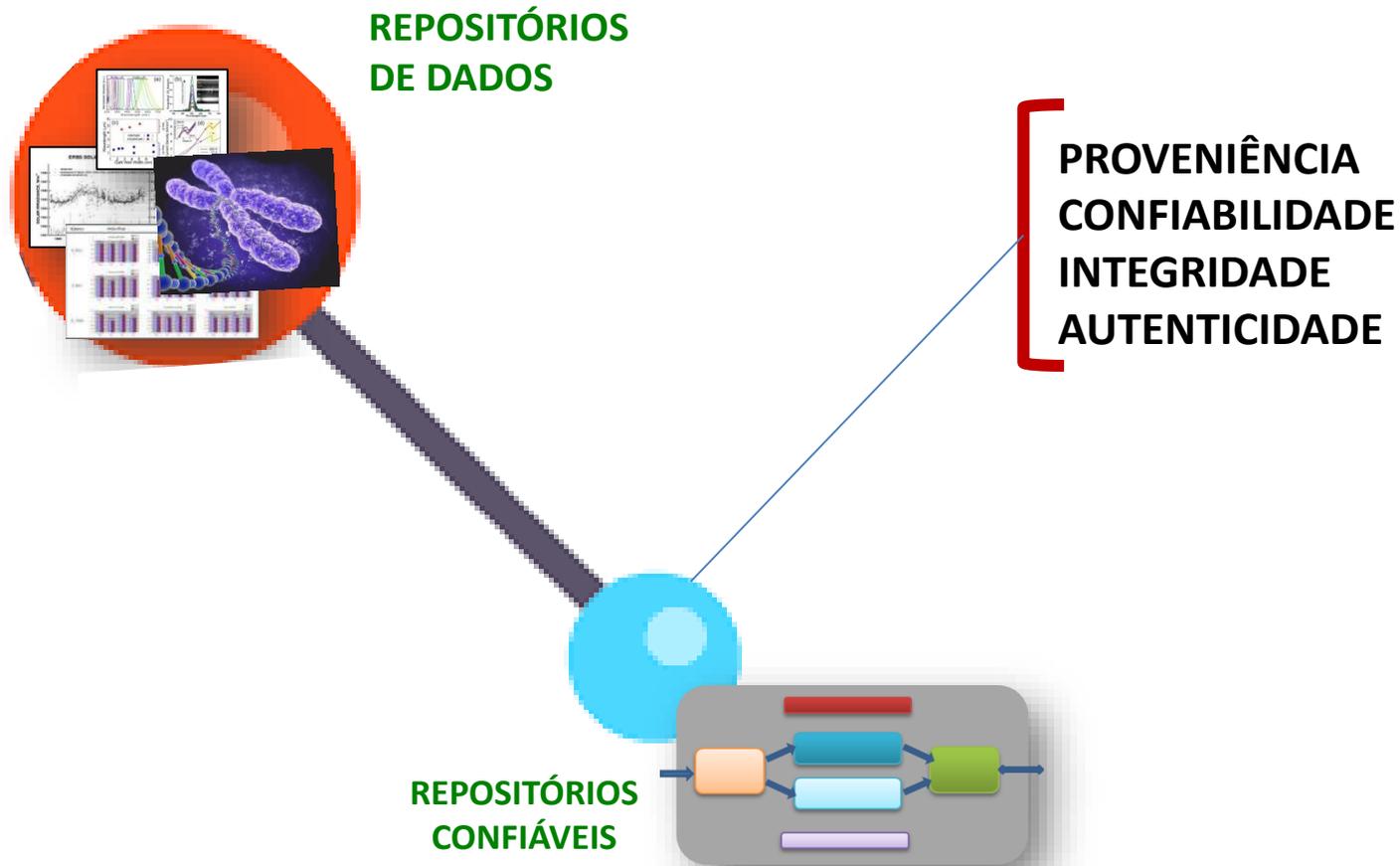
OS DADOS SÃO **ACESSÍVEIS** (direitos e licenças claros)

OS DADOS ESTÃO EM **FORMATOS ADEQUADOS** PARA O USO

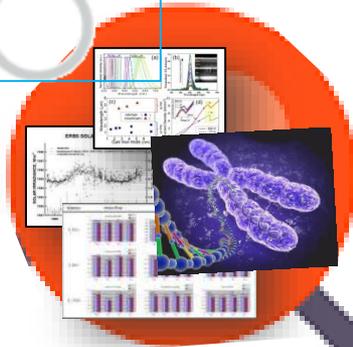
OS DADOS SÃO **CONFIÁVEIS**

OS DADOS SÃO **IDENTIFICADOS** DE MANEIRA ÚNICA E PERSISTENTE DE FORMA QUE POSSAM REFERENCIADOS

# INTEGRAÇÃO REPOSITÓRIO DE DADOS x REPOSITÓRIO CONFIÁVEL



# INTEGRAÇÃO DATAVERSE X ARCHIVEMATICA



PROVENIÊNCIA  
CONFIABILIDADE  
INTEGRIDADE  
AUTENTICIDADE

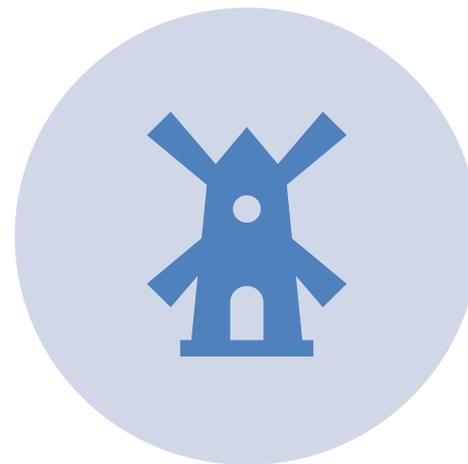
REPOSITÓRIOS  
CONFIÁVEIS



# REPOSITÓRIO ACESSADO



TEM QUE SER CONFIÁVEL

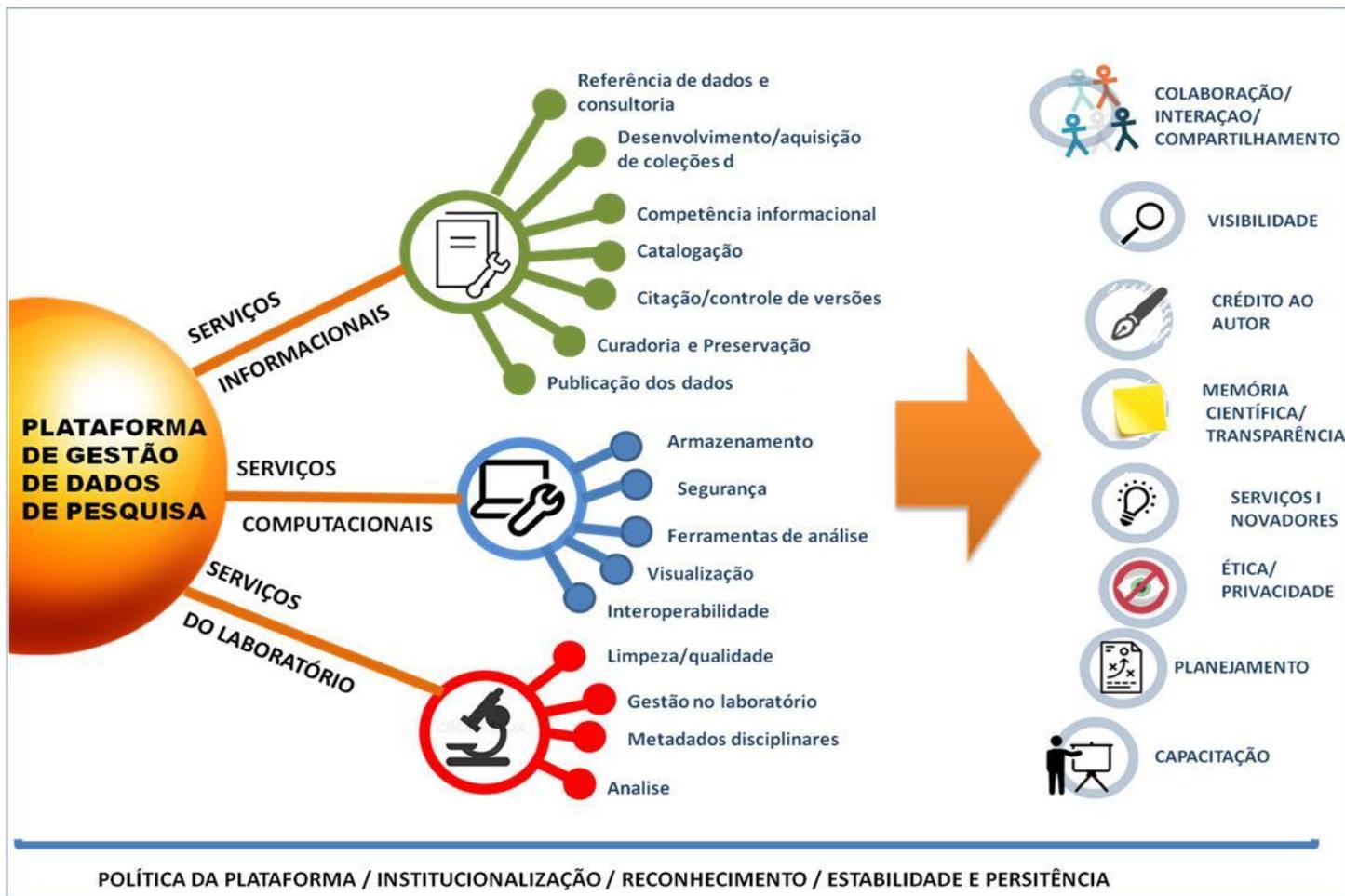


**TEM QUE TER SERVIÇOS**

**PLATAFORMAS  
DE GESTÃO DE  
DADOS DE  
PESQUISA  
Necessitam  
de  
SERVIÇOS**



Figura 1 – Modelo de avaliação de plataforma de gestão de dados de pesquisa



Fonte: Elaborado pelos autores



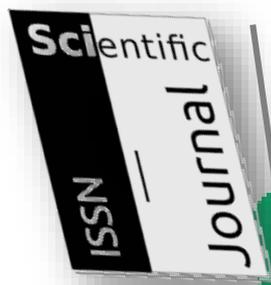
- Usar padrões internacionais genéricos, sem esquecer dos modelos disciplinares
  - Metadados
  - Taxonomias
  - Ontologias

Os dados só atingirão seu potencial máximo de reuso, gerando inovação, se puderem transitar entre plataformas

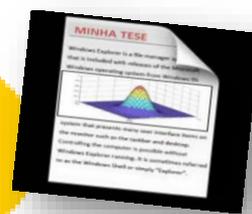
# REPOSITÓRIO INTEROPERÁVEL



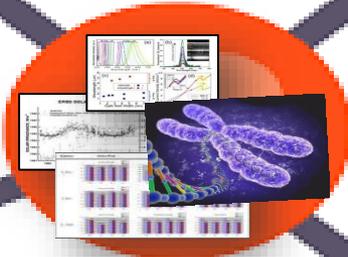
# OUTRAS INTEROPERABILIDADES POSSÍVEIS



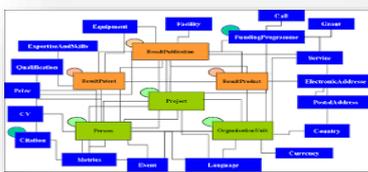
SISTEMAS DE PERIÓDICOS



REPOSITÓRIOS INSTITUCIONAIS



REPOSITÓRIOS DE DADOS



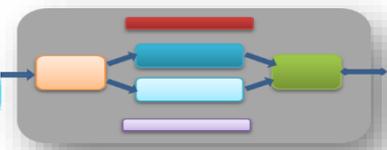
SISTEMAS CRIS  
GESTÃO DE PESQUISA  
GESTÃO ACADÊMICA  
PLATAFORMA LATTES



CLOCKSS



REPOSITÓRIOS CONFIÁVEIS



# OS 10 MANDAMENTOS DO BOM REPOSITÓRIO



# As cinco leis dos dados de pesquisa



1ª Lei: A cada dado o seu pesquisador



2ª Lei: A cada pesquisador o seu dado



3ª Lei: Dados são para serem reusados



4ª Lei: Poupe o tempo do pesquisador



5ª Lei: A geração de dados cresce em ritmo e volume exponencial

# 1ª Lei: A cada dado o seu pesquisador



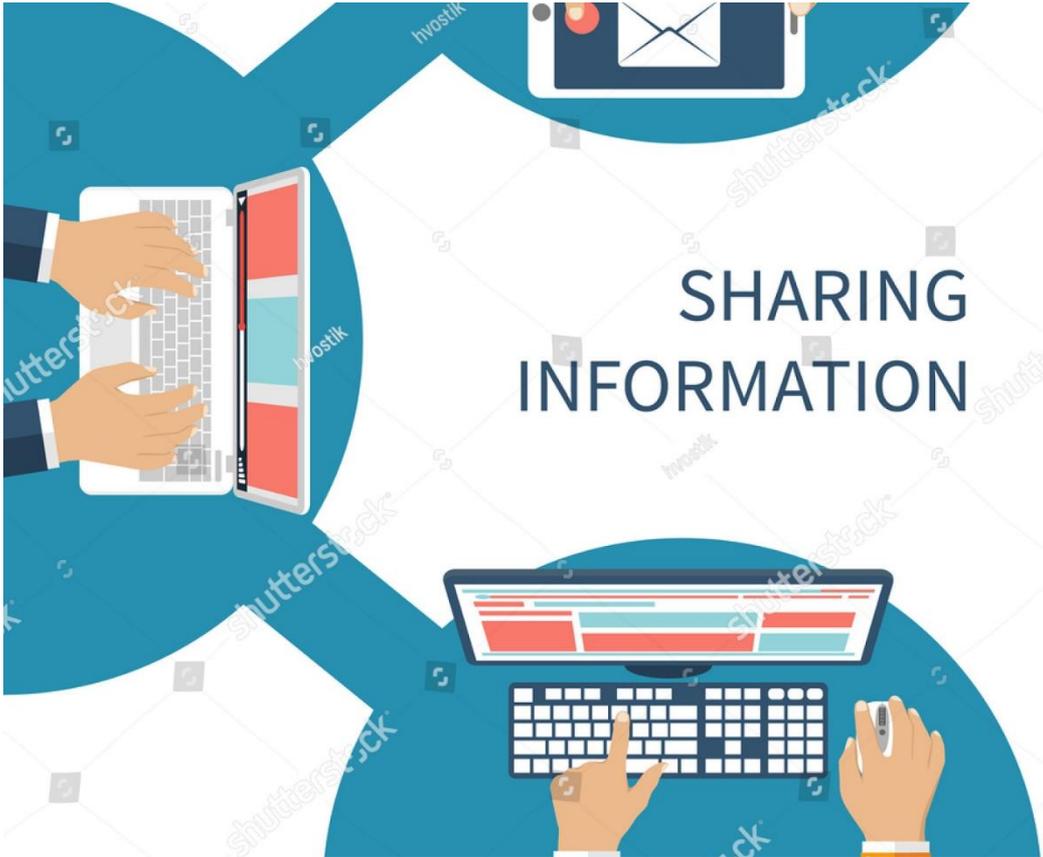
Considerando que repositórios são formados por vários conjuntos de dados, pode-se enunciar:

**A cada repositório o seu pesquisador!**

**1 - Bons repositórios são iniciados por um ou mais pesquisadores.**

**2 - Bons repositórios devem gerar uma comunidade com senso de pertencimento.**

**3 - Bons repositórios devem ser construídos voltados para um domínio -**

An illustration centered around the text 'SHARING INFORMATION'. It features a network of blue lines connecting various digital devices: a smartphone at the top showing an envelope icon, a laptop on the left with hands typing, a desktop monitor and keyboard at the bottom with hands using a mouse, and a tablet on the right. The background is white with a faint grid of nodes and lines, and the text 'SHARING INFORMATION' is prominently displayed in the center in a bold, blue, sans-serif font.

## SHARING INFORMATION

Deve surgir do desejo de uma comunidade pesquisadores, que pode ser mínima, e não da vontade da biblioteca ou da instituição.

Devem ser moldados para atenderem às especificidades de sua comunidade

**1 Bons repositórios são iniciados por um ou mais pesquisadores**



## 2 - Bons repositórios devem gerar uma comunidade com senso de pertencimento

- Devem ter uma ligação orgânica com a comunidade de pesquisa, sendo mais eficazes quando esta comunidade se sente proprietária do mesmo e faz dele um instrumento para desenvolvimento e registro de seu conhecimento.
- Pesquisadores devem sentir orgulho de ter os dados publicados nesse repositório, tanto quanto de uma revista qualis A, tendo em vista que representam uma importante ferramenta de comunicação científica em sua comunidade.

### 3 - Bons repositórios devem ser construídos voltados para um domínio

The GenBank logo features the word "GenBank" in a bold, yellow, sans-serif font. Below the text are several diagonal yellow and black stripes. To the right of the stripes is a vertical strip of a colorful DNA double helix structure.

GenBank

- Criar repositórios para uma determinada comunidade significa construir repositórios disciplinares ou voltados para um projeto ou ainda para a solução de um problema específico
- Repositórios disciplinares podem ter serviços criados especificamente para a comunidade.



Biodiversity  
Literature  
Repository



Download from  
**Dreamstime.com**  
This watermarked comp image is for previewing purposes only.

id 56297656  
© Rawpixelimages | Dreamstime.com

- Para cada pesquisador deve haver o seu conjunto de dados ou o seu próprio repositório.
- Pesquisadores se reúnem em grupos e redes de pesquisa, em comunidades de interesse, atuando também como um coletivo, cujo objetivos são sempre mais claros do que os objetivos institucionais.
- 4) Bons repositórios devem ser **construídos para um propósito específico.**
- 5) Bons repositórios devem ser **atrativos aos pesquisadores.**

## 2ª Lei: A cada pesquisador o seu dado



#### **4) Bons repositórios devem ser construídos para um propósito específico.**

- Não é possível abarcar tudo, todos os domínios, todos os tipos de objetos, todos os usuários.
- É preciso um planejamento anterior à construção que permita o repositório a atingir um objetivo.
- Um repositório de grupo de pesquisa não vai ser configurado da mesma forma que um repositório que vise atender um projeto.



## 5) Bons repositórios devem ser atrativos aos pesquisadores.

- Pesquisadores devem querer depositar e reusar coisas do repositório.
- Um repositório atrativo é aquele que tem serviços interessantes para os pesquisadores.
  - **Tradicionais** - como busca e alertas -
  - **Inovadores** - de interesse da comunidade. (ferramentas de análise de dados, modelagem, estatísticas de uso dos dados, consultoria para elaboração de Plano de Gestão de Dados, serviços de referência para encontrar outros dados e mecanismos que auxiliem a citação)



## 3ª Lei: Dados são para serem reusados

- Se não for para o reuso não há sentido preservar, curar e gerenciar dados.
- Dados de qualidade são aqueles dados que conseguem atingir o seu objetivo final que é ser reusado em novas pesquisas ou na construção de novos conhecimentos
- Repositórios devem ser construídos visando tornar dados reusáveis.
  - 6) Bons repositórios devem ser interativos e interoperáveis.
  - 7) Bons repositórios devem dar visibilidade aos dados.
  - 8) Bons repositórios devem oferecer mecanismos de recompensa ao pesquisador

## 6) Bons repositórios devem ser interativos e interoperáveis.

- interativo - espaço para troca de ideia, compartilhamento e incentivo ao trabalho colaborativo, colocando em contato parceiros potenciais.
- interoperável - usa protocolos, normas e padrões que permitem que os dados sejam intercambiados entre sistemas.
- A interoperabilidade é uma condição *sine qua non* para o reuso de dados.





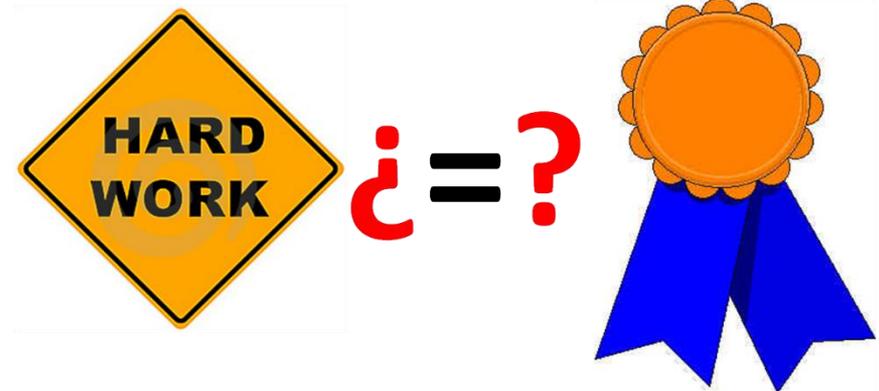
## 7) Bons repositórios devem dar visibilidade aos dados

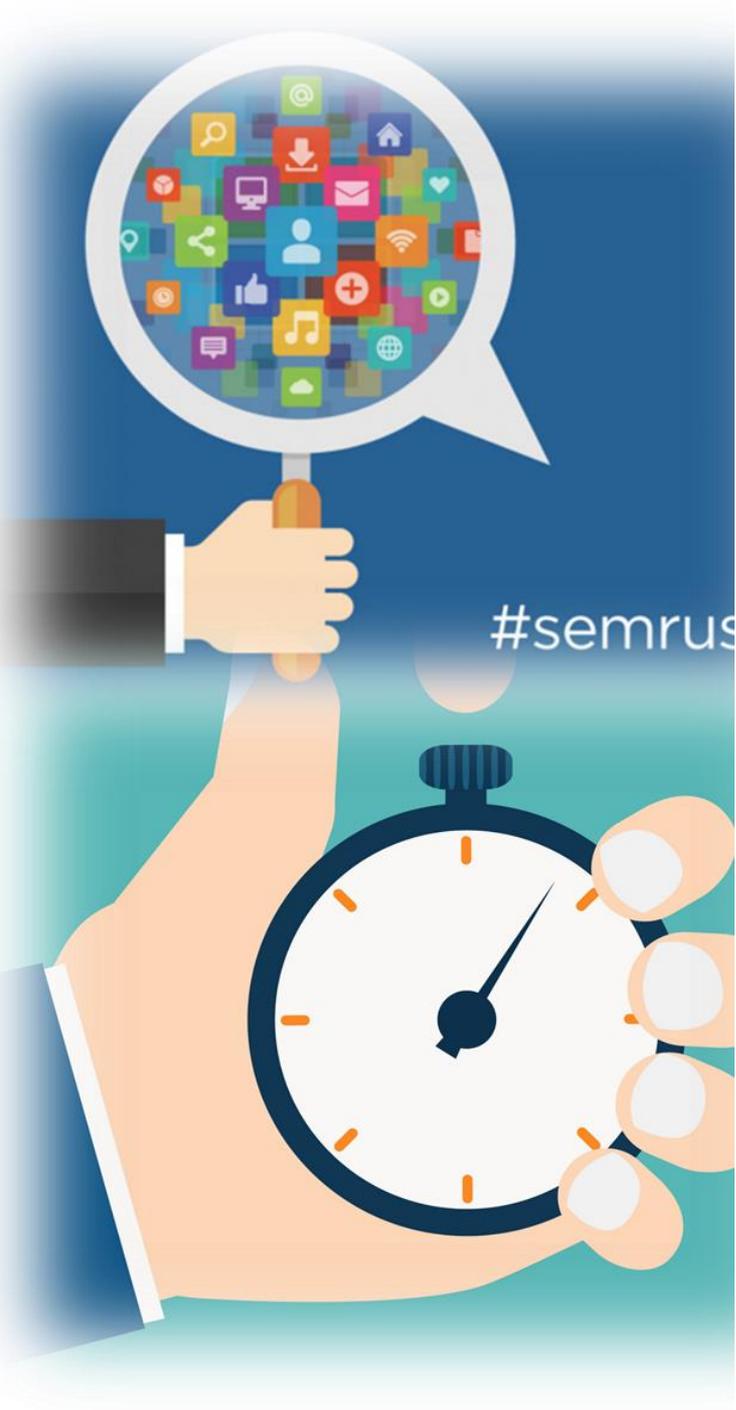
- 1) Para que os dados sejam reusáveis, eles precisam ser encontrados.
  - 2) Para serem encontrados, os dados precisam estar em um repositório que os tornem visíveis.
  - 3) Para serem visíveis, as coleções precisam estar disponíveis não somente para seres humanos, mas também para as máquinas criando pontos de interoperabilidade.
- Essa visibilidade pode acontecer tanto a partir da disponibilização dos metadados para indexação por grandes buscadores e participação em redes de repositórios quanto através de mecanismos de divulgação, como cursos, seminários, alertas, boletins, entre outros.

## 8) Bons repositórios devem oferecer mecanismos de recompensa ao pesquisador

- O repositório será eficiente e seus dados reusados à medida que o pesquisador encontre vantagens no uso desse repositório.
- Um mecanismo de recompensa interessante ao pesquisador é a citação.
- O pesquisador espera e que seus dados sejam citados, comentados e que os artigos baseados nesses dados também se tornem objeto de atenção dos seus pares.
- O repositório deve identificar, disseminar, formatar referência padronizada das coleções de dados e fornecer feedback sobre o uso dos dados em outros projetos.

**DataCite**  
FIND, ACCESS, AND REUSE DATA





## 9) Bons repositórios devem permitir busca precisa e também integrada, poupando o tempo do pesquisador

---

- Dados podem ter seus significados preservados não apenas através de uma representação consistente, mas também através da linkagem com conteúdos relacionados.
- Investimento no uso de padrões internacionais que possibilitem a comunicação com o usuário e a recuperação precisa e integrada de dados e informações que muitas vezes estão em sistemas distribuídos.
- Recuperação integrada - seja através de protocolos e padrões de interoperabilidade, seja a partir do apontamento de links para outros recursos - se torna condição essencial para **POUPAR O TEMPO** do pesquisador

# 4ª Lei: Poupe o tempo do pesquisador

9) Bons repositórios devem permitir busca precisa e também integrada, poupando o tempo do pesquisador.



## 5ª Lei: A geração de dados cresce em ritmo e volume exponencial

- O repositório é um lugar de conhecimento.
- O repositório enquanto morada do conhecimento deve ser um espaço apropriado para acompanhar toda a dinamicidade desse conhecimento.



## 10 Bons repositórios devem considerar a evolução do conhecimento e das tecnologias e apoiar a preservação contínua de seus dados.

- Repositórios devem ser construídos para evolução.
- Repositórios precisam ser vivos, acompanhando a dinamicidade do conhecimento e a modernidade tecnológica.
- Repositórios precisam se valer de estratégias e padrões de preservação para os dados de valor contínuo, garantindo que eles possam ser acessados e compreendidos agora e no futuro.



# Dicas para ter um bom repositório:

Um bom repositório requer planejamento!

1) Identifique quem são seus usuários

2) Faça uma análise de domínio e um estudo dos padrões de comportamento da comunidade

3) Conheça os fluxos de pesquisas

4) Identifique as tecnologias adotadas nas pesquisas e na gestão de dados

5) Levante os padrões já existentes

6) Crie um projeto com objetivos e pensando em recursos para sustentabilidade



# À GUIZA DE CONCLUSÃO



O PLANEJAMENTO  
ANTECEDE A  
TECNOLOGIA



A GESTÃO  
ANTECEDE A  
ABERTURA



É PRECISO  
CONSIDERAR AS  
PRÁTICAS  
DISCIPLINARES JÁ  
EXISTENTES. ISSO  
VALE PARA AS  
TECNOLOGIAS E  
PADRÕES TAMBÉM



É PRECISO MUITA  
PROXIMIDADE COM  
A COMUNIDADE  
CIENTÍFICA.  
COMPREENDER  
BEM OS OBJETIVOS  
E AS  
METODOLOGIAS DE  
PESQUISA

