

# I Seminário Gestão de Dados de Pesquisa UNIFESP e UFABC



## Planos de Gestão de Dados

---

Prof. Fabiano Couto Corrêa da Silva  
[fabianocc@gmail.com](mailto:fabianocc@gmail.com)


# Agenda:

1. Definição de dados científicos

2. Ciclo de vida dos dados

3. Infraestruturas disponíveis

4. Planos de Gestão de Dados



**O que são e que  
importância tem os  
dados científicos?**



# BIG SCIENCE

GRANDES INSTRUMENTOS  
CUSTOS ALTOS  
LONGA DURAÇÃO  
MUITOS COLABORADORES  
PESQUISA DISTRIBUÍDA

# SMALL SCIENCE

PEQUENOS INSTRUMENTOS  
CUSTOS BAIXOS  
CURTA DURAÇÃO EQUIPES  
PEQUENAS PESQUISA LOCAL

Quais  
Dados?



ESTAÇÃO ANTÁRTICA  
COMANDANTE FERRAZ



COMPARTILHAMENTO DE  
DADOS VIA SATÉLITE



LEVANTAMENTO DE  
DADOS EM ÁGUAS  
PROFUNDAS



OBSERVAÇÃO DO  
COMPORTAMENTO  
DOS ANIMAIS



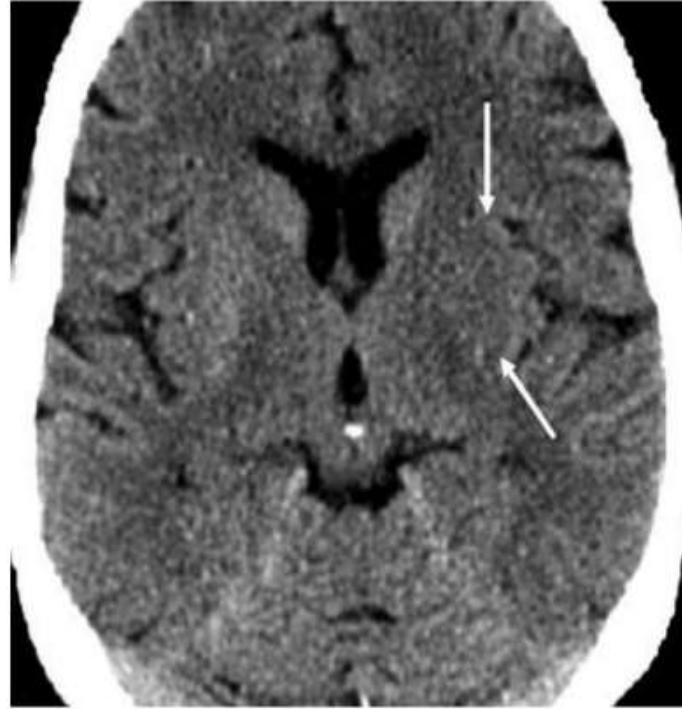
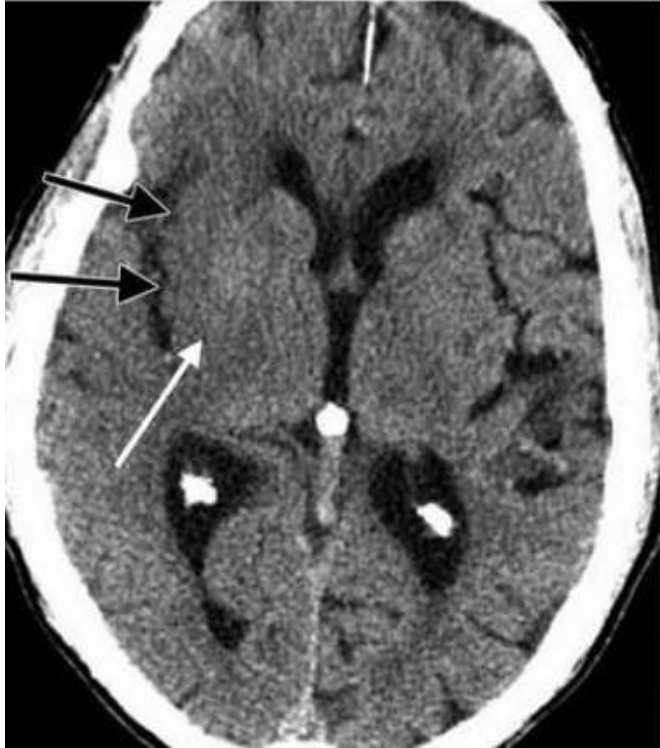
## Matriz de dados registrados por um especialista em AVC isquêmico

Neste caso, um médico especialista avalia a gravidade de seus pacientes em uma escala de 0 a 6. As filas correspondem ao número de pacientes (41 observações) que foram avaliados e as colunas ao valor do curso de cada um dos pacientes em 6 articulações do corpo.

Número de datos	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5	Variable 6
1	1	0	0	0	0	0
2	2	0	0	0	0	0
3	1	2	0	0	0	0
4	4	6	1	1	4	5
5	0	0	5	4	1	1
6	1	0	0	0	1	1
7	0	0	0	0	1	1
8	0	0	0	0	0	0
9	0	0	0	0	0	0
...	...	...	...	...	...	...
<b>41</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>2</b>	<b>2</b>



# Imagem de um acidente vascular cerebral isquêmico



## Exemplo de processamento de dados primários:

```
0711070500276000  
0711070600276000  
0711070700277003  
0711070800282017  
0711070900285000  
0711071000293000  
0711071100301000  
0711071200304000
```



Date	time	air temp C	precip mm
11-Jul-07	5:00	27.6	000
11-Jul-07	6:00	27.6	000
11-Jul-07	7:00	27.7	003
11-Jul-07	8:00	28.2	017
11-Jul-07	9:00	28.5	000
11-Jul-07	10:00	29.3	000
11-Jul-07	11:00	30.1	000
11-Jul-07	12:00	30.4	000

---

## Dados científicos se tornam parte da infraestrutura mundial de pesquisa



---

### **PESQUISADORES**

Desempenhas um papel vital na integridade dos dados científicos

### **AGÊNCIAS FINANCIADORAS DE PESQUISA**

Garantem que os pesquisadores se comprometam com a preservação digital desde a coleta até a disponibilização dos dados.

### **PERIÓDICOS CIENTÍFICOS**

Os periódicos exigem cada vez mais que os dados que sustentam a pesquisa sejam depositados em um banco de dados ou repositório acessível.

### **INSTITUIÇÕES DE PESQUISA**

Financiam / desenvolvem infraestruturas para a gestão e os serviços de dados para facilitar o uso compartilhado dentro de domínios específicos.

Nos últimos anos, várias organizações científicas nacionais e internacionais emitiram declarações e políticas que evidenciam a **necessidade de um imediato arquivo dos dados.**





DataSearch <sup>Beta</sup>

Thompson Reuters  
[Data Citation  
Index](#)

Como encontrar  
o repositório adequado?



O re3data é um serviço oferecido pela DataCite, uma organização global composta por **20 países.**



**Como reutilizar os dados científicos?**

# O poder dos dados:



Enter what you want to calculate or know about

É um mecanismo de busca científica? Uma calculadora? Um dicionário?  
**NÃO É um mecanismo de conhecimento computacional**

**AAGCTAGCTAGC**

**{{1.3, 2.2}, {2.1, 5.8}}...**

**define dogs**

WolframAlpha computational intelligence

**AAGCTAGCTAGC**

Input Interpretation:  
AAGCTAGCTAGC (genome sequence)

Length:  
12 base pairs

Amino acid sequence: [All reading frames](#) | [Use single letters](#)

(5'-3' frame 1)  
AAG | CUA | GCU | AGC  
Lys | Leu | Ala | Ser

Exact matches to reference human genome: [Show genes](#) | [Show positions](#) | [None](#)  
chromosome 1 (24 matches)

positions: 15 488 301 | 21 329 593 | 28 969 883 | ...

genes: FHAD1 | EIF4G3 | YTHDF2 | ...

Input Interpretation:  
fit: data: {{1.3, 2.2}, {2.1, 5.8}, {3.7, 10.2}, {4.2, 11.8}}  
model: linear function

Least-squares best fit:  
 $3.19383x - 1.52256$

Plot of the least-squares fit:

Plot of the residuals:

Computed by: [Wolfram|Mathematics](#) | Download as: [PDF](#) | [Live Mathematics](#)

Input Interpretation:  
dogs: definitions

Definitions: [Show examples](#)

Verb:

dog	go after with the intent to catch
-----	-----------------------------------

Noun:

domesticated animal	a member of the genus <i>Canis</i> (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds
gent	informal term for a man
sausage	a smooth-textured sausage of minced beef or pork usually smoked; often served on a bread roll
stop	a hinged catch that fits into a notch of a ratchet to move a wheel forward or prevent it from moving backward
support	metal supports for logs in a fireplace
unpleasant woman	a dull unattractive unpleasant girl or woman
villain	someone who is morally reprehensible

(6 meanings)



## • Alguns recursos da Wolfram:

- Não é o Google, então não trate como o Google
- É como um monte de dados misturados com análise computacional
- Obtêm os dados de bases públicas
- Armazena 10 trilhões de dados, +50000 algoritmos e modelos
- As pesquisas são feitas por linguagem natural, embora seja necessário saber a maneira correta de inserir a pergunta
- Apresenta e prepara os dados de maneira gráfica e a URL da resposta pode ser copiada e compartilhada

Exemplo: é possível visualizar a estrutura 3D de uma molécula

 **WolframAlpha**™ computational knowledge engine

VX space filling structure 

Input interpretation: *Mathematica form*

2-methylpentyl	space filling structure
S-2-diethylaminoethylethylphosphonothiolate	

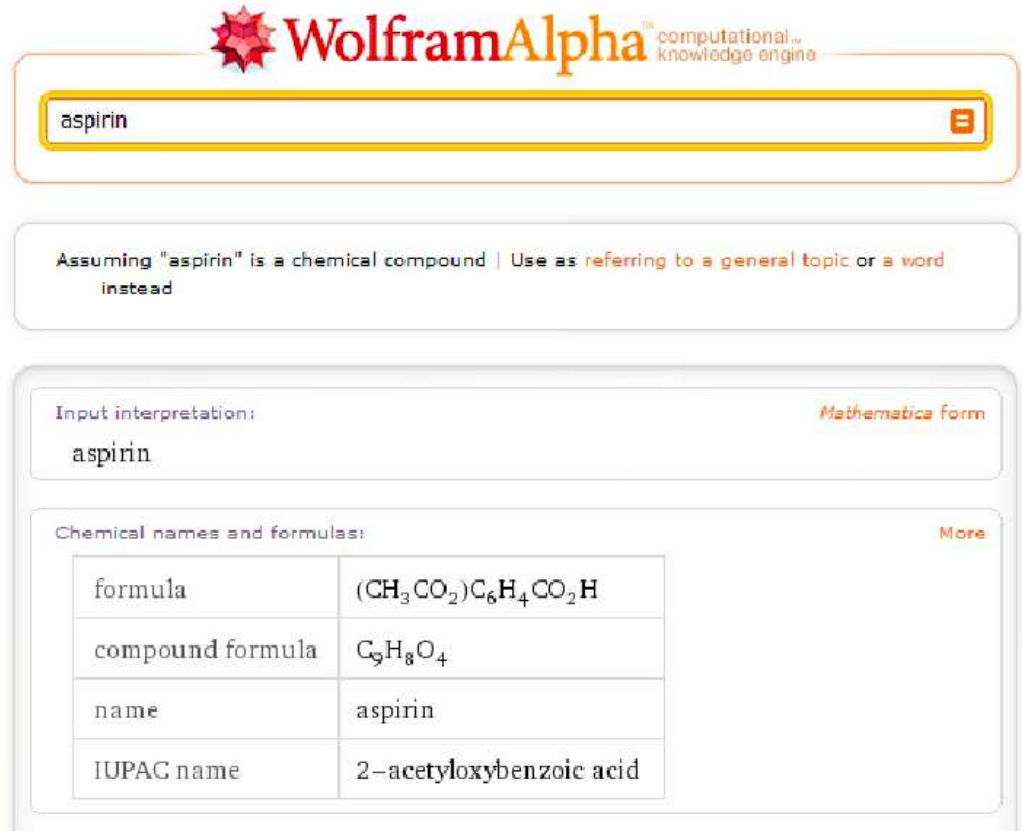
Result:



Computed by: [Wolfram Mathematica](#)    [Source information »](#)    Download as: [PDF](#) | [Live Mathematica](#)

Exemplo:

Conhecer a fórmula da Aspirina



The image shows a screenshot of the WolframAlpha search engine interface. At the top, the WolframAlpha logo is displayed with the tagline "computational knowledge engine". Below the logo is a search input field containing the text "aspirin". Underneath the search bar, a message states: "Assuming 'aspirin' is a chemical compound | Use as referring to a general topic or a word instead". The main results section is titled "Input interpretation:" and shows "aspirin" with a "Mathematica form" link. Below this, a section titled "Chemical names and formulas:" contains a table with four rows of chemical data. A "More" link is visible to the right of the table.

formula	$(\text{CH}_3\text{CO}_2)\text{C}_6\text{H}_4\text{CO}_2\text{H}$
compound formula	$\text{C}_9\text{H}_8\text{O}_4$
name	aspirin
IUPAC name	2-acetyloxybenzoic acid

# Os repositórios de dados + Wolfram Alpha

NCBI Single Nucleotide Polymorphism

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Books

Search for SNP on NCBI Reference Assembly

Search Entrez SNP for Go

BUILD 130

Have a question about dbSNP? Try searching the SNP FAQ Archived

Go

Submitted SNP(ss) Details: ss32475721

FASTA sequence (Legend)

>gn[dbSNP[ss32475721]allelePos=5]len=101[axid=9606]alleles=A/G[mt]=Genomic

ATTTGAAAGA GTTGGCTGTT CACATGAGA GACTGTTTSS TGATATGCTT A  
TCCTTAACCA GCTGAGTCTT GGTATGCACCA ATCGTGAACA TCGATATGAT

Submitted Frequency for ss32475721

Population ID Class	Sample (n)	Major Allele Freq.	Minor Allele Freq.	Estimated Heterozygosity +/- std. err.	Genotype Freq.	Submitted Heterozygosity	Submission Batch	Submitter
EUROPE	100	A=0.980	G=0.020				041213POP	UNAV-GEN

GENERAL HUMAN VARIATION

Search, Annotate, Submit NEW

Annotate and Submit Batch Data with Clinical Impact NEW

SNP SUBMISSION DOCUMENTATION

SEARCH RELATED SITES

GGATGCACCA



WolframAlpha computational knowledge engine

GGATGCACCA

Exact matches to reference human genome: Show genes | Show positions | More

chromosome 1 (171 matches)

0 Mb 30 Mb 60 Mb 90 Mb 120 Mb 150 Mb 180 Mb 210 Mb 240 Mb

positions	938460	1614397	1667676	...
genes	ISG15	RP11345P.44	SLC35B2	...

# Como utilizaremos os dados científicos no futuro?

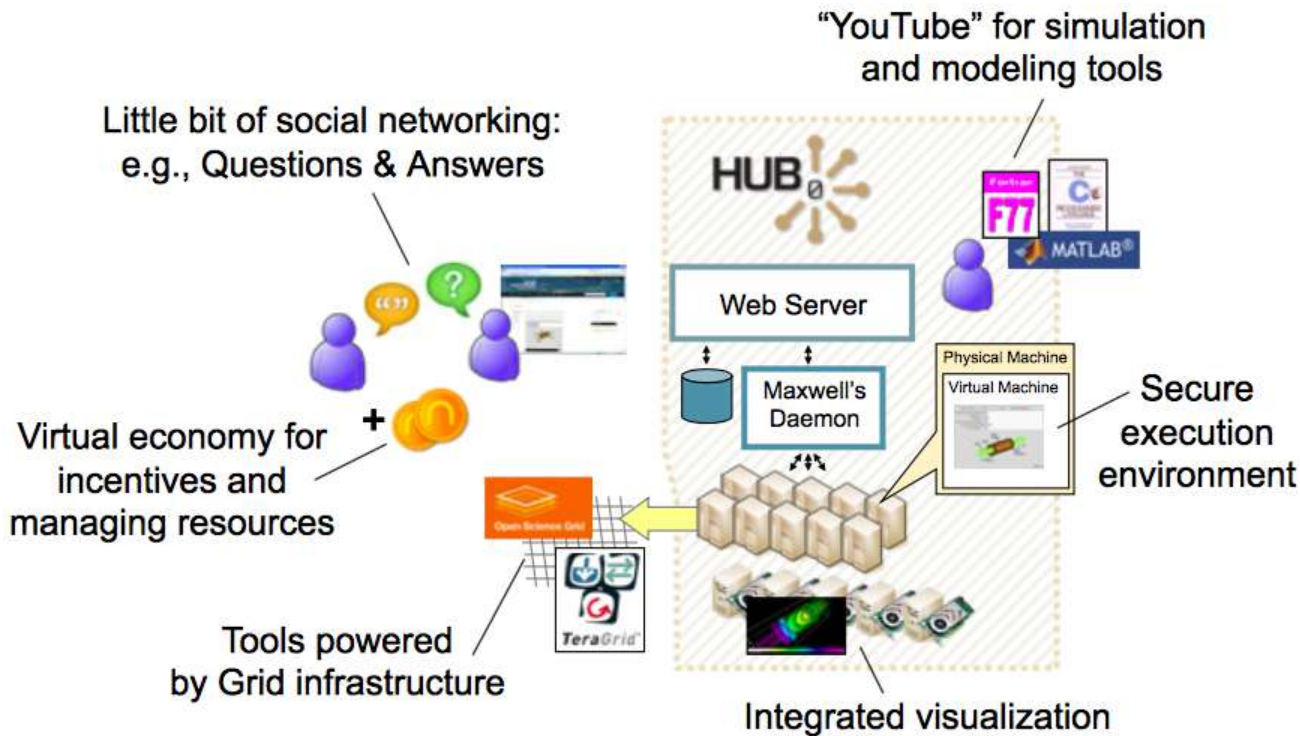
The image shows a banner for the nanoHUB website. The background is a 3D visualization of a carbon nanotube, with atoms represented as spheres in blue, red, and yellow, connected by bonds. The word "SIMULATE" is prominently displayed in the center in large, white, sans-serif capital letters. Below it, the tagline "HUNDREDS OF CUTTING-EDGE TOOLS TO PROPEL YOUR RESEARCH" is written in smaller, white, sans-serif capital letters. In the top left corner, the nanoHUB logo is visible, consisting of a stylized 'n' inside a circle with other atoms around it, followed by the text "nanoHUB". In the top right corner, there is a navigation menu with the following items: "RESOURCES", "EXPLORE", "NANOHUB-U", "PARTNERS", "COMMUNITY", "ABOUT", "SUPPORT", "Login", "Sign Up", "Help", and "Search".

**nanoHUB**

RESOURCES EXPLORE NANOHUB-U PARTNERS COMMUNITY ABOUT SUPPORT Login Sign Up Help Search

# SIMULATE

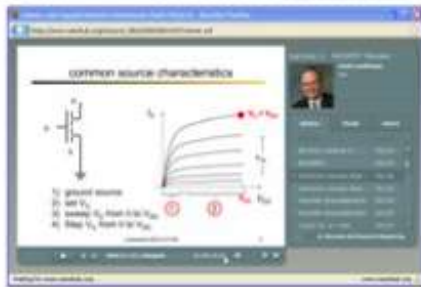
HUNDREDS OF CUTTING-EDGE TOOLS TO PROPEL YOUR RESEARCH



## ECE 612 Nanoscale Transistors (Fall 2006)

**Contributor(s)** Mark Lundstrom  
Purdue University, West Lafayette

**Abstract** Please Note: An updated version of this course is in production

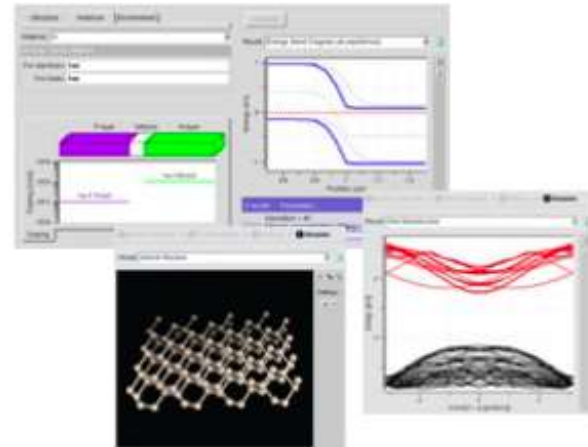


physics of advanced transistors



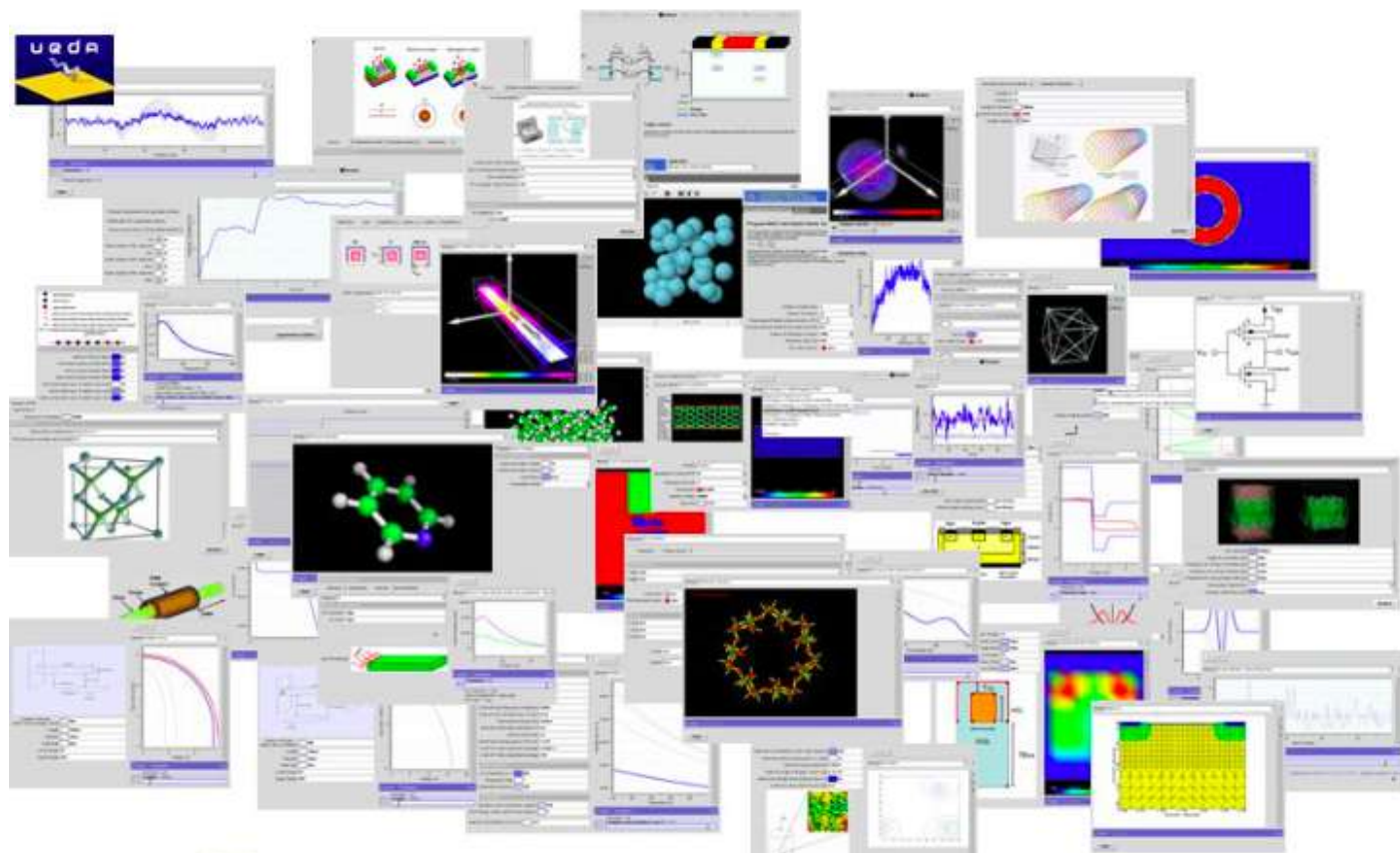
### Teaching Materials:

- Complete courses
- Tutorials and Podcasts
- Homework assignments

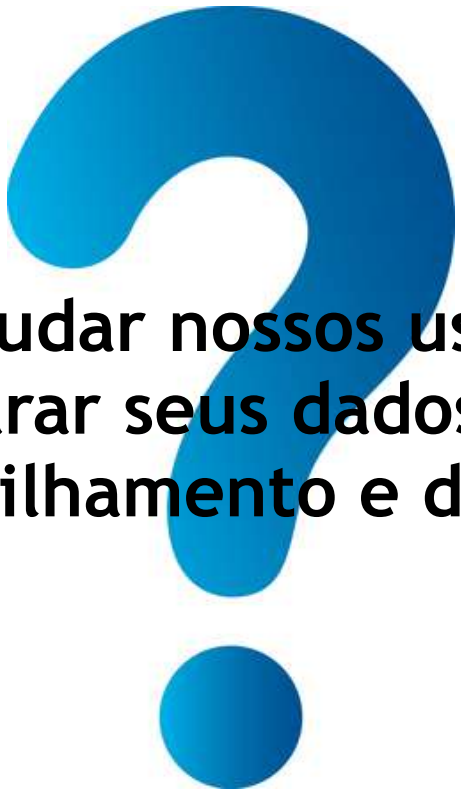


### Simulation tools:

- Demonstrations in class
- Homework assignments
- Class projects







**Como ajudar nossos usuários a  
preparar seus dados para  
compartilhamento e depósito?**

O OSF é um repositório de gerenciamento de projetos gratuito e de código aberto que oferece suporte a pesquisadores durante todo o ciclo de vida do projeto.



<https://cos.io/our-products/osf/>

## mycrocosm. sharing everyday data

a web service that allows you to share snippets of information from the minutiae of daily life in the form of simple statistical graphs.

Microcosm uses [OpenID](#) for authentication.

Your OpenID URL

Please read our [Terms of Service](#) before signing up. [Help](#).  
Check out our [blog](#) for news.



- Vantagens:
  - - Permite trabalhar diretamente com arquivos Excel
  - - Gráficos interativos
  - - Possibilita criar relatórios
  - - Possibilita compartilhamento
  - - Bem incorporado na web
  - - Exportação através de urls

# Painel do Swivel

The screenshot displays the Swivel dashboard interface. At the top, there is a navigation bar with the Swivel logo and icons for Home, Reports, Recent, My Files, Groups, and Public. A search bar labeled "Search Swivel" is positioned on the right. Below the navigation bar, a "My Files" section features buttons for "Upload spreadsheet", "New spreadsheet", "New chart", "New report", and "New conn".

The main content area is divided into two sections: "ALL" and "DRAFTS".

**ALL**

- All
- Reports
- Charts
- Data

**DRAFTS**

- All
- Reports
- Charts
- Data

Three draft items are visible:

- Especialización investigación**  
Saved in Drafts  
Modified 8:19pm
- Especialización investigación**  
Saved in Drafts  
Modified 8:37pm
- Evolución por Quinquenios de la**  
Saved in Drafts  
Modified 8:36pm



**Como adotar uma auditoria sobre a  
Preservação Digital dos dados  
científicos?**



- Verificar onde os dados são gerados, de que tipo são, como são armazenados, gerenciados e compartilhados.
- Identificar problemas: perdas, uso indevido, localizar dados em risco, formatos desatualizados.
- Conhecer as atitudes dos pesquisadores ao criar e compartilhar dados.
- Sugerir procedimentos para o gerenciamento de dados.

# A iniciativa DAF, algumas conclusões das auditorias



- Não existe uma política clara entre os pesquisadores sobre como preservar seus dados
- Os dados são geralmente espalhados por CDs, DVDs, discos rígidos externos, PCs (trabalho e casa).
- Há perdas irrecuperáveis e, às vezes, é difícil identificar a versão mais recente ou encontrar a pessoa responsável
- Pesquisadores não usam nenhuma regra para nomear seus arquivos, nem criam catálogos de coleções de dados
- Não levam em conta sua preservação a longo prazo e, embora pesquisadores pensem a respeito, não conhecem as ferramentas e os procedimentos para lidar com o problema.

# A iniciativa DAF: formulário de controle



- Home
- STAFF
- List Audits
- AUDITOR
- New Audit**
- List Users

## Audit Stage 1 (New Audit)

Required fields are marked with an asterisk.



### Organisational Details

Audit Name	<input type="text"/>	Must be unique within the system
Organisation Identifier	<input type="text"/>	? Username must contain only letters, numbers and underscores
* Institution	<input type="text" value="Institution..."/>	
Organisation Address	<input type="text"/>	



Qual a melhor escola da Cidade de São Paulo considerando a estrutura e nota do Índice de Desenvolvimento da Educação Básica (IDEB)?

**dados.gov.br**

PORTAL BRASILEIRO DE DADOS ABERTOS

\_ escolar



6823 registros para SP



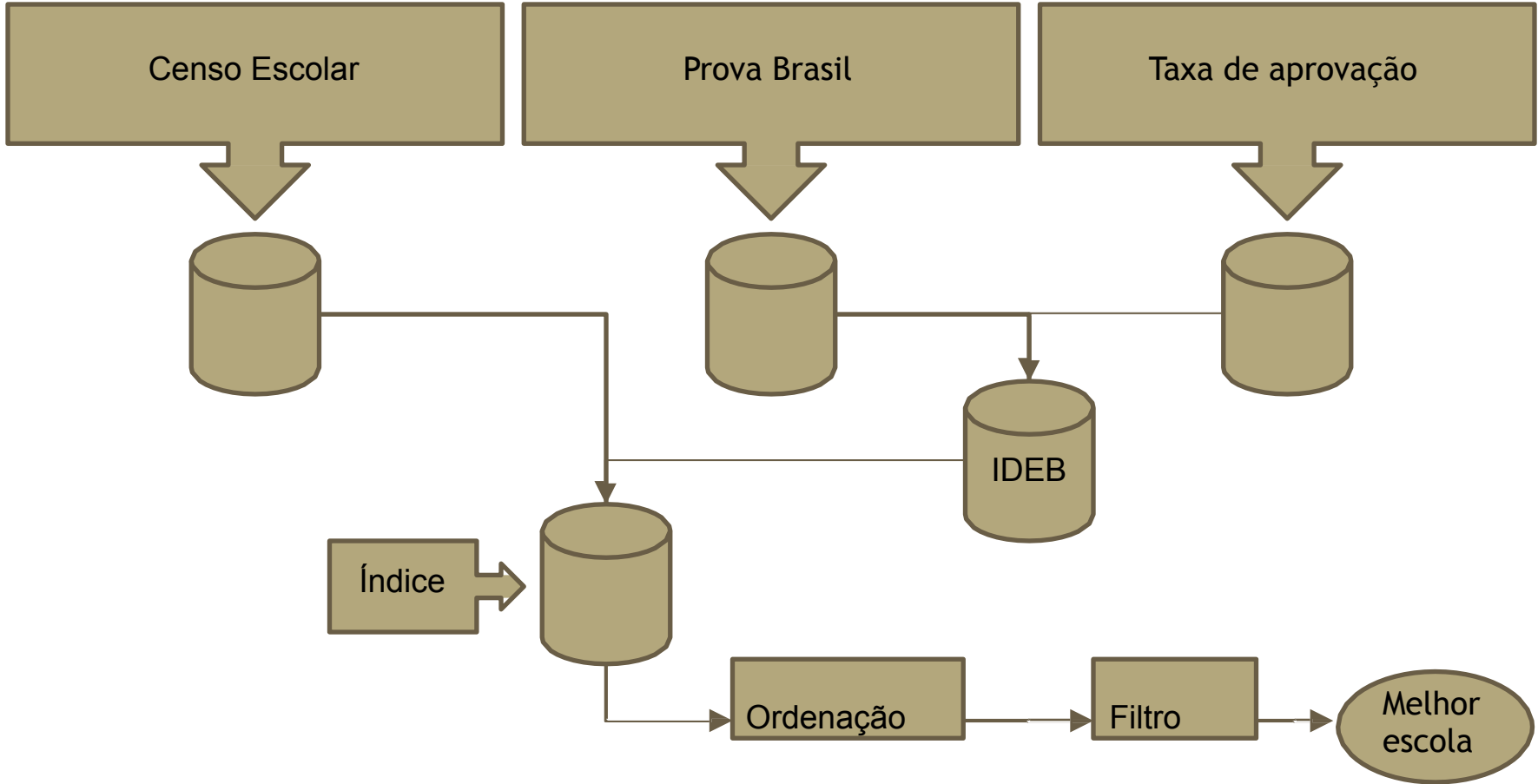
Prova Brasil

1547 registros para SP



\_ de rendimento 3139 registros em  
SP





# Exemplos de publicadores de dados abertos



The screenshot shows the Data.gov.uk website. At the top, there is a navigation bar with links for Home, Data, Apps, and Interact, along with a search bar. Below this is a secondary navigation bar with categories like Datasets, Map Search, Data Requests, Publishers, Public Roles & Salaries, OpenSpending, Spend Reports, and Site Analytics. The main content area features a search bar and a list of datasets, including 'Bona Vista' and 'English Department'.

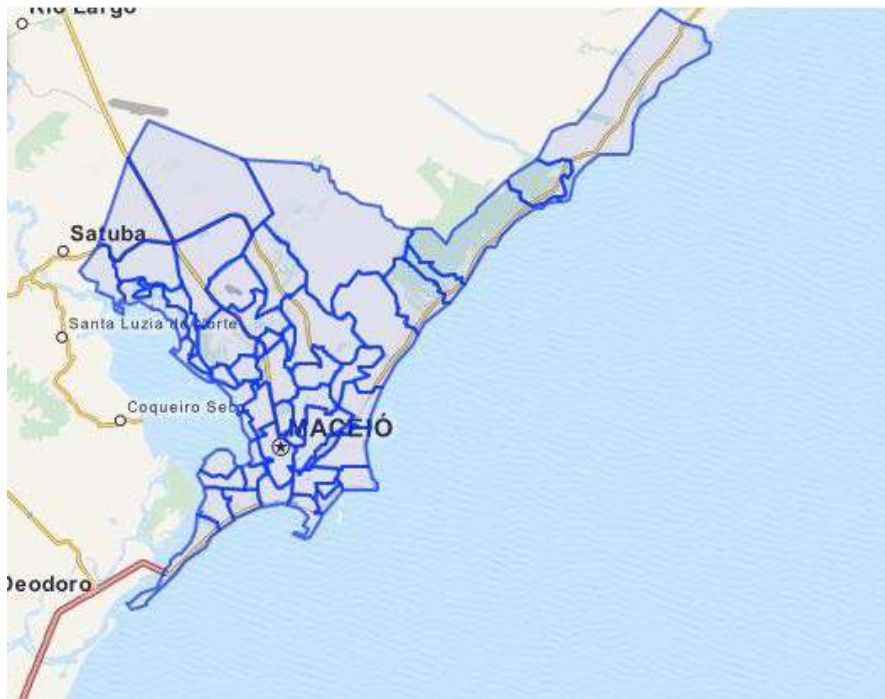


The screenshot shows the 'Dados Abertos' website for the Brazilian Federal Government. It features a green header with the text 'Dados Abertos GOVERNO FEDERAL'. Below the header, there are navigation links for 'Participe', 'Serviços', 'Legislação', and 'Canais'. The main content area includes a search bar and a list of datasets, such as 'Postos e delegacias do Departamento de Polícia' and 'Agências da Previdência Social - APS'.



The screenshot shows the 'Alagoas em Dados e Informações' website. It features a blue header with the text 'ALAGOAS EM DADOS E INFORMAÇÕES'. Below the header, there are navigation links for 'Início', 'Institucional', 'Estatísticas e Indicadores', 'Economia e Sociedade', and 'Geoprocessamento'. The main content area includes a search bar and a list of datasets, such as 'Pop. Residente' and 'População Censitária Urbana'.

# Exemplo

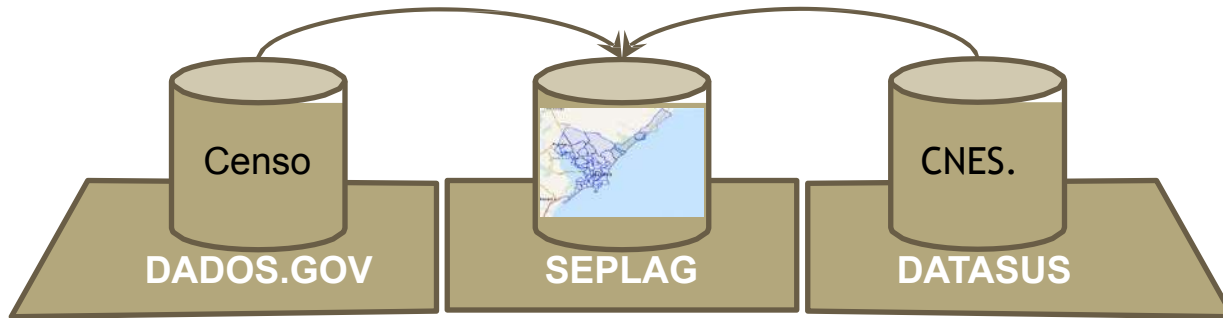


Qual bairro apresenta o maior índice de habitantes por escola?

Qual o bairro que tem mais estabelecimentos de saúde?

Quais bairros de Maceió não possuem delegacias?

É preciso conectar outras fontes



# Conversão de dados

File Edit View Reasoner Tools Refactor Window Help

locality (<http://nees.com.br/linkn/onto/locality/>)

Active Ontology Entities Data Properties Individuals

Class hierarchy Class hierarchy (inferred)

Class hierarchy:

- Thing
  - Address ≡ 'Postal Address'
  - DateTime
  - decimal
  - Locality
    - Bairro
    - Continent ≡ Continent
    - District
    - Estado ≡ State
    - MetropolitanRegion
    - Município ≡ City
    - Pais ≡ Country
    - Rua
    - Sector
    - Subdistrict
  - Resource
  - string
  - Thing
  - Time

File Edit View Reasoner Tools Refactor Window Help

locality (<http://nees.com.br/linkn/onto/locality/>)

Active Ontology Entities Data Properties Individuals Object Properties

Object property hierarchy:

- topObjectProperty
  - boundingbox
  - CEP
  - composesAddress
  - displayName
  - hasCentroidLatitude
  - hasCentroidLongitude
  - hasID
  - hasLatitude
  - hasLocalityAcronym
  - hasLocalityCode
  - hasLocalityName
  - hasLocation
  - hasLongitude
  - hasNumber
  - hasPolygon
  - isComposedOf
  - isLocatedIn
  - neighboring

File Edit View Reasoner Tools Refactor Window Help

locality (<http://nees.com.br/linkn/onto/locality/>)

Active Ontology Entities Data Properties Individuals

Data property hierarchy:

- 'File Size'
- 'Founding Date'
- Free
- 'Guideline Date'
- hasCentroidLatitude
- hasCentroidLongitude
- hasDemographicDensity
- hasID
- hasLatitude
- hasLocalityAcronym
- hasLocalityCode
- hasLocalityName
- hasLongitude
- hasNumber
- hasPerimeter
- hasPolygon
- hasPolygonArea
- hasPopulation
- 'Is Available Generically'

# Conversão de dados

127.0.0.1:3333/project?project=2429341072053

Google refine Maceio - Escolas (Enderecos) Permalink

Open... Export Help

Facet / Filter Undo / Redo 4

541 rows

Extensions: Freebase RDF

Show as: rows records Show: 5 10 25 50 rows « first < previous 1 - 10 next > last »

Nome\_Escola

case sensitive  regular expression

All	ANO_CENSO	PK_COD_ENTID	Nome_Escola	REDE	ID_DEPENDENC	Dependencia_A	CATESCPRIVAD	Localizacao	REGIAO	FK_COD
1.	2012	27034348	ASSOCIACAO COMUNITARIA RZ	Privada	4	Privada	Comunitaria	Urbana	Nordeste	
2.	2012	27034372	CENTRO ESPECIALIZADO PROFIS E EDUC FLORESER - CEPEF APAE-MACEIO	Privada	4	Privada	Filantropica	Urbana	Nordeste	
3.	2012	27034550	ASSOCIACAO CENTRO ESPECIALIZADO LOURDINHA VIEIRA	Privada	4	Privada	Filantropica	Urbana	Nordeste	
4.	2012	27034585	CENTRO DE ESTUDOS PSICOPEDAGOGICOS LTDA	Privada	4	Privada	Particular	Urbana	Nordeste	
5.	2012	27034607	CENTRO DE FORMACAO PROFISSIONAL GUSTAVO PAIVA	Privada	4	Privada	Particular	Urbana	Nordeste	
6.	2012	27034674	ESCOLA ESTADUAL PROFESSORA LAURA DANTAS SANTOS DA SILVA	Publica	2	Estadual		Urbana	Nordeste	
7.	2012	27034755	ESCOLA DE ENS FUND DR HENRIQUE EQUELMAN	Publica	3	Municipal		Urbana	Nordeste	
8.	2012	27034780	COLEGIO ATHENEU	Privada	4	Privada	Particular	Urbana	Nordeste	
9.	2012	27034798	COLEGIO BATISTA ALAGOANO	Privada	4	Privada		Urbana	Nordeste	
10.	2012	27034828	COLEGIO DA IMACULADA CONCEICAO	Privada	4	Privada	Particular	Urbana	Nordeste	



# Conversão de dados

The screenshot shows a web browser window with a URL of 127.0.0.1:3333/project?project=2429341072053. The page title is "Google refine" and the content is "Maceio - Escolas (Endereços)". A dialog box titled "RDF Schema Alignment" is open, displaying instructions and configuration options for mapping grid data to an RDF schema.

**RDF Schema Alignment**

The RDF schema alignment skeleton below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the skeleton. Configure the skeleton by specifying which column to substitute into which node.

**Base URI:** <http://nees.com.br/linkn/onto/> [edit](#)

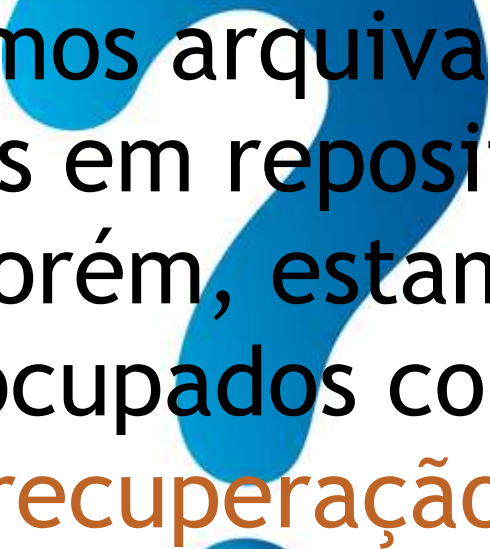
**RDF Skeleton** | RDF Preview

Available Prefixes: loc bdg owl rdf xsd osm rdfs foaf [+ add prefix](#) [manage prefixes](#)

<b>PK_COD_ENTIDADE URI</b> <input type="checkbox"/>	<input checked="" type="checkbox"/> <a href="#">-&gt; bdg:hasSchoolCode-&gt;</a>	<input type="checkbox"/> <b>PK_COD_ENTIDADE cell</b>
<input checked="" type="checkbox"/> <a href="#">osm:School</a>	<input checked="" type="checkbox"/> <a href="#">-&gt; foaf:name-&gt;</a>	<input type="checkbox"/> <b>Nome_Escola cell</b>
<a href="#">add rdf.type</a>	<input checked="" type="checkbox"/> <a href="#">-&gt; bdg:hasAdministrativeDependence-&gt;</a>	<input type="checkbox"/> <b>Dependencia_Administrativa cell</b>
	<input checked="" type="checkbox"/> <a href="#">-&gt; bdg:hasAddress-&gt;</a>	<input type="checkbox"/> <b>PK_COD_ENTIDADE URI</b> <input type="checkbox"/>
		<input checked="" type="checkbox"/> <a href="#">loc:Address</a>
		<a href="#">add rdf.type</a>
		<input checked="" type="checkbox"/> <a href="#">-&gt; loc:isComposedOf-&gt;</a> <input type="checkbox"/>
		<input checked="" type="checkbox"/> <a href="#">-&gt; loc:hasLongitude-&gt;</a> <input type="checkbox"/>
		<input checked="" type="checkbox"/> <a href="#">-&gt; loc:hasLatitude-&gt;</a> <input type="checkbox"/>
		<input checked="" type="checkbox"/> <a href="#">-&gt; loc:isComposedOf-&gt;</a> <input type="checkbox"/>

[Add another root node](#) [Save](#)

The background shows a data table with columns: REGIAO, FK\_COD\_ENTIDADE, and FK\_COD\_ENTIDADE URI. The data rows all contain "Nordeste" in the REGIAO column.



Estamos arquivando os  
dados em repositórios.  
Porém, estamos  
preocupados com sua  
recuperação ?

# Identificadores persistentes

O quê é um identificador persistente?

*É um método de resolução de endereços (URL) que procura garantir o acesso a objetos na Internet, mesmo quando eles mudam de local (URL de acesso).*

*Handle: [hdl.handle.net/123456789/1234](https://hdl.handle.net/123456789/1234)*

*DOI: [dx.doi.org/10.4100/jhse.2010.52.15](https://dx.doi.org/10.4100/jhse.2010.52.15)*

*PURL: [purl.org/net/example/purlName](https://purl.org/net/example/purlName)*

# Identificadores persistentes

Algumas alternativas disponíveis

**10045/13546**

↑ prefijo ↑ sufixo ↑

**Handle**

<http://purl.org/net/example/myFirstPURL>



Crossref

Código:10.20396

Sigla do título do periódico

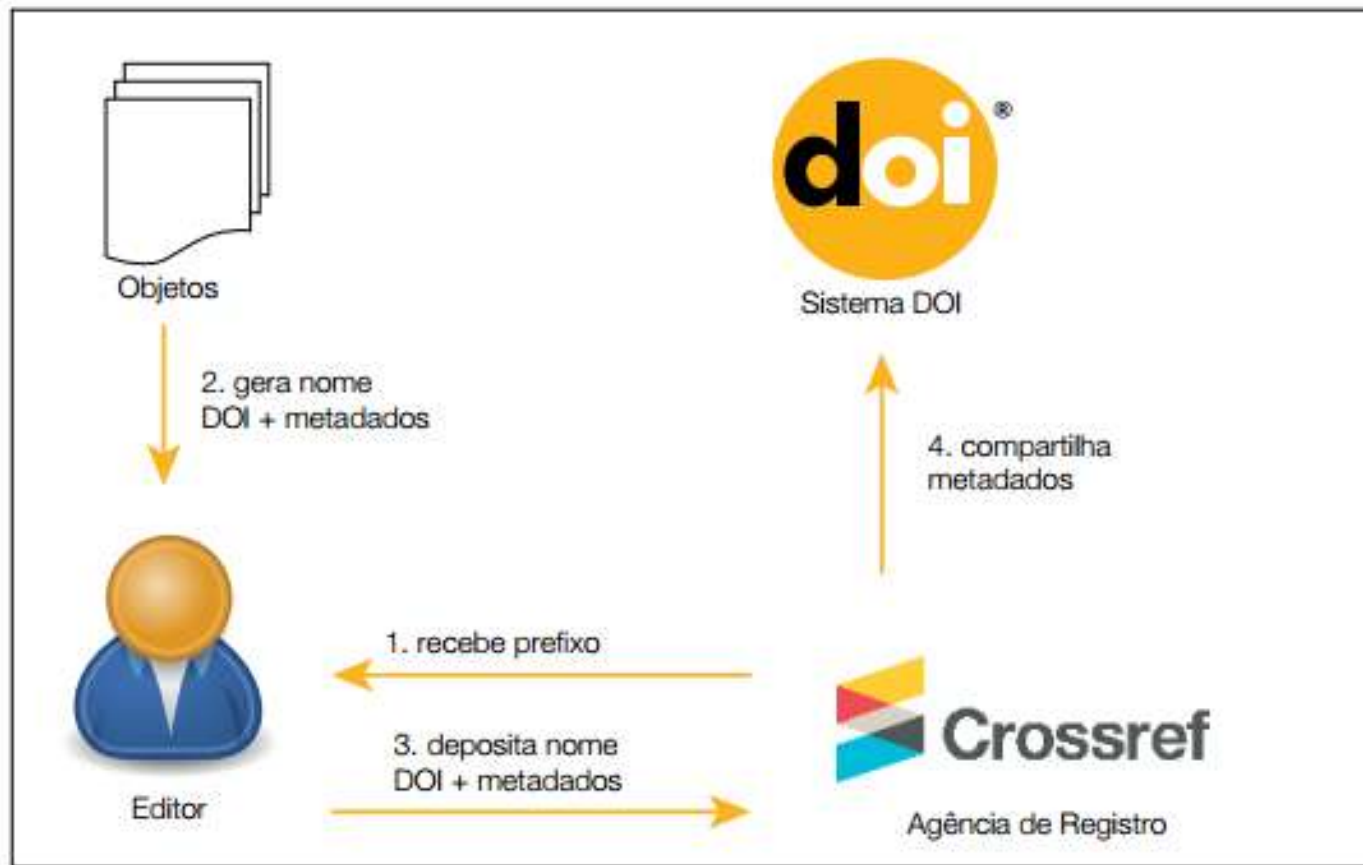
DOI: <http://dx.doi.org/10.20396/rdbci.v14i1.8640985>

volume

número

ID do artigo no OJS

# Como funciona o DOI?



Os metadados que estamos utilizando **não são** universais.

Os números **são**.

A única coisa que falta no "**canivete suíço**" dos identificadores persistentes é demonstrar o estado da arte dos dados que registram.

*Vamos fazer um teste:*

**crossref**  
metadata search

[Help](#) - [Example Queries](#)

<http://search.crossref.org>

Português

Inglês

Castelhano

Tailandês

Italiano

Grego

Francês

Mandarim

Alemão

Japonês

Holandês

Norueguês

Polaco

Hebreu

Romano

Russo

Turco

Coreano ...

# Como as bibliotecas podem organizar tudo isso? (e auxiliar os pesquisadores)





Primeiro desenvolva um **Plano de gestão de dados**



Seleção da  
mostra



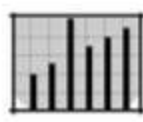
Coleta dos  
dados



Preparação  
dos dados



Análise  
exploratória



Definição  
dos  
indicadores



Visualização  
gráfica



Divulgação dos  
resultados

- **1. Revise os requisitos do órgão financiador**
- **2. Identifique os dados:** tipos, procedência, volume, formatos e ficheiros
- **3. Defina como os dados serão organizados e gerenciados:** nomes dos arquivos, controle das versões, software necessário
- **4. Explique como os dados serão documentados:** identifique os dados que serão processados, consulte se existem formatos ou esquemas de metadados; identifique ferramentas que possibilitem sua gestão
- **5. Descreva os processos que garantam a qualidade dos dados**

- **6. Prepare uma estratégia para o depósito (durante o processo) e preservação dos dados (repositório)**
- **7. Defina as políticas de dados do projeto:** questões sobre a propriedade intelectual e como os dados sensíveis e pessoais serão tratados.
- **8. Descreva como os dados serão divulgados:** onde, o que, quando serão divulgados. Se você publicar os dados em um repositório, como informações suplementares do artigo ou como um “data paper”.
- **9. Atribua papéis e responsabilidades para as pessoas e organizações que participam do projeto**
- **10. Prepare um orçamento realista:** gerenciamento de dados custa tempo e dinheiro em termos de software, hardware, serviços e pessoal.

## Open data: 5 Estrelas (Tim Berners-Lee)



Torne seus recursos disponíveis na Web (tanto faz o formato) sob uma licença aberta<sup>1</sup>



Torne seus recursos disponíveis como dados estruturados (ex. excel no lugar de imagem escaneada)



Utilize formatos não-proprietários (ex. CSV e não excel)

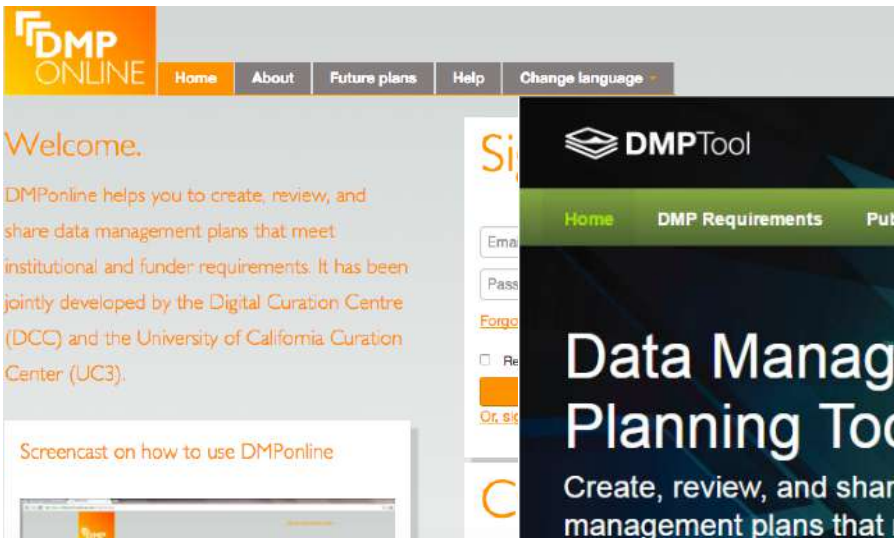


Utilize URLs para identificar recursos. Isso vai ajudar as pessoas a apontarem para eles



Conecte seus dados com outros para prover contexto (dados linkados)

Existem ferramentas na forma de um portal da web que facilitam a tarefa de escrever um Plano de Gestão de Dados:



**DMP ONLINE** Home About Future plans Help Change language

Welcome.

DMPonline helps you to create, review, and share data management plans that meet institutional and funder requirements. It has been jointly developed by the Digital Curation Centre (DCC) and the University of California Curation Center (UC3).

Screencast on how to use DMPonline

<https://dmponline.dcc.ac.uk/>



**DMPTool** Home DMP Requirements Public DMPs News Help Contact Us About Log In

**Data Management Planning Tool**

Create, review, and share data management plans that meet institutional and funder requirements.

Video sobre cómo usar PGDonline

<https://dmptool.org>



**PGDonline** Inicio Acerca de Hoja de ruta Ayuda Cambiar idioma

Ha cerrado la sesión satisfactoriamente.

Bienvenido/a.

PGDonline ha sido desarrollado por el Digital Curation Centre y adaptado por el Consorcio Madroño para ayudarle a escribir planes de gestión de datos.

Conectar

Registrarse

<http://dmp.consorciomadrono.es>

# O que deve ser levado em conta?

Quem possui os dados?

Quais requisitos são impostos por outras pessoas?

Quais dados devem ser mantidos e por quanto tempo?

Como os dados digitais devem ser preservados?

Existem considerações éticas?

Como os dados são acessados?

Como os dados devem ser oferecidos?

Como os custos serão tratados?

# Resumindo...

## **Criar o Plano de Gestão de Dados**

- DMP Tool. <https://dmp.cdlib.org/>
- PGD Online <http://dmp.consorciomadrono.es>
- DMPonline. <http://dmponline.dcc.ac.uk/>

## **Publicar dados abertos**

1. Verifique se você tem um repositório temático adequado. Consulte [re3data.org](http://re3data.org)
2. Publicar dados no repositório apropriado
3. Outros repositórios multidisciplinares (Zenodo, Dryad, Dataverse, Figshare, Mendeley Data ...)





**I Seminário Gestão de Dados de Pesquisa UNIFESP e UFABC**



## **Planos de Gestão de Dados**

---

Prof. Fabiano Couto Corrêa da Silva  
[fabianocc@gmail.com](mailto:fabianocc@gmail.com)