

Gestão de dados científicos de pesquisa em Ciências Sociais Aplicadas

DOI:
10.5281/zenodo.7795068

Janete Saldanha Bach Estevão

 <https://orcid.org/0000-0001-9011-5837>



SIDIP2023
V SEMINÁRIO DE INTEGRAÇÃO E
DICAS INSTRUMENTAIS PARA PESQUISA

Gestão de Dados Científicos de Pesquisa em Ciências Sociais Aplicadas

 17h

Transmissão via Canal do PPGGI no YouTube e Facebook



Janete Saldanha Bach Estevao
Palestrante
SiBi/UFPR



Logan Nobre
Mediador
PPGGI/UFPR



PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO DA INFORMAÇÃO





MINI BIO JANETE BACH

Dra. Janete Saldanha Bach Estevão, - atua no GESIS - Instituto Leibniz de Ciências Sociais, na Alemanha desde 2021. Pesquisadora no consórcio KonsortSWD da NFDI no departamento "Tecnologias do Conhecimento para as Ciências Sociais", equipe FAIR Data and Human Information Interaction, trabalhando no consórcio KonsortSWD da National Research Data Infrastructure (NFDI). É doutora e mestre em Estudos Tecnologia e Sociedade (UTFPR) e bacharel em Biblioteconomia (UFPR). Sua experiência em pesquisa é em Ciência Aberta, especialmente em gerenciamento de dados de pesquisa e reutilização de dados nas Ciências Sociais. Possui experiência em gestão da informação e do conhecimento, gestão documental e de processos e tecnologia em empresas nacionais e multinacionais. Na área de Tecnologia possui a certificação CDIA+ – Comptia. Na área de privacidade, é especialista em proteção de dados, capacitada como DPO, certificada pela Exin. Também é certificada pela The IAPP como CIPM. Atuou entre 2012 e 2021 com a capacitação em pesquisa científica para estudantes da graduação e da pós-graduação na Universidade Federal do Paraná.



Agenda

Contexto dos dados de pesquisa na Ciência aberta

Os **porquês** dos dados de pesquisa

Plano de gestão de dados + **Coleta** ou **reuso** de dados

Plano de gerenciamento de dados (DMP)

Formatos de arquivos e metadados

Criação de **documentação**

Depósito em portais de dados

Cópia de **segurança de dados**, ferramentas e *backup*

Licenças, direitos autorais e de propriedade intelectual

Dados confidenciais, **personais**, sensíveis e privados

Identificadores **Persistentes**

Reuso de dados – **melhores práticas**

Contexto: dados na Ciência Aberta

Exigências de compartilhamento de dados

A pressão regulatória aumenta a disponibilidade de dados

Dados que existem não são facilmente encontrados

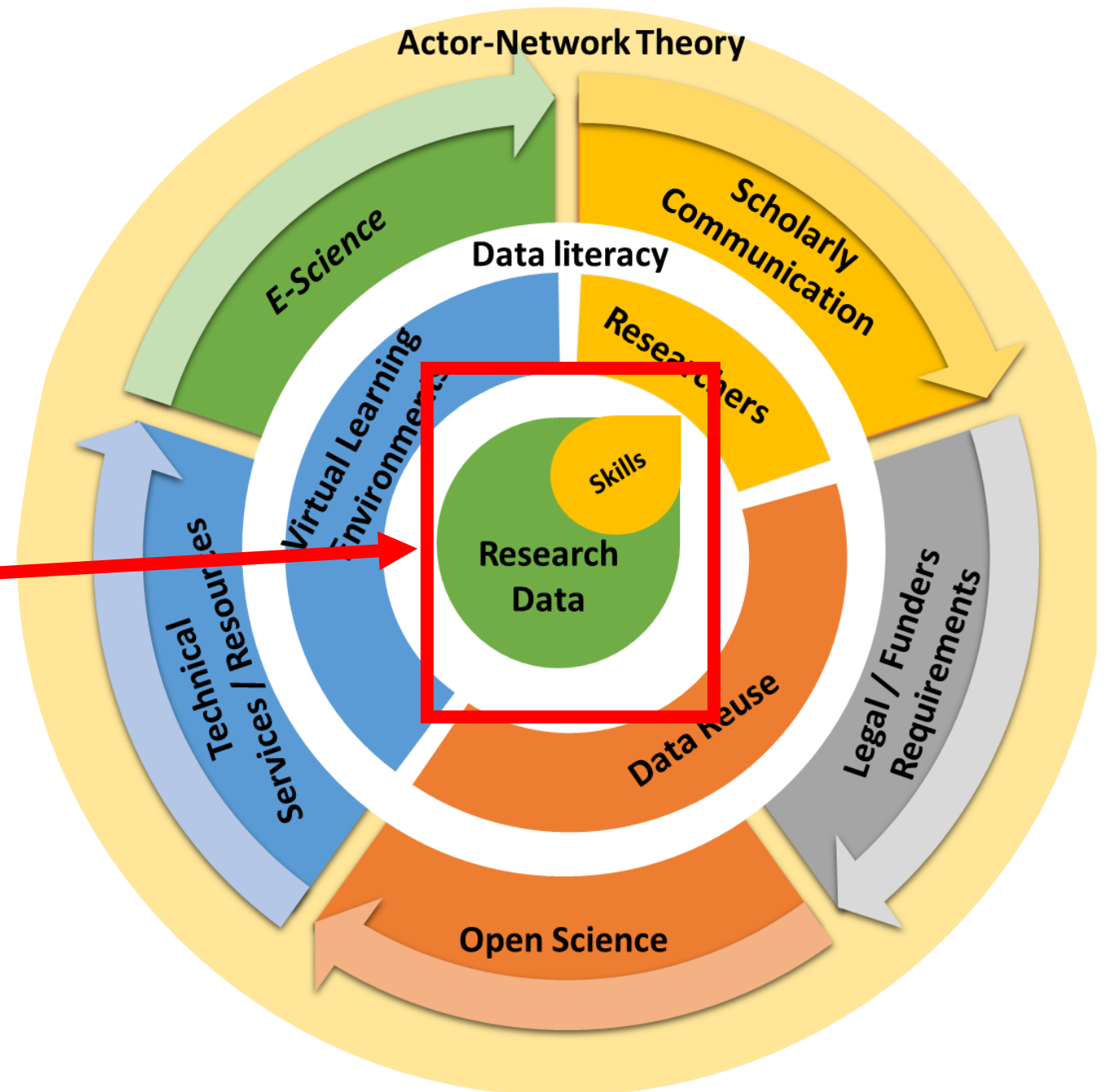
Quando encontrados não são bem documentados
(metadados, documentação, *codebooks*, etc.)

Barreiras de acesso e embargo

Pesquisadores(as) precisam de letramento em dados, o que expõe o potencial inexplorado de reutilização de dados

Habilidades para pesquisadores

Habilidades para gestão de dados pesquisa no ecossistema científico





O que são os dados de pesquisa?

Os dados de pesquisa referem-se ao “[material de fato registrado](#), comumente aceito na comunidade científica como **necessário para validar os resultados da pesquisa**”.

- [arquivos](#) que apóiam seu projeto de pesquisa, estudo ou publicação, como **planilhas, documentos, imagens, vídeos ou áudio**.
- [todos os dados](#), documentos e objetos **analógicos relacionados** à pesquisa que devem **ser digitalizados** no decorrer de um determinado processo de pesquisa

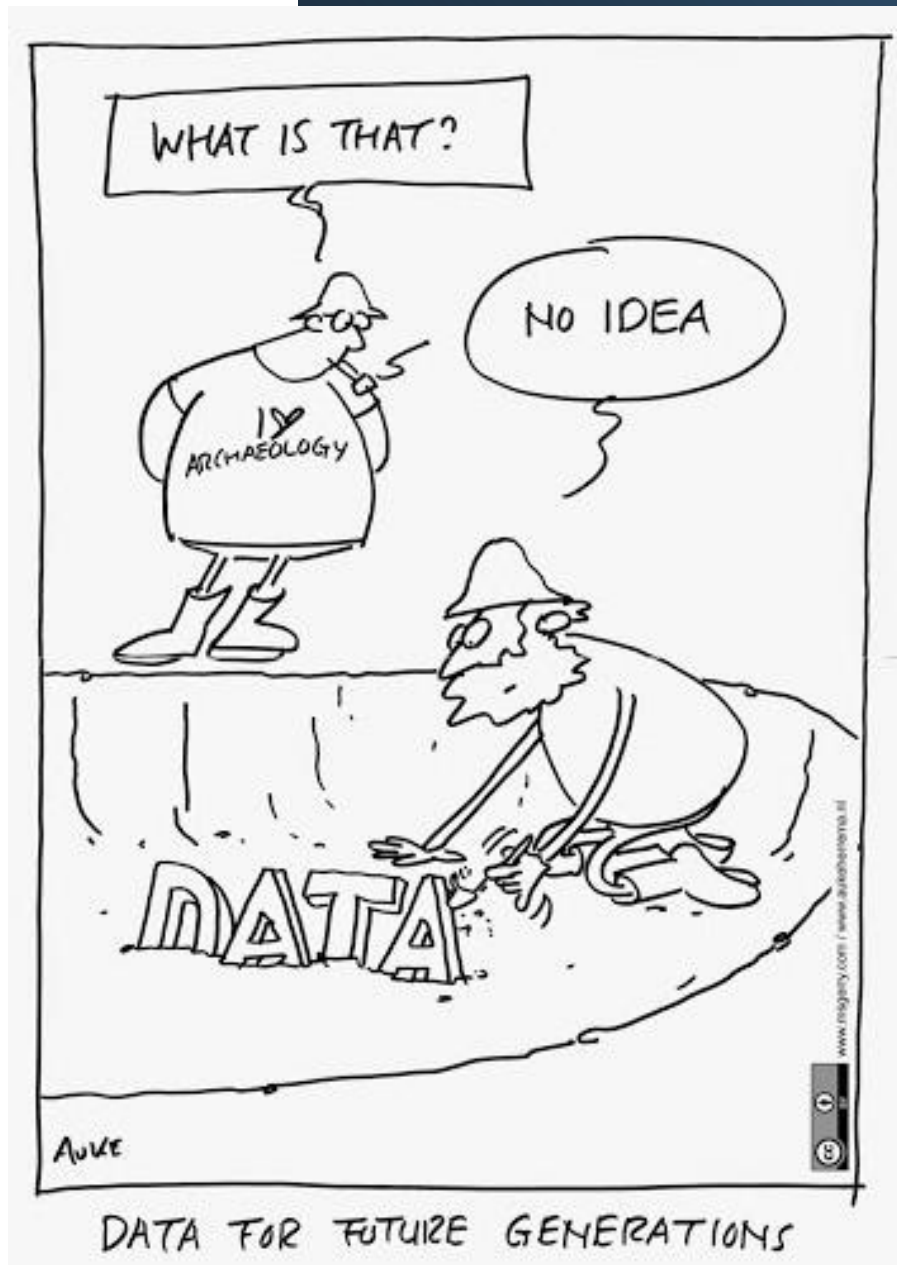
(The Engineering and Physical Sciences Research Council, Nature, Freie Universität Berlin).

Publication and data (Cartoon by [Auke Herrema for the Research Data Alliance.](#))



PUBLICATIONS AND DATA

Os dados de pesquisa têm características e particularidades disciplinares



What is data (Cartoon by [Auke Herrema for the Research Data Alliance.](http://www.researchdataalliance.com))

- Os dados cobrem uma ampla gama de **tipos** de registros e podem ser estruturados e armazenados em vários **formatos** de arquivos
- Todos os dados de pesquisa são incluídos na conceituação, independentemente do formato em que são criados
- Pesquisadores em quase todas as disciplinas agora criam dados em formato digital (HEY; TANSLEY; TOLLE, 2009).



Information Types. (Cartoon by [Auke Herrema](#) for the [Research Data Alliance](#).)

E os dados de pesquisa
nas Ciências Sociais
Aplicadas?

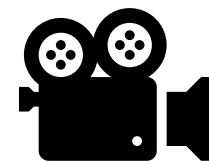
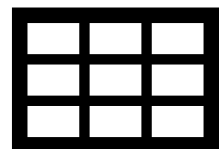


O quê?

Dados nas
Ciências
Sociais
Aplicadas

Evidências de fenômenos sociais, políticos, econômicos e culturais para entender a realidade e resolver problemas da sociedade.

Os dados podem ser **qualitativos** ou **quantitativos**.



Como?

Metodologia e Técnicas de pesquisa

Dados são **obtidos** por meio de pesquisas e questionários, entrevistas, observação, análise de documentos e registros, análise de conteúdo (textos, imagens), experimentos e estudos de caso, dados administrativos e estatísticos.

De onde?

Fontes
primárias ou
secundárias

Coletar novos dados

Reusar dados

E por que eu tenho que gerenciar os dados?



- Mudanças **políticas** das agências de fomento e publicações científicas e nas universidades ;
- Maior regulação como requisito de **contrapartida** ao financiamento público de pesquisas



Data Sharing (Cartoon by [Auke Herrema](#) for the Research Data Alliance.)

- Produz **novos recursos de informação** viabilizada pela retroalimentação da atividade científica;
- conjuntos de dados são agora disponibilizados por outros pesquisadores, que **antes eram inacessíveis**
- **Credibilidade, transparência, prestação de contas**



Scene from the past? (Cartoon by [Auke Herrema](http://www.maggie.com/) for the Research Data Alliance.)

Vamos conversar?



Go to **www.menti.com** and
use the code **2702 5912**

[https://www.menti.com/
al8gtmpp7a22](https://www.menti.com/al8gtmpp7a22)



Como eu gerencio os
dados de pesquisa nas
Ciências Sociais Aplicadas?

Principais entregas

Plano de gestão de dados |

+

Coletar novos dados |

Reusar dados |



Gerenciamento de dados de pesquisa – melhores práticas

- 1. Plano de gerenciamento de dados (DMP)**
2. Formatos de dados, metadados, normas relacionadas
3. Criação de documentação e metadados
4. Depósito em portais de dados

Plano de gerenciamento de dados (DMP)



Orientação para a Elaboração do Plano de Gestão de Dados

<https://www.youtube.com/watch?v=VwLNaQiLjrU>

Plano de Gestão de dados

<https://www.youtube.com/watch?v=D-NjUp77pZ4>



Guia de Preparação e Arquivamento de Dados de Ciências Sociais

<https://www.icpsr.umich.edu/web/pages/deposit/guide/>

Check list:  | D | C | C

[https://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP Checklist 2013.pdf](https://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP%20Checklist%202013.pdf)

Planos online  

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

<https://www.dcc.ac.uk/guidance/how-guides/develop-data-plan>

Principais entregas

Plano de gestão de dados |

+

Coletar novos dados |

E

OU

Reusar dados |



Gerenciamento de dados de pesquisa – melhores práticas

1. Plano de gerenciamento de dados (DMP)
- 2. Formatos de dados, metadados, normas relacionadas**
3. Criação de documentação e metadados
4. Depósito em portais de dados

Formatos de archivo



Papel dos formatos
para dados FAIR
(documentação,
padrões)

Problemas sobre formatos de arquivos e seus padrões abertos X proprietários:

- a. obsolescência
- b. perdas nas conversões
- c. Dados não serem mais acessíveis
- d. Custo da licença para acesso aos dados

- Como converter formatos de arquivo?
- Como exportar arquivos para um formato diferente?
- Como salvar os arquivos para preservar o original (proprietário) juntamente com uma opção em formato aberto?

Identificação de formatos PRONOM

- <https://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=new>

The screenshot shows the National Archives website's PRONOM search page. The browser address bar displays the URL: nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=new. The page features the National Archives logo and a search bar. A navigation menu includes links for About us, Education, Records, Information management, and Archives sector. The breadcrumb trail reads: You are here: Home > Information management > Digital preservation > PRONOM > Search: Search by format. The main heading is 'The technical registry PRONOM'. A search bar is labeled 'Search : By format'. Below this, there are tabs for 'Simple search', 'File format', 'PRONOM Unique Identifier', 'Software', 'Vendor', 'Lifecycles', and 'Migration Pathways'. The 'File format' tab is active. Under '1. File formats', there are two search options: one for file extensions and one for file format names. Under '2. Compatible software', there is a search option for software based on file extension and processing type. The page also includes a 'Help : search by format' link.

nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=new

The National Archives

About us Education Records Information management Archives sector

You are here: [Home](#) > [Information management](#) > [Digital preservation](#) > [PRONOM](#) > Search: Search by format

The technical registry
PRONOM

Welcome : About Add an entry
Search ? Help Information resources

? Help : search by format

Search : By format

Simple search File format PRONOM Unique Identifier Software Vendor Lifecycles Migration Pathways

1. File formats

Enter a file extension and click 'search' to find all file formats with that extension. Leave the file extension blank to find all file formats in the database.

*. Search >

To search for a particular file format, enter the name of the file format and then click 'search'. Leave the file format name blank to find all file formats in the database.

Search >

2. Compatible software

Enter a file extension and click 'search' to find software which can files with that extension.

*. Search >

Gerenciamento de dados de pesquisa – melhores práticas

1. Plano de gerenciamento de dados (DMP)
2. Formatos de dados, metadados, normas relacionadas
- 3. Criação de documentação e metadados**
4. Depósito em de portais de dados

Documentação e Metadados



LETRA	NOME	FRUTA	ANIMAL	OBJETO	CIDADE/ESTADO/PAÍS
L	LÚCIO	LIMÃO	LAGARTIXA	LIXA	LISBOA
J	JOSEFA	JACA	JABUTI	O	JUQUITIBA
E	ERIC	O	ELEFANTE	ESCADA	ESPAÑA
M	MURILDO	MAÇÃO	MACACO	MESA	MANAUS
P	PEDRO	PERA	PATO		

Qual documentação será necessária para que os dados sejam lidos e interpretados corretamente no futuro?

- Nível de projeto
- Nível de arquivo
- Nível do item ou variável
- Qualquer outra informação contextual necessária para que outros possam interpretar

Este guia orienta os cientistas sociais a tornar seus dados de pesquisa acessíveis, compreensíveis, sustentáveis e reutilizáveis.

<https://www.cessda.eu/Research-Infrastructure/Training/Expert-Tour-Guide-on-Data-Management>

	Discipline	MainDiscij	Usefulness	Usefulness Use
1	20	6	2	2
2	49	6	1	1
3	29	3	1	1
4	62	2	2	2
5	62	2	3	3
6	59	2	2	1
7	21	2	3	3
8	33	2	2	2
9	79	1	2	2
10	31	2	4	4
11	58	2	3	3
12	33	2	99	99
14	21	2	3	4
15	59	2	2	2

Project Description

Summary: This study explores the factors that influence the data reuse behaviors of scientists and identifies the generalized patterns that occur in data reuse across various disciplines. An online survey was distributed to the scientists through Qualtrics. The initial email invitation to the survey was sent to 15,703 scientists within academic institutions on October 5, 2015, with a reminder sent on November 10, 2015. The survey closed on November 30, 2015. 1,987 email messages (12.65%) were returned and a total of 13,716 participants (87.35%) received the email invitation to participate in the survey. This research used the National Science Foundation (NSF) STEM discipline codes (2014) for the respondents to indicate their specific academic disciplines based on their current research activities. Of these participants, 1,528 scientists from 94 specific disciplines (as categorized by NSF STEM discipline codes (2014)), completed the survey with less than 5% of missing values (response rate: 11.14%).

Reference:

NSF. (2014, February 9). Crosswalk of NSF Discipline Codes to CASPAR Academic Discipline Codes. Retrieved from <https://ncesdata.nsf.gov/nsf/srs/webcasp/data/gradstud.htm>



Exemplo: documentação está incompleta, o *link* quebrado (não havia um PID) e a tabela de classes de áreas já está atualizada (os códigos usados no dataset já podem ter sido alterados)

Guia para escrever metadados em estilo "readme"
<https://data.research.cornell.edu/content/readme>

Os dados devem
facilmente
interpretados por
pessoas fora do
estudo, por meio
da adequada
documentação

- Dependendo da domínio, os documentos relevantes podem ser:

- a. Arquivo **Leiname**

- b. **Dicionário** de dados

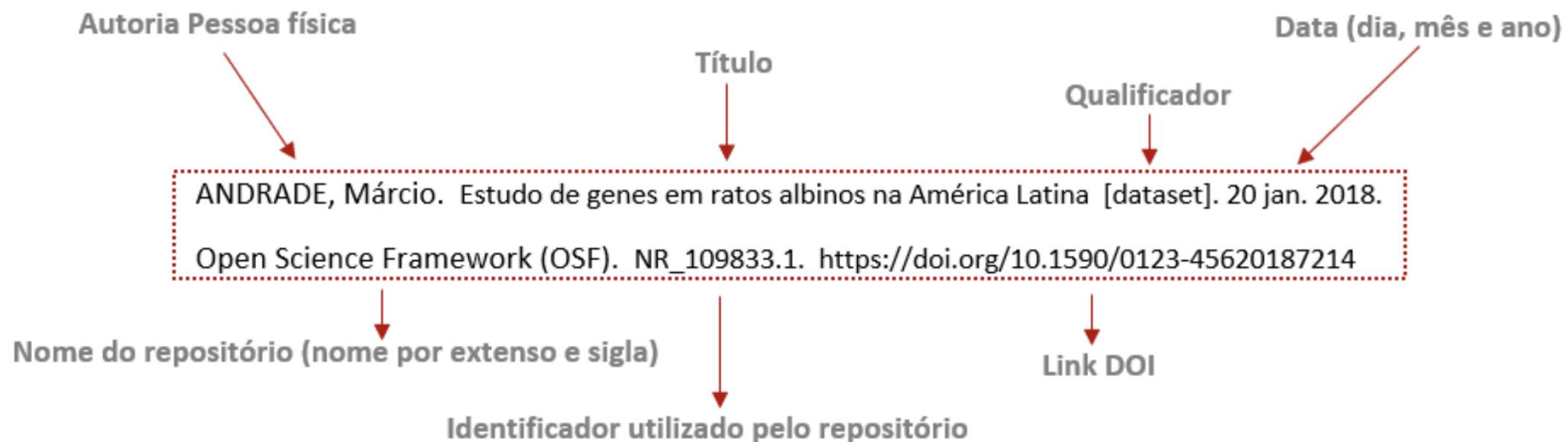
- c. **Codebooks**

- d. Códigos **comentado**

- e. Bloco de **notas** de laboratório/campo

AGÊNCIA DE BIBLIOTECAS E COLEÇÕES DIGITAIS USP. Dados Científicos: como construir metadados, descrição, readme, dicionário-de-dados e mais <https://www.abcd.usp.br/noticias/dados-cientificos-como-construir-metadados-descricao-readme-dicionario-de-dados>

Citação de dados de pesquisa completa:



Os metadados estruturam as descrições de dados, são legíveis por máquinas e por humanos

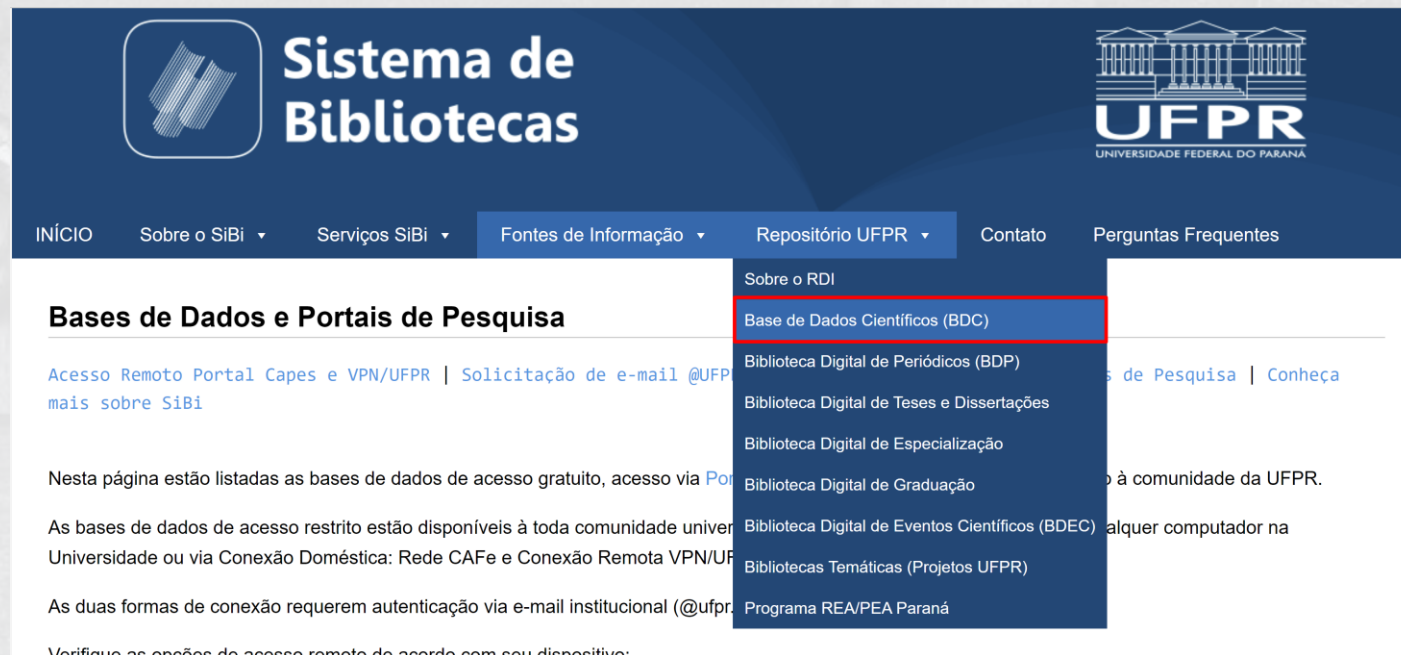
Gerenciamento de dados de pesquisa – melhores práticas

1. Plano de gerenciamento de dados (DMP)
2. Formatos de dados, metadados, normas relacionadas
3. Criação de documentação e metadados
- 4. Depósito em portais de dados**

Repositórios para depositar os dados

[Base de Dados Científicos da Universidade Federal do Paraná \(BDC/UFPR\)](https://bdc.c3sl.ufpr.br/)

<https://bdc.c3sl.ufpr.br/>



The image shows a screenshot of the UFPR Sistema de Bibliotecas website. The header features the logo of the Sistema de Bibliotecas and the UFPR logo. The navigation menu includes links for INÍCIO, Sobre o SiBi, Serviços SiBi, Fontes de Informação, Repositório UFPR, Contato, and Perguntas Frequentes. The 'Repositório UFPR' dropdown menu is open, listing various digital libraries and repositories. The 'Base de Dados Científicos (BDC)' option is highlighted with a red border. Below the menu, there is a section titled 'Bases de Dados e Portais de Pesquisa' with a link to 'Acesso Remoto Portal Capes e VPN/UFPR' and a brief description of the services provided.

Sistema de Bibliotecas

UFPR
UNIVERSIDADE FEDERAL DO PARANÁ

INÍCIO Sobre o SiBi Serviços SiBi Fontes de Informação Repositório UFPR Contato Perguntas Frequentes

Bases de Dados e Portais de Pesquisa

[Acesso Remoto Portal Capes e VPN/UFPR](#) | [Solicitação de e-mail @UFPR](#) | [Mais sobre SiBi](#)

Nesta página estão listadas as bases de dados de acesso gratuito, acesso via Portal de Acesso Remoto e acesso via VPN/UFPR.

As bases de dados de acesso restrito estão disponíveis à toda comunidade universitária via Portal de Acesso Remoto ou via Conexão Doméstica: Rede CAFE e Conexão Remota VPN/UFPR.

As duas formas de conexão requerem autenticação via e-mail institucional (@ufpr.br).

Verifique as opções de acesso remoto de acordo com seu dispositivo:

- Sobre o RDI
- Base de Dados Científicos (BDC)**
- Biblioteca Digital de Periódicos (BDP)
- Biblioteca Digital de Teses e Dissertações
- Biblioteca Digital de Especialização
- Biblioteca Digital de Graduação
- Biblioteca Digital de Eventos Científicos (BDEC)
- Bibliotecas Temáticas (Projetos UFPR)
- Programa REA/PEA Paraná

Repositórios e indexadores

Zenodo <https://zenodo.org/>

- O Zenodo é um repositório de dados de pesquisa de acesso aberto de propósito geral, hospedado pelo CERN (Suíça) que fornece um lugar para os pesquisadores depositarem conjuntos de dados. Pesquisadores em qualquer área temática podem fazer upload de arquivos de até 50 GB. Ele tem uma integração com GitHub para tornar o código hospedado no GitHub editável.

Depósito de dados

“Dados valiosos de pesquisa **não são úteis** para a comunidade científica e para o futuro se os criadores de dados originais são as únicas pessoas a ter qualquer **informação sobre os dados**. Se eles se mudarem para outras organizações ou para outras tarefas, ou se aposentar, todas as informações desaparecerão”.

<https://www.fsd.tuni.fi/en/services/data-management-guidelines/why-are-research-data-managed-and-reused/>

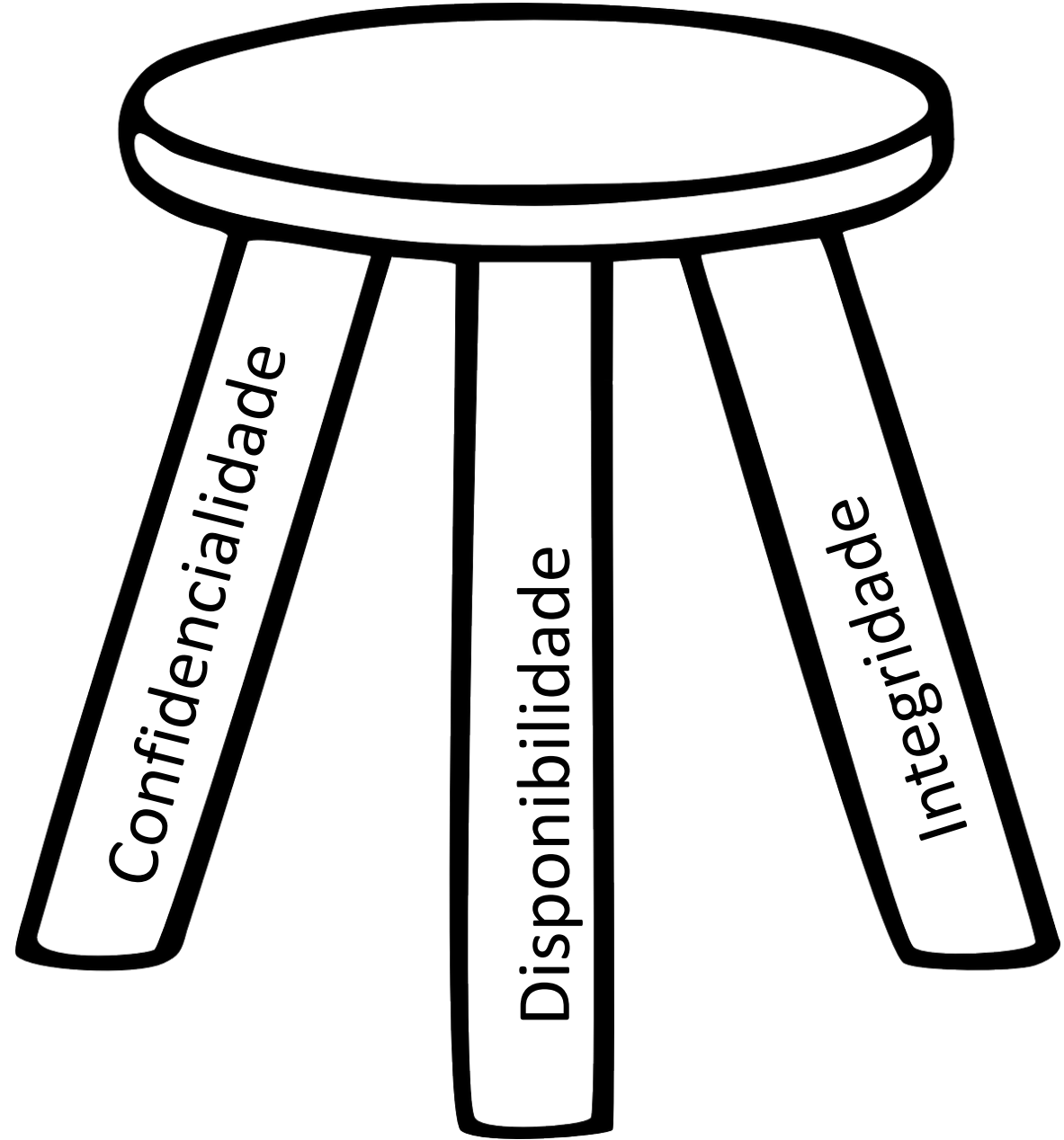
Gerenciamento de dados de pesquisa – melhores práticas

5. Cópia de segurança de dados, ferramentas e serviços de backup

- 6. Uso responsável dos dados (citações, direitos autorais, restrições de dados)
- 7. Manuseio de dados pessoais e sensíveis (Leis de proteção de dados e conformidade com a LGPD)
- 8. Identificadores persistentes

Tríade da segurança da informação

Confidencialidade
Integridade
Disponibilidade



Algumas medidas de segurança

- a. Impedir o **acesso não autorizado** por meio de métodos de verificação confiáveis (senhas, autenticação de dois fatores)
- b. Armazenar arquivos em um local **separado** de outros dados de pesquisa
- d. **Criptografia** (disco completo, pastas, arquivos)
- e. Sempre manter **cópias de segurança** – backup - e **TESTAR** as cópias (teste de *restore*)

Sempre seguir a política do repositório, por exemplo: [ICPSR](#)

Gerenciamento de dados de pesquisa – melhores práticas

5. Cópia de segurança de dados, ferramentas e serviços de backup

6. Uso responsável dos dados (citações, direitos autorais, restrições de dados)

7. Manuseio de dados pessoais e sensíveis (Leis de proteção de dados e conformidade com a LGPD)

8. Identificadores persistentes

Licenças, direitos
autorais e questões
de propriedade
intelectual

Qual tipo de objeto digital deve e pode ser licenciado:

- Dados
- Texto
- Softwares
- Códigos
- AV data (áudio e vídeo)
-

Licença

Entender os
tipos de **licenças**



Direitos autorais e direitos de propriedade intelectual



Webconferência Integridade da Pesquisa e o Plágio Acadêmico

<https://www.youtube.com/watch?v=gW0gbYmL4>

Direitos autorais, patentes, marcas registradas, direitos de desenho industrial, segredos comerciais, Direitos de bancos de dados

Open Data Commons Open Database License (ODbL)

- Attribution and Share-Alike for Data/Databases

Diferentes tipos de licenças

Linceça Creative Commons

- CC0 - No Rights Reserved
- Attribution CC BY
- Attribution ShareAlike CC BY-SA
- Attribution-NoDerivs CC BY-ND
- Attribution-NonCommercial CC BY-NC
- Attribution-NonCommercial-ShareAlike CC BY-NC-SA
- Attribution-NonCommercial-NoDerivs CC BY-NC-ND

<https://creativecommons.org/choose/>

Gerenciamento de dados de pesquisa – melhores práticas

5. Cópia de segurança de dados, ferramentas e serviços de backup

6. Uso responsável dos dados (citações, direitos autorais, restrições de dados)

**7. Manuseio de dados pessoais e sensíveis
(Leis de proteção de dados e conformidade
com a LGPD)**

8. Identificadores persistentes

Dados
confidenciais,
pessoais, sensíveis
e privados

Legislações de proteção de dados é uma crescente no mundo .

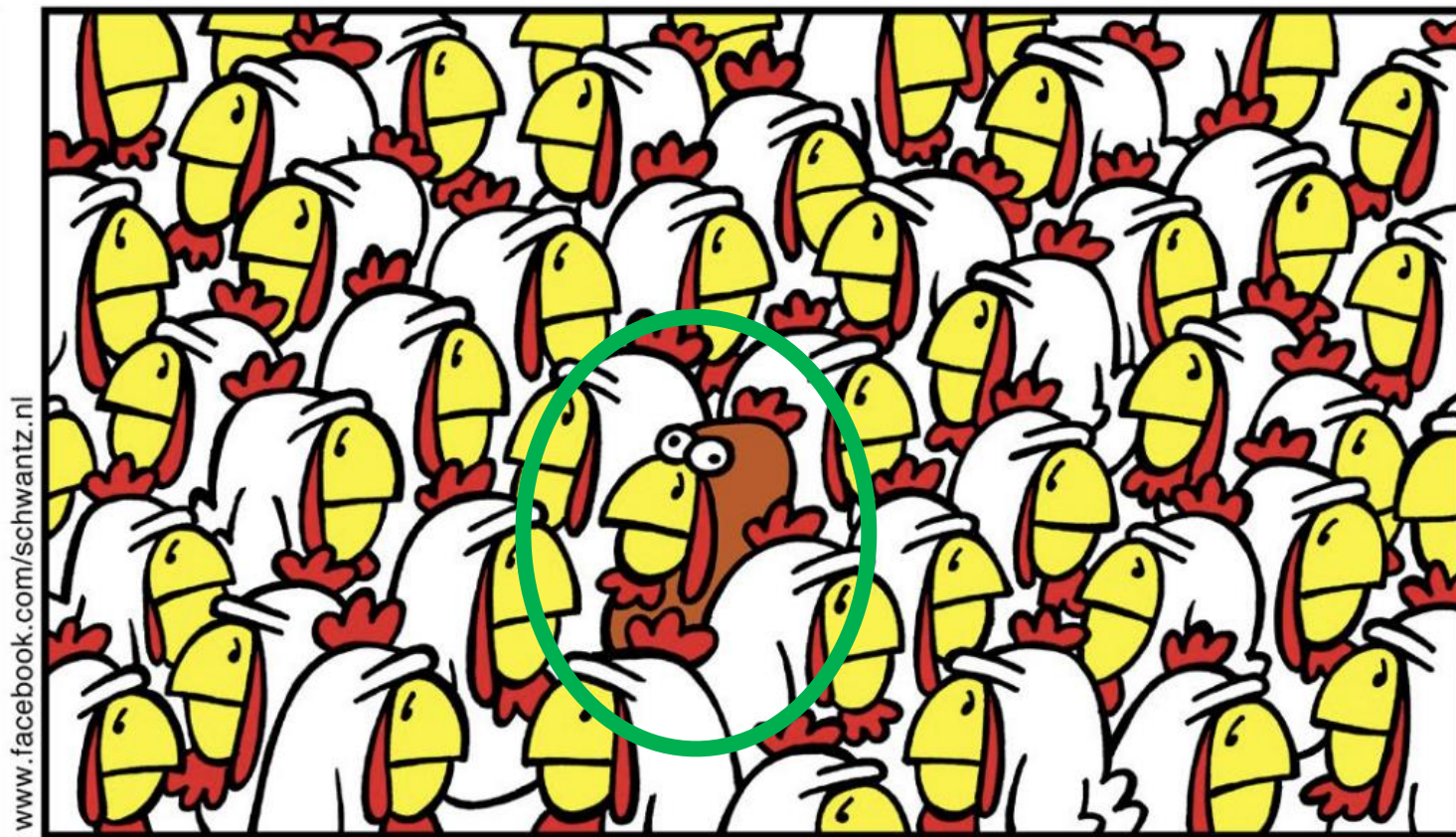
Quais os tipos de dados e quais as legislações aplicáveis?

Apenas **SE**:

Dado pessoal: dados que identificam uma **pessoa natural**

Dados anonimizado não é dado pessoal, portanto, é irrelevante sob o aspecto da privacidade

NÃO SE PREOCUPE ! SÓ GUARDAREMOS INFORMAÇÕES
GERAIS SOBRE VOCÊ, NADA QUE POSSA IDENTIFICÁ-LO!



Dados
confidenciais,
pessoais, sensíveis
e privados

Dados pessoais: informação relacionada a pessoa natural identificada ou identificável.

Dados pessoais sensíveis: informação sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a **uma pessoa natural.**

Algumas medidas proteção de dados

- a. Impedir o **acesso não autorizado** por meio de métodos de verificação confiáveis (senhas, autenticação de dois fatores)
- b. **Anonimização** de dados pessoais
- c. Ferramenta de anonimização: [amnesia](#)

Gerenciamento de dados de pesquisa – melhores práticas

5. Cópia de segurança de dados, ferramentas e serviços de backup
6. Uso responsável dos dados (citações, direitos autorais, restrições de dados)
7. Manuseio de dados pessoais e sensíveis (Leis de proteção de dados e conformidade com a LGPD)

8. Identificadores persistentes



PIDs

Identificadores
Persistentes



Identificadores
Persistentes são a
espinha dorsal dos
princípios FAIR

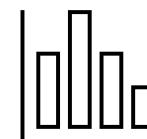
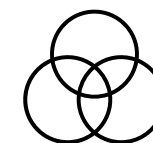
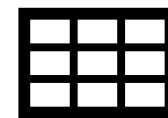
Objetos digitais podem ter o
nomes iguais ou semelhantes, tais como
versões ou autores diferentes.

Diferentes entidades e níveis de
atributos que podem ter por um PID, por
exemplo:

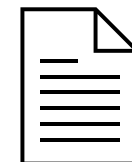
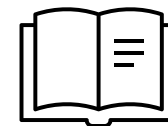
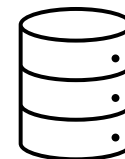
a. Pessoas



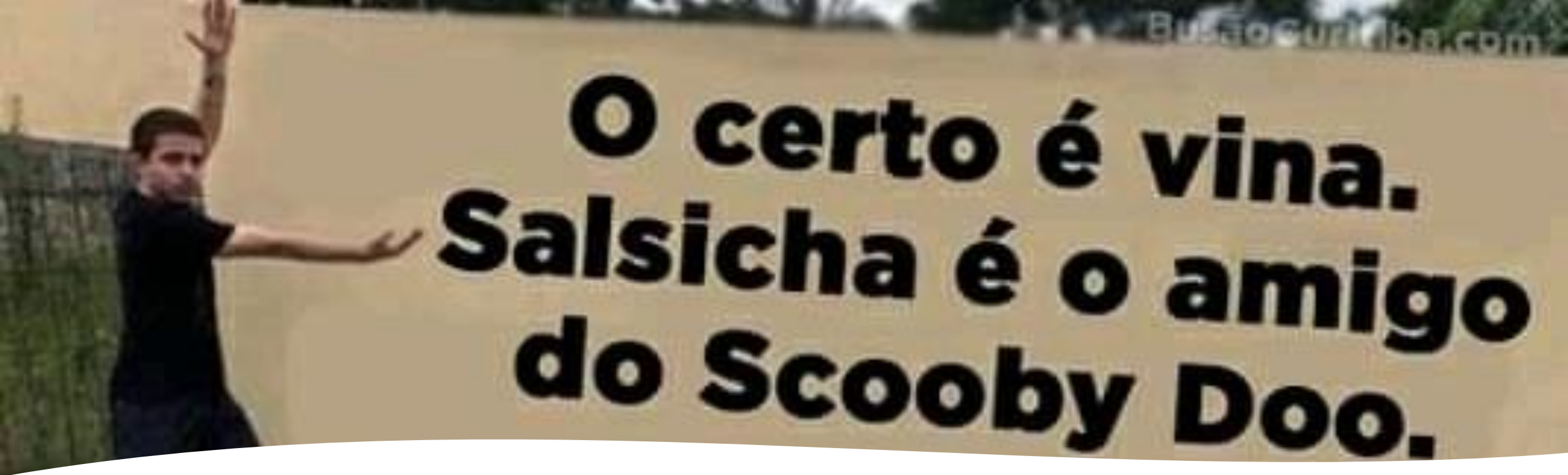
b. Instituições



c. Artigos

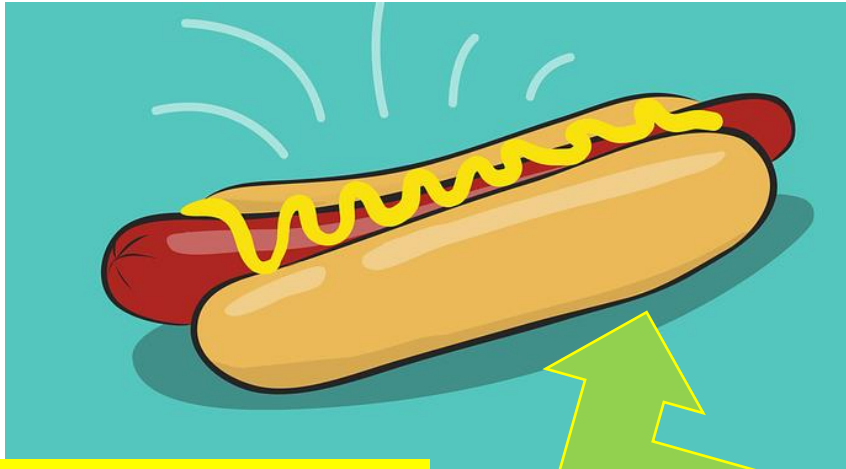


d. **Dados**



Identificadores Persistentes são a espinha dorsal dos princípios FAIR

O PID serve para **desambiguar** objetos digitais e para a localização única na web



PREFIX 1 > / < SUFFIX 2 >



PREFIX 1 > / < SUFFIX 3 >



PREFIX 1 > / < SUFFIX 2 >

Identificadores Persistentes são a espinha dorsal dos princípios FAIR

Pode ser resolvido por navegadores web, etc., são **acionáveis**.

<https://doi.org/10.25490/a97f-egyk>

Eles têm sintaxes diferentes mas o mesmo propósito

(< **PREFIX** > / < **SUFFIX** > (**11239/123456745**)

< **PREFIX** > / < **SUFFIX** > # < **FRAGMENT-ID** >

- DOI / Handle
- ORCID
- ROR
- RAID
- ...

What are persistent identifiers (PIDs)?

<https://support.orcid.org/hc/en-us/articles/360006971013-What-are-persistent-identifiers-PIDs>

Principais entregas

Plano de gestão de dados |

+

Coletar novos dados |

E

OU

Reusar dados |



Reuso de dados de pesquisa – melhores práticas

- Descobrir conjuntos de dados publicados em sua disciplina e a sua reutilização
- Elaborar uma estratégia para encontrar dados apropriados para um projeto de pesquisa específico
- Articular os critérios de seleção de dados – por exemplo, variáveis
- Reconhecer a proveniência dos dados que pretendem utilizar – repositórios confiáveis, estudos, metodologias na geração dos dados
- Identificar as licenças da reutilização de dados
- Fazer a correta citação ao reutilizar dados

Encontrar e reutilizar dados

**O primeiro
passo no
(re)uso de
dados é
encontrá-los**



Descoberta de dados

Desenvolva um planejamento dos dados de pesquisa que você precisa

- Qual é o tema/domínio que você estuda?
- Qual é a sua pergunta de pesquisa?
- Quais são os constructos com os quais você deseja trabalhar?

Localizar recursos de dados apropriados

- Você conhece os recursos de dados apropriados (ou sabe a quem perguntar)?
 - Repositórios de dados
 - Mecanismos de busca ou (meta)agregador de dados , exemplos:
 - Pesquisa Google Data Search
 - DataCite
 - OpenAIRE Schoexplorer

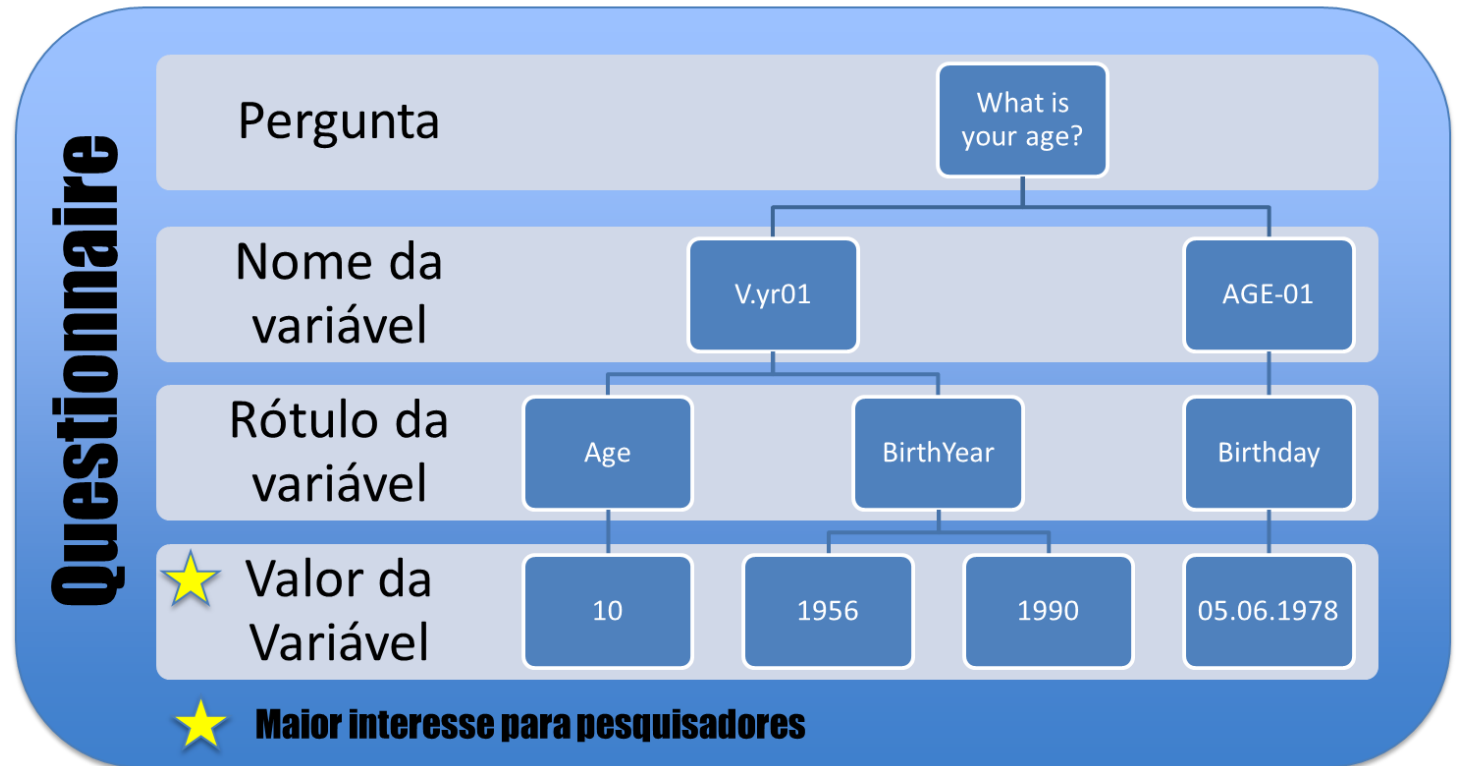
Pesquisar os recursos de dados

Selecione os candidatos de dados

Avaliar a qualidade dos dados

Dados de Pesquisa nas Ciências Sociais (surveys)

Estrutura de dados nas Ciências Sociais (mais comum de dados tabulados)



Nível de granularidade

Maior ambiguidade

PID

Nível da pesquisa

PID

Dataset

Códigos ou rótulos

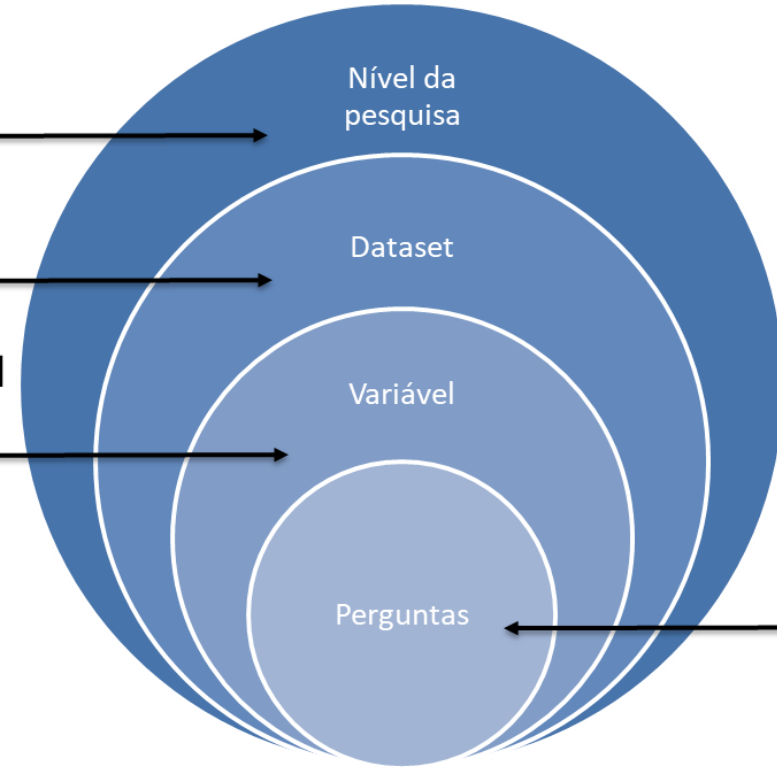
Variável

Perguntas

Códigos ou rótulos

Menor ambiguidade
Mais especificidade

Estrutura de dados nas Ciências Sociais (mais comum de dados tabulados)



Artigo com dados disponíveis



Preprint

This is a preprint, currently under review at *Quantitative Science Studies*. Please cite as:

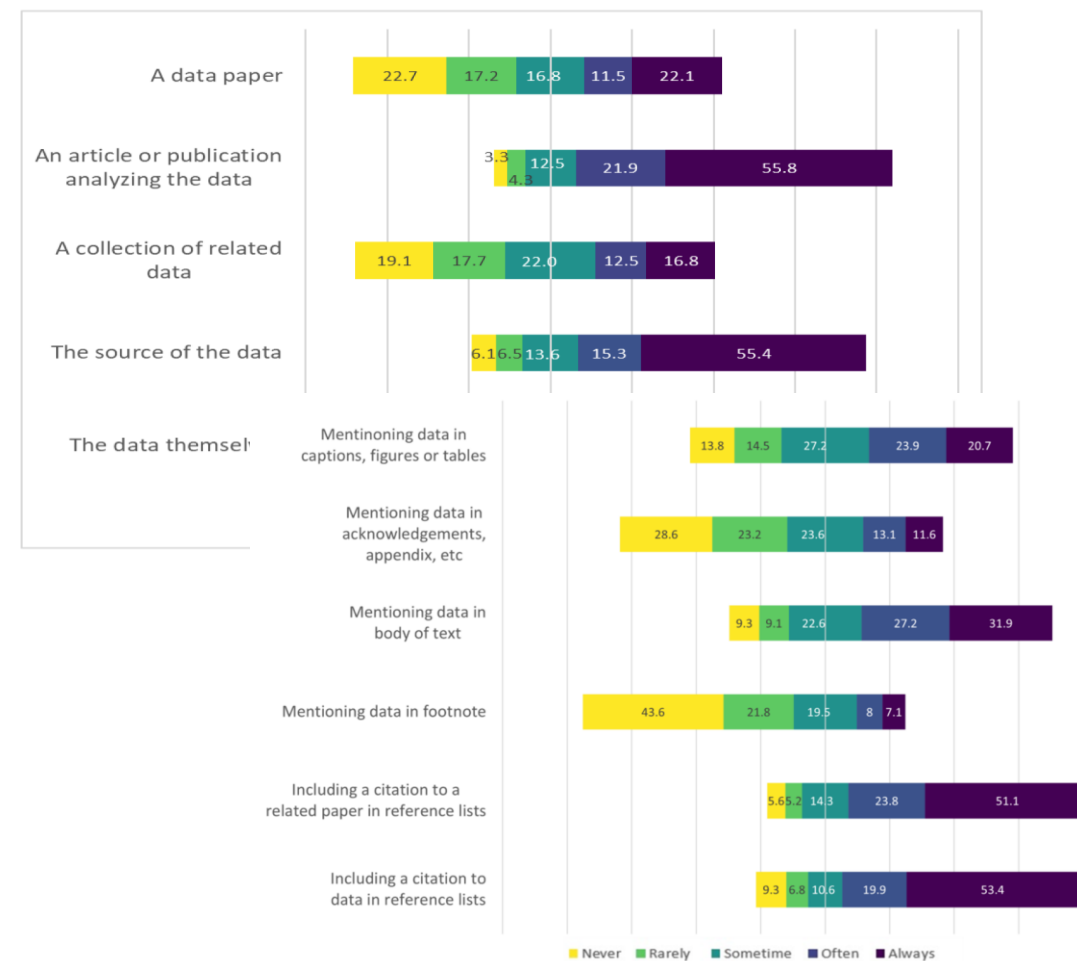
Gregory, K., Ninkov, A., Ripp, C., Roblin, E., Peters, I., & Haustein, S. (2023). *Tracing data: A survey investigating disciplinary differences in data citation*. Zenodo. <https://doi.org/10.5281/zenodo.7555266>

Tracing data: A survey investigating disciplinary differences in data citation

Kathleen Gregory¹, Anton Ninkov², Chantal Ripp³, Emma Roblin⁴, Isabella Peters⁵,
& Stefanie Haustein⁶

¹kathleen.gregory@uottawa.ca

2492 pesquisadores de várias áreas



Preprint (10.5281/zenodo.7555266) under review at Quantitative Science Studies

Documentação dos dados disponíveis



Codebook

Variable: Q26_FOS

Label: OECD Fields of Research and Technology classification

Question text: With which disciplinary domain do you most identify?

Variáveis de interesse

Q5.1	Q6.1
Q5.2	Q6.2
Q5.3	Q6.3
Q5.4	Q6.4
Q5.5	Q6.5
	Q6.6

460 pesquisadores das Ciências Sociais

Variable: Q26_FOS

Label: OECD Fields of Research and Technology classification

Question text: With which disciplinary domain do you most identify?

Response categories	Code	Frequency
Social sciences	5.00	3
Psychology	5.01	98
Economics and business	5.02	128
Education sciences	5.03	47
Sociology	5.04	51
Law	5.05	5
Political science	5.06	34
Social and economic geography	5.07	18
Media and communications	5.08	12
Other social sciences	5.09	64

N 460 2492

(18,5 %)

Localização dos dados disponíveis



Dataset

Variável do domínio

Texto da pergunta

Variable: Q26_FOS

Label: OECD Fields of Research and Technology classification

Question text: With which disciplinary domain do you most identify?

Variable: Q5.2

Label: The source of the data

Question text: When you reuse data, what do you usually cite or reference?

Note: The source of the data (e.g. a data repository or database)

Universe: Responded yes to Q1

Response categories	Code	Frequency
Never	1	135
Rarely	2	130
Sometimes	3	272
Often	4	307
Always	5	1123
N/A	6	59
Missing	999	466
N		2026

RespondentID	Q1	Q3	Q4.1	Q4.2	Q4.3	Q4.4	Q4.5	Q4.6	Q4.7	Q4.8	Q4.9	Q5.1	Q5.2	Q5.3
270278019	2	3	2	3	2	1	1	1	2	2	1	3	3	1
270278014	1	999	999	999	999	999	999	999	999	999	999	999	999	999
270283052	2	3	2	4	2	3	1	1	1	2	5	6	6	6
270278026	2	2	4	4	4	4	3	3	4	3	2	5	5	3
270278011	2	3	2	1	2	3	2	1	1	1	3	3	5	3
270283052	1	999	999	999	999	999	999	999	999	999	999	999	999	999
270278030	2	3	3	4	4	4	1	3	3	3	3	5	2	2
270278028	2	2	4	2	5	3	1	1	2	3	3	5	3	3
270278026	2	3	5	5	5	3	5	5	3	4	3	2	3	2

Total of 2,026 repostas válidas
460 pesquisadores das Ciências Sociais

Ninkov, Anton Boudreau, Ripp, Chantal, Gregory, Kathleen, Peters, Isabella, & Haustein, Stefanie. (2023). A dataset from a survey investigating disciplinary differences in data citation (Version v1) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.7555363>

Nome da variável

Descrição da variável

Código de respostas

Localização das variáveis

		G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	
1																				
2																				
	1	Q4.4	Q4.5	Q4.6	Q4.7	Q4.8	Q4.9	Q5.1	Q5.2	Q5.3	Q5.4	Q5.5	Q6.1	Q6.2	Q6.3	Q6.4	Q6.5	Q6.6	Q7_Q1!	Q8
	11	1	2	4	5	3	4	1	1	1	5	1	2	5	2	4	1	4	2	2
	13	2	2	1	1	1	2	1	3	3	5	1	2	5	2	2	2	2	2	2
	16	1	1	1	4	2	4	2	5	2	5	2	5	5	3	2	2	2	2	2
	17	1	1	3	2	2	3	3	3	2	3	1	4	3	3	3	2	2	2	2
	25	2	2	4	5	4	3	3	5	3	4	5	2	2	3	4	2	4	2	2
	29	3	3	5	5	4	5	4	5	4	5	4	5	5	2	4	3	3	2	2
	33	1	1	1	3	3	4	4	5	1	5	3	5	5	1	5	3	4	2	2
	39	1	3	3	4	4	4	5	5	5	5	5	5	5	2	4	2	4	2	2
	51	1	1	1	2	1	4	5	5	1	3	1	5	5	4	4	3	4	2	2
	52	1	1	4	4	4	3	5	5	5	5	3	4	4	5	5	4	4	2	2
	59	4	4	3	2	2	3	5	5	5	5	4	4	5	1	1	2	2	2	2
	64	2	2	2	3	4	3	5	5	2	4	2	5	5	4	4	2	3	2	2
	66	2	1	4	4	4	4	1	5	1	5	1	5	5	3	3	3	3	2	2
	70	2	1	1	3	2	5	3	3	2	5	3	2	3	2	1	1	2	2	2
	79	4	1	5	5	5	5	3	5	3	5	1	5	5	2	5	2	4	2	2
	81	4	2	2	3	4	4	4	4	2	2	1	5	3	1	5	3	4	2	2
	84	2	2	5	5	5	5	4	5	5	3	2	5	5	4	2	2	2	2	2

Contagem de frequência

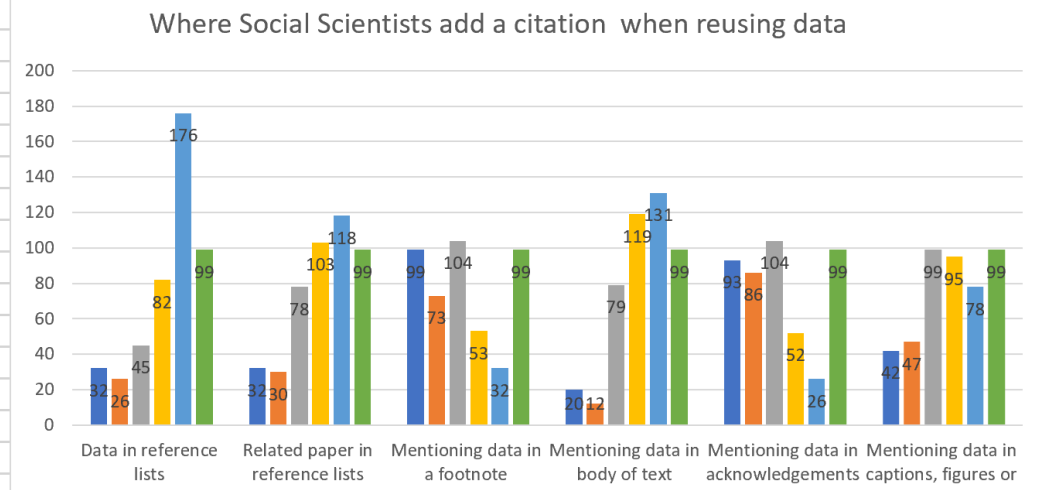
Variáveis de interesse

- Q5.1 Q6.1
- Q5.2 Q6.2
- Q5.3 Q6.3
- Q5.4 Q6.4
- Q5.5 Q6.5
- Q6.6

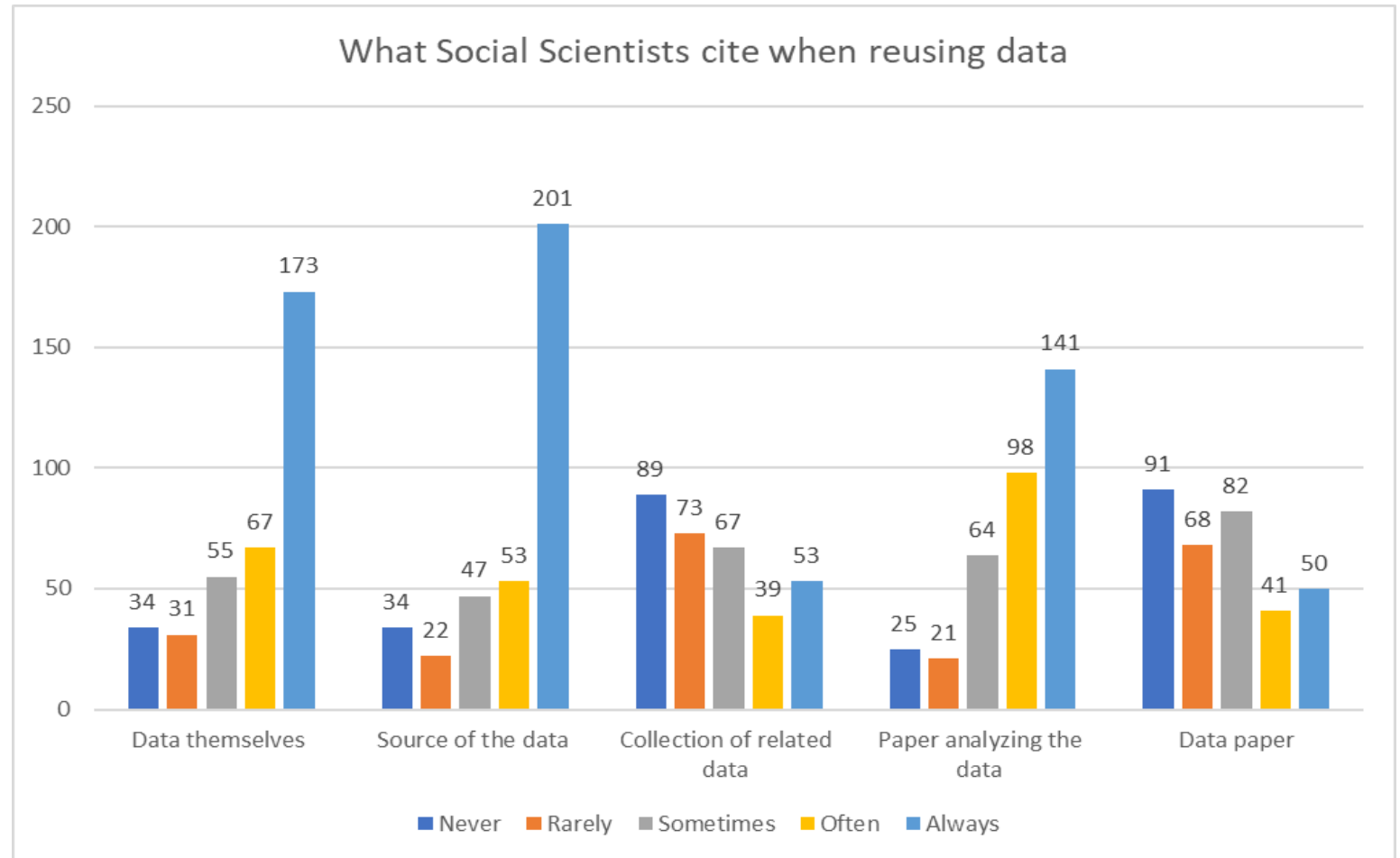
M	N	O	P	Q
Q5.1	Q5.2	Q5.3	Q5.4	Q5.5
1	1	1	5	1
1	3	3	5	1
2	5	2	5	2
3	3	2	3	1
3	5	3	4	5
4	5	4	5	4
4	5	1	5	3

R	S	T	U	V	W
Q6.1	Q6.2	Q6.3	Q6.4	Q6.5	Q6.6
2	5	2	4	1	4
2	5	2	2	2	2
5	5	3	2	2	2
4	3	3	3	2	2
2	2	3	4	2	4
5	5	2	4	3	3

	A	B	C	D	E	F	G	H	J	K	L	M	N	O	P
	Data in reference lists	Related paper in reference lists	Mentioning data in a footnote	Mentioning data in body of text	Mentioning data in acknowledgements	Mentioning data in captions, figures or tables		Response Schema	Data in reference lists	Related paper in reference lists	Mentioning data in a footnote	Mentioning data in body of text	Mentioning data in acknowledgements	Mentioning data in captions, figures or tables	
1															
2	2	5	2	4	1	4		Never	32	32	99	20	93	42	
3	2	5	2	2	2	2		Rarely	26	30	73	12	86	47	
4	5	5	3	2	2	2		Sometime	45	78	104	79	104	99	
5	4	3	3	3	2	2		Often	82	103	53	119	52	95	
6	2	2	3	4	2	4		Always	176	118	32	131	26	78	
7	5	5	2	4	3	3		Missing	99	99	99	99	99	99	
8	5	5	1	5	3	4			0	0	0	0	0	0	
9	5	5	2	4	2	4									
10	5	5	4	4	3	4									
11	4	4	5	5	4	4									
12	4	5	1	1	2	2									
13	5	5	4	4	2	3									
14	5	5	3	3	3	3									
15	2	3	2	1	1	2									
16	5	5	2	5	2	4									
17	5	3	1	5	3	4									
18	5	5	4	2	2	2									
19	5	4	3	5	4	5									
20	999	999	999	999	999	999									
21	999	999	999	999	999	999									
22	5	1	1	5	2	2									
23	1	2	2	3	1	3									
24	2	5	1	5	1	1									
25	2	4	2	4	3	4									
26	5	1	1	5	1	3									
27	2	5	2	2	2	2									
								Sum	460	460	460	460	460	460	460

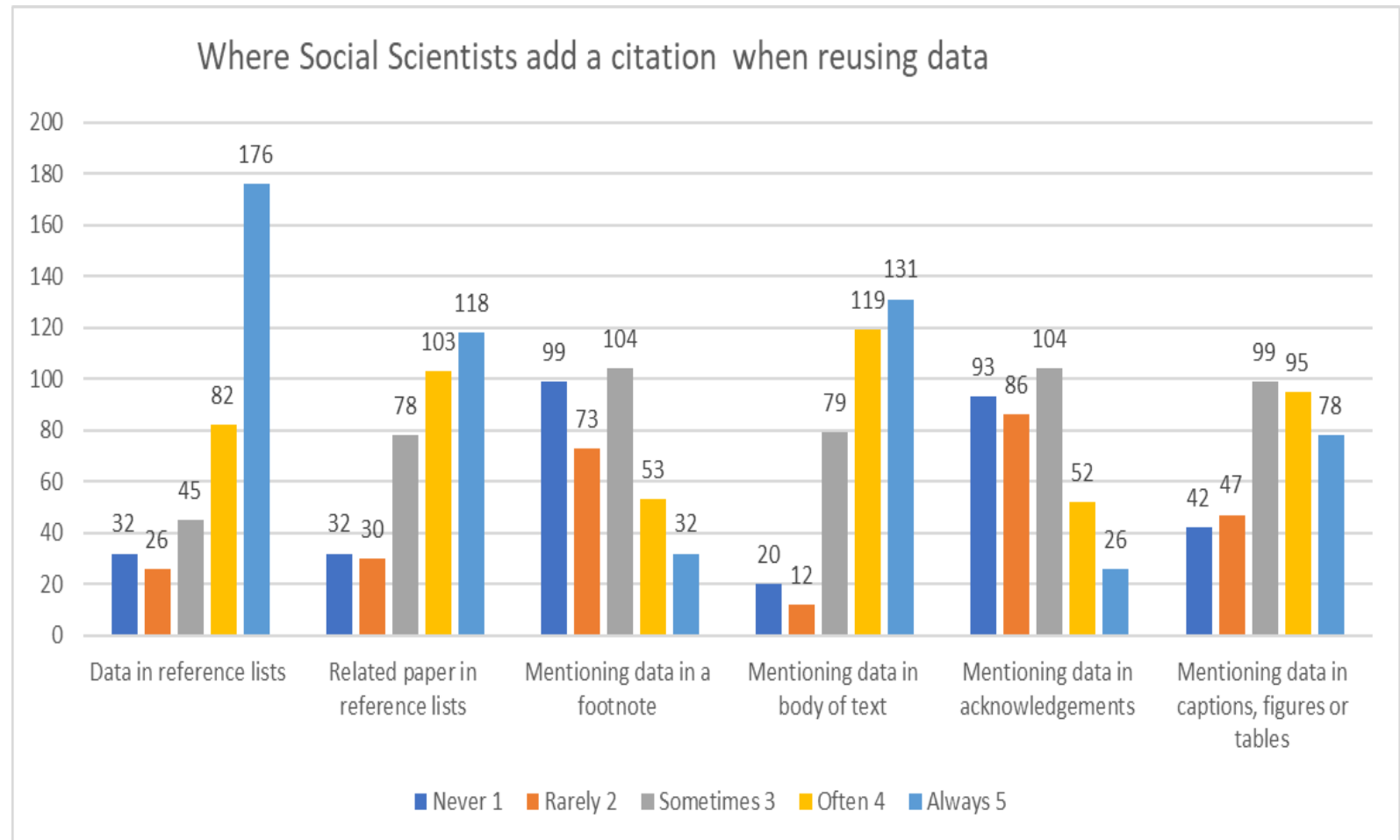


Apresentação dos resultados



Bach, Janete Saldanha, Klas, Claus-Peter, & Mutschke, Peter. (2023, March 28). Use cases and benefits of persistent identifiers for dataset elements to foster reliable data citation. 9th Conference on Social and Economic Data - 2023 (KSWD 2023), Berlin, Germany. Zenodo. <https://doi.org/10.5281/zenodo.7741688>

Apresentação dos resultados



Bach, Janete Saldanha, Klas, Claus-Peter, & Mutschke, Peter. (2023, March 28). Use cases and benefits of persistent identifiers for dataset elements to foster reliable data citation. 9th Conference on Social and Economic Data - 2023 (KSWD 2023), Berlin, Germany. Zenodo. <https://doi.org/10.5281/zenodo.7741688>



**Localizar, acessar
e obter os dados**

Repositórios e indexadores

Registry of Research Data Repositories <https://www.re3data.org>

- O re3data.org é um (indexador) registro global de repositórios de dados de pesquisa que abrange repositórios de dados de pesquisa de diferentes disciplinas acadêmicas.

Zenodo <https://zenodo.org/>

- O Zenodo é um repositório de dados de pesquisa de acesso aberto de propósito geral, hospedado pelo CERN (Suíça) que fornece um lugar para os pesquisadores depositarem conjuntos de dados. Pesquisadores em qualquer área temática podem fazer upload de arquivos de até 50 GB. Ele tem uma integração com GitHub para tornar o código hospedado no GitHub editável.

Harvard Dataverse <https://dataverse.harvard.edu/dataverse.xhtml>

- Dataverse é um repositório de dados que é amplamente utilizado dentro das Ciências Sociais. Os dados podem ser encontrados através da aplicação de esquemas de metadados específicos de disciplina e identificadores de objetos digitais (DOIs). Os dados se tornam reutilizáveis especificando as licenças de reutilização relevantes.

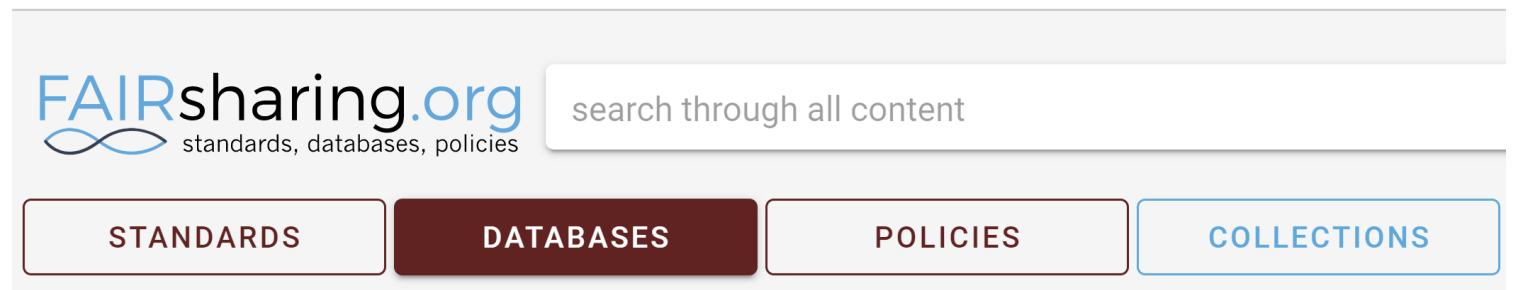


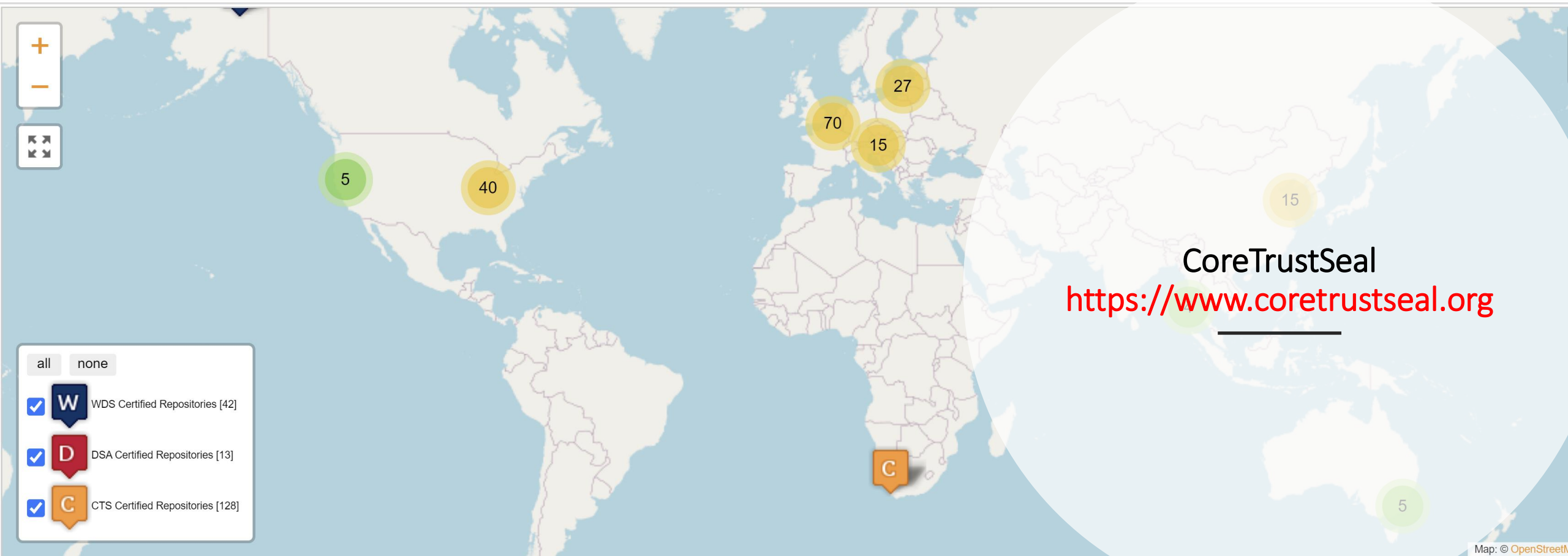
Registry of Research Data
Repositories

<https://www.re3data.org>



[https://fairsharing.org/search?fairs
haringRegistry=Database](https://fairsharing.org/search?fairs
haringRegistry=Database)

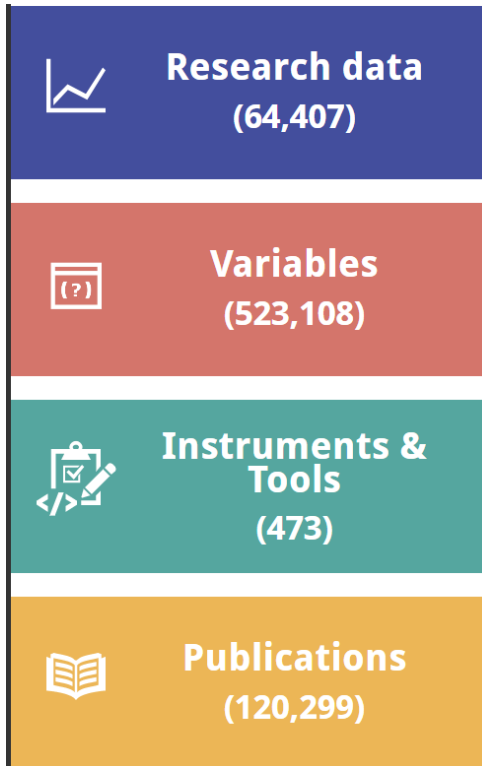




CoreTrustSeal
<https://www.coretrustseal.org>

all none

- WDS Certified Repositories [42]
- DSA Certified Repositories [13]
- CTS Certified Repositories [128]



gesis
Leibniz-Institut
für Sozialwissenschaften

GESIS Search
(<https://search.gesis.org>)

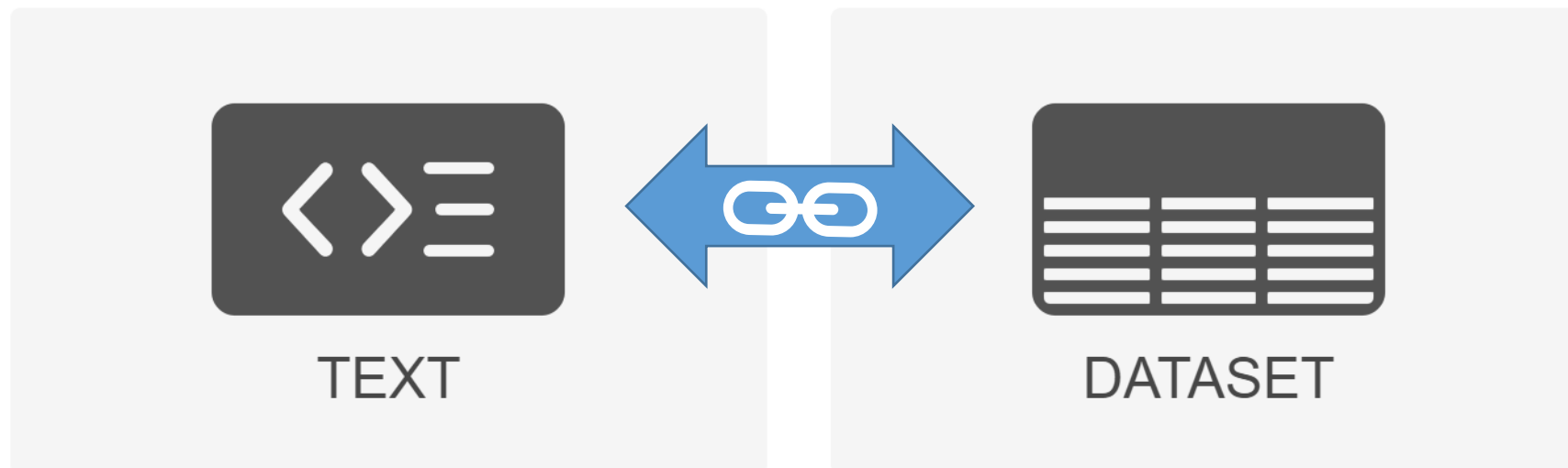
- **GESIS – Leibniz Institute for the Social Sciences**

OpenAIRE Scholexplorer

(<http://scholexplorer.openaire.eu/#/>)



- Implementação do API Scholix para intercâmbio de links acadêmicos

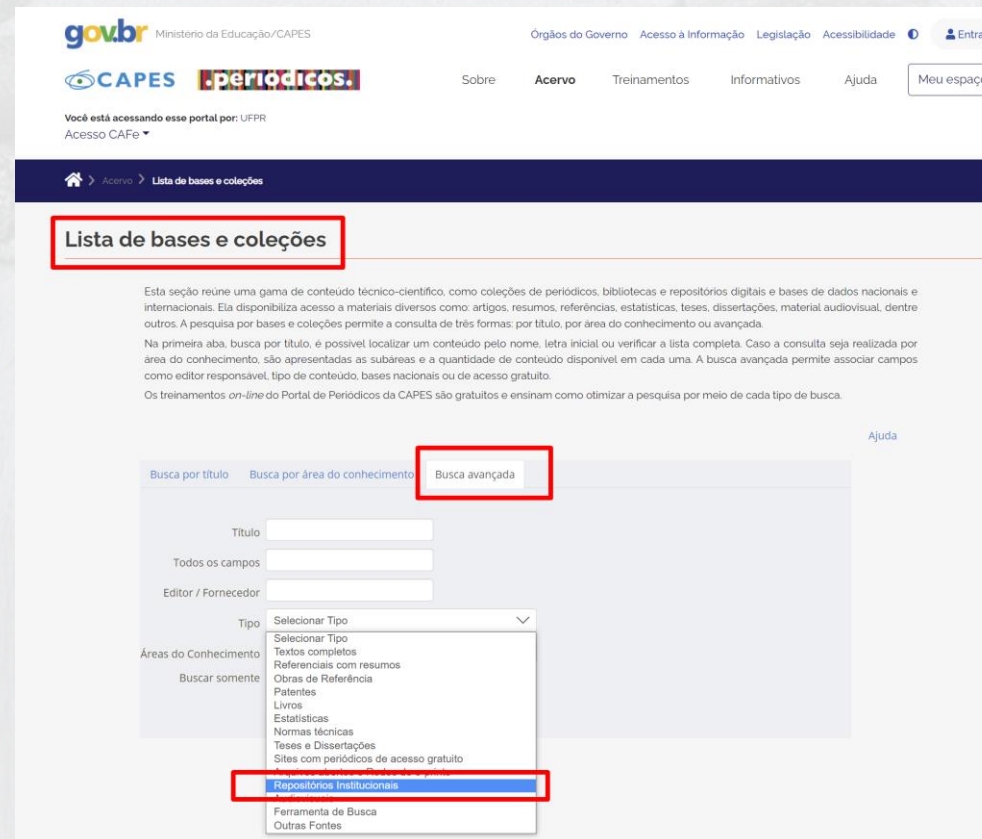


SCHOLIX

Repositórios e indexadores

Portal de periódicos da Capes, lista de repositórios institucionais

- <https://www-periodicos-capes-gov-br>



The screenshot shows the website interface for the CAPES Periodicos portal. At the top, there is a navigation bar with the logo 'gov.br' and 'Ministério da Educação/CAPES'. Below it, the 'CAPES PERIÓDICOS.' logo is displayed. The main content area features a dark blue header with the text 'Lista de bases e coleções' highlighted in a red box. Below this, there is a paragraph of text describing the section's purpose and a search interface. The search interface includes three tabs: 'Busca por título', 'Busca por área do conhecimento', and 'Busca avançada', with 'Busca avançada' highlighted in a red box. The search form contains fields for 'Título', 'Todos os campos', and 'Editor / Fornecedor'. A dropdown menu for 'Tipo' is open, showing various search options, with 'Repositórios Institucionais' highlighted in a red box. The page also includes a 'Ajuda' link in the top right corner.

Repositórios e indexadores



Portal Brasileiro de Publicações
e Dados Científicos em Acesso Aberto

Portal Brasileiro de Publicações e dados científicos em acesso aberto

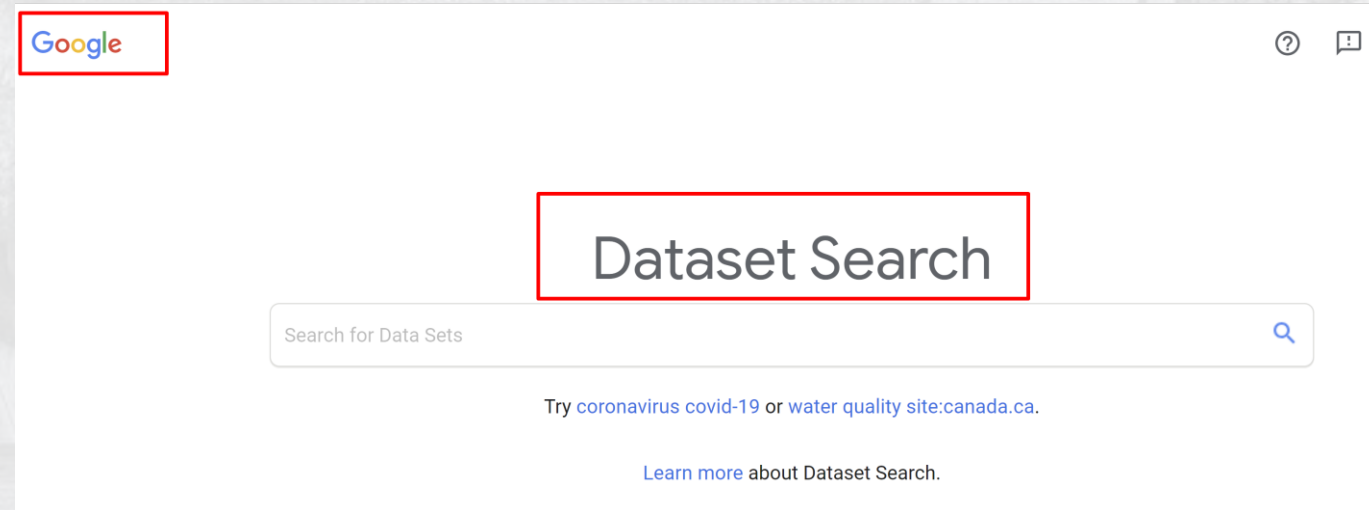
- <https://oasisbr.ibict.br/vufind/Search/Advanced>.

The screenshot shows the advanced search page of the Oasisbr portal. At the top, there is a navigation bar with links for 'gov.br', 'CORONAVÍRUS (COVID-19)', 'ACESSO À INFORMAÇÃO', 'PARTICIPE', 'LEGISLAÇÃO', and 'ÓRGÃOS DO GOVERNO'. Below this is the 'ibict' logo and the text 'Instituto Brasileiro de Informação em Ciência e Tecnologia', along with a secondary navigation bar containing 'INÍCIO', 'SOBRE', 'FONTES COLETADAS', 'INDICADORES', and 'CONTATO'. The main content area is titled 'Busca / Avançada'. On the left, under 'Dicas de Busca', there are two links: 'Ajuda com a busca avançada' and 'Ajuda com Operadores de busca'. The central 'Busca avançada' section contains three search input fields, each with a 'Todos os campos' dropdown menu and a close button. To the right, there is a 'correspondência da busca:' section with a 'TODOS os termos' dropdown. Below the search fields are two links: 'Adicionar campo de busca' and 'Adicionar outros campos de busca'. A green 'Buscar' button and a 'Limpar' button are positioned at the bottom of the search area. The 'Limitar a' section includes two dropdown menus: 'Idioma:' with a list of languages (Albanês, Alemão, Alemão, Bósnio, Catalão, Chinês, Chinês, ENG, enk, Eslovaco) and 'Tipo de documento:' with a list of document types (Artigo, Artigo (review), Artigo (working paper), Artigo de conferência, Capítulo de livro, Conjunto de dados, Dissertação, Livro, Outros, Relatório). The 'Conjunto de dados' option is highlighted with a red box.

Repositórios e indexadores

Google Dataset Search

(<https://datasetsearch.research.google.com>)



Onde encontrar dados Oficiais

Instituto Brasileiro de Geografia e Estatística IBGE. Site: <https://www.ibge.gov.br>

Instituto de Pesquisa Econômica Aplicada (IPEA). Site: <https://www.ipea.gov.br>

DataSUS do Ministério da Saúde. Site: <http://datasus.saude.gov.br>

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Site: <http://www.inep.gov.br>

Portal Brasileiro de Dados Abertos. Site: <https://dados.gov.br/home>

Onde encontrar dados

Portais internacionais

ICPSR (<https://www.icpsr.umich.edu/web/pages/ICPSR/index.html>)

GESIS Search (<https://search.gesis.org>)

B2FIND data search (<http://b2find.eudat.eu>)

BASE (<https://www.base-search.net>)

DataCite Metadata Search (<https://search.datacite.org>)

Europäisches Datenportal (<https://www.europeandataportal.eu/de>)

Forschungsdaten.info Social and Behavioral Sciences

<https://www.forschungsdaten.info/wissenschaftsbereiche/sozial-und-verhaltenswissenschaften/>

Acesso aos dados

Nem todos os dados
estão disponíveis,
ainda que sejam
encontráveis

A ideia de OPEN é *as
open as possible, as
closed as necessary*



Disponibilidade

- Uma vez que o usuário encontre os dados necessários, precisa saber como eles podem ser acessados, possivelmente incluindo **autenticação** e autorização.
- Nem todos os dados estão acessíveis, mas os metadados devem estar.





Categorias de acesso

As categorias de acesso aos dados podem ser:

- Acesso aberto
- Acesso restrito
- Embargo
- Acesso fechado

Categorias de acesso

- Acesso aberto **exige citação**
- **Restrições** de acesso podem exigir um **contrato ou acordo de compartilhamento** de dados a ser assinado.
- **Embargo** significa que o acesso é **fechado temporariamente**.
- O acesso fechado significa que os dados **não são acessíveis**, exceto talvez aos órgãos reguladores.

Regras de acesso x uso

Categoria 0 liberados para todos.

Categoria A para pesquisa e ensino acadêmico.

Categoria B pesquisa acadêmica e ensino com permissão para publicação.

Categoria C Somente com autorização por escrito do depositante.

Category 0	Data and documents are released for everybody.
Category A	Data and documents are released for academic research and teaching.
Category B	Data and documents are released for academic research and teaching, if the results are not published. If any publication or further processing of the results is planned, permission must be obtained by the Data Archive.
Category C	Data and documents are only released for academic research and teaching after the data depositor's written authorization. For this purpose the Data Archive obtains written permission with specification of the user and the analysis intention.

Restrições de acesso podem ocorrer em função de:

- Presença de informações **personais** não anonimizadas;
- Dados **sensíveis**, como posto de vista **políticos**, atividades **religiosas**;
- Identificação de **biodiversidade**, por exemplo, a localização de espécies raras e ameaçadas de **extinção**;
- Sujeitos à **Propriedade intelectual** e comercial;
- **Acordo de confidencialidade**, onde o acesso e compartilhamento de dados é restrito para as partes contratantes

Busque uma imagem no google e filtre por tipo de licença

← → ↻ 🏠 google.com/search?q=research+data&tbm=isch&ved=2ahUKEwipybemueD5AhXZNOwKHQSPCNEQ2-cCegQIABAA&oq=...

Google 🔊 📷 🔍

🔍 Todas **🖼️ Imagens** 📰 Notícias 📍 Maps 📺 Vídeos ⋮ Mais Ferramentas

Tamanho ▾ Cor ▾ Tipo ▾ Tempo ▾ **Direitos de uso ▾**

- ✓ Todas
- Licenças Creative Commons**
- Licenças comerciais e outras

nfdi data analysis practical sustainable sustainable research journal springe

A circular diagram showing the Research Data Management cycle. At the center is 'Researcher'. Surrounding it are 'Planning', 'Collecting', 'Analyzing', 'Publishing', 'Presenting & Sharing', and 'Re-using'. Arrows connect these stages in a clockwise direction. External entities like 'Funders', 'Research Community', 'Repositories', and 'Publishers' are also indicated.

A Guide to Research Data Management - Lab...
labfolder.com

A pyramid diagram titled 'SUCCESSFUL DATA'. The pyramid is divided into three horizontal layers: 'Reusable' (top), 'Reproducible' (middle), and 'Reviewed' (bottom). To the left of the pyramid is a thumbs-up icon and the word 'Trusted'.

Research Data | Open Science | Elsevier
elsevier.com

A photograph of a person sitting at a desk with a laptop. The laptop screen displays the word 'DATA' with several arrows pointing to the right. There are books on the desk next to the laptop.

5 ways sharing your research data could help enhance ...
springernature.com

A circular diagram illustrating the Research Data lifecycle. The cycle includes stages: 'CREATING DATA', 'ANALYSING DATA', 'PRESERVING DATA', 'GIVING ACCESS TO DATA', and 'RE-USING DATA'. Each stage has a brief description of what it involves. The cycle is supported by 'Existing 3rd party data' and 'Patent data'.

Research Data | AIMS
aims.fao.org

Vantagens do reuso de dados

Encontrar e utilizar dados:

Pronk, TE. 2019. The Time Efficiency Gain in Sharing and Reuse of Research Data. *Data Science Journal*, 18: 10, pp. 1–8. DOI: <https://doi.org/10.5334/dsj-2019>

- Reutilizar dados economiza recursos (<http://doi.org/10.5334/dsj-2019-010>)
- Se os dados adequados estiverem disponíveis, há menos necessidade de gastar tempo e dinheiro para coletar material novo
- Os dados de grandes pesquisas frequentemente incluem material que não foi analisado na pesquisa original
- Reutilizar dados ajuda a evitar a duplicação da coleta de dados
- Reutilizar dados pode minimizar a coleta em pessoas de difícil acesso ou vulneráveis

Reuso de
dados é
reuso de
conhecimento

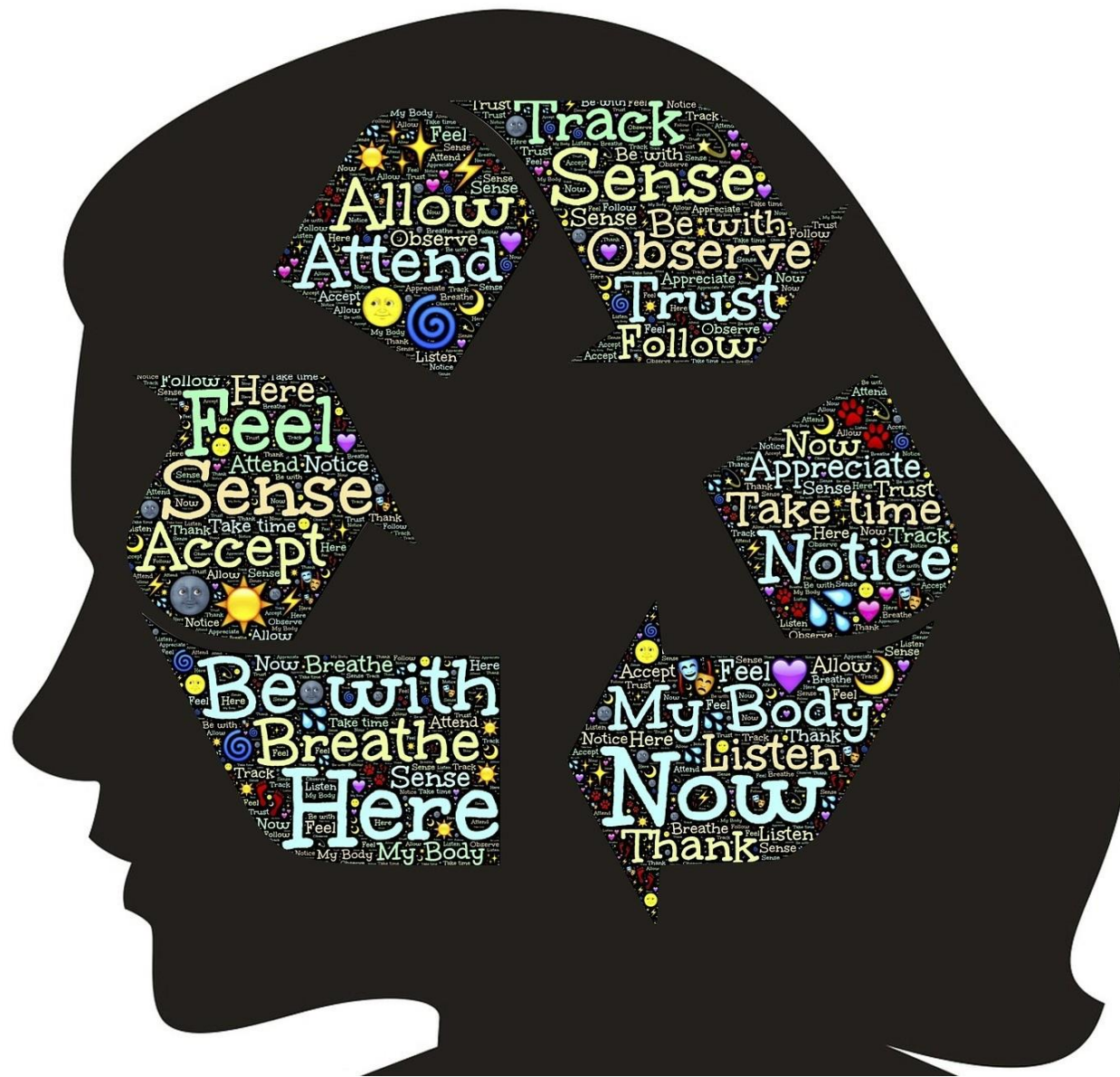


Image by [John Hain](#) from [Pixabay](#)



Obrigada

CONTATO:

Janete Saldanha Bach Estevão
janeteest@gmail.com

REFERENCE:

ESTEVÃO, Janete Saldanha Bach. Gestão de dados científicos de pesquisa em Ciências Sociais Aplicadas. In: *Seminário de integração e dicas instrumentais de pesquisa em Ciências Sociais Aplicadas*, V. 2023. 89 slides. DOI: 10.5281/zenodo.7795068

LICENSE:

Gestão de dados científicos de pesquisa em Ciências Sociais Aplicadas by Janete Saldanha Bach Estevão is licensed under a Creative Commons Attribution 4.0 International Licence.

