



Preparar dados de investigação para publicar no **REPOSITÓRIO DE DADOS** Abertura e partilha de Dados

TÓPICOS & OBJECTIVOS

1. **Gestão e abertura de dados** – boas práticas e requisitos no contexto da Ciência Aberta.
2. **Preparar dados para publicação** – dicas sobre organização e documentação dos dados.
3. **Repositórios e publicação de dados** – conhecer as ferramentas para partilha de dados.



Ciência Aberta

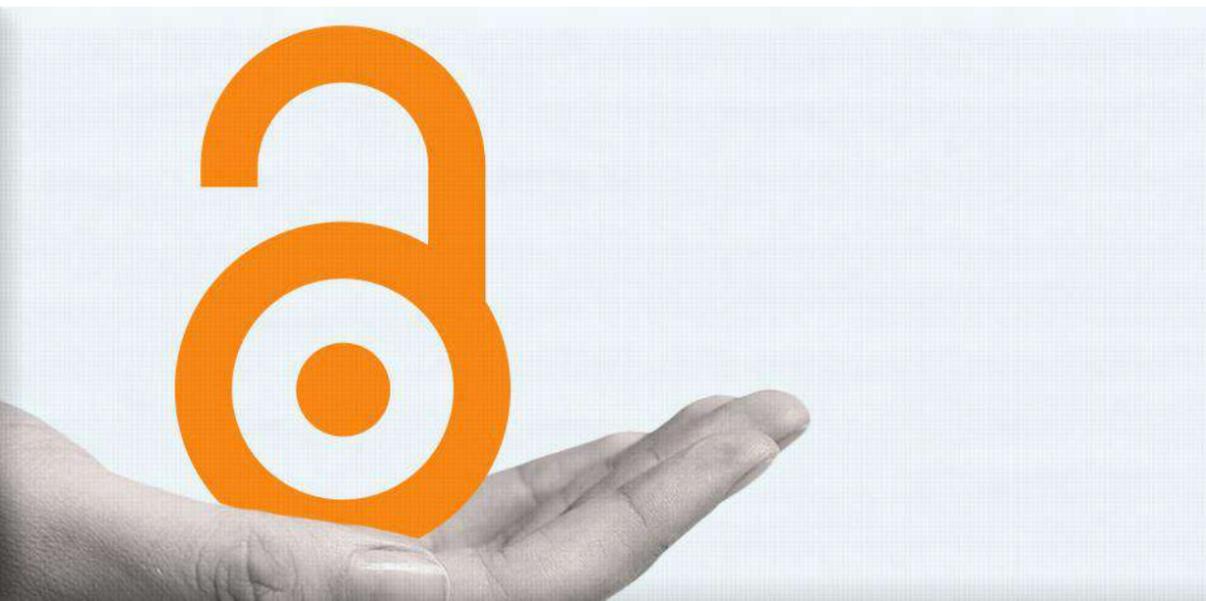
Nos resultados

- Publicações
- Dados
- ...

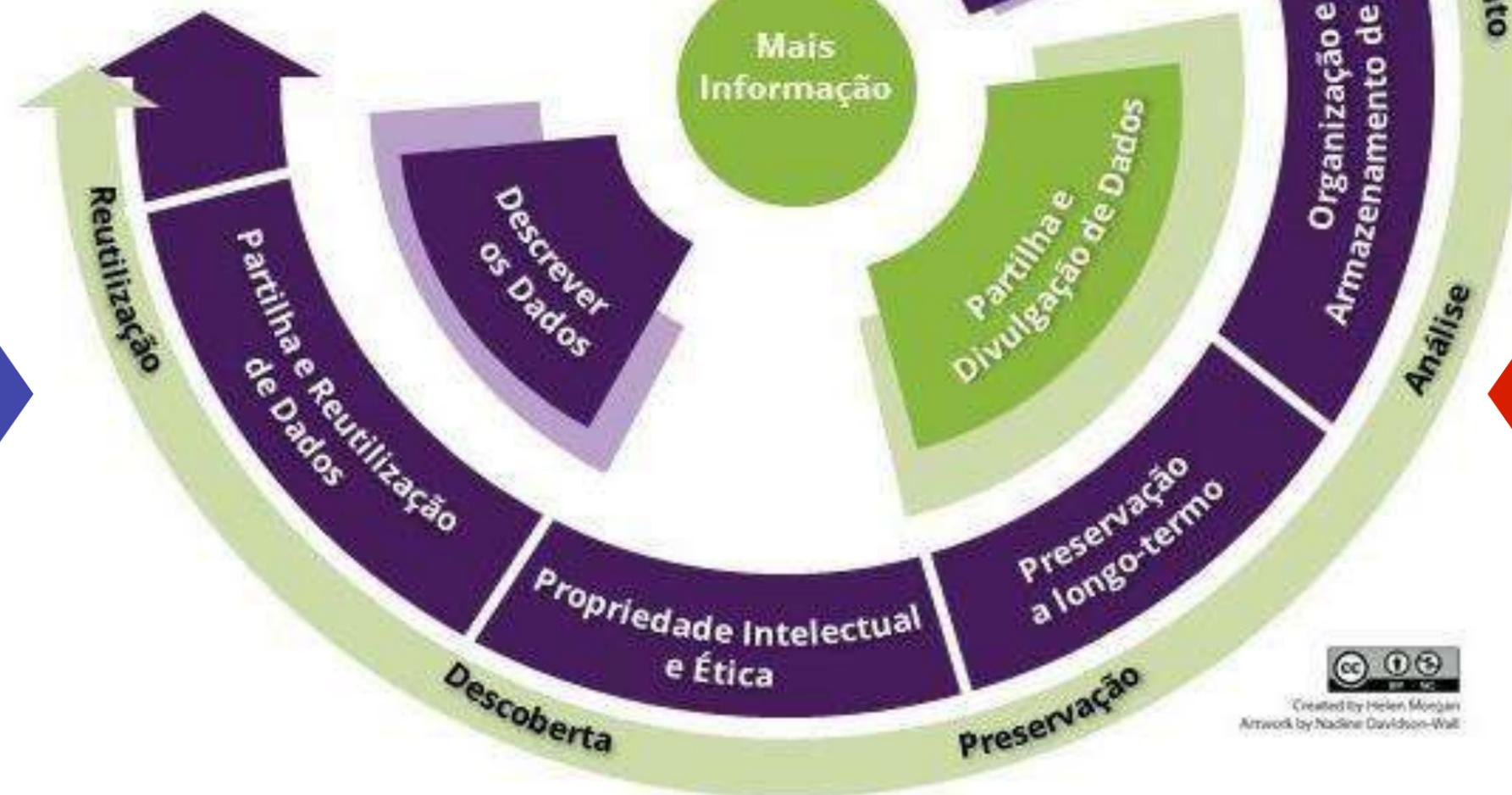
Nos métodos

- Processos
- Ferramentas
- Infraestruturas

Para tornar a ciência mais eficiente, transparente, confiável e reprodutível.



Ciclo de Vida de Dados de Investigação



ABERTURA DE DADOS

GESTÃO DE DADOS

Dados...



“Gestão de Dados responsável é parte da boa investigação”
NWO - Introduction to the pilot Data Management

GESTÃO E ABERTURA DE DADOS DE INVESTIGAÇÃO

Razões?



Tornar o processo de investigação mais fácil



Obter ganhos e créditos por isso



Salvaguardar os dados



Evitar acusações de fraude ou má ciência



Partilhar os dados para reutilização



Cumprir com os requisitos de financiadores



OPEN RESEARCH DATA IN HORIZON 2020

CHALLENGE

Wider access to scientific facts and knowledge helps researchers, innovators and the public find and re-use data, and check research results:

offers better value for EU research funds

encourages research across scientific fields



a public benefit



essential for solving today's complex societal challenges

SOLUTION

Horizon 2020 already mandates open access to all scientific publications



From 2017, research data is open by default, with possibilities to opt out

RESEARCH DATA - OPEN BY DEFAULT

HORIZON 2020 GRANTEEES ARE REQUIRED



INÍCIO SOBRE CIÊNCIA ABERTA PNCA

RESOLUÇÃO DE CONSELHO DE MINISTROS | PRINCÍPIOS ORIENTADORES

Política Nacional de Ciência



Home Funding Research Innovation Skills Public Engagement New

RCUK Common Principles on Data Policy

Home > Research > RCUK Common Principles on Data Policy

Making research data available to users is a core part of the Research Councils' remit and is undertaken in a variety of coherent approaches across the research base. These RCUK common principles on data policy provide an overarching data policy.

Principles

- Publicly funded research data are a public good, produced in the public interest, which should be made available in a transparent and responsible manner.
- Institutional and project specific data management policies and plans should be in accordance with the principle that the long-term value should be preserved and remain accessible and usable.
- To enable research data to be discoverable and effectively re-used by others, sufficient information should be provided to researchers to understand the research and re-use potential of the data. Published research data should be made available in a format that is suitable for re-use.
- RCUK recognises that there are legal, ethical and commercial constraints on release of research data. Where necessary, appropriate release of data, research organisation policies and practices should ensure that these constraints are met.



Grants & Funding NIH's Central Resource for Grants and Funding Information

Entire Site Search this Site

HOME ABOUT GRANTS FUNDING POLICY & COMPLIANCE NEWS & EVENTS

Home > Policy & Compliance > Policy & Guidance > NIH Data Sharing Policy and Implementation Guidance

Policy & Compliance

THE ROYAL SOCIETY

Home Fellows Events Grants, Schemes & Awards Topics & policy Journals Collections

Data sharing and mining

Open data policy

To allow others to verify and build on the work published in Royal Society journals, it is a condition of publication that authors make available the data, code and research materials supporting the results in the article.

Datasets and code should be deposited in an appropriate, recognised, publicly available repository. Where no data-specific repository exists, authors should deposit their datasets in a general repository such as [Dryad](#) or [Figshare](#).

To encourage best practice in data sharing, *Biology Letters*, *Proceedings B* and *Royal Society Open Science* have Dryad data deposition integrated into the journal submission system. For all its science journals, the Society will cover the cost of depositing up to 20GB of data with Dryad. In addition, we deposit all supplementary material into the Figshare repository on the author's behalf.

Exceptions to the sharing of data, code and materials may be granted at the discretion of the editor, especially for sensitive information such as human subject data or the location of endangered species. Authors must disclose upon submission of the manuscript any restrictions on the availability of data, code and research materials.



Funding What we do About us News

Scheme finder Managing a grant

Policy on data, software and materials management and sharing

As a charity, Wellcome works to ensure that the results of the research we fund are made available to the public in a way that maximises their long-term value. Key documents such as the [UK concordat on open research data \(2016\)](#) articulate this.

Um grande número de financiadores apresenta requisitos para a abertura e gestão de dados que resultem de projetos financiados

- Proprietary Data
- Methods for Data Sharing
- Data Documentation
- Funds for Data Sharing
- Review Considerations

There is international consensus on the need to share and preserve research datasets in a way that maximises their long-term value. Key documents such as the [UK concordat on open research data \(2016\)](#) articulate this.

Requisitos de Dados Abertos no H2020 e HE



1. Plano de Gestão de Dados



2. Depositar Dados em Repositórios confiáveis



3. Documentar os dados para validação



4. Assegurar abertura dos dados



**Horizon Europe (HORIZON)
Euratom Research and Training Programme
(EURATOM)**

General Model Grant Agreement

(HE MGA — Multi & Mono)

Version 1.0 DRAFT
25 February 2021

Disclaimer
This document is aimed at assisting applicants. It shows the full range of provisions that may be applied to this type of agreement, and is provided for information purposes only. The legally binding agreement will be that which is signed by the parties in the system.



Open science: research data management

The beneficiaries must manage the digital research data generated in the action ('data') responsibly, in line with the FAIR principles and by taking all of the following actions:

- establish a data management plan ('DMP') (and regularly update it)
- as soon as possible and within the deadlines set out in the DMP, deposit the data in a trusted repository; if required in the call conditions, this repository must be federated in the EOSC in compliance with EOSC requirements
- as soon as possible and within the deadlines set out in the DMP, ensure open access — via the repository — to the deposited data, under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC 0) or a licence with equivalent rights, following the principle 'as open as possible as closed as necessary', unless providing open access would in particular:
 - be against the beneficiary's legitimate interests, including regarding commercial exploitation, or
 - be contrary to any other constraints, in particular the EU competitive interests or the beneficiary's obligations under this Agreement; if open access is not provided (to some or all data), this must be justified in the DMP
- provide information via the repository about any research output or any other tools and instruments needed to re-use or validate the data.

Metadata of deposited data must be open under a Creative Commons Public Domain Dedication (CC 0) or equivalent (to the extent legitimate interests or constraints are safeguarded), in line with the FAIR principles (in particular machine-actionable) and provide information at least about the following: datasets (description, date of deposit, author(s), venue and embargo); Horizon Europe or Euratom funding; grant project name, acronym and

**Establish a DMP &
regularly update it**

**Ensure open data
via repository**

**As open as possible
as closed as necessary**

**Metadata CC 0 in
line with FAIR**

POLÍTICA DE GESTÃO E PARTILHA DE DADOS FCT



PRINCÍPIOS

- 1** Os dados resultantes de investigação financiada pela FCT **devem ser disponibilizados em acesso aberto, admitindo-se exceções devidamente justificadas**, de acordo com o princípio “tão aberto quanto possível, tão fechado quanto necessário”.
- 2** **A gestão dos dados de investigação é um requisito e uma obrigação dos beneficiários de financiamento** de I&D da FCT.
- 3** Os dados resultantes de investigação financiada pela FCT **devem ser geridos de acordo com os princípios FAIR**, para que sejam localizáveis, acessíveis, interoperáveis e reutilizáveis.



REQUISITOS

- 1** Os instrumentos de financiamento aprovados e financiados, devem submeter **nos primeiros 6 meses**, após a data de início, um **Plano de Gestão de Dados de Investigação**, elaborado com **base no modelo disponibilizado pela FCT**.
- 2** Depositar os dados resultantes da investigação e projeto num **repositório de dados de investigação confiável e reconhecido pela FCT**.



REQUISITOS

- 3** Fornecer informações - através do repositório - sobre as **ferramentas e instrumentos à disposição e necessários para a validação dos resultados.**
- 4** Incluir **referência ao financiamento** de acordo com as especificações previstas.
- 5** **A título excepcional**, os beneficiários não têm de garantir o **acesso aberto a partes específicas dos seus dados de investigação.**

PUBLICAR DADOS



Universidade do Minho
Serviço de Documentação e Bibliotecas

GARANTIR QUE OS DADOS TÊM UM PAPEL + PROEMINENTE NOS RESULTADOS DE INVESTIGAÇÃO

- Asseguro boas práticas de gestão de dados (armazenamento e documentação)
- Deposito dados em repositórios confiáveis com identificadores persistentes.
- Associo dados às publicações.
- Publico os resultados dos meus projetos em coleções organizadas e para reuso.

nature materials

Editorial | Published: 18 December 2019

Data take centre stage

Nature Materials 19, 1(2020) | Cite this article

999 Accesses | 42 Altmetric | Metrics

We are updating our editorial policies to further encourage authors to make their data publicly accessible. Publishing Extended Data figures and source data online will also ensure that data are given a more prominent role.

A Data availability statement for accepted articles has been requested by the Nature Research journals for the past three years¹. In this statement, our authors declare how the data behind their published research can be accessed by interested readers, and disclose any potential restriction limiting data sharing. This initiative, aiming at increasing the reproducibility of our papers, has been seamlessly accepted by researchers, who have promptly added this declaration to their manuscripts submitted to the Nature titles. However, in the vast Data take centre stage. *Nat. Mater.* 19, 1 (2020). <https://doi.org/10.1038/s41563-019-0574-2>



OPEN RESEARCH DATA IN HORIZON 2020

CHALLENGE

Wider access to scientific facts and knowledge helps researchers, innovators and the public find and re-use data, and check research results:

offers better value for EU research funds

encourages research across scientific fields



a public benefit



essential for solving today's complex societal challenges

SOLUTION

Horizon 2020 already mandates open access to all scientific publications



From 2017, research data is open by default, with possibilities to opt out

As desculpas para não abertura de dados de pesquisa
Fifty Shades of – No

Too expensive	Demasiado caro
There's no business case	Não tem interesse
There's no commercial value	Não tem valor comercial
It's private	É privado
It's secret	É secreto
It's our data	São nossos dados
We have invested a lot of money in this	Nós investimos muito dinheiro nisso
Link enough data and one will arrive at sensitive private information	Se ligarem os dados chegarão a informações sensíveis e confidenciais
It's not data, it's information	Não são dados, são informações
It will never work	Nunca vai funcionar
We don't know how to do this	Nós não sabemos como fazer isso
We don't have the right people to do this	Nós não temos as pessoas certas para
We need the money	Nós precisamos do dinheiro
It's not ours, and we don't know who's data it is	Não são nossos, e não sabemos a quem pertencem os dados
No idea what the quality of the data is	Não faço ideia de qual é a qualidade de
We don't know where to find it	Nós não sabemos onde encontrá-los
It's not our job	Esse não é o nosso trabalho
It isn't in the right format	Não estão no formato certo
I am not <u>authorized</u>	Eu não estou autorizado
Who is going to use this anyway	Ninguém vai usar isso de qualquer ma
People are going to misuse it	As pessoas vão abusar disso
Image damage for the minister	Vão provocar danos de imagem para o
We are not ready for this	Nós não estamos prontos para isso
Image loss for Government	Vão provocar perda de imagem para o governo
The data file is too big	O dataset é muito grande
Not enough bandwidth	A largura de banda é insuficiente
This is a first step, we will see what we can do later	Este é um primeiro passo, vamos ver o podemos fazer mais tarde
We can't find it	Não conseguimos encontrar os dados
We have no access	Não temos acesso aos dados
It is out of date / too old	Estão desatualizados
We have it only on paper	Nós temos os dados apenas em papel
We don't know if it's legal	Não sabemos se é legal
Management says no	A gerência diz que não
We never did this before	Nós nunca fizemos isso antes
No value in it	Nenhum valor nele
No time / no resources	Sem tempo / sem recursos
We will open up (but adapt 90%)	Vamos abrir (mas adaptar 90%)
It's incorrect	É incorreto

São os nossos dados

Não sabemos como fazer isso

O Dataset é muito grande

Não sabemos se é legal

Os dados são sensíveis

Não temos autorização

As desculpas para não abertura de dados de pesquisa
Fifty Shades of – No

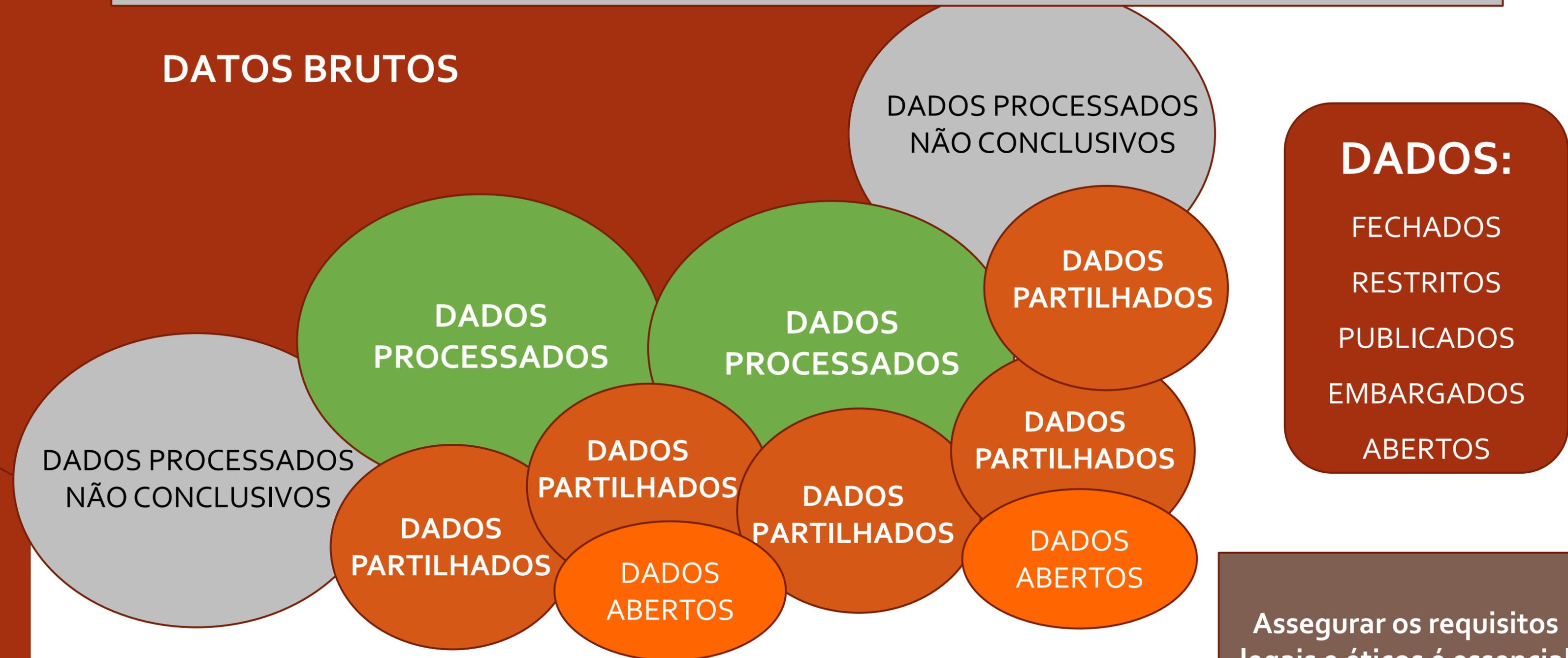
Commercially sensitive	Os dados são comercialmente sensíveis
It is dangerous when linked	Será perigosos quando forem interligados
People are going to make the wrong conclusions	As pessoas vão tirar conclusões erradas
start a wrong discussion	Isso vai começar um debate errado
whether we have it or we don't	Não podemos dizer se temos os dados ou não
data is wrong, and people will tell us	Sabemos que os dados estão errados, e as pessoas vão dizer-nos onde estão erradas, então iremos desperdiçar recursos inserindo as correções que as pessoas nos enviam
g, then we'd waste resources	
corrections people send us	
will charge us a fortune to do an	Nossos fornecedores de TI nos cobrarão uma fortuna para fazer um extrato de dados ad hoc
act	
careful whit existing contracts	Temos que ter cuidado com os contratos existentes
Our website cannot hold files this large	Nosso site não pode manter conjuntos de dados tão grandes
we don't have authorization from	Não são nossos e não temos autorização do proprietário dos dados
published the data (but it's	Nós já publicamos os dados (mas são indecifráveis / inúteis)
able)	
People may download and cache the data and it will be out of date when they reuse it	As pessoas podem baixar e armazenar em cache os dados e ficarão desatualizados quando eles forem reutilizá-los
it regularly	Nós não coletamos esses dados regularmente
they will want to download it, which	Muitas pessoas vão querer baixá-los, o que fará com que nossos servidores falhem
ervers to fail	
get upset	As pessoas ficarão insatisfeitas
It's very sensitive information	
dy for this	
ing to use it and we will make it	

<https://philarcher.org/diary/2015/50shadesofno/>



Diferentes níveis de processamento e publicação de dados

DADOS BRUTOS



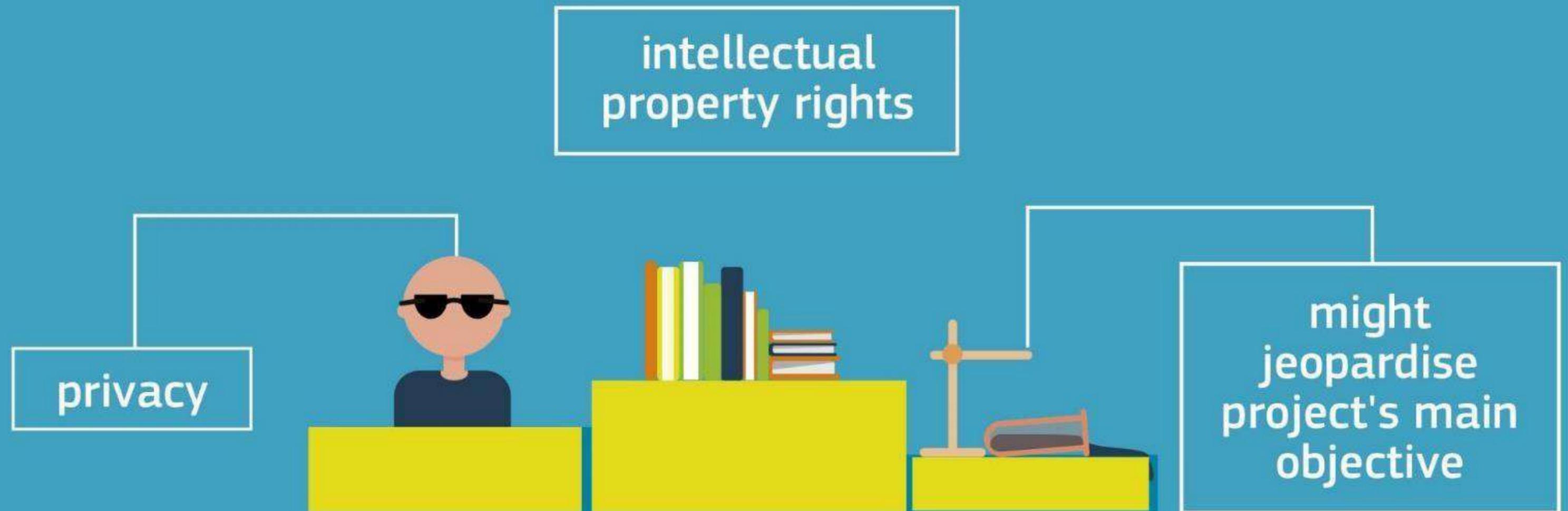
DADOS:

- FECHADOS
- RESTRITOS
- PUBLICADOS
- EMBARGADOS
- ABERTOS

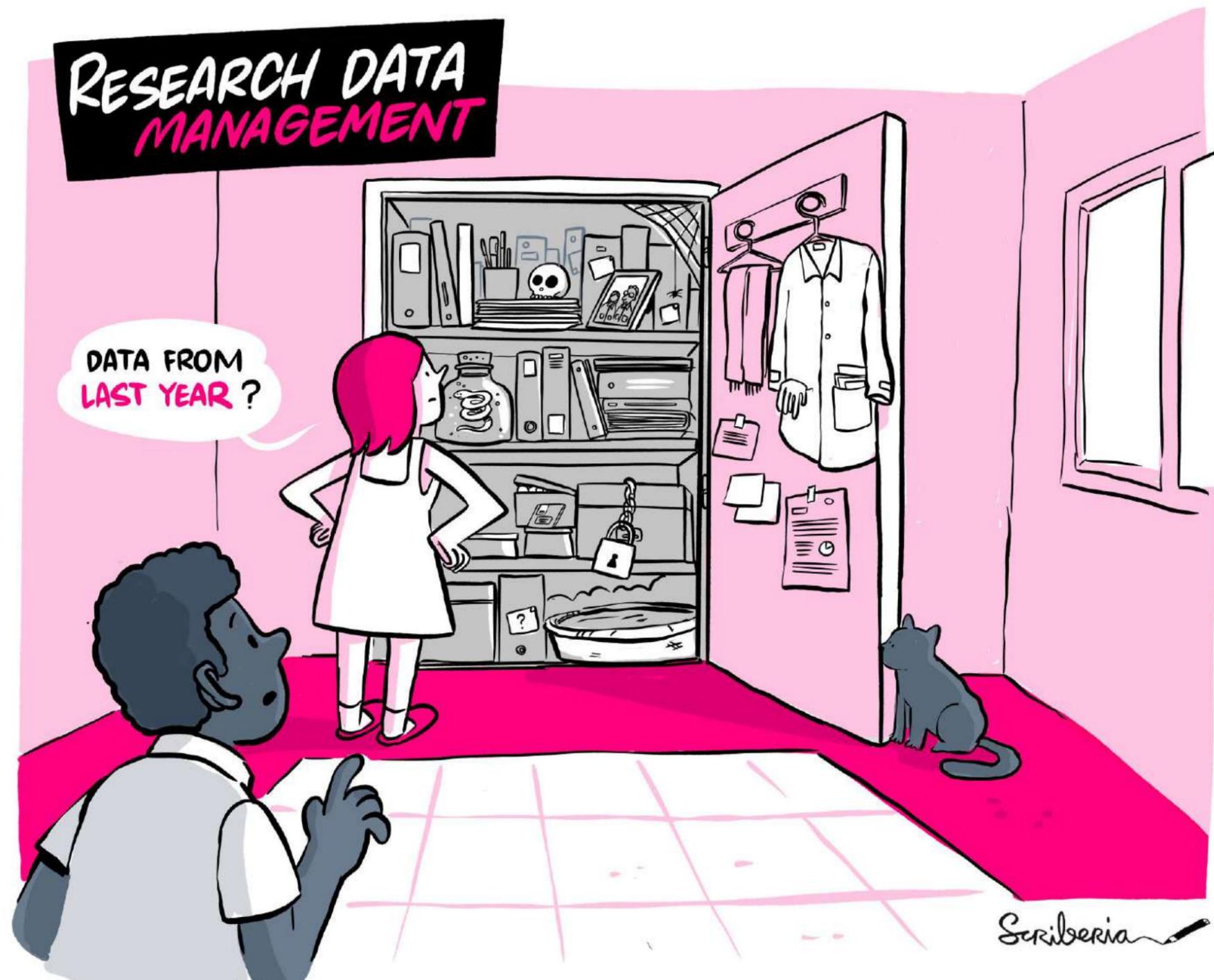
Assegurar os requisitos legais e éticos é essencial

AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY

Grantees have the right to **opt-out**, but need to say **why**







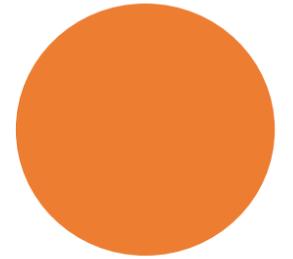
**Preparar
dados:**
organização
e formatos

Preparar os dados

Organização

Os ficheiros poderão ser organizados em pastas e por tipologia de dados:

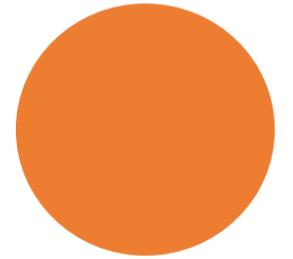
- Conjuntos de dados
- Texto
- Imagens
- Modelos
- Som
- Outras atividades decorrentes da investigação
(entrevistas, inquéritos, grupos de discussão, ou outro material - dados, documentação, publicações)



Preparar os dados

Atribuição de nomes aos ficheiros

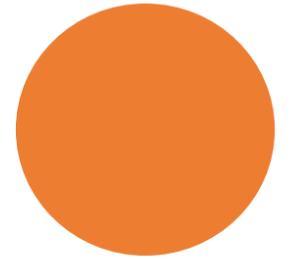
- Criar nomes significativos e curtos
- Evitar a utilização de espaços e caracteres especiais
- Não utilizar nomes genéricos que possam entrar em conflito quando alterada a localização
- Deverão permanecer para além da pessoa que os criou
- Deverão incluir informação descritiva que ajuda na sua identificação



Preparar os dados

Atribuição de nomes aos ficheiros

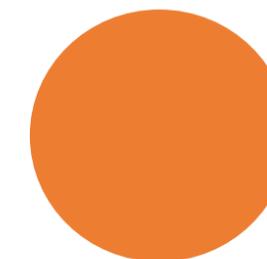
- Usar linha (_) em alternativa aos pontos finais (.) ou espaços
- Se incluir datas faça-o de forma consistente (mais comum: **AAAA-MM-DD** ou **AAAA-MM** ou **AAAA-AAAA**)
- Tenha em **atenção** o uso de maiúsculas, pois os sistemas podem assumir informações diferentes
- Poderá usar as extensões dos ficheiros se o ajudar na sua identificação



Preparar os dados

Atribuição de nomes aos ficheiros
(o que podem conter)

- acrónimos
- as iniciais dos investigadores
- número de versões
- estado do ficheiro
- data



Preparar dados: Formatos

Boa prática: no momento de escolha por um ou vários formatos a utilizar tenha em consideração estes exemplos, pois terão menos probabilidade de se tornarem obsoletos. Fonte: [EPFL Fast Guides on Research Data Management](https://www.epfl.ch/research/data-management/guides/)

Tipo de dados	APROPRIADO	ACEITÁVEL	NÃO APROPRIADO
Dados tabulares com metadados extensos	.csv - .hdf5	.txt - .html - .tex - .por	
Dados tabulares com metadados mínimos	.csv - .tab - .ods - SQL	.xml if appropriate DTD - .xlsx	.xls - .xlsb
Dados textuais	.pdf - .txt - .odt - .odm - .tex - .md - .htm - .xml	pptx - PDF with embedded forms - .rtf	.doc - .ppt
Código	.m - .R - .py - .iypnb - .rstudio - .rmd - NetCDF	.sdd	.mat - .rdata
Dados de imagem digital	.tif - .png - .svg - .jpeg	jpg - .jp2 - .tif - .tiff - .pdf - GIF - BMP	.indd - .ait - .psd
Dados de áudio digital	.flac - .wav - .ogg	.mp3 - .mp4 - .aif	
Dados de vídeo digital	.mp4 - .mj2 - .avi - .mkv	.ogm - .webm	.wmv - .mov
Dados geoespaciais	MetCDF, tabular GIS attribute data, shp - .shx - .dbf - .prj - .sbx - .sbn - PostGIS - .tif - .tfw - GeoJSON	/.mdb/.mif/	
Dados vetoriais e matriciais	.dwg - .dxf - .x3d - .x3dv - .x3db - .pdf - PDF3D		
Dados genéricos	.xml - .json - .rdf		

Versões e cópias de segurança

Existem **diferentes tipologias de backups**, aconselhe-se na sua instituição/centro para avaliar da existência de política de backups

Quando iniciar?

- Assim que inicia processo de recolha de dados necessitará de fazer cópias de segurança

Sugestão: guarde **SEMPRE 1 cópia dos seus dados:**

- Antes de os limpar
- Outra depois de os analisar (poderá precisar de repetir a experiência)

3 > 2 > 1 --> cópias de segurança

Não guarde todos os dados originais e cópias de segurança no mesmo local

Acautelar que tem:

- 3 cópias de ficheiros importantes
- 2 diferentes tipos de suporte de armazenamento
- 1 cópia num local remoto



Versões e cópias de segurança

- **Se tem todos os dados online** considere ter uma cópia dos dados offline
- **Guarde versões de alterações importantes** (recorrer a convenções de nomeação de ficheiros para numerar as versões)
- **Anote as alterações** numa tabela dedicada ao registo de versões (existem softwares que fazem este tipo de trabalho)

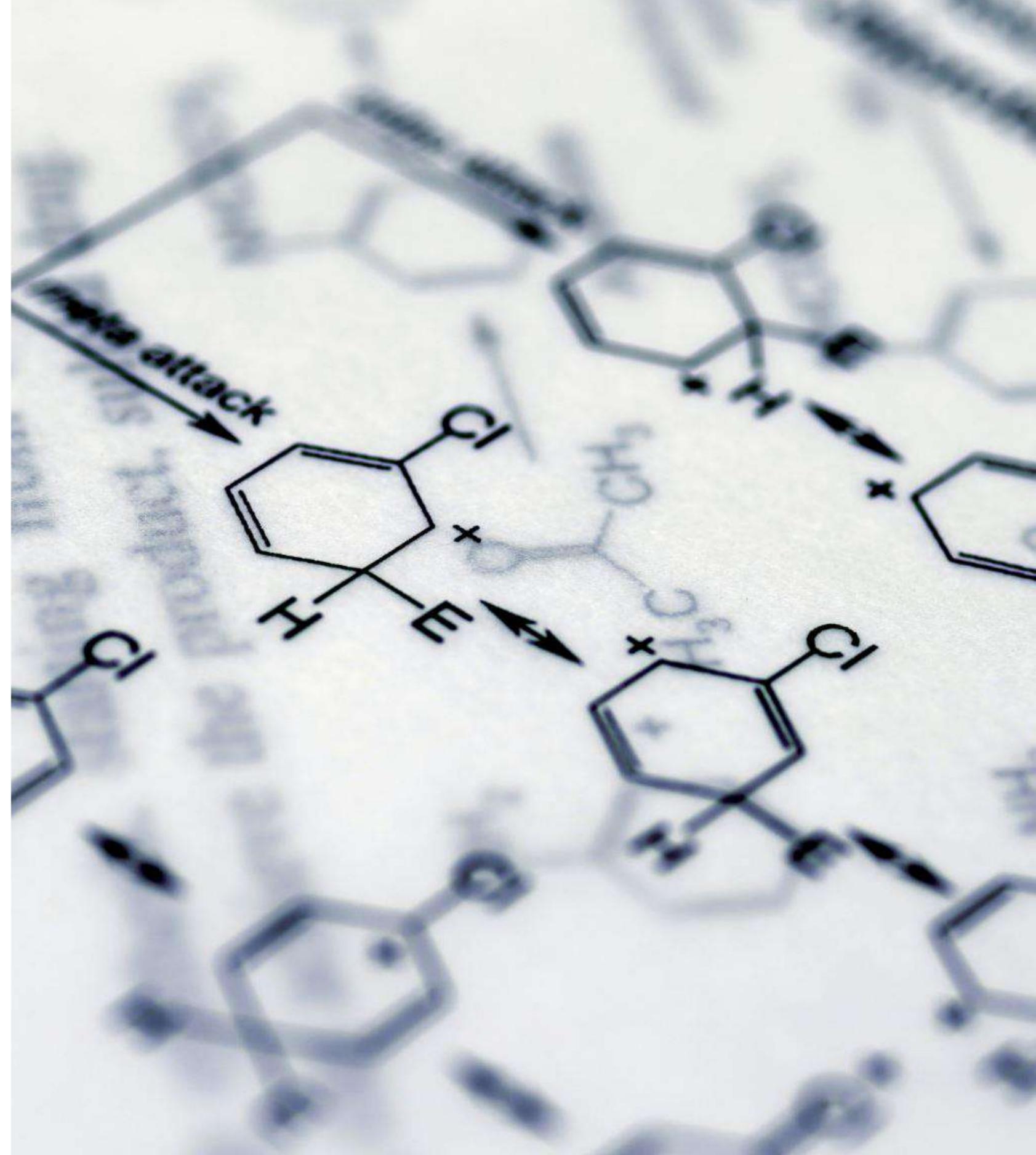


Preparar dados:

Documentação

A documentação refere-se ao ato de descrever os conjuntos de dados que ocorre a vários níveis:

- Descrição do processo de recolha dos dados por parte do investigador, que poderá ser feita num caderno de laboratório
- Descrição dos dados (quantidade, formato, software utilizado para a sua leitura)
- Descrição das alterações realizadas nos dados ao longo do processo, muitas vezes utilizado para registar e criar um histórico de todas as utilizações e edições que os dados sofreram durante um período de tempo



Preparar dados: Documentação

- Uma **documentação correta** dos dados assegura que estes sejam **pesquisáveis e inequivocamente compreendidos**, pelos utilizadores atuais e futuros (incluindo o próprio investigador).
- A **função de documentação** dos dados dependerá igualmente do momento em que estes se encontram - **dentro do ciclo de vida dos dados**.



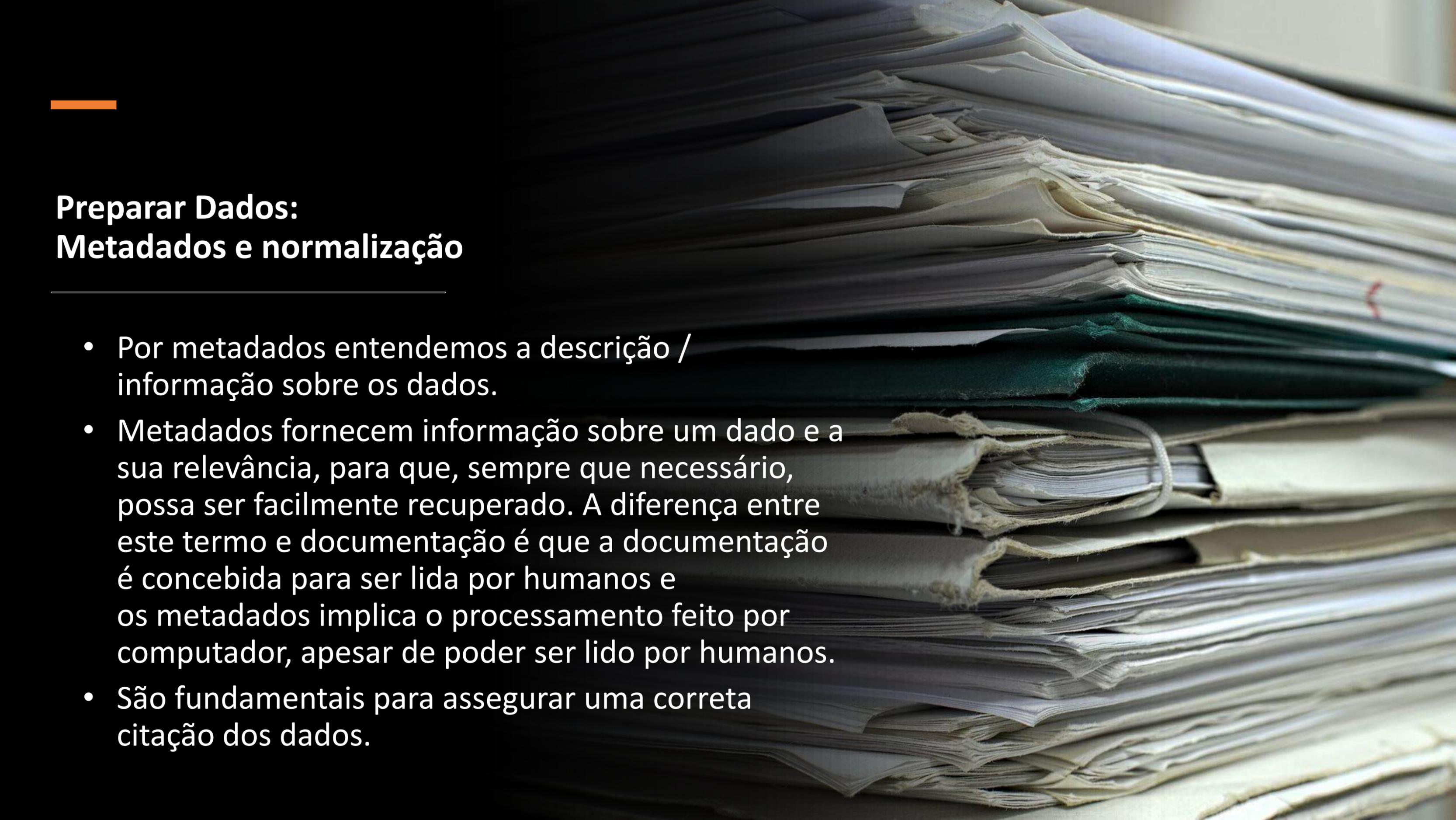
Preparar dados: Documentação

BOA PRÁTICA

É recomendado o carregamento de um ficheiro em formato .txt (**ficheiro README**) no qual é colocada toda a informação necessária para a **correta compreensão do Dataset**.

README.txt > Deve conter a seguinte informação:

- Título do Dataset, DOI, informações de contacto
- Métodos (Procedimentos metodológicos)
- Informação geral dos dados e ficheiros
- Informação específica sobre dados (ao nível do ficheiro)
- Termos de Reutilização / Licenças



Preparar Dados: Metadados e normalização

- Por metadados entendemos a descrição / informação sobre os dados.
- Metadados fornecem informação sobre um dado e a sua relevância, para que, sempre que necessário, possa ser facilmente recuperado. A diferença entre este termo e documentação é que a documentação é concebida para ser lida por humanos e os metadados implica o processamento feito por computador, apesar de poder ser lido por humanos.
- São fundamentais para assegurar uma correta citação dos dados.

Descrição de dados no repositório - Dataverse

Plantas interpretadas da análise das unidades domésticas de Complutum
Version 1.1

Magalhães, F., 2020, "Plantas interpretadas da análise das unidades domésticas de Complutum", <https://doi.org/10.34622/datarepositorium/TYRY0C>, Repositório de Dados da Universidade do Minho, V1

Dataset Metrics
7 Downloads

Description
As plantas apresentadas foram elaboradas no âmbito do trabalho para elaborar um corpus das domus de Bracara Augusta, que servisse de base para realizar análises comparativas com outros centros urbanos do NO peninsular, bem como com o resto da Hispânia. Para o efeito, foi indispensável desenhar e analisar as planimetrias das diferentes casas que se caracterizam pela diversidade de soluções formais, construtivas e funcionais. Entre os exemplares de casas que foram valorizados desta análise destacamos as de Complutum, Iuliobriga, Tongobriga, Augusta Emerita e Conimbriga, tendo-se produzido plantas interpretadas para efeito de ilustração da nossa análise das unidades domésticas, algumas das quais são objeto de modelos 3D. The plans presented were elaborated in the scope of the work to elaborate a corpus of the domus of Bracara Augusta, which would serve as a basis to carry out comparative analyzes with other urban centers in the NW peninsular, as well as with other regions of Hispania. For this purpose, it was essential to design and analyze the planimetry of the different houses that are characterized by the diversity of formal, constructive and functional solutions. Among the examples of houses that were valued we highlight those of Complutum, Iuliobriga, Tongobriga, Augusta Emerita and Conimbriga, having produced plants interpreted for the purpose of illustrating our analysis of domestic units, some of which are the subject of 3D models. (2019)

Subject
Arts and Humanities

Keyword
Complutum, Arquitetura privada, Domus, Casa de Baco, Casa de los Grifos, Casa de Leda, Casa de Marte, Casa del Átrio, Casa de la Lucerna de la Máscara Trágica, Vida cotidiana

Related Publication
Magalhães, F. (2019). A domus romana no Noroeste Peninsular. Arquitetura, construção e sociabilidades. Tese de Doutoramento em Arqueologia, Área do conhecimento em Arqueologia da Paisagem e do Povoamento. Universidade do Minho. handle: 1822/64109

Files | Metadata | Terms | Versions

Terms of Use
Waiver
Our Community Norms as well as good scientific practices expect that proper credit is given via citation. Please use the data citation above, generated by the Dataverse.
CC0 - "Public Domain Dedication"

Metadados descritivos
(título, autor, DOI,
resumo, palavras-chave,
assunto, ...)

Metadados
administrativos
(licenças, direitos de
acesso e termos de
uso, ...)

Descrição de dados no repositório - Zenodo

License required ▼

Access right *

- Open Access
- Embargoed Access
- Restricted Access
- Closed Access

Required. Open access uploads have considerably higher visibility on Zenodo.

License *

Creative Commons Attribution 4.0 International

Required. Selected license applies to all of your files displayed on the top of the form. If you want to upload some files with a different license, please do so in separate uploads. If you cannot find the license you're looking for, include a relevant LICENSE file in your upload. Other licenses available (*Other (Open)*, *Other (Attribution)*, etc.). The supported licenses in the list are harvested from spdx.org. If you think that a license is missing from the list, please contact us.

Funding

Zenodo is integrated into reporting lines for research funded by the European Commission via [OpenAIRE](#). Specify grants which have funded your research, and we will let the funding agency know!

Grants

European Commission (EU)

Optional. OpenAIRE-supported projects only. For other funding acknowledgements, please use the *Additional Notes* field.
Note: a human Zenodo curator will need to validate your upload - you may experience a delay before it is available in OpenAIRE.

[+ Add another grant](#)

Metadados
administrativos
(licenças, direitos de
acesso e termos de uso,
financiamento, ...)

Descrição de dados



Nikon D5100



f/5.3



90.0 mm



1/60



3600



Flash
(desligado,
não disparado)



Ocultar EXIF

JFIFVersion - 1.02

X-Resolution - 96 dpi

Y-Resolution - 96 dpi

Make - NIKON CORPORATION

Software - PhotoScape

Date and Time (Modified) -
2013:10:07 11:07:50

YCbCr Positioning - Co-sited

ISO Speed - 3600

Sensitivity Type - Recommended
Exposure Index

Exif Version - 0230

Date and Time (Original) - 2013:10:07
11:07:50

Date and Time (Digitized) -
2013:10:07 11:07:50

Metadados
administrativos
(dados técnicos sobre
os formatos, ...)

Preparar Dados: Organização e Formatos

Checklist

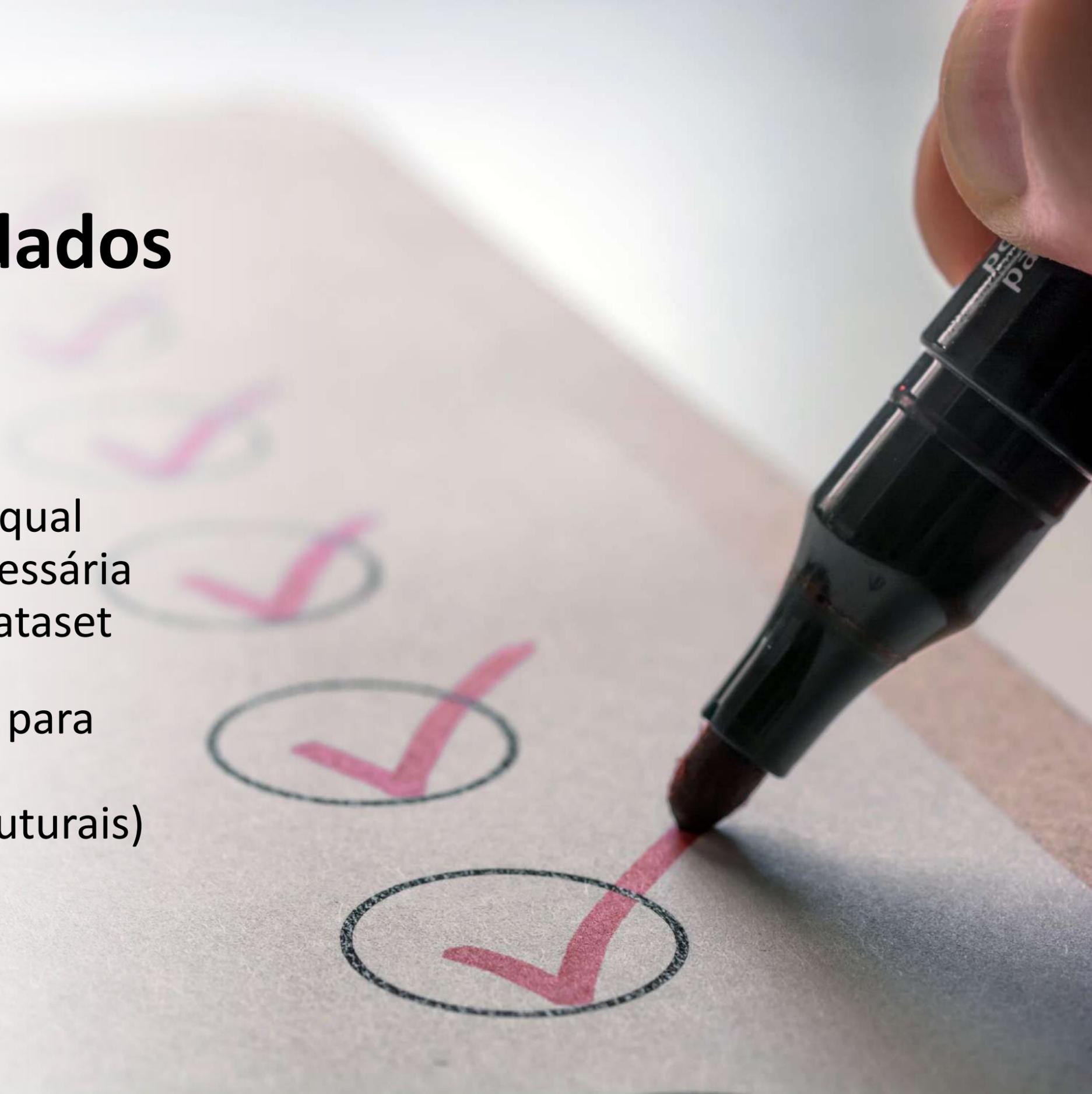
- ✓ Organizar os **ficheiros em pastas** por tipologia de dados
- ✓ Atribuir **nomes** aos ficheiros seguindo as boas práticas
- ✓ Utilizar **formatos normalizados**



Preparar Dados: Documentar e Metadados

Checklist

- ✓ Criar um **ficheiro README.txt** no qual é colocada toda a informação necessária para a correta compreensão do Dataset
- ✓ Reunir os **metadados** necessários para a correta descrição dos dados (Descritivos, Administrativos, Estruturais)





PARTILHA E PUBLICAÇÃO DE DADOS

Data Journals

Partilha e descrição do seu conjunto de dados de forma detalhada

Potenciar a atribuição e citação dos dados e facilitar o seu reuso – boas práticas de Ciência

Dados nem sempre falam por si, precisam sempre de contexto

Multiplicidade de políticas dos editores e assegurar Propriedade Intelectual

Serviço de Revistas para material suplementar

Cumprir requisitos do editor

Dados disponíveis junto dos resultados publicados

Pode ser dispendioso e risco com direitos sobre os dados

Acesso fechado e improvável que faculte preservação

Repositório de dados institucional

Aceitar vários tipos de dados, garantir acesso a longo prazo

Podem confiar mais facilmente e não haverá custos

Pode não oferecer acesso sustentável a longo prazo

Pode não ter metadados disciplinares

Repositório genérico

Alcance de um público mais vasto.

Aceita vários de tipos, adequado para dados interdisciplinares

Normalmente apenas metadados simples estão disponíveis

Sem controle editorial sobre a qualidade dos materiais depositados

Repositório disciplinar

Oferece especialização e experiência em gestão de dados

Provável que aceite conjuntos de dados completos

É provável que sejam selectivos no tipo de dados que aceitam

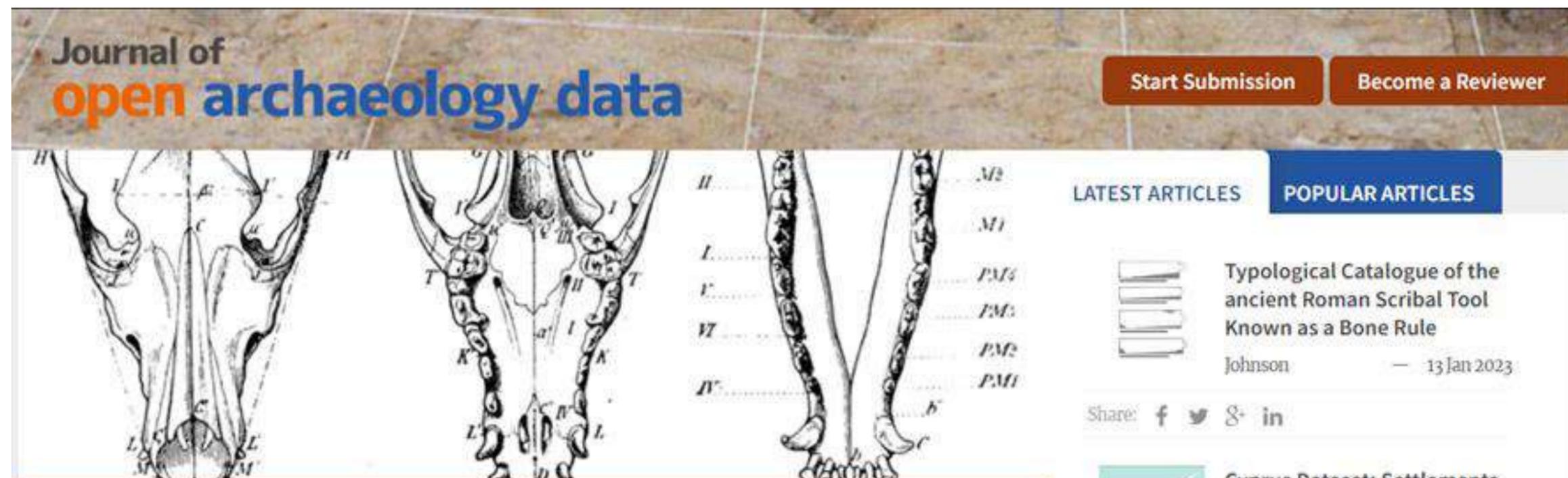
Requer planeamento e padrões elevados, pode implicar custos

Data Journals

The screenshot displays the MDPI website's 'Data' journal page. At the top, the MDPI logo is on the left, and navigation links for Journals, Topics, Information, Author Services, Initiatives, and About are in the center. On the right, there are 'Sign In / Sign Up' and 'Submit' buttons. Below the navigation is a search bar with fields for 'Title / Keyword', 'Author / Affiliation / Email', 'Data', and 'All Article Types', along with 'Search' and 'Advanced' buttons. The breadcrumb trail reads 'Journals / Data / Aims & Scope'. On the right side, there are two circular badges: 'Tracked for Impact Factor' and 'CITESCORE 4.8', along with social media icons. The main content area is titled 'About Data' and features a circular diagram with three nodes: 'DEPOSIT DATA in an online repository.', 'PUBLISH DATA DESCRIPTOR including the methods and description of the dataset and get credit for your work.', and 'PUBLISH ARTICLE based on the datasets in other journals, and allow others to reuse your dataset.' Below the diagram is the section 'Rationale for Data' with the text: 'One of the current hot topics in science is data: how can datasets be used in scientific and scholarly research in a more reliable...'. On the left side of the page, there is a 'data' logo, 'Submit to Data' and 'Review for Data' buttons, and a 'Journal Menu' with a list of links including Data Home, Aims & Scope, Editorial Board, Reviewer Board, Topical Advisory Panel, Instructions for Authors, Guidelines for Reviewers, Special Issues, Topics, Sections & Collections, Article Processing Charge, Indexing & Archiving, Editor's Choice Articles, Most Cited & Viewed, Journal Statistics, Journal History, Journal Awards, and Editorial Office.

<https://www.mdpi.com/journal/data/about>

Data Journals



Journal of
open archaeology data

Start Submission Become a Reviewer

LATEST ARTICLES POPULAR ARTICLES

Canan Çakırlar & E.E. Scheele
Van Giffen's Dogs: Cranial Osteometry of Iron Age to Medieval Period Dogs from the...

Follow on Twitter Like on Facebook Follow Via RSS

About this Journal

The *Journal of Open Archaeology Data* (JOAD) features peer reviewed data papers describing archaeology datasets with high reuse potential. We work with a number of specialist and institutional data repositories to ensure that the associated data are professionally archived, preserved, and openly available. Equally importantly, the data and the papers are citable, and reuse is tracked.

JOAD data papers are relatively quick to create, and provide you with a peer-reviewed publication to gain credit for your data. **Submit a paper today!**

Typological Catalogue of the ancient Roman Scribal Tool Known as a Bone Rule
Johnson — 13 Jan 2023

Cyprus Dataset: Settlements from 11000 BCE to 1878 CE
Crawford & Vella — 20 Dec 2022

Presenting the AfriArch Isotopic Database
Goldstein et al. — 08 Dec 2022

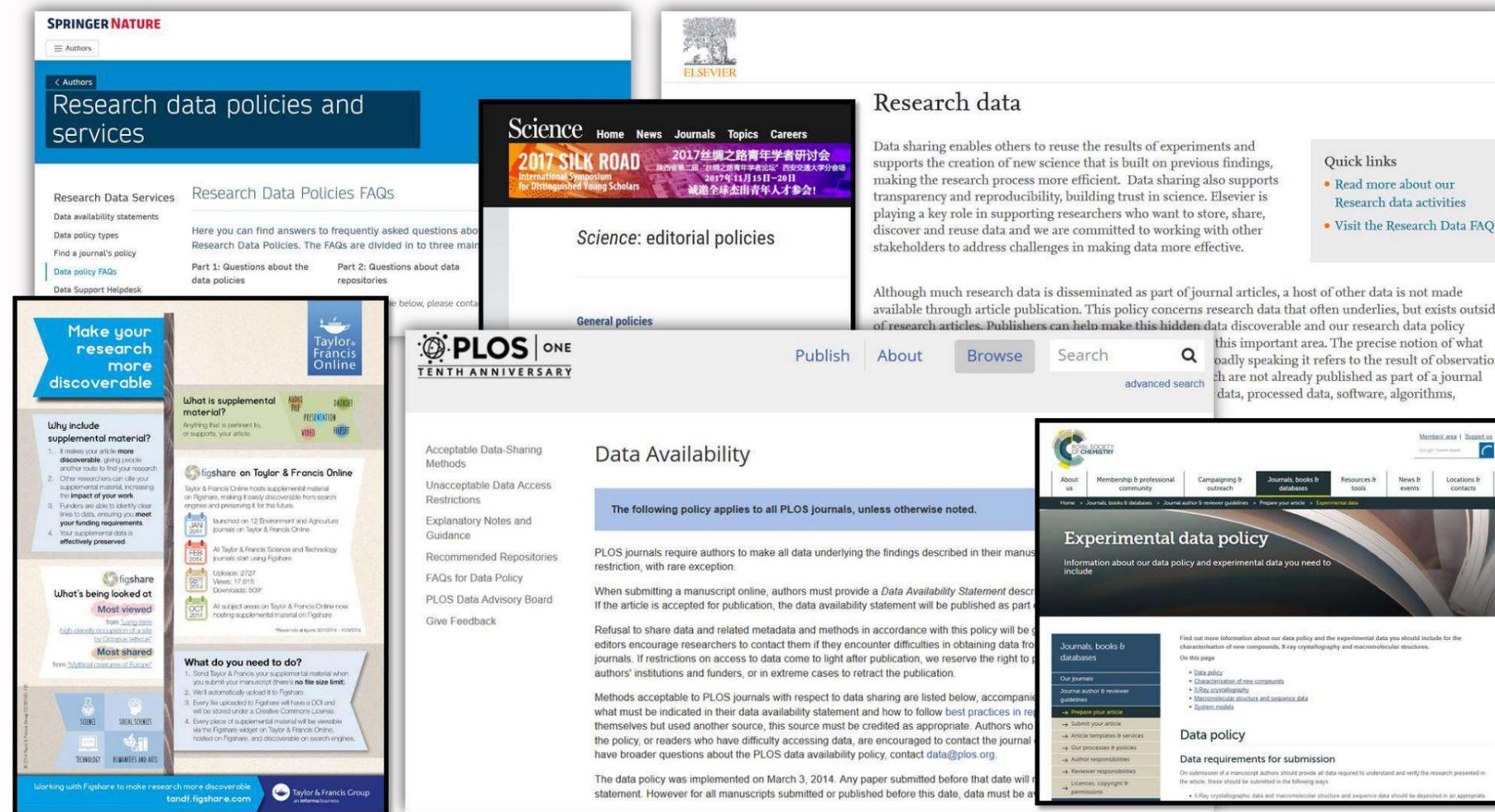
Prehistoric Landscape Use in the Central Alaska Range: Lithic Analysis Dataset from the Upper Susitna River Basin, Alaska
Blong — 01 Aug 2022

<https://openarchaeologydata.metajnl.com/>

REQUISITOS DOS EDITORES PARA A DISPONIBILIZAÇÃO DE DADOS



Universidade do Minho
Serviço de Documentação e Bibliotecas



Quais os 3 tipos de requisitos:

- Enviar dataset para o editor.
- Depositar dataset num repositório confiável.
- Providenciar a informação de contacto.

Um número significativo de revistas científicas apresenta uma política específica de disponibilização dos dados ou de arquivo de dados. Os requisitos geralmente encontram-se no site das revistas.

ALERTA! providenciar serviços nas instituições ou comunidades: dados devem permanecer na administração dos investigadores e instituições.

Repositório de dados institucional

Aceitar vários tipos de dados, garantir acesso a longo prazo

Podem confiar mais facilmente e não haverá custos

Pode não oferecer acesso sustentável a longo prazo

Pode não ter metadados disciplinares



Repositório de dados institucional

Aceitar vários tipos de dados, garantir acesso a longo prazo

Podem confiar mais facilmente e não haverá custos

Pode não oferecer acesso sustentável a longo prazo

Pode não ter metadados disciplinares

The screenshot displays the IPB Institutional Data Repository website. At the top, the logo for Instituto Politécnico de Bragança is visible, along with navigation links for Search, User Guide, Support, Sign Up, and Log In. The main content area shows the repository name and a search bar. Below the search bar, there are filters for Dataverses (14), Datasets (29), and Files (5,004). The search results are displayed in a list format, showing the title of the dataset, the date, and the authors. The first result is "Base de dados de dieta alimentar, Moçambique / Database of food diet, Mozambique" with authors including Edson, Manhiça, Agnaldo, Venâncio, Armando, Cabo, Paula, Rodrigues, Paula, and Macuamule. The second result is "Base de dados de dieta alimentar, Angola / Database of food diet, Angola" with authors including Fernandes, Lafayete, Neto, Fernando, Neto, Felizardo, Lucamba, Zeida, Cabo, Paula, Rodrigues, Paula, Afonso, and Sandra. The third result is "Base de dados de análise de micotoxinas / Database of mycotoxin analysis" with authors including Matusse, Cláudio, Lucamba, Zeida, Venâncio, Armando, Afonso, Sandra, Macuamule, Custódia, and Rodrigues, Paula.

<https://dados.ipb.pt/>

10 RAZÕES PARA DEPOSITAR DADOS DE INVESTIGAÇÃO EM REPOSITÓRIOS

1

Promover uma ciência mais inclusiva, transparente e aberta

2

Contribuir para o cumprimento dos princípios FAIR nos dados de investigação ao longo do seu ciclo de vida (Findable, Accessible, Interoperable, Reusable)

3

Beneficiar do serviço de arquivo e preservação digital de forma segura e transparente

4

Contribuir para uma maior disseminação do trabalho científico, do acesso aos dados que validam as publicações, potenciando a sua reprodutibilidade e reutilização

5

Beneficiar da atribuição de Identificadores Persistentes (PIDs), promovendo o acesso aos (meta)dados

6

Disponibilizar os dados em acesso aberto, respeitando as orientações CE "*as open as possible, as closed as necessary*", garantindo o acesso aos metadados descritivos dos dados depositados, independentemente dos termos de acesso

7

Facilitar a disponibilização dos dados aos revisores, aquando da submissão para publicação

8

Aumentar e facilitar a citação e atribuição da autoria dos dados

9

Potenciar a visibilidade e acesso aos dados, contribuir para a sua reutilização, complementando e agilizando os processos de investigação

10

Cumprir os requisitos das agências financiadoras de ciência

Repositório genérico

Alcance de um público mais vasto.

Aceita vários de tipos, adequado para dados interdisciplinares

Normalmente apenas metadados simples estão disponíveis

Sem controle editorial sobre a qualidade dos materiais depositados



DRYAD



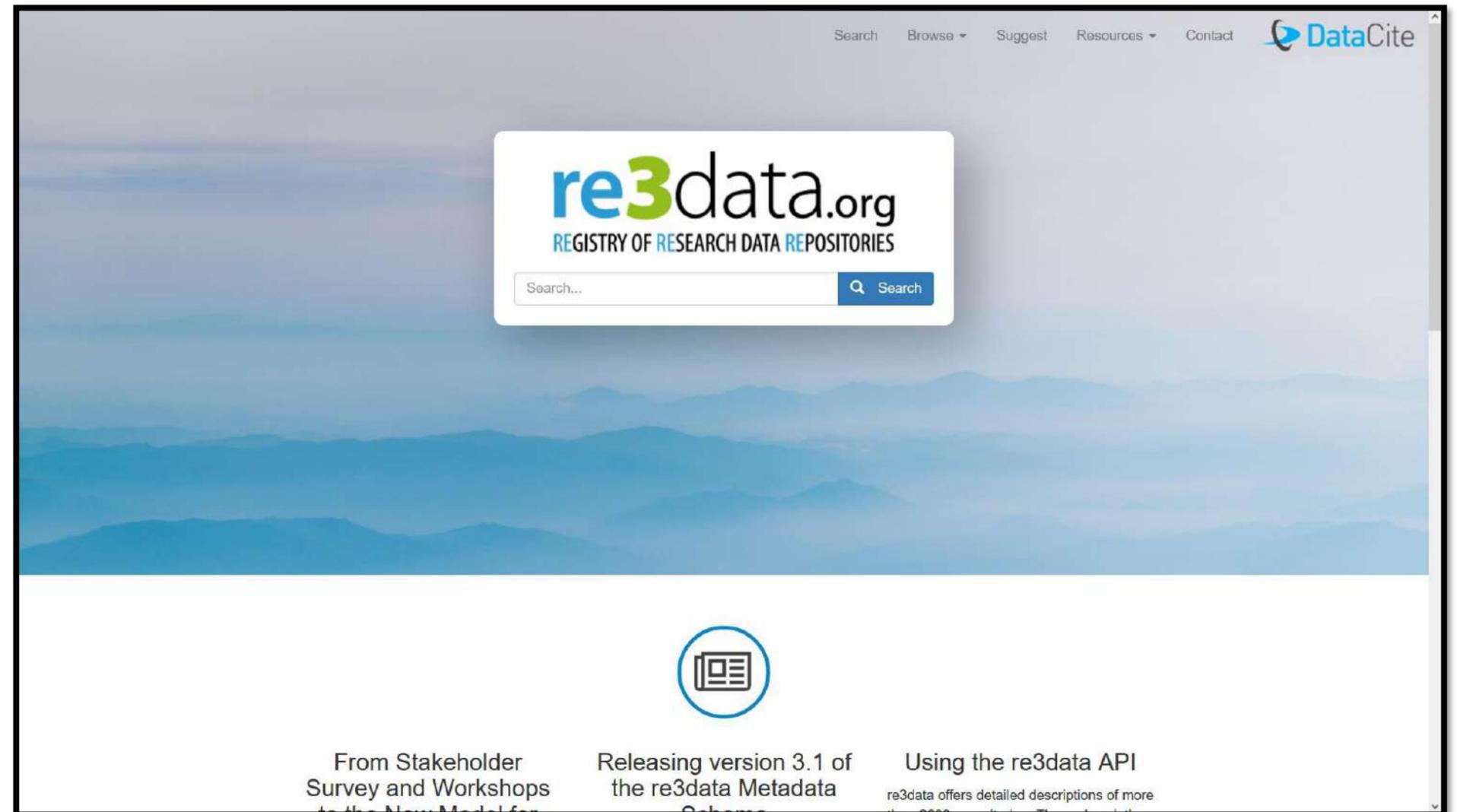
Repositório disciplinar

Oferece especialização e experiência em gestão de dados

Provável que aceite conjuntos de dados completos

É provável que sejam selectivos no tipo de dados que aceitam

Requer planeamento e padrões elevados, pode implicar custos



Repositórios disciplinares

**PORTULAN
CLARIN**



Infraestrutura de Investigação para a Ciência e Tecnologia da Linguagem

Repositório Bancada Apoio Alcance

<https://portulanclarin.net/>



Global Biodiversity Information Facility
PORTUGAL

<https://www.gbif.pt/>

APIS

Arquivo Português de Informação Social

<http://www.apis.ics.ulisboa.pt/>

EMBL's European Bioinformatics Institute

EMBL-EBI

Unleashing the potential of big data in biology

European Molecular Biology Laboratory

<https://www.ebi.ac.uk/>

REPOSITÓRIOS CONFIÁVEIS

Instrumentos de Certificação para auditoria e certificação de repositórios.

CoreTrustSeal (CTS)

Nestor Seal;

ISO 16363: 2013





Repositório de dados institucional

Aceitar vários tipos de dados, garantir acesso a longo prazo

Podem confiar mais facilmente e não haverá custos

Pode não oferecer acesso sustentável a longo prazo

Pode não ter metadados disciplinares

The screenshot displays the website for the Instituto Politécnico de Bragança's Institutional Data Repository. The header includes the IPB logo and navigation links for Search, User Guide, Support, Sign Up, and Log In. The main content area shows the repository name, a metrics bar indicating 110 Downloads, and a search bar with a search button and a link to 'Advanced Search'. On the left, there are filters for 'Dataverses (14)', 'Datasets (29)', and 'Files (5,004)', along with filters for 'Dataverse Category', 'Publication Year', 'Author Name', and 'Subject'. The main results area shows '1 to 10 of 43 Results' with a 'Sort' dropdown. Three search results are visible, each with a document icon, title, date, project name, and a list of authors. The first result is 'Base de dados de dieta alimentar, Moçambique / Database of food diet, Mozambique' from Jan 23, 2024, by FCT-Aga Khan Project MYCOTOX-PALOP, with authors including Mussalama, Edson; Manhiça, Agnaldo; Venâncio, Armando; Cabo, Paula; Rodrigues, Paula; Macuamule, Custódia. The second result is 'Base de dados de dieta alimentar, Angola / Database of food diet, Angola' from Jan 23, 2024, by FCT-Aga Khan Project MYCOTOX-PALOP, with authors including Fernandes, Lafayette; Neto, Fernando; Neto, Felizardo; Lucamba, Zeida; Cabo, Paula; Rodrigues, Paula; Afonso, Sandra. The third result is 'Base de dados de análise de micotoxinas / Database of mycotoxin analysis' from Jan 19, 2024, by FCT-Aga Khan Project MYCOTOX-PALOP, with authors including Matusse, Cláudio; Lucamba, Zeida; Venâncio, Armando; Afonso, Sandra; Macuamule, Custódia; Rodrigues, Paula.

<https://dados.ipb.pt/>

DATAVERSE É UM REPOSITÓRIO

localizar, partilhar, citar e publicar dados

DATAVERSE É UM SOFTWARE

Plataforma de software repositório

DATAVERSE É UMA COMUNIDADE

Para promover a gestão e partilha de dados a nível mundial

<https://dataverse.org/>

Open source research data repository software



Researchers

Enjoy full control over your data. Receive *web visibility, academic credit, and increased citation counts*. A personal Dataverse collection is easy to set up, allows you to display your data on your personal website, can be branded uniquely as your research program, makes your data more discoverable to the research community, and satisfies data management plans. *Want to set up your personal Dataverse collection?*



Journals

Seamlessly manage the submission, review, and publication of data associated with published articles. Establish an *unbreakable link* between *articles in your journal* and *associated data*. Participate in the open data movement by using a Dataverse collection as part of your journal data policy or list of repository recommendations. *Want to find out more about journal Dataverse collections?*



Institutions

Establish a research data management solution for your community. Federate with a growing list of Dataverse repositories worldwide for increased discoverability of your community's data. Participate in the drive to set norms for sharing, preserving, citing, exploring, and analyzing research data. *Want to install a Dataverse repository?*



Developers

Participate in a vibrant and growing community that is helping to drive the norms for sharing, preserving, citing, exploring, and analyzing research data. Contribute code extensions, documentation, testing, and/or standards. *Integrate research analysis, visualization and exploration tools*, or other research and data archival systems with the Dataverse Project. *Want to contribute?*

DATAVERSE REPOSITORIES - A WORLD VIEW



View more Metrics

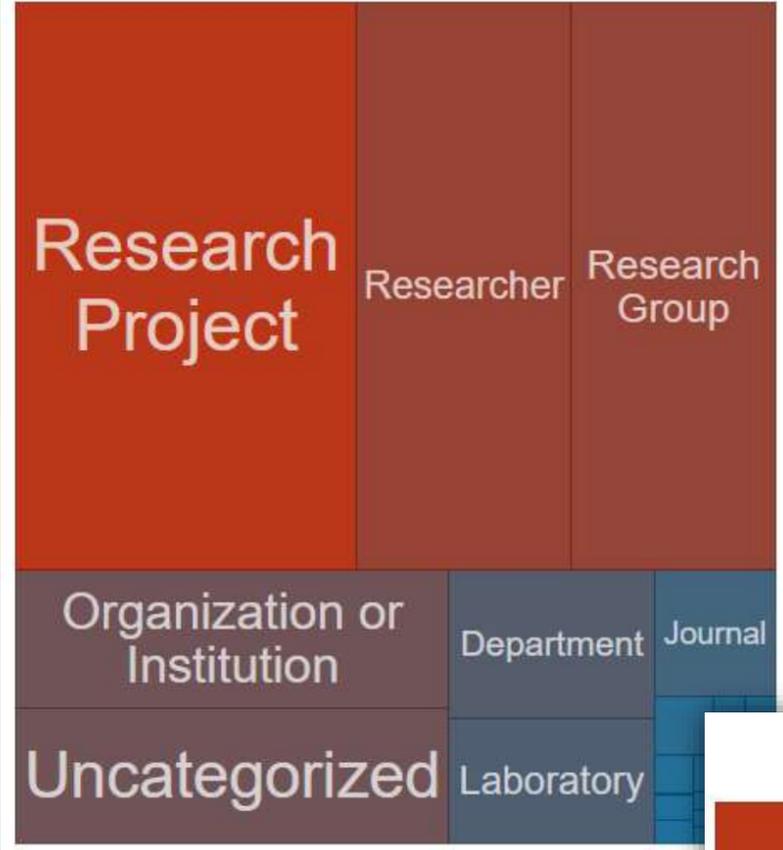
DATAVERSE REPOSITORIES - A WORLD VIEW

105 Installations



Dataverses by Category

Total: 15.8k



Datasets by Most Common Subject

Total: 374k (98% of 381k)





Universidade do Minho
Serviço de Documentação e Bibliotecas

REPOSITÓRIO CONFIÁVEL

Registado em
diretórios



Agregado em serviços
de referência



Dataset Search

Pesquisar conjuntos de dados



Certificado de
confiabilidade

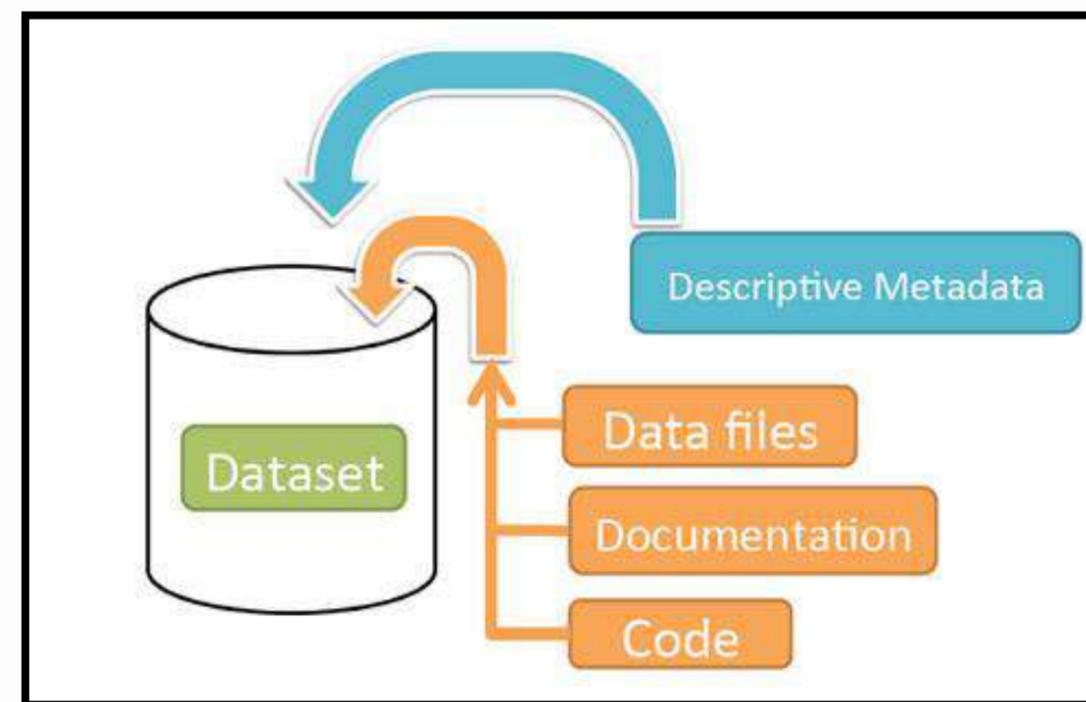
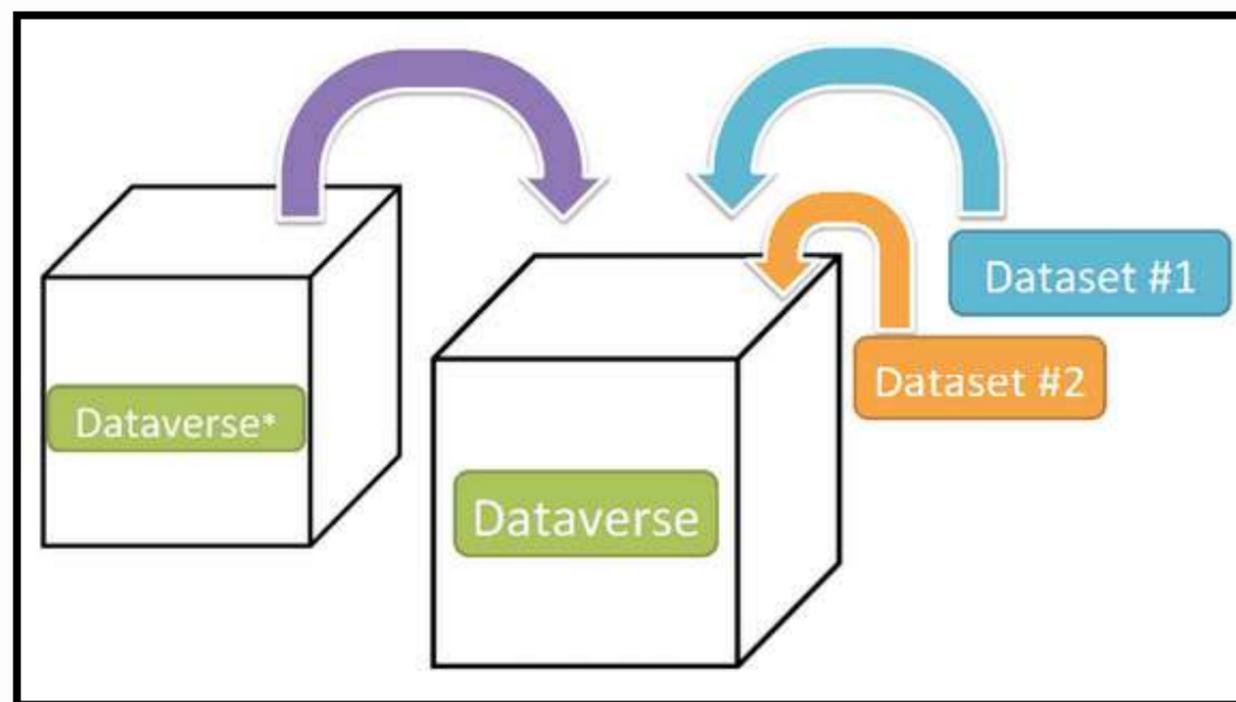


como funciona na prática

DATAVERSE = coleção / repositório / comunidade

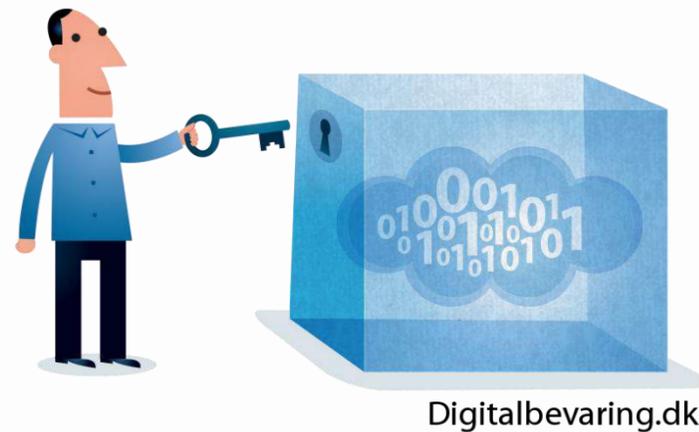
DATASET = registo de conjunto de dados com ficheiros

FILE = ficheiros que constituem os conjuntos de dados

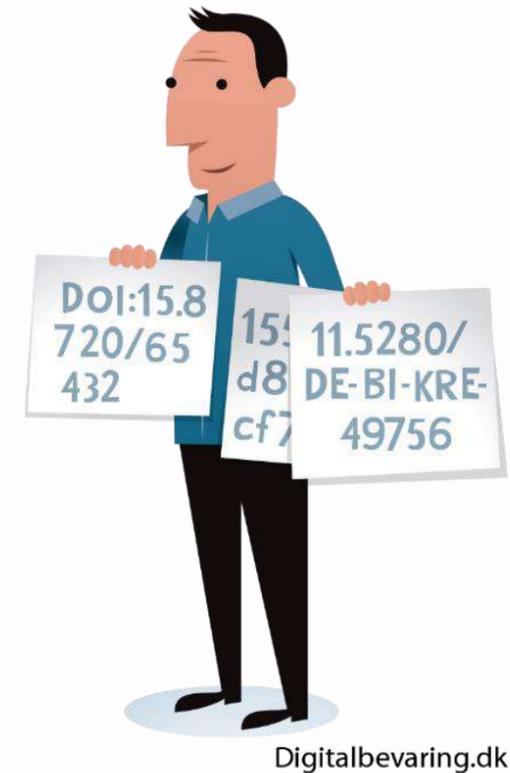
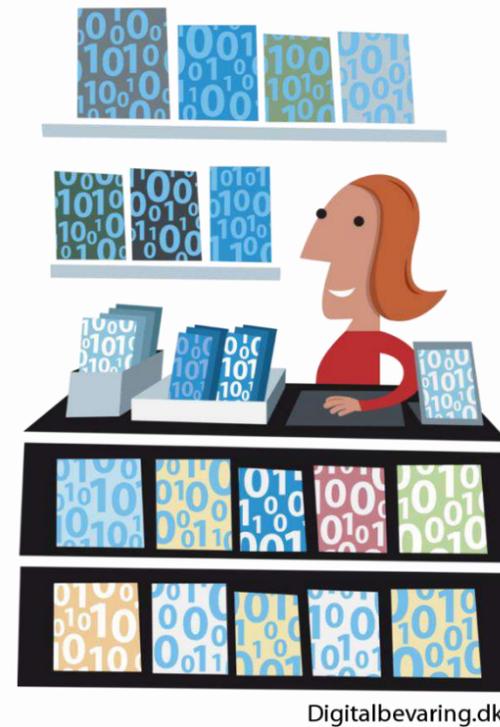


Destques do Dataverse

GESTÃO SEGURA DE DADOS



REGISTO DAS MUDANÇAS



PARTILHA EFECTIVA

ACESSO A LONGO PRAZO E PRESERVAÇÃO

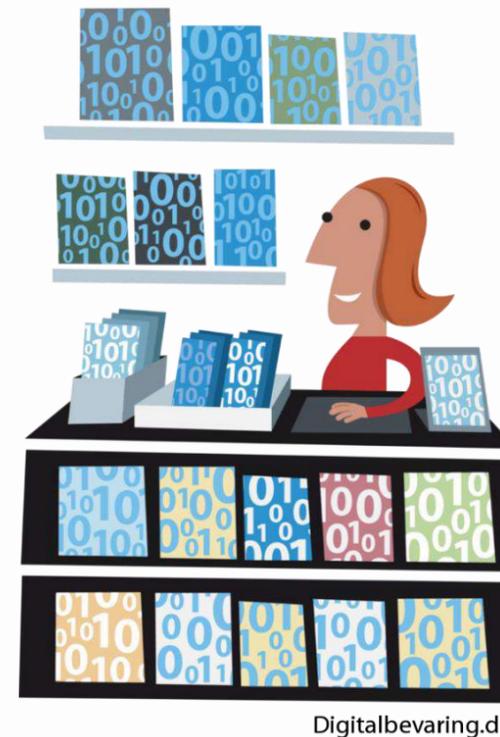
Destques do Dataverse

ORGANIZAÇÃO E COMPATIBILIDADE



ECONOMIA DE TEMPO

AUMENTA A VISIBILIDADE DA INVESTIGAÇÃO



CUMPRE OS REQUISITOS DO FINANCIAMENTO

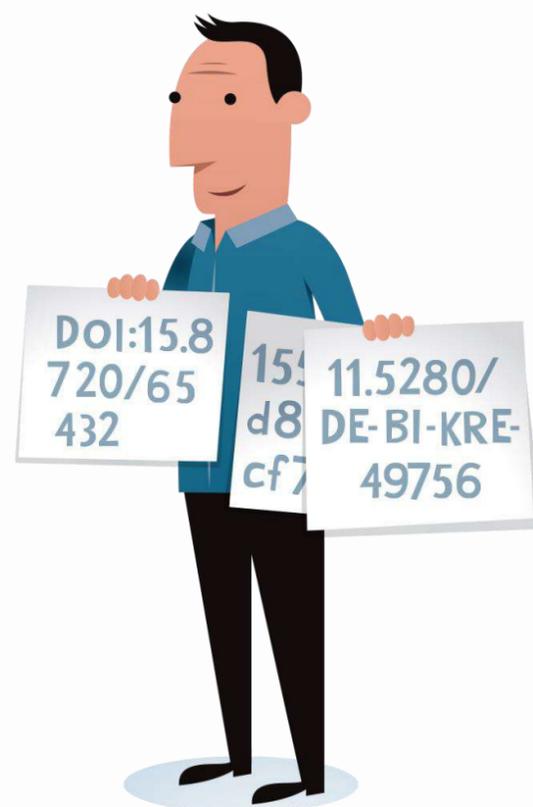
Depósito de ficheiros em coleções organizadas no ciclo de vida da investigação



O QUE VEM RESOLVER OU APOIAR:

- No processamento dos dados posso depositar e documentar os dados (*atenção, não é serviço de box*)
- Ter uma coleção organizado com datasets de um projeto específico
- Processar dados e documentar e depois partilhar em equipa para acesso limitado.
- Apoiar a execução de um plano de dados
- ...

Registo de identificadores persistentes (DOI)

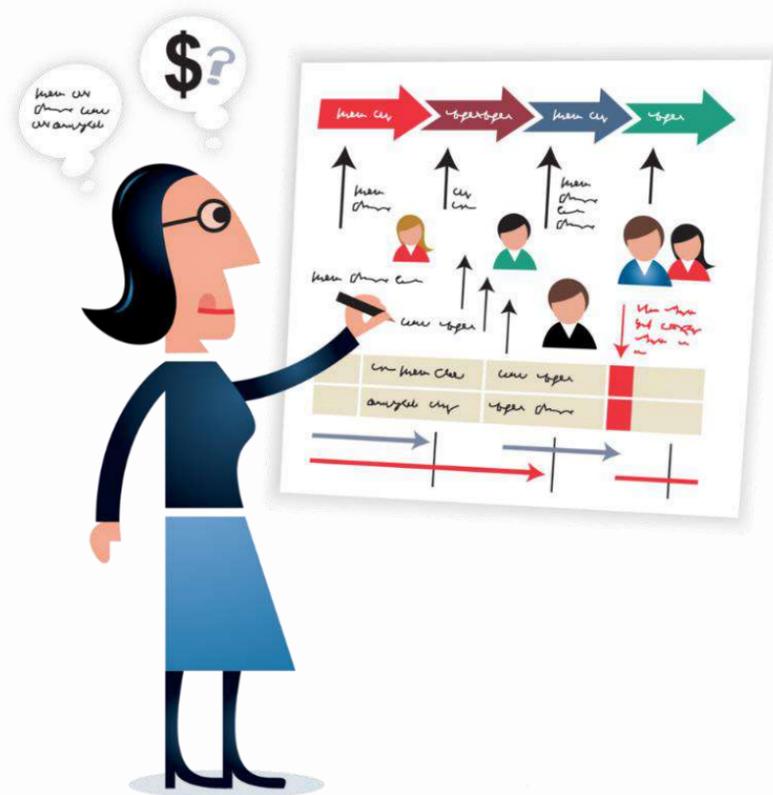


Digitalbevaring.dk

O QUE VEM RESOLVER OU APOIAR:

- Promove a citação de dados
- Apoia a publicação de dados ligados a artigos e resultados em conferências (ex. em poster ou apresentações)
- Requisito básico dos princípios dos dados FAIR (findable & accessible)
- Requisito para report em projetos financiados (FCT, CE)
- ...

Apoiar o cumprimento dos requisitos de financiadores



Digitalbevaring.dk

O QUE VEM RESOLVER OU APOIAR:

- Com facilidade podemos incluir registos de dados nos reports dos projetos (H2020, HE, FCT...).
- Incluir a referência ao repositório da elaboração de um plano de gestão de dados.
- Indicar que a documentação dos datasets é devidamente assegurada pelos metadados no repositório
- ...

Facilitar a disponibilização de dados aos editores



Digitalbevaring.dk

O QUE VEM RESOLVER OU APOIAR:

- O repositório apoia o cumprimento dos requisitos de editores para a disponibilização de dados.
- Permite enviar ou disponibilizar dados para processo de revisão por pares.
- Facilita ligação dos dados aos resultados presentes nos artigos publicados.
- ...

Documentar os datasets com metadados gerais e disciplinares



Digitalbevaring.dk

O QUE VEM RESOLVER OU APOIAR:

- Documentar de forma adequada os datasets gerados na investigação.
- Usar standards de metadados para descrever os conjuntos de dados.
- Poder referenciar o DataRepositóriUM no plano de gestão de dados ou numa proposta de projeto (FCT ou CE).
- ...

Registrar e licenciar os dados para referência, reuso e citação

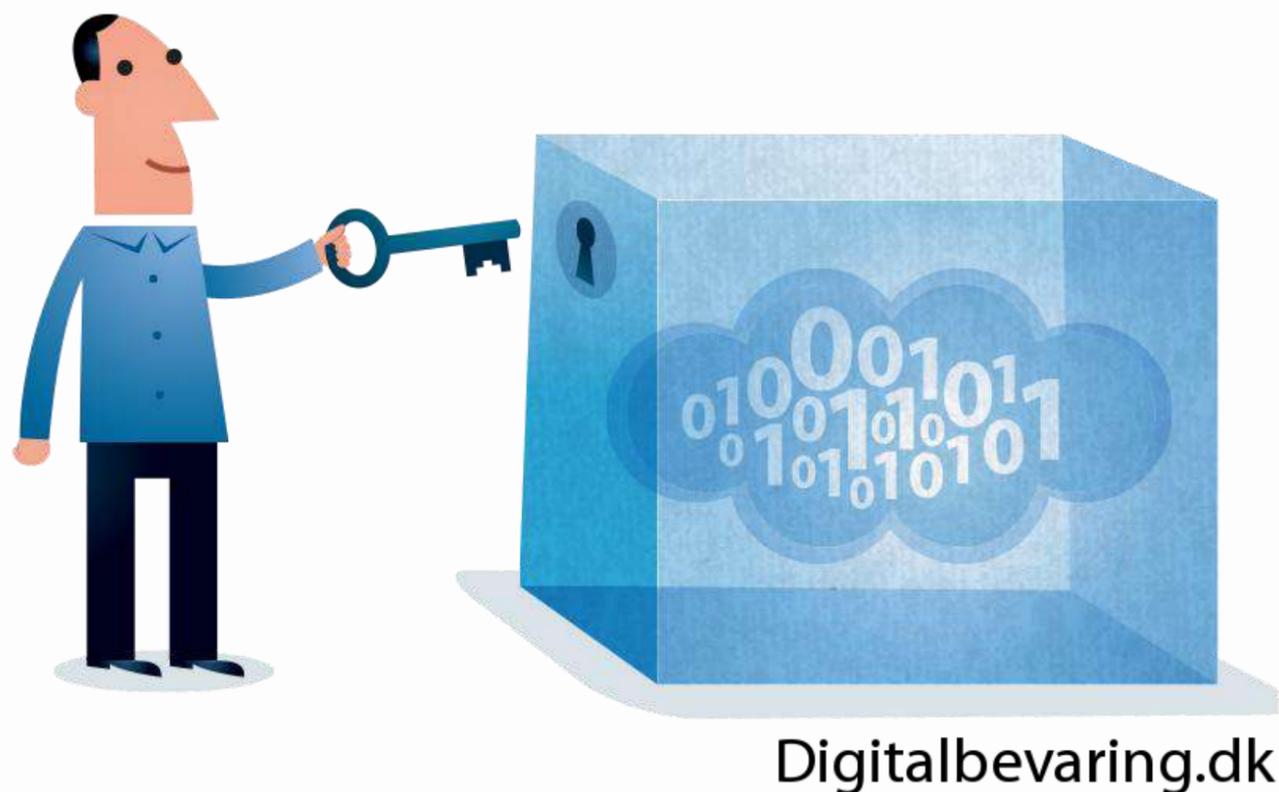


Digitalbevaring.dk

O QUE VEM RESOLVER OU APOIAR:

- Associar licenças para reuso.
- Promover a referenciação e a citação dos dados
- Explicitar os termos de uso.
- Apoiar os investigadores a cumprirem requisitos de reutilização incluídos nas diretrizes de financiadores.
- ...

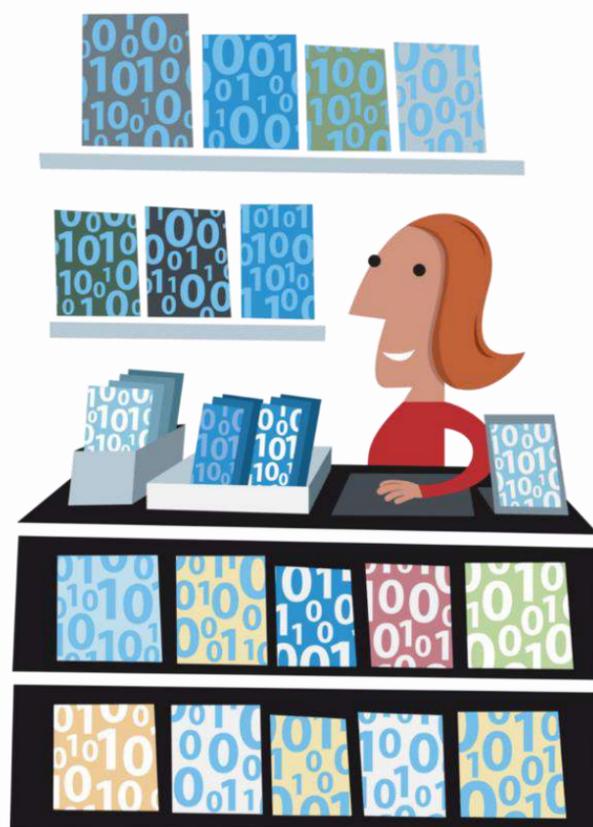
Permitir acesso controlado aos dados ou sob identificação prévia



O QUE VEM RESOLVER OU APOIAR:

- Facilitar de forma clara o princípio de *“As Open as possible, as closed as necessary”*
- Apoiar os investigadores a cumprirem requisitos mesmo que necessitem ter dados com acesso limitado.
- ...

Depositar e publicar dados como parte integrante de teses e dissertações



Digitalbevaring.dk

O QUE VEM RESOLVER OU APOIAR:

- Facilitar a documentação dos dados ao longo do ciclo de vida de investigação numa tese de doutoramento.
- Apoiar a gestão dos dados por via da disponibilização de uma coleção para organizar os datasets.
- Registrar os dados e depois ligar ao produto final (tese)
- ...

Preservação de ficheiros e datasets



Digitalbevaring.dk

O QUE VEM RESOLVER OU APOIAR:

- Repositório confiável de dimensão institucional
- Assegura padrões de sustentabilidade alinhados com requisitos de financiadores
- ...

Facilitar a disseminação dos resultados de investigação



Digitalbevaring.dk

O QUE VEM RESOLVER OU APOIAR:

- Ter um link (URL) para a coleção dos dados de um projeto.
- Poder incluir imagens (logos da instituição, centro ou do financiador) nas páginas das coleções de dados.
- Divulgar os DOI dos conjuntos de dados em conferências ou em reuniões de grupo.
- ...

Licenças para reuso

OPEN SOURCE

CREATIVE COMMONS

Choose an open source license

An open source license protects contributors and users. Businesses and savvy developers won't touch a project without this protection.

Which of the following best describes your situation?

- I need to work in a community.**
Use the **license preferred by the community** you're contributing to or depending on. Your project will fit right in. If you have a dependency that doesn't have a license, ask its maintainers to **add a license**.
- I want it simple and permissive.**
The **MIT License** is short and to the point. It lets people do almost anything they want with your project, including to make and distribute closed source versions. **Babel**, **.NET Core**, and **Rails** use the MIT License.
- I care about sharing improvements.**
The **GNU GPLv3** also lets people do almost anything they want with your project, *except* to distribute closed source versions. **Ansible**, **Bash**, and **GIMP** use the GNU GPLv3.

What if none of these work for me?

- My project isn't software.**
There are licenses for that.
- I want more choices.**
More licenses are available.
- I don't want to choose a license.**
Here's what happens if you don't.

The content of this site is licensed under the Creative Commons Attribution 3.0 Unported License. About Terms of Service Help improve this page Curated with <3 by GitHub, Inc. and You!

<https://choosealicense.com/>

LICENSES

MOST FREE

- ATTRIBUTION CC BY**
This license lets you distribute, remix, tweak, and build upon the original work, even commercially, as long as you credit the original creation. This is the most accommodating of licenses offered.
- ATTRIBUTION-SHAREALIKE CC BY-SA**
This license lets you remix, tweak, and build upon the original work even for commercial purposes, as long as you credit the original work and license your new creations under the identical terms. This license is often compared to "copyleft" free and open source software licenses. All new works based on the work should carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia.
- ATTRIBUTION-NODERIVS CC BY-ND**
This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to the original work.
- ATTRIBUTION-NONCOMMERCIAL CC BY-NC**
This license lets you remix, tweak, and build upon the original work non-commercially. Your new works must be non-commercial and acknowledge the original work, but you don't have to license your derivative works on the same terms.
- ATTRIBUTION-NONCOMMERCIAL-SHAREALIKE CC BY-NC-SA**
This license lets you remix, tweak, and build upon the original work non-commercially, as long as you credit the original work and license your new creations under the identical terms.
- ATTRIBUTION-NONCOMMERCIAL-NODERIVS CC BY-NC-ND**
This license is the most restrictive of the six main licenses, only allowing you to download the original work and share it with others as long as you credit the original work. You can't change the original work in any way or use it commercially.

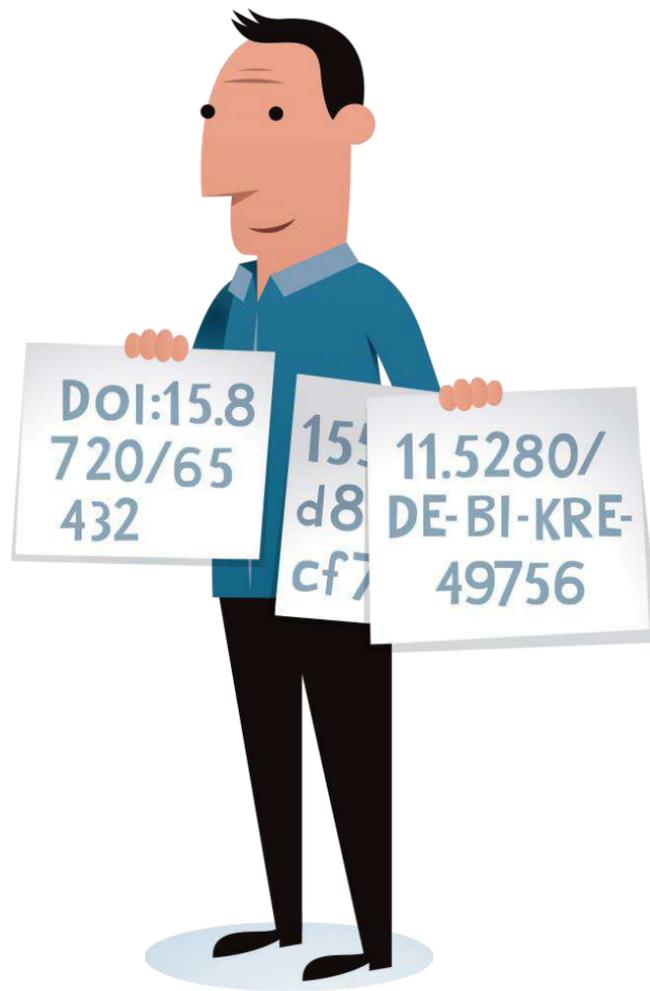
LEAST FREE

<https://creativecommons.org/>



Repositórios de dados registam licenças

QUESTÕES



Digitalbevaring.dk

PUBLICAÇÃO DE DADOS

Obrigado!

Pedro Principe

pedro.principe@usdb.uminho.pt

<https://openaccess.sdum.uminho.pt>