

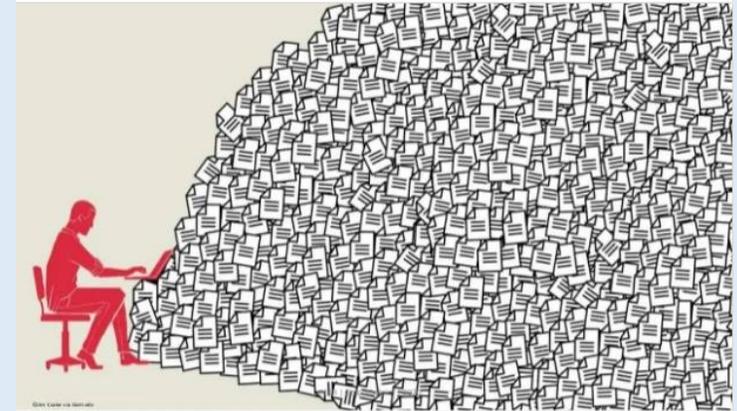
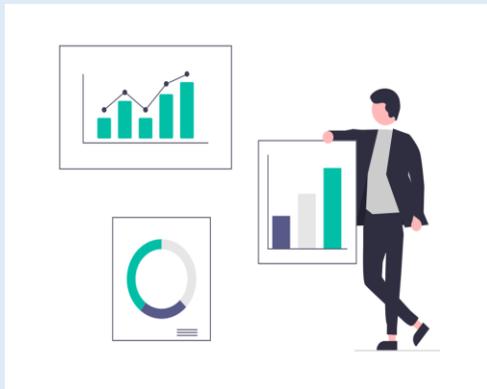
Dados de investigação: o que fazer com eles ?

Conteúdos

- Dados de investigação:
 - Conceito
 - Tipo de dados
 - Princípios FAIR
- Planos de gestão de dados de investigação:
 - O que são
 - Porquê elaborar um PGDI
 - Ferramentas para a elaboração de um plano gestão de dados de investigação
- Importância de depositar os dados de investigação
 - Repositórios

Dados de Investigação

São todos e quaisquer dados resultantes do processo de investigação científica, logo necessários para a validação dos resultados científicos.



Dados de Investigação

- **Dados em bruto** - capturados com recurso a instrumentos tais como telescópios, smartphones, satélites.
- **Visualizações, modelos e algoritmos:** os investigadores também produzem recursos digitais tais como modelos e algoritmos para os auxiliar a analisar, visualizar e apresentar dados em bruto de forma significativa.
- **Imagens, áudio e ficheiros de vídeo:** imagens digitais são consideradas dados, assim como quaisquer ficheiros de vídeo ou áudio capturados no decorrer de uma investigação, tais como as entrevistas gravadas.
- **Qualquer coisa!** essencialmente, dados de investigação podem ser tudo aquilo que os investigadores produzem ou trabalham durante o decurso da sua investigação.

Tipos de dados de investigação

- Classificados de acordo com:
 - A) Grau de processamento
 - B) Proveniência
 - C) Dimensão
 - D) Tipo de Investigação



A) Grau de processamento

- **Em Bruto** - dados de investigação obtidos directamente do processo de investigação, instrumento ou metodologia científica
- **Processados** - dados resultantes da interpretação, processamento ou transformação de dados em bruto

B) Proveniência

- **Primários** - gerados pelos próprios investigadores no decorrer do seu trabalho.
- **Secundários** - dados disponíveis (abertos) reutilizados por outras pessoas que não os seus produtores. Para serem usados necessitam de ter associada documentação de contextualização.

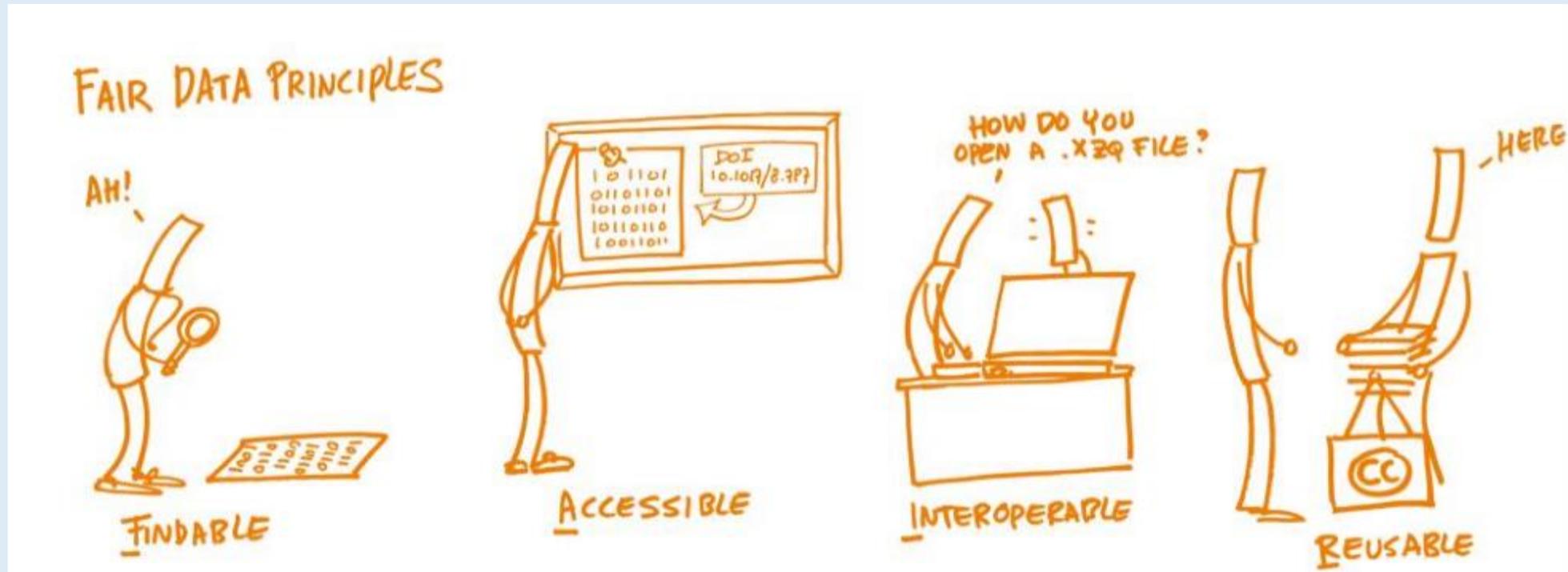
C) Dimensão

- **Big data:** dados (datasets) de grande dimensão, muitas vezes provenientes de instrumentos específicos.
- **Long tail data:** datasets de pequenas dimensões.

D) Tipo de Investigação

- **Dados de observação:** capturados em tempo real, geralmente únicos e insubstituíveis (ex.: imagens cerebrais, dados de inquéritos);
- **Dados experimentais:** recolhidos a partir de equipamentos de laboratório, podendo ser reproduzíveis (ex.: cromatogramas, micro-ensaios);
- **Dados de simulação:** gerados a partir de modelos de teste, onde os modelos e metadados podem ser mais importantes do que os dados resultantes do modelo (p. ex.: modelos económicos e climáticos);
- **Dados derivados ou compilados:** resultam do processamento ou de combinação de dados em bruto, podendo ser reproduzíveis;
- **Dados de referência ou canónicos (estáticos ou orgânicos):** coleções de pequenos conjuntos de dados (revistos por pares), podendo ser publicados e curados (ex.: banco de dados de genes).

Princípios FAIR para os dados de investigação



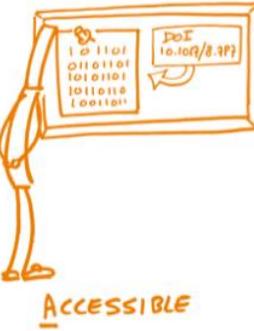
Diretrizes para aplicação dos princípios FAIR à gestão e reutilização de dados
<https://doi.org/10.5281/zenodo.3937183>

Localizáveis



- L1. (meta)dados associados a identificadores globais, únicos e persistentes.
- L2. dados descritos com metadados pormenorizados.
- L3. (meta)dados registados ou indexados em recursos de pesquisa.
- L4. metadados especificando o identificador dos dados.

Acessíveis



- A1. (meta)dados são recuperáveis através do seu identificador usando um protocolo de comunicações normalizado.
- A2. protocolo aberto, gratuito e de implementação universal.
- A3. protocolo permite procedimentos de autenticação e autorização, quando necessário.
- A4. metadados acessíveis, mesmo quando os dados já não estão disponíveis.

Reutilizáveis



- R1. meta(dados) têm uma grande diversidade de atributos precisos e relevantes.
- R1.1. (meta)dados são disponibilizados com uma licença de uso de dados clara e acessível.
- R1.2. (meta)dados têm associada a sua proveniência.
- R1.3. (meta)dados seguem as normas relevantes na comunidade disciplinar.

Interoperáveis



- I1. (meta)dados utilizam uma linguagem formal, acessível, partilhada e largamente aplicável para representação do conhecimento.
- I2. (meta)dados utilizam vocabulários que seguem os princípios FAIR.
- I3. (meta)dados contêm referências qualificadas a outros (meta)dados

Princípios FAIR para os dados de investigação

PREMISSAS:

- São princípios orientadores de alto nível.
- Indicam as opções de implementação e não sugerem nenhuma tecnologia, padrão ou solução de implementação específica.
- Os princípios FAIR não são uma norma, atuam como um guia para quem publica ou administra dados para ajudá-los a avaliar se suas opções de implementação específicas.

Princípios FAIR para os dados de investigação

NÃO ESQUECER!

- FAIR é um **conjunto de princípios**, não é um standard!
- Seguir os princípios FAIR significa que os dados dever ser partilhados com todos? **NÃO**.
- Os dados podem ser FAIR, **mas não complementarmente abertos**. Por exemplo, os dados podem atender aos princípios da FAIR mas serem partilhados sob certas restrições.
- Os dados abertos podem não ser FAIR. Por exemplo, dados públicos podem não ter documentação suficiente sobre licenciamento.

Requisitos de Dados Abertos no H2020 & Horizonte Europa



1. Plano de Gestão de Dados



2. Depositar Dados em Repositórios confiáveis



3. Documentar os dados para validação



4. Assegurar abertura dos dados



Plano de Gestão de Dados de Investigação

(Data Management Plans ou DMPs)

O que são:

- Os planos de gestão de dados (PGD) são documentos formais que descrevem como lidar com dados de pesquisa durante o processo de investigação e após este projeto estar concluído.

Plano de Gestão de Dados de Investigação

(Data Management Plans ou DMPs)

- Descrevem o ciclo de vida dos dados gerados ou recolhidos no âmbito da investigação. são documentos fundamentais para compreensão dos dados e das circunstâncias em que foram gerados, o que poderá permitir a sua reutilização/ replicação.
- Estratégias de gestão de dados adequadas asseguram vários **fatores de valor acrescentado** ao processo de investigação, nomeadamente **a salvaguarda dos dados**, algo muito crítico para qualquer projeto de investigação, independentemente da sua dimensão.

No âmbito do Horizonte Europa: Gestão e abertura de dados obrigatórias

- Gestão de dados de acordo com os **princípios FAIR**
- Apresentação de um Plano de Gestão de Dados nos **primeiros 6 meses do projeto** e **atualização do mesmo**, sempre que houver alterações
- **Depósito dos dados num repositório certificado**
- **Dados abertos** se possível e assim que possível, com atribuição de uma **licença CC BY**, de acordo com o princípio “as open as possible, as closed as necessary”

Gestão dos dados de investigação: importância

- Cumprir os requisitos do financiador;
- Assegurar a integridade e reprodutibilidade da investigação;
- Aumentar a eficiência da investigação;
- Assegurar que os dados e registos de investigação são precisos, completos, autênticos e confiáveis;
- Economizar tempo e recursos a longo prazo;
- Aprimorar a segurança dos dados e minimizar o risco da perda de dados;
- Prevenir a duplicação de esforços, permitindo que outros utilizem os seus dados;
- Cumprir com as práticas conduzidas na indústria e comércio;
- Possibilitar a comparação dos dados nos vários momentos do ciclo de investigação.

“Open Science will *become the new modus operandi* in Horizon Europe”



<https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/home>

https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/aga_en.pdf



https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/agr-contr/general-mga_horizon-auratom_en.pdf

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

- *Os PGDs são documentos vivos que evoluem e adaptam-se à medida que a investigação segue o seu curso.*

CICLO DE VIDA DOS DADOS



Ferramentas para a elaboração de planos de gestão de dados



- DMPonline
<https://dmponline.dcc.ac.uk/>
- ARGOS
<https://argos.openaire.eu/>
- OpenDMP
<https://gitlab.eudat.eu/dmp/Open/RE-EUDAT-DMP-service-pilot>
- Data Stewardship Wizard
<https://ds-wizard.org>
- RDMO - RDM Organiser
<https://rdmorganiser.github.io/en/>
- DMPTool
<https://dmptool.org>
- ezDMP
<https://ezdmp.org>
- easyDMP
<https://easydmp.eudat.eu>
- DMP ARIADNEplus
<https://vast-lab.org/dmp/ariadneplus/form/>

Ferramentas para a elaboração de planos de gestão de dados



Ferramenta desenvolvida pelo Digital Curation Centre para auxiliar na criação de Planos de Gestão de Dados.

DMPonline disponível em <https://dmponline.dcc.ac.uk> é uma ferramenta que integra uma variedade de financiadores de projectos, facilitando aos investigadores a organização do processo de recolha, edição, atualização e partilha de dados com outros investigadores das diferentes versões do projeto e em cada uma das fases, sendo possível a exportação dos dados em diferentes formatos.

Ferramentas para a elaboração de planos de gestão de dados



Serviço OpenAIRE para criar, ligar e partilhar Planos de Gestão de Dados

É um espaço colaborativo para a criação de Planos de Gestão de Dados, é uma ferramenta online de suporte a processos automatizados para criação, gestão, partilha e ligação de PGD com os resultados de investigação a que correspondem. Para além, da ligação às entidades financiadoras e projectos de investigação, o ARGOS permite a utilização de modelos pré-definidos para a criação de um PGD, bem como a criação de novos modelos parametrizados de acordo com as especificações da instituição, projecto, entidade financiadora, entre outros.

<https://argos.openaire.eu>

10 Razões para depositar em Repositórios



Grupo de Trabalho
Formação e competências
para Gestão de Dados FAIR



FÓRUM
GESTÃO DE DADOS
DE INVESTIGAÇÃO

Repositórios e Depósito de Dados

Repositórios de Dados

(algumas dicas para encontrar um repositório de dados):

- 1-Use um **repositório disciplinar, se existir**;
- 2-Em **alternativa** use o **repositório institucional**, ou outro que garanta a preservação a longo prazo;
- 3-Use um repositório generalista, como o **Zenodo**, criado e gerido pelo CERN;
- 4-Ou pesquise no **portal re3data.org** por um repositório que se ajuste às suas necessidades e reúna as condições para o seu depósito;
- 5 –Na maior parte das vezes é o próprio financiador que determina a forma como este processo deverá decorrer. Em todo o caso, **a partilha dos dados através do seu depósito num repositório de dados aberto, potencia a sua partilha e comunicação dos mesmos.**

Onde partilhar e depositar dados

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

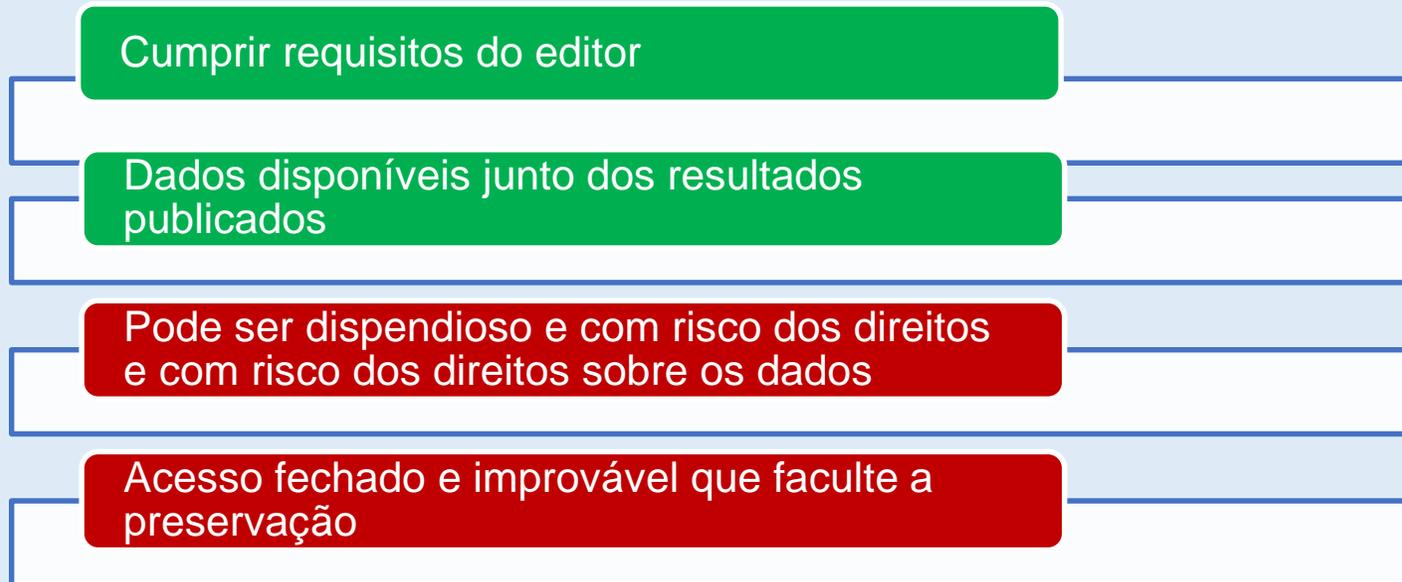
Oferece especialização e experiência em gestão de dados

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

Serviço de revistas para material suplementar, tais como dados de investigação



3 tipos de requisitos:

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

Alerta!

Providenciar serviços nas instituições ou comunidades: os dados devem permanecer na administração dos investigadores e instituições

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.

Repositório genérico

Alcança um público mais vasto

Aceita vários tipos de dados, adequado para dados interdisciplinares

Normalmente, apenas os metadados simples estão disponíveis

Sem controle editorial sobre a qualidade dos materiais depositados

Serviços de repositórios catch-all



Zenodo

Criado e desenvolvido por investigadores com o intuito que possam participar e usufruir na Ciência Aberta.

Repositório aberto multidisciplinar, que permite aos investigadores a partilha e a preservação dos seus resultados de investigação, independentemente do formato e tamanho, e ainda a atribuição de PIDs.

É uma excelente opção para o investigador que não tenha acesso a um repositório adequado ou a sua instituição não disponha de um repositório de dados.

O Projeto OpenAIRE está na vanguarda do acesso Aberto e nos contínuos movimentos associados ao Acesso Abertos aos Dados, com a chancela da União Europeia para apoiar na criação de uma política sólida de Dados Abertos, com a visão de abarcar todos os repositórios de investigação financiados pela EU.

<https://zenodo.org/>

Serviços de repositórios catch-all

Figshare



Assume-se como um repositório onde os utilizadores podem partilhar os resultados da sua investigação e projectos, possibilitando a sua pesquisa, partilha e citação de forma apropriada

<https://figshare.com/>

Dryad



A visão deste repositório é tornar os dados de investigação disponíveis de forma aberta, integrada e passível de serem reutilizados e citados para gerar novo conhecimento

<https://datadryad.org/stash>

- Etc...

Repositório de Dados Institucional

Aceita vários tipos de dados, garante o acesso aos mesmos a longo prazo

Confiável e sem custos

Podem não oferecer acesso sustentável a longo prazo

Podem não ter metadados disciplinares

Repositório disciplinar

Oferece especialização e experiência na gestão de dados

Provável aceitar conjuntos de dados completos

selectivos no tipo de dados que aceitam

Requer planeamento e padrões elevados, podendo implicar custos



- O Re3data é um diretório internacional de repositórios de dados de investigação de diferentes áreas científicas. Permite o armazenamento e a pesquisa dos dados de investigação, a sua partilha, o seu acesso e aumento da sua visibilidade.
- É uma ferramenta útil para investigadores, agências de financiamento, instituições académicas e de investigação científica e editoras.
- <https://www.re3data.org/>

Repositórios confiáveis

Instrumentos de Certificação para auditoria e certificação de repositórios:

✓ **CoreTrustSeal (CTS)**

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



✓ **Nestor Seal**



https://www.langzeitarchivierung.de/Webs/nestor/EN/Home/home_node.html

✓ **ISO 16363: 2013**

<http://www.iso16363.org/>

Repositórios confiáveis

- Repositórios generalistas (Zenodo, FigShare ou Dryad) apesar de não terem o selo da Core Trust Seal, seguem outras normas de confiabilidade para os investigadores, quer sejam os princípios FAIR dos dados, ou outras normas para sistemas de gestão de segurança de informação (ISSO 27001).

Links úteis

Licenciamento



Licenças Creative Commons-Atribuição de licenças de partilha de conteúdos na web.

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



Ciência Aberta- Legislação e mandatos

<https://www.ciencia-aberta.pt/legislacao-e-mandatos>

Links úteis

Políticas de Acesso Aberto da Fundação para a Ciência e Tecnologia

<https://www.fct.pt/acessoaberto/index.phtml.pt>

Amnesia- Serviço OpenAIRE para a anonimização de dados sensíveis

<https://amnesia.openaire.eu/>

Guia: How do I license my research data? Saiba mais sobre licenças para dados de investigação e como as aplicar.

<https://www.openaire.eu/how-do-i-license-my-research-data>

Guia: Can I reuse someone else's research data? Saiba mais sobre como reutilizar dados de investigação

<https://www.openaire.eu/can-i-reuse-someone-else-research-data>

Science Europe-Practical Guide to The International Alignment of Research Data Management

<https://www.scienceeurope.org/our-resources/practical-guide-to-the-international-alignment-of-research-data-management/>

Links úteis

REPOSITÓRIO	INSTITUIÇÃO RESPONSÁVEL	ÁREA TEMÁTICA	LINK
Antimicrobial Combination Networks	Universidade do Minho	Biologia	http://www.sing-group.org/antimicrobialCombination/
DataRepositoriUM	Universidade do Minho	Geral	https://datarepositorium.uminho.pt/
European Archive of Historical Earthquake Data	Múltiplos organismos internacionais	Ciências da Terra	https://www.emidius.eu/AHEAD/
INTEGRALL - The Integron Database	Universidade de Aveiro	Biologia	http://integrall.bio.ua.pt/
KiMoSys - Kinetic Models of biological Systems	Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa	Biologia	https://www.kimosys.org/
Perdigão Field Experiment	Universidade do Porto	Ciências da Terra	https://perdigao.fe.up.pt/
Portulan Clarin Repository	Universidade de Lisboa/ Universidade de Évora	Ciências e tecnologias da linguagem	https://portulanclarin.net/repository/search/
Repositório Dados Científicos	RCAAP/FCT	Geral	https://dados.rcaap.pt/
Arquivo Português de Informação Social	Instituto de Ciências Sociais da Universidade de Lisboa	Ciências Sociais	http://www.apis.ics.ulisboa.pt/en/

Questões



Muito Obrigada!

Patrícia Dias de Melo
p.pequeno@fct.unl.pt