



Re.Data

Rede para a Gestão de
Dados de Investigação

DIRETRIZES PARA A CURADORIA DE METADADOS EM REPOSITÓRIOS DE DADOS DE INVESTIGAÇÃO

Versão 1

Julho 2025

Apoio

FCCN serviços
digitais
fct

fct Fundação
para a Ciência
e a Tecnologia

Financiamento


PRR
Plano de Recuperação
e Resiliência



REPÚBLICA
PORTUGUESA



Financiado pela
União Europeia
NextGenerationEU



Autoria: Ana Inácio, André Vieira, Bruno Almeida, Clara Boavida, Clarisse Pais, Cátia Carvalho, Daniel Alves, Mariana Alves Pereira
Re.Data - Rede para a Gestão de Dados de Investigação
Data: Julho 2025
DOI: <https://doi.org/10.5281/zenodo.16639937>

As Diretrizes para a Curadoria de Metadados em Repositórios de Dados de Investigação tiveram na sua base os seguintes documentos:

Cooper, Alexandra, Steeleworthy, Michael, Paquette-Bigras, Ève, Clary, Erin, MacPherson, Erin, Gillis, Louise, Wilson, Lee, & Brodeur, Jason. (2021). *Dataverse Curation Guide* (Version 01). <https://doi.org/10.5281/zenodo.5579820>

Curtin, Lisa, Herzog, Jessica, Richardson, Joshua, Prabhakaran Nair Sheelakumari, Lekshmi, & Steltenpohl, Crystal (2025) *Best Practices for Data Submission in Generalist Repositories: A Checklist*. <https://doi.org/10.5281/zenodo.14278907>



DIRETRIZES

PARA A CURADORIA DE METADADOS EM REPOSITÓRIOS DE DADOS DE INVESTIGAÇÃO

A curadoria de metadados em repositórios de dados de investigação visa assegurar que estes têm qualidade e que são consistentes, permitindo que sejam partilhados, encontrados, compreendidos e reutilizados de forma eficiente. As diretrizes que aqui se apresentam são fundamentais para a gestão ativa dos dados de investigação ao longo do seu ciclo de vida.

Estas diretrizes são destinadas a profissionais que gerem metadados em repositórios de dados de investigação e contêm instruções relativas às seguintes tarefas:

- Preparação de dados para preservação
- Formatação e organização de ficheiros
- Seleção de estruturas de metadados
- Criação de metadados e descrições
- Utilização de ferramentas e recursos de metadados
- Curadoria de dados
- Adoção de esquemas de identificadores e de vocabulários controlados

DIRETRIZES

ORGANIZADAS EM TRÊS NÍVEIS DE CURADORIA:

NÍVEL
 1

Os dados são
localizáveis?

Através deste nível de curadoria estão garantidos os requisitos mínimos para que o conjunto de dados possa ser localizado.

NÍVEL
 2

Os dados são
utilizáveis?

Através deste nível de curadoria potencia-se a descoberta e garante-se a usabilidade do conjunto de dados.

NÍVEL
 3

Os dados são
preservados e
reprodutíveis?

Através deste nível de curadoria assegura-se a reprodutibilidade e preservação do conjunto de dados.

ANTES DE COMEÇAR

Obtenha informações completas e detalhadas do Projeto de Investigação junto do Investigador/a Responsável ou da Equipa de Investigação para uma correta gestão do conjunto de dados.

Conheça os regulamentos e as políticas da sua instituição em matéria de segurança, preservação e partilha de dados.

Se necessário, consulte o/a Encarregado/a de Proteção de Dados, a Comissão de Ética, o Departamento Jurídico, os Serviços de Apoio à Investigação ou os Serviços de Informação e Documentação da sua Instituição.

Tenha em consideração os requisitos da Entidade Financiadora do Projeto de Investigação.

Analise a política de partilha de dados das Editoras que irão publicar os resultados do Projeto de Investigação.

Refleta sobre a melhor forma de organizar o processo descritivo relativo ao trabalho de curadoria. O documento a produzir deverá sistematizar toda a informação sobre as alterações realizadas ao conjunto de dados, as razões que conduziram a estas ações e sua proveniência.

Nível 1

OS DADOS SÃO
LOCALIZÁVEIS?

NÍVEL 1

OS DADOS SÃO LOCALIZÁVEIS?

DEVERÁ CONFIRMAR QUE ESTÃO REUNIDOS TODOS OS ELEMENTOS NECESSÁRIOS À PUBLICAÇÃO DO CONJUNTO DE DADOS E QUE EXISTE INFORMAÇÃO QUE PERMITE A SUA LOCALIZAÇÃO. DADOS E METADADOS DEVEM SER LEGÍVEIS E FÁCEIS DE ENCONTRAR TANTO POR HUMANOS COMO POR MÁQUINAS.



O conjunto de dados foi submetido na coleção adequada?

Confirme que o/a Investigador/a Responsável ou a Equipa de Investigação depositaram o conjunto de dados na coleção apropriada do Repositório de dados.



A documentação de apoio é fornecida?

Assegure-se de que existe documentação de apoio suficiente para permitir a correta interpretação do conjunto de dados.



O QUE É DOCUMENTAÇÃO DE APOIO?

Entende-se por documentação de apoio toda a informação que permite contextualizar e descrever o conjunto de dados, assim como os materiais necessários para compreender e reutilizar esses dados, tais como:

- Ficheiro README
- Dicionário de dados ou Codebook
- Artigo de dados
- Manual de utilização
- Nota metodológica



Os ficheiros descritos na documentação de apoio estão incluídos no conjunto de dados?

Compare os ficheiros existentes no conjunto de dados com a informação descrita na documentação de apoio e vice-versa. Caso a informação não seja coincidente contacte o/a Investigador/a Responsável ou a Equipa de Investigação para se proceder a uma harmonização.

NÍVEL 1

OS DADOS SÃO LOCALIZÁVEIS?



Os campos obrigatórios dos metadados estão devidamente preenchidos e são precisos?

Assegure que o conjunto de dados é descrito com exatidão.



Os campos de preenchimento obrigatório variam consoante o Repositório de dados.

MAIS INFORMAÇÃO

Cornell University. Metadata tips

<https://guides.library.cornell.edu/ecommons/metadata>

Dataverse North Metadata Best Practices Guide

<https://dx.doi.org/10.14288/1.0388724>

RDA Metadata Standards Catalog

<https://rdamsc.bath.ac.uk>

Datacite Mandatory Properties

<https://datacite-metadata-schema.readthedocs.io/en/4.6/properties/overview/#table-1-datacite-mandatory-properties>

University of North Carolina. Metadata for Data Management

<https://guides.lib.unc.edu/metadata/standards>



Existem riscos associados à divulgação e partilha dos dados?

- Avalie se há limitações à partilha de dados relacionadas com questões legais ou éticas, nomeadamente a existência de dados pessoais ou sensíveis ou se existem questões relacionadas com licenciamento, propriedade intelectual e/ou direitos de autor.
- Certifique-se que existe consentimento para o acesso e reutilização dos dados e que o mesmo é do seu conhecimento. O consentimento deve ser informado, livre, específico a cada uso e inequívoco.

MAIS INFORMAÇÃO

OpenAIRE. How do I know if my research data is protected?

<https://www.openaire.eu/how-do-i-know-if-my-research-data-is-protected>

Nível 2

OS DADOS SÃO
UTILIZÁVEIS?

NÍVEL 2

OS DADOS SÃO UTILIZÁVEIS?

DEVERÁ GARANTIR QUE OS DADOS ESTÃO BEM DESCRITOS E QUE A DOCUMENTAÇÃO DE APOIO NECESSÁRIA ESTÁ DISPONÍVEL PARA UMA CORRETA UTILIZAÇÃO E COMPREENSÃO DOS DADOS.



A documentação de apoio é detalhada, correta e completa?

Reveja a documentação de apoio e se considerar que está incompleta, incoerente ou que não permite a compreensão e interpretação correta do conjunto de dados, corrija-a em colaboração com o/a Investigador/a Responsável ou a Equipa de Investigação.



QUE INFORMAÇÃO DEVE CONSTAR NA DOCUMENTAÇÃO DE APOIO?

- Informação de contexto sobre os dados: objetivo da investigação, circunstância e forma como os dados foram recolhidos ou produzidos, etc.
- Registo da forma como os dados foram modificados ou processados.
- Procedimentos padronizados e documentados sobre decisões, etapas, procedimentos e fluxos de trabalho (protocolos, cadernos de laboratório, etc.).
- Descrição do sistema utilizado na nomenclatura e estruturação dos ficheiros.
- Informação sobre confidencialidade e quaisquer limitações ao uso e partilha do conjunto de dados.
- Explicação dos nomes de descritores e variáveis, informação sobre valores permitidos e unidades de medida, códigos e classificações, se aplicável.
- Descrição dos requisitos necessários para executar qualquer código incluído (sistema operativo, software, etc.).



MAIS INFORMAÇÃO

Cornell University. Writing READMEs for Research Data
<https://data.research.cornell.edu/data-management/sharing/readme/>

University of British Columbia. Creating a README for your dataset
<https://zenodo.org/records/4058972>

Data Documentation Initiative Codebook
<https://ddialliance.org/ddi-codebook>

NÍVEL 2

OS DADOS SÃO UTILIZÁVEIS?



Os ficheiros abrem corretamente e o conteúdo apresenta-se como esperado?

- Descarregue os ficheiros e aceda ao seu conteúdo.
- Apresente soluções para as questões relacionadas com ficheiros em formato proprietário.



Quando se identificam ficheiros em formato proprietário e não é possível a sua disponibilização em formato aberto, assegure que a documentação inclui instruções sobre como podem ser acedidos, como por exemplo que software deve ser utilizado.



A nomeação e estruturação de ficheiros e pastas é adequada?

- Confirme se a estrutura de pastas e ficheiros é lógica e revela de forma clara o seu conteúdo.
- Certifique que as pastas e os ficheiros são nomeados de forma simples e compreensível, de preferência adotando normas e práticas partilhadas pela comunidade, devendo adotar designações concisas e consistentes.

MAIS INFORMAÇÃO

Dataverse NO. Prepare your Data

<https://site.uit.no/dataverseno/deposit/prepare>

DATAACC Project. Naming files and managing versions: good habits

<https://www.dataacc.org/en/best-practices/establishing-data-management-plan/naming-files-managing-versions-good-habits/>

Harvard University. File Naming Conventions

<https://datamanagement.hms.harvard.edu/plan-design/file-naming-conventions>

UK Data Service. Formatting Data. Organising

<https://ukdataservice.ac.uk/learning-hub/research-data-management/format-your-data/organising/>

NÍVEL 2

OS DADOS SÃO UTILIZÁVEIS?



Os metadados são ricos, detalhados, precisos e completos?

Assegure que o conjunto de dados é descrito com rigor e precisão para que possa ser devidamente encontrado, compreendido e agregado por diferentes plataformas.



ALGUMAS DICAS

- Utilizar metadados disciplinares sempre que possível.
- Inserir cada termo no seu campo próprio.
- Usar linguagem rigorosa e específica.
- Garantir que o título do conjunto de dados é coerente com o seu conteúdo.
- Preencher campos como Assunto ou Palavra-chave com termos que descrevam inequivocamente o conjunto de dados.
- Selecionar termos provenientes de vocabulários controlados, sempre que possível no âmbito do domínio científico.



MAIS INFORMAÇÃO

Dataverse North Metadata Best Practices Guide
<https://dx.doi.org/10.14288/1.0388724>

RDA Metadata Standards Catalog
<https://rdamsc.bath.ac.uk>

University of North Carolina. Metadata for Data Management
<https://guides.lib.unc.edu/metadata/standards>

CESSDA VS - Vocabulary Services
<https://vocabularies.cessda.eu/>



Os links para publicações, conjuntos de dados e outros recursos relacionados estão incluídos?

Confirme se os Identificadores Persistentes das publicações ou recursos relacionados - tais como Planos de Gestão de Dados, código, instrumentos de investigação, entre outros - estão associados ao conjunto de dados.



Os identificadores persistentes estão incluídos?

Confirme a atribuição de Identificadores Persistentes ao conjunto de dados e metadados sempre que aplicável.



O QUE SÃO IDENTIFICADORES PERSISTENTES?

Um identificador persistente, também intitulado PID, é uma denominação (referência) única e estável de um recurso digital por meio da atribuição de um código que pode ser referenciado de forma persistente e explícita na Internet, como por exemplo ORCID (identificador de autor), DOI ou Handle (identificador de publicação), entre outros.



Nível 3

OS DADOS SÃO
PRESERVADOS E
REPRODUTÍVEIS?

NÍVEL 3

OS DADOS SÃO PRESERVADOS E REPRODUTÍVEIS?

DEVERÁ ASSEGURAR QUE FORAM DESENVOLVIDAS AS AÇÕES NECESSÁRIAS QUE GARANTEM A PRESERVAÇÃO DOS DADOS E QUE ESTES SÃO PASSÍVEIS DE SER REUTILIZADOS E APLICADOS EM NOVOS CONTEXTOS DE INVESTIGAÇÃO.



Em caso de depósito de código, este está bem depositado e apresenta-se como esperado?

- É recomendada a opção de publicação de código em repositórios específicos, que permitam desenvolvimento permanente.
- Se o código foi produzido especificamente para leitura e análise de um conjunto de dados, ambos devem estar depositados no mesmo repositório de dados. Esta opção deve ser registada na documentação referente ao processo de curadoria.

EXEMPLOS DE REPOSITÓRIOS DE CÓDIGO

Bitbucket	https://bitbucket.org	GitLab	https://gitlab.com
GitHub	https://github.com	Source Forge	https://sourceforge.net



O conjunto de dados submetido ou depositado contém dados sensíveis?

- Assegure que não são disponibilizados dados sensíveis e elementos que permitam a identificação de indivíduos.
- Reitere perante o/a Investigador/a Responsável ou a Equipa de Investigação a sua responsabilidade ao nível do cumprimento de termos de uso, consentimentos informados e acordos de partilha de dados.

MAIS INFORMAÇÃO

DARIAH ELDAH Consent Form Wizard

<https://consent.dariah.eu>

OpenAIRE. Sensitive Data

<https://www.openaire.eu/sensitive-data-guide>



O QUE SÃO DADOS SENSÍVEIS?

Entende-se por dados sensíveis toda a informação que permite identificar de forma direta e indireta um indivíduo, grupo ou comunidade, tais como dados pessoais, dados biológicos, dados confidenciais, dados de segurança, entre outros.

NÍVEL 3

OS DADOS SÃO PRESERVADOS E REPRODUTÍVEIS?



O conjunto de dados contém dados ou código de outras fontes?

- Verifique se o conjunto de dados contém dados ou código de terceiros e certifique que o/a Investigador/a Responsável ou a Equipa de Investigação têm as devidas autorizações para partilha.
- Assegure que são respeitados os termos de uso definidos pelos/as autores/as dos dados ou código quando se verifica a existência de outras fontes.

MAIS INFORMAÇÃO

OpenAIRE. How do I know my research data is protected?

<https://www.openaire.eu/how-do-i-know-if-my-research-data-is-protected>

OpenAIRE. Can I reuse someone else's research data?

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



Os formatos dos ficheiros são abertos ou são bem documentados?

- Garanta que os dados são disponibilizados em formato aberto, assegurando a possibilidade de uso generalizado e preservação dos dados a longo prazo.
- No caso de existirem ficheiros em formatos proprietários, verifique se estes podem ser transformados em formatos não proprietários sem que ocorra perda de informação. Se não for possível transformar os ficheiros de formatos proprietários, a documentação de apoio deverá fornecer toda a informação necessária sobre o formato e conteúdo dos ficheiros, o software necessário para apresentar ou manipular os ficheiros e, se disponível, uma alternativa gratuita para aceder aos ficheiros.

MAIS INFORMAÇÃO

Cornell University. Recommended File Formats

<https://guides.library.cornell.edu/ecommons/formats>

DANS. File formats

<https://dans.knaw.nl/en/file-formats/>

Excel Archival Tool

<https://github.com/mcgrory/ExcelArchivalTool>

Fórum GDI. Como preparar dados para depositar e publicar num repositório

<https://zenodo.org/records/7082551>

Open Data Handbook. File formats

<https://opendatahandbook.org/guide/en/appendices/file-formats/>

OpenAIRE. Data formats for preservation

<https://www.openaire.eu/data-formats-preservation-guide>

UK Data Services. File formats

<https://ukdataservice.ac.uk/learning-hub/research-data-management/format-your-data/file-formats/>

NÍVEL 3

OS DADOS SÃO PRESERVADOS E REPRODUTÍVEIS?



Os termos de licenciamento e utilização do conjunto de dados foram revistos?

Reveja os termos de uso e as licenças atribuídas ou, se necessário, forneça informação de licenciamento adequada ao conjunto de dados.

MAIS INFORMAÇÃO

Creative Commons

<https://creativecommons.org>

OpenAIRE. How do I license my research data?

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

Choose a License

<https://creativecommons.org/chooser>

Open Data Commons

<https://opendatacommons.org>

Choose an open source license

<https://choosealicense.com>



O conjunto de dados está em conformidade com os princípios FAIR?

Avalie se o conjunto de dados está alinhado com os princípios FAIR:

F Findable (Localizáveis)	Os dados e metadados são legíveis e fáceis de encontrar tanto por humanos como por máquinas
A Accessible (Acessíveis)	Os dados são disponibilizados através de protocolos normalizados e de um processo claramente definido. Os metadados devem estar acessíveis, mesmo quando o conjunto de dados não está disponível
I Interoperable (Interoperáveis)	Os dados são estruturados e organizados utilizando métodos normalizados por forma a serem utilizados em sucessivas aplicações e fluxos de trabalho e combinados com outros conjuntos de dados e sistemas
R Reusable (Reutilizáveis)	Os dados são descritos com um nível de pormenor e é disponibilizada uma licença clara e acessível que permite a sua reprodução e aplicação em novos contextos

MAIS INFORMAÇÃO

OpenAire. How to make your data FAIR

<https://www.openaire.eu/how-to-make-your-data-fair>

ARDC FAIR Data Self-Assessment Tool

<https://ardc.edu.au/resource/fair-data-self-assessment-tool>

FORCE11. The FAIR Data Principles

<https://force11.org/info/guiding-principles-for-findable-accessible-interoperable-and-re-usable-data-publishing-version-b1-0>

FAIRsharing

<https://fairsharing.org>

FAIR-Impact

<https://fair-impact.eu/>

F-UJI

<https://www.f-uji.net>

RECURSOS

IMPORTANTES

Fórum GDI. Guia para a curadoria de dados em repositórios

<https://zenodo.org/records/8340747>

DataverseNO. Prepare your Data

<https://site.uit.no/dataverseno/deposit/prepare/>

DataverseNO. Curation Guidelines

<https://site.uit.no/dataverseno/admin-en/curatorguide/>

Purdue University. Data Curation Profile Toolkit

<https://docs.lib.purdue.edu/dcptoolkit/>

DRYAD. Good Data Practices

https://datadryad.org/best_practices

DataONE. Best Practices of Data Management

<https://dataoneorg.github.io/Education/bestpractices/>

Data Curation Network. Data Curation Primers

<https://datacurationnetwork.org/outputs/data-curation-primers/>

Data Curation Network. The CURATE(D) Steps

<https://datacurationnetwork.github.io/CURATED/>

ANTES DE TERMINAR

Esteja atento à publicação de novas versões do conjunto de dados e mantenha o registo do processo de curadoria atualizado.

Sempre que forem integrados novos membros na Equipa de Investigação garanta que lhes são transmitidos os procedimentos relativos à prática de Gestão de Dados de Investigação.

Sensibilize o/a Investigador/a Responsável e a Equipa de Investigação a partilhar os DOI, ou outro identificador persistente atribuído ao conjunto de dados, no CV, redes sociais, páginas web institucionais, perfis profissionais, sistemas de gestão de informação científica, entre outros.