

SEMINÁRIO

INTRODUÇÃO À GESTÃO DE DADOS DE INVESTIGAÇÃO

PRINCÍPIOS E BOAS PRÁTICAS

JOANA RODRIGUES

(FACULDADE DE LETRAS DA UNIVERSIDADE PORTO
DEPARTAMENTO DE CIÊNCIAS DA COMUNICAÇÃO
E DA INFORMAÇÃO)

jsrodrigues@letras.up.pt

O que são dados?

Dados são a base de toda a investigação científica — não apenas “resultados”, mas **matéria-prima do conhecimento**.



Dados = representações de observações, medições ou factos, recolhidos para análise.

O que são dados?

Tipos de dados:

- **Quantitativos vs Qualitativos**

↪ valores numéricos (ex: temperatura: 20°C) vs qualidades ou características (ex: cor dos olhos: castanho)

- **Estruturados vs Não Estruturados**

↪ dados organizados em MySQL vs anotações de observações

- **Experimentais:** recolhidos através de experiências controladas (ex: testar um medicamento em laboratório)
- **Observacionais:** recolhidos sem interferência (ex: registar temperaturas ao longo do tempo)
- **Simulados:** gerados por modelos computacionais que imitam a realidade (ex: modelação do tráfego numa cidade)



O que são dados?

Exemplos:

- Entrevistas
- Imagens de manuscritos
- Bases de dados linguísticas
- Medições laboratoriais
- Registos de utilização de apps
- Dados meteorológicos
- Dados de censos populacionais
- Sequências de ADN utilizadas em genética
- Dados de sensores de veículos
- Vídeos
- Documentos de arquivo (cartas, registos históricos, tratados)
- Textos literários e obras históricas

Dados como infraestrutura

- **Dados constituem uma infraestrutura global**
 - Sensores, mapas, estatísticas, saúde, resultados de investigação
 - Sociedade e economia dependentes de dados
- **Dados de investigação**
 - COVID-19 demonstrou como os dados e a sua partilha são essenciais
- **Construir e manter infraestrutura global**
 - Desafios: políticos, disciplinares, técnicos, legais

Dados de investigação (research data)

Dados de investigação são todos os dados produzidos ou utilizados durante o processo científico que servem para validar resultados.

Inclui:

- dados brutos
- dados processados
- código
- documentação

Conjuntos de dados (Datasets) vs Bases de Dados (Databases)

Conjunto de dados (dataset)

- Coleção de dados sobre um tema específico
- Normalmente estático
- Pode ser um ficheiro (CSV, Excel, etc.) ou vários
- O foco está no conteúdo

Exemplo: lista de alunos com as notas num ficheiro do Microsoft Excel

Base de Dados (database)

- Sistema organizado para armazenar, gerir e atualizar dados
- Pode conter vários datasets
- Permite consultas, alterações e relações entre dados
- O foco está na estrutura + gestão

Exemplo: sistema em MySQL



Conjuntos de dados (Datasets) vs Bases de Dados (Databases)

Diferença chave

Dataset → conjunto de dados (conteúdo isolado)

Database → sistema que guarda e organiza um ou vários datasets



Exemplo:

Dataset = uma folha de Excel

Database = todo o sistema que guarda várias folhas de Excel e permite pesquisar nelas



O que é a Gestão de Dados de Investigação (GDI)/Research Data Management (RDM)?

Gestão de Dados de Investigação (GDI) é o conjunto de práticas que garantem que os dados são organizados, documentados, preservados e partilhados de forma eficaz ao longo do seu ciclo de vida.



Não é uma tarefa extra — é parte integrante da investigação.

Contexto atual

Gerir dados já não é opcional — é parte do processo científico.

Políticas e tendências:

- **European Commission** → Open Science

https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science_en

- **Horizon Europe** → exige gestão de dados

https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_en

- **European Open Science Cloud** → partilha de dados

<https://digital-strategy.ec.europa.eu/pt/policies/open-science-cloud>

Importância da gestão de dados para a ciência



Partilha



Reutilização dos dados



Citação dos dados



Menos dados/pesquisas duplicados



Comparação de resultados pelos pares



Reprodutibilidade da ciência



Transparência



Impacto científico



Colaboração



Problemas comuns

Gerir dados é importante e exige método e disciplina.

- **Dados perdidos ou inacessíveis** – *já recolhi estes dados há tanto tempo, sei lá onde os guardei.*
- **Ficheiros sem organização** – *tenho os dados todos guardados na mesma pen, sem organização, agora não os consigo agrupar para publicar.*
- **Falta de documentação** – *estou no terceiro ano do projeto, já não me lembro desses dados que recolhi no 1º ano.*
- **Dependência de pessoas** – *esses dados estavam no computador do outro investigador.*

O que dizem as grandes instituições

Algumas instituições que promovem a GDI:

- European Commission
- Horizon Europe
- Organisation for Economic Co-operation and Development (OECD)
- UNESCO

Mensagens a destacar:

- Dados devem ser **geridos desde o início**
- Devem ser **documentados e preservados**
- Devem ser **tão abertos quanto possível, tão fechados quanto necessário**
- Necessidade de **planos de gestão de dados**

Princípios gerais de uma boa gestão de dados

Não existe um padrão rígido de princípios ou orientações, mas o estabelecimento de alguns hábitos pode beneficiar o processo científico.

Exemplos:

- Organização consistente
- Documentação clara
- Versionamento
- Segurança
- Sustentabilidade (pensar no futuro dos dados)

Princípios FAIR – o centro conceptual

FAIR = princípios para tornar os dados úteis para humanos e máquinas

F – Findable (Encontráveis)

- identificadores (DOI)
- metadados

A – Accessible (Acessíveis)

- dados disponíveis (mesmo que com restrições)

I – Interoperable (Interoperáveis)

- formatos standard
- vocabulários controlados

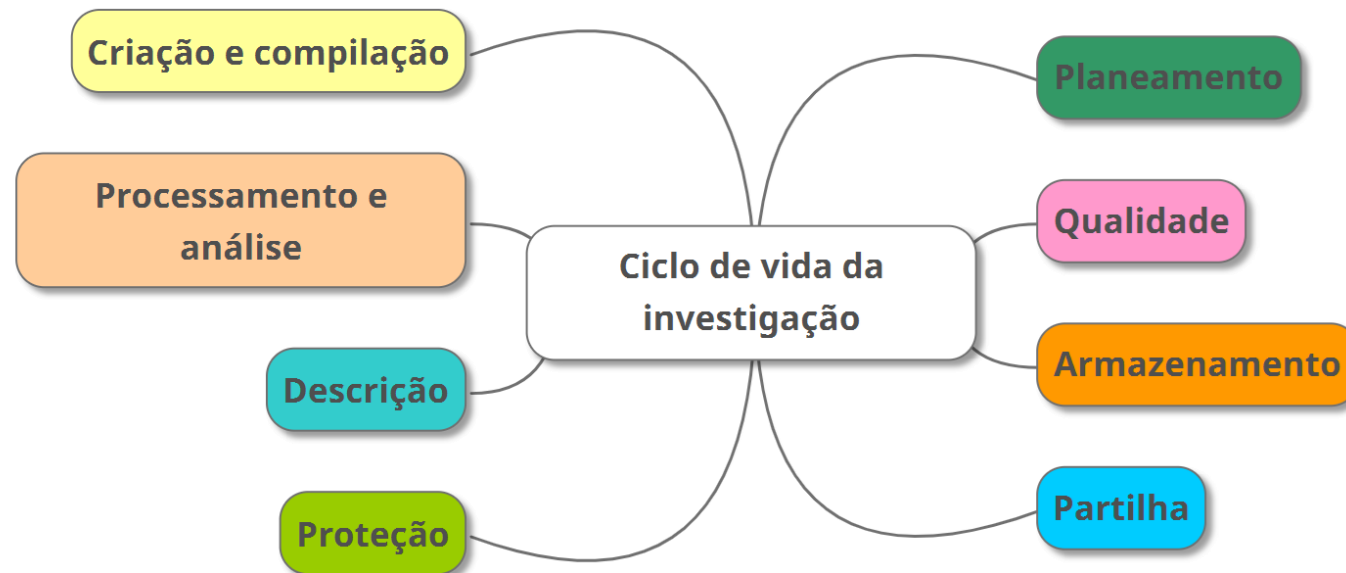
R – Reusable (Reutilizáveis)

- documentação clara
- licenças



Ligar os princípios FAIR ao ciclo de vida da investigação

FAIR não é um passo final → aplica-se ao longo de todo o ciclo de vida



Adaptado dos modelos do DDI, DataONE e UK Data Archive



Ligar os princípios FAIR ao ciclo de vida da investigação

Planeamento

O que fazer:

- Criar um Plano de Gestão de Dados
- Definir:
 - tipos de dados (texto, imagem, bases de dados, etc.)
 - formatos (preferir abertos: CSV, TXT, XML, TIFF...)
 - organização de ficheiros (estrutura de pastas)
- Identificar:
 - questões éticas (consentimento, anonimização)
 - direitos de autor
- Escolher repositório(s) onde os dados poderão ser depositados

Benefícios:

- Evita perda de tempo mais tarde
- Facilita candidaturas a financiamento (ex: Horizon Europe)
- Reduz risco de problemas legais/éticos

Ligar os princípios FAIR ao ciclo de vida da investigação

Criação e compilação

O que fazer:

- Usar convenções de nomes consistentes
 - ex: entrevistas_2024_porto_v1.wav
- Criar dados já em formatos sustentáveis
- Registrar contexto:
 - quem recolheu
 - quando
 - como
- Usar identificadores únicos para ficheiros/datasets

Benefícios:

- Dados mais fáceis de entender (no momento e ao longo do tempo)
- Menos erros e duplicações
- Base sólida para FAIR



Ligar os princípios FAIR ao ciclo de vida da investigação

Processamento e Análise

O que fazer:

- Documentar:
 - scripts (R, Python, etc.)
 - decisões analíticas
 - procedimentos de extração e análise
- Implementar versionamento
 - ex: v1, v2, final
- Manter dados brutos separados dos dados processados
- Registrar workflows (passagem de dados brutos → resultados)

Benefícios:

- Reprodutibilidade
- Transparência científica
- Facilita colaboração

Ligar os princípios FAIR ao ciclo de vida da investigação

Qualidade

O que fazer:

- Validar dados:
 - valores em falta
 - inconsistências
- Normalizar formatos
- Rever manualmente (especialmente em dados qualitativos)
- Documentar limitações e alterações dos dados

Benefícios:

- Maior confiança nos resultados
- Evita erros em publicações
- Dados mais reutilizáveis

Ligar os princípios FAIR ao ciclo de vida da investigação

Descrição

O que fazer:

- Criar:
 - ficheiro README
 - dicionário de dados
- Descrever através de modelos de metadados

Benefícios:

- Dados tornam-se encontráveis (Findable)
- Outros conseguem compreender e reutilizar
- Aumenta impacto científico

Ligar os princípios FAIR ao ciclo de vida da investigação

Armazenamento

O que fazer:

- Regra 3-2-1:
 - 3 cópias (1 original + 2 cópias de segurança)
 - 2 suportes diferentes (dois armazenamentos distintos. Exemplo: HDD + SSD)
 - 1 fora do local (uma cópia noutra local. Exemplo: cloud)
- Usar armazenamento institucional sempre que possível
- Definir níveis de acesso (especialmente para dados sensíveis) e de segurança
- Automatizar backups

Benefícios:

- Evita perda de dados (um dos maiores problemas reais)
- Garante continuidade de projetos
- Protege contra falhas técnicas

Ligar os princípios FAIR ao ciclo de vida da investigação

Partilha

O que fazer:

- Depositar dados em repositórios (institucionais ou temáticos)
- Atribuir DOI (identificador persistente)
- Definir licença (ex: CC-BY)
- Tornar os dados citáveis
 - dados devem ser citados como artigos
 - incluir referência em publicações

Dados como requisito de publicação:

- Muitas revistas exigem: data availability statement e acesso aos dados subjacentes

Benefícios:

- Maior visibilidade (datasets também são citados)
- Aumento de citações dos artigos
- Cumprimento de requisitos de revistas e financiadores
- Colaboração

Ligar os princípios FAIR ao ciclo de vida da investigação

Proteção

O que fazer:

- Aplicar anonimização/pseudonimização
- Garantir conformidade com RGPD
- Definir quem pode aceder aos dados
- Usar termos de consentimento informado (quando aplicável)
- Recorrer a Comissões de Ética e a Encarregados de Proteção de Dados – DPO (quando aplicável)
- Desenvolver uma Privacy Impact Assessment (PIA)* (quando aplicável)

Benefícios:

- Evita problemas legais
- Protege participantes
- Permite partilha segura

Ligar os princípios FAIR ao ciclo de vida da investigação

Erros comuns:

- Só pensar em dados no final
- FAIR como “checklist”
- Falta de documentação
- Ignorar dados sensíveis
- Querer os dados só para si

Ligar os princípios FAIR ao ciclo de vida da investigação

Mensagens importantes:

- A boa gestão de dados não acontece no fim do projeto — começa no planeamento e acompanha todo o ciclo de investigação.
- Cada fase do ciclo de vida dos dados é uma oportunidade para aumentar a qualidade, o impacto e a credibilidade da investigação.

Planear, descrever e partilhar

Planeamento

Descrição

Partilha

Planeamento → Planos de Gestão de Dados (PGD)/ Data Management Plan (DMP)

O que é um PGD?

- É um documento/guia que descreve como os dados serão geridos ao longo do ciclo de vida do projeto de investigação.
- Um PGD não é burocracia — é uma ferramenta de planeamento científico.
- Permite ter uma estrutura flexível e adaptada ao projeto e ao contexto dos dados.

Planeamento → Planos de Gestão de Dados (PGD)/ Data Management Plan (DMP)

O que é um PGD?

- Requisito essencial para a gestão de dados no contexto Horizonte Europa.
- Chave para a gestão e partilha de dados.
- Documento vivo (de trabalho).

Componentes de um PGD

1. Tipos de dados

- que dados vão ser produzidos/recolhidos
- formatos (ex: áudio, texto, imagens...)

2. Documentação e metadados

- como os dados serão descritos
- modelos a usar

3. Armazenamento e backup

- onde os dados vão estar durante o projeto
- nomenclatura para dados e conjuntos de dados
- políticas de segurança

4. Questões éticas e legais

- dados pessoais
- consentimento informado
- direitos de autor

5. Partilha e acesso

- se, como e quando os dados serão partilhados
- possíveis restrições
- licenças

6. Preservação a longo prazo

- onde os dados serão depositados
- por quanto tempo



Ferramentas úteis

- **DMPonline** (Digital Curation Centre)

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

- **Argos** (OpenAIRE)

<https://argos.openaire.eu/home>

- **FCT** (FCT - unidade de serviços digitais FCCN)

<https://blueprint.polen.fccn.pt/home>

Planos de Gestão de Dados

Erros comuns:

- Fazer o PGD “à pressa” só para candidatura
- Não atualizar ao longo do projeto
- Ser demasiado genérico

Benefícios:

- Melhor organização desde o início
- Menos problemas éticos/legal
- Alinhamento com requisitos da European Commission
- Poupança de tempo no fim do projeto

Descrição → Metadados

O que são metadados?

- Dados sobre dados — descrevem conteúdo, contexto, estrutura e proveniência.
- Sem metadados, os dados praticamente “não existem” ou estão “invisíveis”.

Descrição → Metadados

1

INFORMAÇÃO NUTRICIONAL

Porção: 25 g (17 unidades)

	Quantidade por porção	%VD(*)
Valor energético	126kcal = 529 KJ	6
Carboidratos	13 g	4
Proteínas	1,0 g	1
Gorduras totais	7,8 g	14
Gorduras saturadas	3,4 g	15
Gorduras trans	0 g	
Fibra alimentar	0,8 g	3
Sódio	150 mg	8

*Valores Diários de referência com base em uma dieta de 2000 kcal ou 8400 kJ. Seus valores diários podem ser maiores ou menores, dependendo de suas necessidades energéticas.

Tabela 1: exemplo de informações nutricionais de acordo com a atual legislação

Livros da Índia

MEÇÃO DE RESPONSABILIDADE:
Autor: **Barahona, António, 1939-**

PUBLICAÇÃO:
Lisboa : Imprensa Nacional-Casa da Moeda , imp. 1984

COLEÇÃO:
Biblioteca de autores portugueses

ASSUNTO:
Poesia portuguesa

NOTAS GERAIS:
Ex. 2 no depósito

CDU:
821.134.3-1

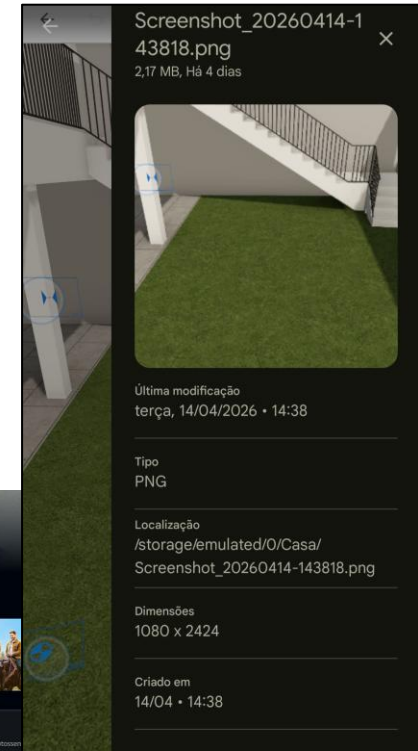
DESCRIÇÃO FÍSICA:
119, [9] p. ; 24 cm.

REGISTO:
13354

COTA:
BMV A 821.134.3-1 BAR

BMV-DEP A 821.134.3-1 BAR

DOCUMENTOS:



ABERTO	
Título:	Describing Data in Image Format: Proposal of a Metadata Model and Controlled Vocabularies
Data de publicação:	2022-09-09
Resumo:	Research data management (RDM) includes people with different needs, specific scientific contexts, and diverse requirements. The description is a big challenge in the domain of RDM. Metadata plays an essential role, allowing the inclusion of essential information for the interpretation of data, enhances the reuse of data and its preservation. The establishment of metadata models can facilitate the process of description and contribute to an improvement in the quality of metadata. When we talk about image data, the task is even more difficult, as there are no explicit recommendations to guide image management. In this work, we present a proposal for a metadata model for image description. To validate the model, we followed an experiment of data description, where eleven participants described images from their research projects, using a metadata model proposed. The experiment shows that participants do not have formal practices for describing their imagery data. Yet, they provided valuable contributions and recommendations to the final definition of a metadata model for image description, to date nonexistent. We also developed controlled vocabularies for some descriptors. These vocabularies aim to improve the image description process, facilitate metadata model interpretation, and reduce the time and effort devoted to data description. (c) 2022 Joana Rodrigues and Carla Teixeira Lopes Published with license by Taylor & Francis Group, LLC.
DOI:	10.1080/19386389.2022.2117511
URI:	https://hdl.handle.net/10216/152743
Tipo de Documento:	Artigo em Revista Científica Internacional
Condições de Acesso:	openAccess
Aparece nas coleções:	FEUP - Artigo em Revista Científica Internacional

Update: New response to your poll Reunião XXIII Jornadas de Ciência da Informação (Ciência da Informação)



Doodle <mailer@doodle.com>
para mim



de: **Doodle** <mailer@doodle.com>
para: joanasousarodrigues.14@gmail.com
data: 16/04/2026, 10:25
assunto: Update: New response to your poll Reunião XXIII Jornadas de Ciência da Informação
enviado por: doodle.com
assinada por: doodle.com
segurança: Criptação padrão (TLS) Saiba mais
: Sem categoria definida, mas importante segundo a Google.



Descrição → Metadados

Tipos de metadatos:

- **Descritivos** → explicam o conteúdo do recurso
título, autor, palavras-chave...
- **Técnicos** → indicam como o ficheiro foi criado e como funciona
formato, tamanho, software...
- **Administrativos** → usados para gerir o recurso
direitos, permissões, datas de acesso...
- **Proveniência** → mostram a origem e o histórico do recurso
nome do criador, histórico de alterações...



Descrição → Metadados

Research data management in image format - Survey

These datasets were the result of research on research data management in image format. Based on the data collected, it was possible to study the practices and habits in the management of research data in image format. The data were collected within the scope of a Master's dissertation in Information Science from the Faculty of Engineering of the University of Porto.

Data and Resources

- dedosfonte_questionario**
File with responses collected through the questionnaire. [Explore](#)
- dedosanalise_percentagem_mediana**
General tendency in the overall set of answers and comparison between 4... [Explore](#)
- kruskal-wallis_frequenciauso**
Results of Kruskal-Wallis method. Used to detect differences among rates of... [Explore](#)
- kruskal-wallis_arealotificao**
Results of Kruskal-Wallis method. Used to detect differences among domains. In... [Explore](#)
- multiproportionaltest_perguntarespostaaberta**
We used a Chi-square test for equality of proportions. This file present the... [Explore](#)
- multiproportionaltest_perguntaselecao**
We used a Chi-square test for equality of proportions. This file present the... [Explore](#)

Image Management | Research Data Management

Additional Info

Field	Value
Author	Miguel Fernandes, Joana Rodrigues, Carla Teixeira Lopes
Last Updated	June 4, 2020, 15:49 (UTC)
Created	June 4, 2020, 10:26 (UTC)
CiteAs	FERNANDES, M., RODRIGUES, J., LOPES, C.T. Research data management in image format - Survey [dataset]. 04 jun 2020. INESC TEC research data repository. DOI: https://doi.org/10.25747/7ma9-9132
DOI	https://doi.org/10.25747/7ma9-9132
do.Coverade.Temporal	20/02/2020 - 26/03/2020
do.Coverage.Spatial	University of Porto research community, namely IS3 and INESC-TEC, and in research units outside the University of Porto funded in 2019 by the Foundation for Science and Technology. Namely, Centre for Informatics and Systems of the University of Coimbra, Cardiovascular Centre of the University of Lisbon, Centre for Philosophical and Humanistic Studies of Universidade Católica Portuguesa and Center for Mathematics and Applications of Universidade Nova de Lisboa.
do.Date	April 2020
do.Format	*.xlsx
do.Format.Extent	9,50 MB
do.Language	PT, EN
do.Publisher	INESC-TEC/FEUP
do.Relation	Miguel Fernandes, Joana Rodrigues and Carla Lopes. Management of research data in image format: an exploratory study on current practices, 24th International Conference on Theory and Practice on Digital Libraries (paper accepted and waiting publication)
do.Type	Digital document - Spreadsheet with results obtained and statistical analysis
ddi.InstrumentName	Google Forms
ddi.InstrumentType	Questionnaire
ddi.Software	The data were analyzed using SPSS

Ultra-high-accuracy X-ray spectroscopy of transition metal oxides and rare earths

POLEN DataHub > Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa > LIBPhys-UNL > Ultra-high-accuracy X-ray spectroscopy of transition metal oxides and rare earths >

Cadmium

Access Dataset | Contact Owner | Share

Dataset Metrics | 15 Downloads

Description | Cadmium (2023-05-31)

Subject | Exact Sciences - Physical Sciences - Atomic Physics

Keyword | X-ray interaction with matter, X-ray emission, Auger Emission, Fluorescence Yields, Line Energies, Line Widths, Spectral Shapes, Satellite Lines, Shake-Up, Shake-Off, Atomic spectra

Related Publication | Ho et al., 2021 *Intensity Ratio of K β K α in Selected Elements from Mg to Cu, and the Chemical Effects of Cr K α 1,2 Diagram Lines and Cr K β K α Intensity Ratio in Cr Compounds*. *International Journal of Molecular Sciences*, 24(6), 6670. doi: 10.3390/ijms24066670

License/Data Use Agreement | CC BY 4.0

Files | Metadata | Terms | Versions

This dataset has been configured to use **English** as the language for all metadata entries.

Citation Metadata

Persistent Identifier	doi: 10.57979/OKSRDJ
Publication Date	2028-03-26
Title	Cadmium
Author	Guerra, Mauro Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa ORCID 691D-6398-3993 Machado, Jorge Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa ORCID E21E-1215-6A7C Pinheiro, Daniel Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa ORCID D61C-7D90-52C8 Baptista, Gonçalo Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa ORCID B817-C368-8962 Godinho, César Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa ORCID 181D-118A-02D0 Fernandes, André Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa ORCID BA14-D34D-71D8 Amaro, Pedro Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa ORCID 2F12-1F5A-5C2D Sampaio, Jorge Miguel Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa ORCID A81A-DF21-3B11 Marques, José Pires Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa ORCID 0317-318E-6C47 Parente, Fernando Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa ORCID F919-AF9F-93E4 Santos, José Paulo Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa ORCID 941F-953C-9FDC
Point of Contact	Use email button above to contact. Mauro Guerra (Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa)
Description	Cadmium (2023-05-31)
Subject	Exact Sciences - Physical Sciences - Atomic Physics
Keyword	X-ray interaction with matter X-ray emission Auger Emission Fluorescence Yields Line Energies Line Widths Spectral Shapes Satellite Lines Shake-Up Shake-Off Atomic spectra
Related Publication	Ho et al., 2021 <i>Intensity Ratio of KβKα in Selected Elements from Mg to Cu, and the Chemical Effects of Cr Kα1,2 Diagram Lines and Cr KβKα Intensity Ratio in Cr Compounds</i> . <i>International Journal of Molecular Sciences</i> , 24(6), 6670. doi: 10.3390/ijms24066670 https://doi.org/10.3390/ijms24066670

Published April 2, 2026 | Version v1

RESCUE - esm-hist - 1850-1949 - EC-Earth3-ESM-1 - Ocean - thetao

Franco, Ana C.

Project: Horizon Europe - RESCUE

Experiment:

- Name:** esm-hist (emission driven historical run)
- Description:** Simulation with historical forcing conditions (1850 - 2014)

Model: EC-Earth3-ESM-1

Period: 1850–1949

Members: 1 (r1i1p1f1)

Grid type: gn (native grid)

Resolution:

- Atmosphere:** T255L91 (~80 km), 91 vertical levels
- Ocean:** ~1° (ORCA1 grid), 75 vertical levels

Frequency:

- Omon (monthly ocean variables)

Variables:

- thetao: Sea Water Potential Temperature
- units: degC

Fundings:
Produced with funding from the European Union's Horizon Europe Research and Innovation Framework Programme under Grant Agreement

Files

Name	Size
thetao_Omon_EC-Earth3-ESM-1_esm-hist_r1i1p1f1_gn_185001-185012.nc Checksum: md5:c749d961e6e6011e1739f1579d8ef0d PID: http://hdl.handle.net/11304/381d15a5-1e2f-4f84-abc7-a0ad5fdd30b	169.6 MB
thetao_Omon_EC-Earth3-ESM-1_esm-hist_r1i1p1f1_gn_185101-185112.nc Checksum: md5:da738a596a8a8d29edc23095108bf4c PID: http://hdl.handle.net/11304/381d15a5-1e2f-4f84-abc7-a0ad5fdd30b	169.7 MB



Onde procurar modelos de metadados?

1. Catálogos de standards

- **Digital Curation Centre (DCC)**

Um dos melhores diretórios de standards por área científica

<https://www.dcc.ac.uk/guidance/standards/metadata>

- **Research Data Alliance (RDA)**

Grupos de trabalho e recomendações práticas (muito alinhado com FAIR)

<https://www.rd-alliance.org/working-groups/>

- **OpenAIRE Guidelines**

Úteis para projetos financiados pela FCT/UE (compatibilidade garantida)

<https://guidelines.openaire.eu/en/latest/>



Onde procurar modelos de metadados?

2. Standards genéricos

- **DataCite Metadata Schema**

Muito usado em repositórios

Focado em citação e identificação de datasets

Define campos core como autor, título, data, DOI, etc

<https://schema.datacite.org/>

- **Dublin Core**

Simple e transversal

Muito utilizado para documentação básica

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

Onde procurar modelos de metadados?

3. Standards no domínio das SSH

- **Data Documentation Initiative (DDI)**

Amplamente utilizado nas SSH

Aplicado a dados de investigação nas ciências sociais, comportamentais, económicas e da saúde

<https://ddialliance.org/>

- **Text Encoding Initiative (TEI)**

Aplicado a dados provenientes de textos e edições digitais

<https://tei-c.org/support/learn/introducing-the-guidelines/>

- **CIDOC CRM**

Padrão internacional/conceptual reference model (ISO 21127) de ontologia semântica para documentação e interoperabilidade de património cultural

<https://cidoc-crm.org/>

Como escolher o melhor modelo de metadados?

IMPORTANTE: Não escolher “o melhor” — escolher o **mais adequado ao nosso caso!**

Perguntas-chave:

- Vou depositar onde? (Zenodo, Polen BluePrint, repositório institucional, etc.)
- A minha área tem um standard próprio? (ciências sociais, biologia, química, etc.)
- Preciso de interoperabilidade entre sistemas?

Metadados

Erros comuns:

- Metadados mínimos ou inexistentes
- Usar apenas nomes de ficheiros como “descrição”
- Falta de normalização

Benefícios:

- Dados encontráveis (FAIR)
- Interpretação e reutilização por outros investigadores
- Integração em projetos maiores
- Maior visibilidade

Partilha → Repositórios

Princípios gerais

- *Tão abertos quanto possível, tão fechados quanto necessário.*
- Atribuição de identificadores persistentes.
- Os metadados devem ser acessíveis mesmo que os dados não estejam publicamente disponíveis.
- Condições de licenças bem definidas.



ACESSO ABERTO

Partilha → Repositórios

Onde partilhar dados?

Repositórios genéricos:

- Zenodo
- Figshare
- CKAN
- B2SHARE

Repositórios institucionais:

- Universidades/centros de investigação

Repositórios temáticos/disciplinares:

- Depende da área

Partilha → Repositórios

No programa Horizonte Europa, a União Europeia exige que os resultados científicos financiados com fundos públicos sejam acessíveis a todos, em acesso aberto.

Os repositórios escolhidos devem:

- Ser confiáveis e sustentáveis
- Garantir preservação e integridade dos dados a longo prazo
- Suportar identificadores persistentes, metadados interoperáveis e acesso aberto

Partilha → Repositórios

O que NÃO é um repositório?

NÃO são válidos	São repositórios válidos
<ul style="list-style-type: none">- O site da editora- Uma plataforma de rede social (ex: Academia.edu, ResearchGate)- A página pessoal ou institucional do autor- O site do projeto	<ul style="list-style-type: none">- Um repositório institucional- Um repositório temático (área científica) ou um repositório generalista



Partilha → Repositórios + contexto português

O contexto português está muito alinhado com os princípios do Horizonte Europa, nomeadamente em relação à política de acesso aberto.

Infraestruturas no contexto português:

- Fundação para a Ciência e a Tecnologia (FCT)
- Repositórios institucionais (ex: universidades)

Partilha → Repositórios + contexto português

Iniciativa POLEN

- Iniciativa nacional para ciência aberta, promovida pela FCT, através da FCCN
- Promove gestão e partilha de dados
- Alinhada com políticas europeias
- Apoio a investigadores
- Criação de infraestruturas
- Ligação ao ecossistema europeu

Partilha → Repositórios + contexto português

POLEN Sync

- Serviço nacional de armazenamento e sincronização de ficheiros, que permite guardar, organizar e partilhar dados de investigação de forma segura, facilitando o trabalho colaborativo e o cumprimento de boas práticas de gestão de dados.
- O serviço já dispõe de instância para a Universidade do Porto, sendo possível o acesso através de autenticação federada com Ciência Vitae.

<https://polen.fccn.pt/servico/polen-sync/>

<https://sync.polen.fccn.pt/login>



Partilha → Repositórios + contexto português

POLEN DataHub

- Repositório nacional para a publicação, preservação e disseminação de dados de investigação, promovendo a sua partilha e reutilização de acordo com os princípios FAIR (Findable, Accessible, Interoperable, Reusable).
- O acesso pode ser efetuado com autenticação via Ciência Vitae, estando igualmente disponível a entrada institucional para a Universidade do Porto.

<https://datahub.polen.fccn.pt/>

https://datahub.polen.fccn.pt/dataverse/_0000000115037226



Partilha → Repositórios + contexto português

POLEN Blueprint

- Plataforma dedicada à criação, edição e validação de Planos de Gestão de Dados (PGDs), alinhada com os requisitos dos financiadores nacionais e com as políticas de ciência aberta.
- Apoia o planeamento da gestão de dados ao longo de todo o ciclo de vida dos projetos, incluindo a definição de estratégias para a descrição dos dados, adoção de modelos de metadados adequados, armazenamento, preservação e partilha.

<https://blueprint.polen.fccn.pt/home>



Boas práticas de partilha

O que fazer:

- Escolher o repositório adequado
- Atribuir DOI ou outro identificador persistente
- Definir licença clara

Citação de dados:

Datasets devem ser citados como publicações. Incluir, pelo menos:

- autor
- ano
- título
- DOI

Exemplo: *Rodrigues, J., & Teixeira Lopes, C. (2022). Describing data in image format: Proposal of a metadata model and controlled vocabularies [Data set]. INESC TEC.*
<https://doi.org/10.25747/S54M-EH07>

Partilha de dados

Erros comuns:

- Guardar apenas no computador pessoal
- Partilhar sem documentação
- Ignorar licenças

Benefícios:

- Mais citações
- Maior impacto
- Transparência
- Reprodutibilidade da ciência
- Novas colaborações

Como é que tudo se liga?



Para refletir

- De que me servem os meus dados, se ninguém além de mim os pode usar?
- Quantos dados já perdi porque estavam num computador pessoal?
- Qual o sentido de financiar a recolha de dados que já existem?
- Quantos projetos poderiam avançar mais rápido se os dados fossem acessíveis?
- Se ninguém pode verificar os dados e reproduzir as experiências, como validamos os resultados?
- Estamos a produzir conhecimento ou apenas resultados isolados?
- Que legado científico deixam os meus dados?
- Se os meus dados podem ser citados e reutilizados, que impacto e reconhecimento científicos estou a perder ao não os partilhar?

SEMINÁRIO

INTRODUÇÃO À GESTÃO DE DADOS DE INVESTIGAÇÃO

PRINCÍPIOS E BOAS PRÁTICAS

JOANA RODRIGUES

(FACULDADE DE LETRAS DA UNIVERSIDADE PORTO
DEPARTAMENTO DE CIÊNCIAS DA COMUNICAÇÃO
E DA INFORMAÇÃO)

jsrodrigues@letras.up.pt