



9.º FÓRUM
GESTÃO DE DADOS
DE INVESTIGAÇÃO

24 e 25 DE NOVEMBRO 2022
UNIVERSIDADE DE ÉVORA

Um guia para a curadoria de dados em repositórios

Grupo de Trabalho do Fórum GDI
Repositórios de dados: tecnologia, organização e certificação

Objetivos

Identificar **boas práticas** com vista ao depósito de dados em repositórios.

Sistematizar as **tarefas a realizar** pelos curadores de dados em cada uma das etapas do processo de curadoria.

Motivo

A curadoria de dados desempenha um papel central na Gestão de Dados de Investigação (GDI) e traz **desafios** para investigadores, gestores de repositórios e de dados.

Tradução e adaptação do

Dataverse Curation Guide

Dataverse Curation Guide Working Group

New Digital Research Infrastructure Organization (NDRIO) / Portage

Curadoria de dados

Um fluxo de curadoria de dados bem articulado pode **melhorar a qualidade dos dados**, pois cria condições para:

- descrever de forma adequada o seu conteúdo,
- criar uma estrutura coerente,
- fornecer documentação explicativa,
- permitir mecanismos de automação,
- relacionar os conjuntos de dados com outros dados e resultados de investigação.

Três tipos de serviços de curadoria

Curadoria sem mediação

Sem intervenção do serviço de GDI. O investigador cria, deposita e publica os seus conjuntos de dados.

Curadoria semimediada

O serviço de GDI cria uma coleção no repositório ou inicia o depósito de um conjunto de dados e atribui um papel ao investigador. O investigador deposita dados nessa coleção ou conjunto de dados. O conjunto de dados é sinalizado para revisão pelos administradores do repositório ou o depositante solicita que o conjunto de dados seja revisto pela equipa de gestão de dados antes ou depois da sua publicação.

Curadoria mediada

Os dados são enviados pelo investigador para o serviço de GDI, que cria a coleção ou conjunto de dados; os dados são curados pela biblioteca e publicados uma vez aprovados pelo investigador.

Três níveis de curadoria

Nível 1

As etapas mínimas necessárias para publicar com êxito num repositório e tornar o conjunto de dados localizável.

Nível 2

Atividades que melhoram a descoberta dos conjuntos de dados e ajudam a garantir a sua utilização ao longo do tempo.

Nível 3

Ações intensivas de curadoria destinadas a garantir a reprodutibilidade e a preservação dos conjuntos de dados.

Guia 'CURATION'

C

Check

Assegurar que todos os dados e metadados necessários para publicar o conjunto de dados com sucesso estão reunidos e em condições de publicação.

U

Understand

Garantir que o conjunto de dados está bem descrito e que os utilizadores terão uma visão clara a que se referem os dados e como podem ser utilizados.

R

Recomend

Solicitar informações adicionais ao depositante ou sugerir alterações aos metadados e ficheiros de forma a potenciar a descoberta e a utilização dos dados de acordo com os princípios FAIR.

A

Augment

Melhorar a submissão para facilitar a capacidade de descoberta e uso.

T

Transform

Assegurar que o conjunto de dados utiliza o maior número possível de formatos abertos e de uso generalizado.

I

Include

Facilitar a reutilização, bem como a correta atribuição e creditação dos dados, incluindo os identificadores persistentes relevantes e as informações de licenciamento adequadas.

O

Optimize

Avaliar a adequação do conjunto de dados aos princípios FAIR e tomar medidas para melhorar a descoberta, acessibilidade, interoperabilidade e reutilização dos dados.

N

Note Down

Certificar-se de que se faz um registo exato do trabalho de curadoria.

Estrutura do guia de curadoria

C	U	R	A	T	I	O	N
Check	nderstand	ecommend	ugment	ransform	include	ptimize	ote Down
Nível 1	Nível 2 Nível 3	Nível 1	Nível 2	Nível 3	Nível 2 Nível 3	Nível 3	Nível 2

Sim	Não	Levanta questões	N/A	A documentação inclui...
☐	☐	☐	☐	Informação de contexto sobre os dados (como os dados foram recolhidos ou produzidos, o objetivo da investigação).
☐	☐	☐	☐	Descrição das convenções de nomenclatura dos ficheiros e da estrutura dos ficheiros, se importante.
☐	☐	☐	☐	Registo de como os dados foram modificados ou processados.
☐	☐	☐	☐	Informações sobre confidencialidade e quaisquer restrições impostas ao uso dos dados.
☐	☐	☐	☐	Nomes de descritores e variáveis, informações sobre valores permitidos e unidades de medida, códigos e classificações, se aplicável.
☐	☐	☐	☐	Descrição do ambiente computacional necessário para executar qualquer código incluído (sistema operacional, pacotes de software e dependências).

Sim	Não	Levanta questões	N/A	Identifique os formatos e o software usado na criação de ficheiros...
☐	☐	☐	☐	O conjunto de dados contém ficheiros com formatos proprietários ou inapropriados para preservação? Para orientações, consulte os formatos de ficheiros elencados em DataverseNO .
☐	☐	☐	☐	Os ficheiros proprietários podem ser transformados em formatos não-prorietários sem que ocorra perda de informação?
☐	☐	☐	☐	Caso os ficheiros não possam ser transformados, a sua documentação descreve: - O formato e o conteúdo dos ficheiros? - O software ou instrumento que gerou os ficheiros? - O software necessário para apresentar ou manipular os ficheiros? - Uma alternativa freeware para aceder aos ficheiros (se disponível)?
☐	☐	☐	☐	A documentação descreve os dados, as fórmulas usadas, os cabeçalhos de colunas, os nomes de variáveis, etc. por forma a que os dados sejam usáveis.
☐	☐	☐	☐	A documentação está num formato adequado (ex.: PDF/A ou .txt).

Sim	Não	N/A	
☐	☐	☐	Um ficheiro de registo foi criado para documentar a seguinte informação: ☐ DOI, identificador interno do item ou outra forma de relacionar o ficheiro de registo com o conjunto de dados ☐ Identificar problemas descobertos no decorrer do processo de curadoria ☐ Questões e recomendações de alterações significativas para o investigador ☐ Resumo detalhado das respostas do investigador às sugestões e questões ☐ Qualquer alteração efetuada no conjunto de dados durante o processo de curadoria, incluindo: ☐ Alterações aos ficheiros existentes e documentação ☐ Nome de ficheiros adicionados ou removidos do conjunto de dados e qual o motivo ☐ Listagem dos termos alterados, adicionados ou removidos dos <u>metadados</u>

Outros conteúdos do guia

Indicação de guias complementares e recursos adicionais sobre curadoria de dados.

Exemplos de conjuntos de dados que passaram pelo processo de curadoria.

Modelos de mensagens de contacto com os depositantes.

CURATION – guia abreviado.

Guia disponível em:

<https://forumgdi.rcaap.pt/grupos-de-trabalho/gt-repositorios-de-dados/>

OBRIGADO!

Rita Gonçalves (Universidade de Aveiro), Pedro Príncipe (Universidade do Minho),
Filomena Borba (ISCAL), José Carvalho (Universidade de Aveiro),
Paula Moura (Universidade do Minho), Francisca Alves Cardoso (CRIA (NOVA FCSH)),
Salima Rehemtula (IHMT, Universidade NOVA de Lisboa)

forumgdi-repositorios@fccn.pt

Grupo de Trabalho
Repositórios de Dados:
Tecnologia, organização e
certificação FAIR



FÓRUM
GESTÃO DE DADOS
DE INVESTIGAÇÃO