



10.º FÓRUM  
GESTÃO DE DADOS  
DE INVESTIGAÇÃO

14 e 15 DE NOVEMBRO 2023  
INSTITUTO POLITÉCNICO DE SETÚBAL

# Recomendações para aplicação de uma *checklist* de curadoria de metadados em repositórios

CAFÉ COM DADOS

Questões críticas no desenvolvimento de repositórios confiáveis, interoperáveis  
e sustentáveis: casos de estudo

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

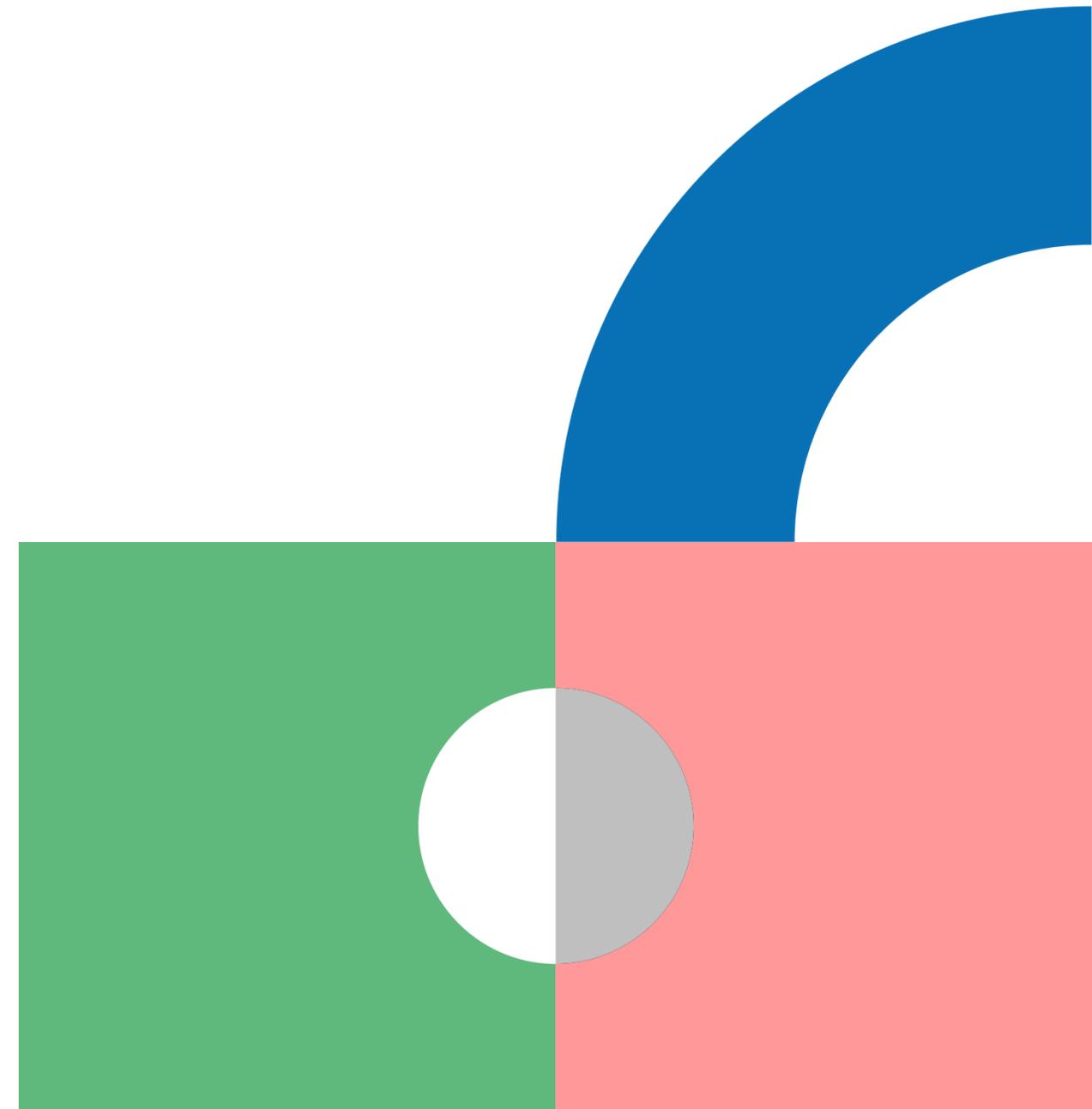
15 DE NOVEMBRO DE 20223



# Curadoria de (meta)dados

## VERTENTES

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



# Definição do tipo de serviço 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 ao investigador as permissões necessárias à sua interação com o repositório. 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.

## TER EM CONTA



- o tipo de dados com os quais se está a trabalhar;
- as políticas, estratégias e práticas institucionais;
- a experiência na atividade de curadoria e a quantidade de tempo que o investigador pode dedicar ao processo;
- a necessidade de saltar etapas ou alterar a ordem pelas quais as mesmas são concluídas, assim como de completar algumas das etapas uma segunda vez;
- as prioridades concorrentes e os recursos locais dedicados ao serviço de curadoria.





# Três níveis de curadoria

## Nível 1

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

- o conjunto de dados foi depositado na coleção adequada e os campos de metadados obrigatórios estão preenchidos e a sua informação é suficiente.

## Nível 2

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

- os campos de metadados recomendados são preenchidos e o conjunto de dados inclui documentação suficiente para permitir que um utilizador com experiência semelhante compreenda e possa utilizar o conjunto de dados.

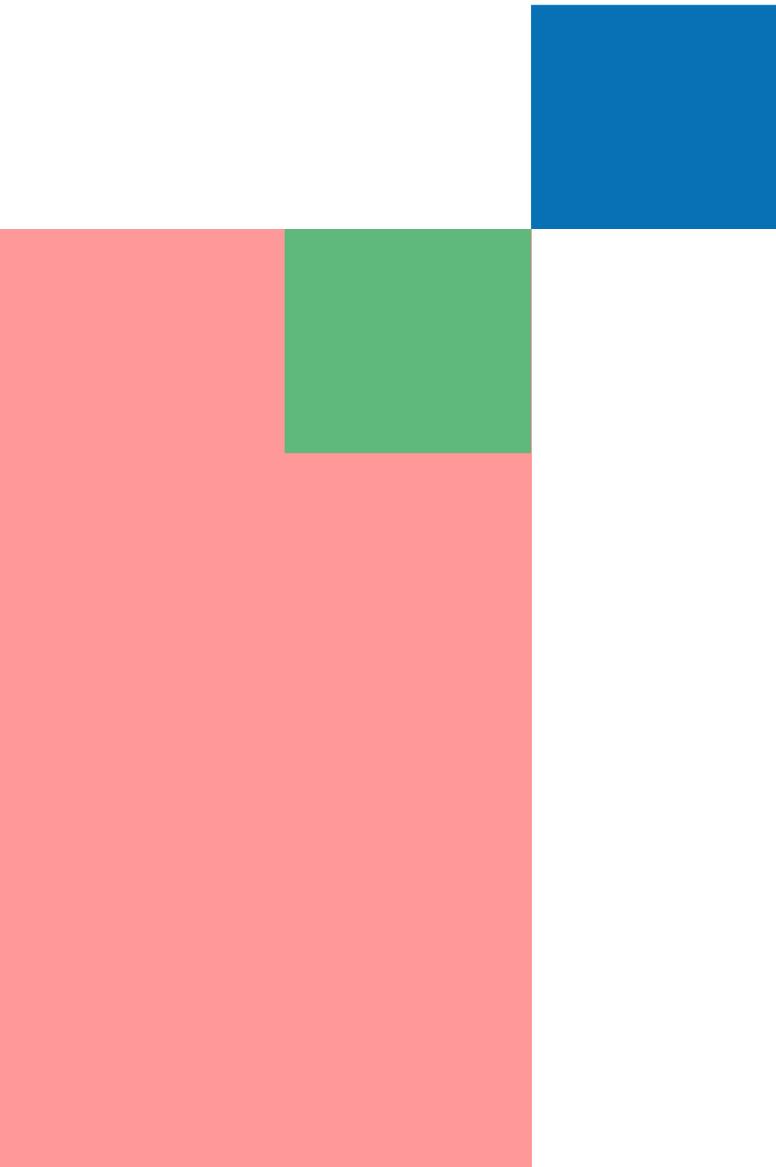
## Nível 3

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

- a documentação de suporte é aprimorada, o conteúdo dos ficheiros e o código são revistos e os ficheiros de dados são transformados em formatos adequados para preservação a longo prazo.

# Checklist

## INCLUIR ITENS PARA VERIFICAÇÃO DOS SEGUINTE OBJETIVOS



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

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

Potenciar a descoberta e a utilização dos dados de acordo com os princípios FAIR.

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

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

Sugerir ao depositante alterações aos metadados e ficheiros.  
Registar as atividades de curadoria.



# Itens a considerar



- O conjunto de dados foi submetido na coleção adequada.
- O investigador verificou que não existem problemas de licenciamento ou de direitos de propriedade intelectual relacionados com o conjunto de dados e garantiu que o mesmo não contém quaisquer informações sensíveis.
- A documentação de apoio está incluída (ex.: ficheiro *Readme*)
  - Disponibilizar modelos e outras orientações para a criação da documentação
- Todos os ficheiros identificados na documentação estão incluídos no conjunto de dados.



# Itens a considerar



- Os campos de metadados obrigatórios estão corretamente preenchidos.
  - Definir conjuntos de campos obrigatórios e recomendados.
  - Disponibilizar ajudas de campo no formulário de submissão no repositório e conteúdos de apoio sobre o preenchimento dos campos.
- Os ficheiros estão nomeados e organizados de forma compreensível.
- Os metadados são ricos, precisos e exaustivos.
  - Metadados de citação, metadados disciplinares, metadados ao nível dos ficheiros



# Itens a considerar



- Os *links* diretos para publicações relacionadas, código, modelos, documentação, dados de origem e outros recursos que estejam ligados ao conjunto de dados estão, sempre que possível, incluídos.
- Os ficheiros associados ao conjunto de dados são, tanto quanto possível, disponibilizados em formatos abertos e de uso comum.
- Quando os ficheiros proprietários não podem ser transformados em formatos não-proprietários (por ex.: por razões de perda de informação), o ficheiro *Readme* identifica o *software* necessário para usar os ficheiros, assim como a versão utilizada do mesmo, e descreve o seu conteúdo.



# Itens a considerar



- O conjunto de dados inclui identificadores persistentes relevantes
  - Atribuição de identificador persistente ao conjunto de dados (ex.: DOI)
  - Inclusão de outros identificadores nos metadados (ex.: identificadores dos autores como o ORCID)
- O conjunto de dados inclui informação apropriada acerca da licença de utilização/termos de uso personalizados.
- O conjunto de dados inclui informação sobre o financiamento.

# Contacto com os depositantes

Solicitar informações adicionais ao investigador e/ou sugerir alterações aos metadados e ficheiros.

Sinalizar as questões sobre metadados ou dados identificadas ao longo da revisão do depósito.

Hierarquizar os pedidos, estabelecendo a diferença entre o que é necessário para publicar o conjunto de dados e o que são sugestões de melhoria / recomendações.

Enviar / disponibilizar diretrizes e conteúdos de apoio que possam ser importantes e úteis aos investigadores durante o processo de depósito.

Considerar o uso de *templates* (ex.: e-mail) como ponto de partida nas comunicações com os depositantes.

# Registo do trabalho de curadoria

Reunir a informação relativa às alterações efetuadas no conjunto de dados, quem as efetuou e qual o motivo das mesmas.

- DOI, ID interno do item ou outra forma de identificar o conjunto de dados
- Questões identificadas no decorrer do processo de curadoria
- Recomendações e sugestões de alteração enviadas para o investigador
- Resumo das respostas do investigador às sugestões e questões
- Qualquer alteração efetuada no conjunto de dados durante o processo de curadoria

# O caso do DUnAs: Repositório de Dados de Investigação da Universidade de Aveiro

<https://dunas.ua.pt/>

Depósito de dados realizado em regime de autoarquivo, pelos investigadores ou membros do projeto designados, mediante a atribuição de permissões pelos Serviços de Biblioteca, Informação Documental e Museologia (SBIDM).

Depósito de dados gerados no âmbito de uma unidade ou projeto de investigação da UA, sendo que pelo menos um dos autores do *dataset* tem afiliação à instituição.

Curadoria semimediada.

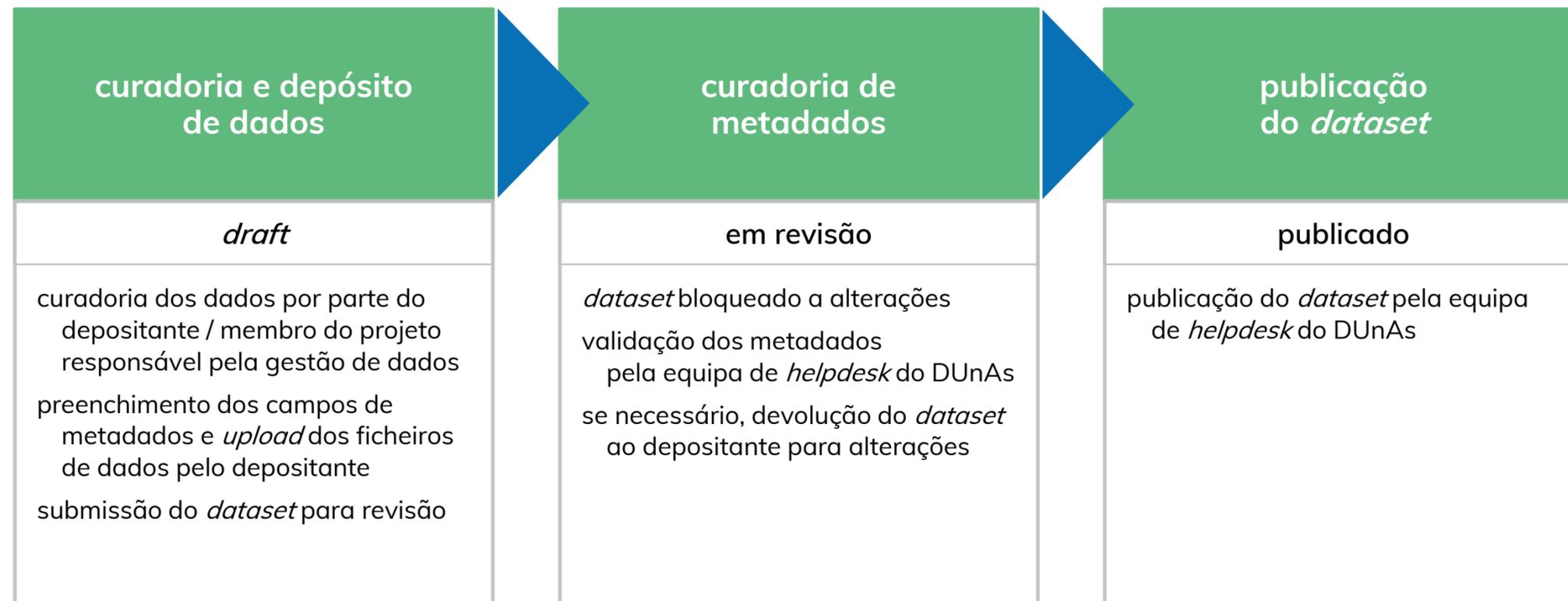


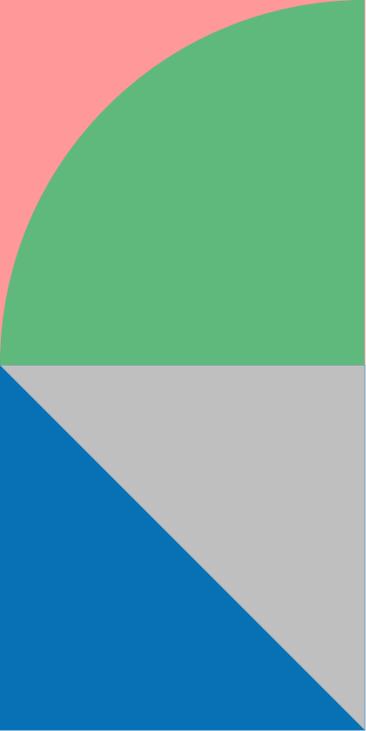
# O caso do DUnAs

## Fluxo de trabalho

A publicação de *datasets* implica uma submissão prévia para revisão por parte dos SBIDM ao nível dos metadados. A permissão de publicação de *datasets* é exclusiva dos SBIDM.

A veracidade e qualidade dos dados, assim como as questões de proteção de dados pessoais e informação confidencial e de direitos de autor, são da responsabilidade dos depositantes.





# Referência bibliográfica

Borba, F., Cardoso, F. A., Carvalho, J., Moura, P., Príncipe, P., Gonçalves, R., & Rehemtula, S. (2023). *Guia para a curadoria de dados em repositórios (1.0)*. Zenodo. <https://doi.org/10.5281/zenodo.8340747>



# Obrigada!

Rita Gonçalves  
Universidade de Aveiro

