



Scientific Electronic Library Online

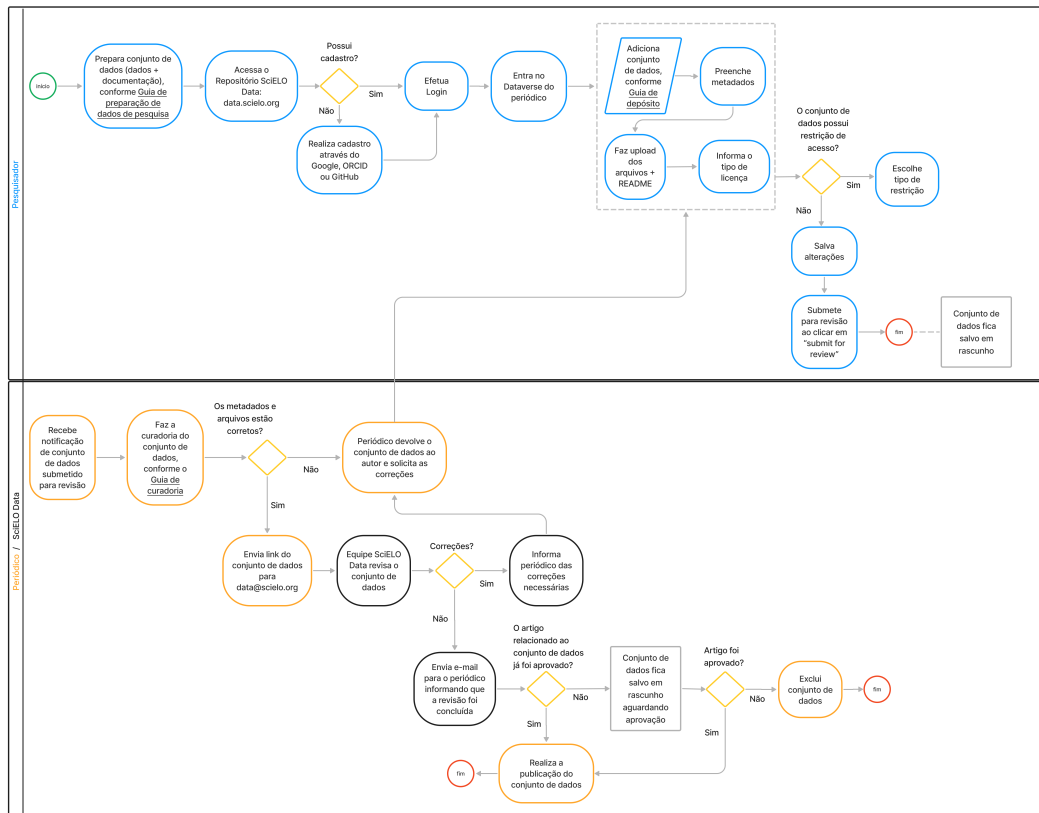
## Guia de curadoria de dados de pesquisa para equipes editoriais

Abril de 2023



This is an Open Access document distributed under the terms of the Creative Commons Attribution License (**CC-BY**), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Fluxograma SciELO Data



De acordo com o *CoreTrustSeal*<sup>1</sup> (2019), curadoria de dados é:

*Gerenciar e promover o uso de dados desde sua concepção para assegurar que são adequados para fins contemporâneos e disponíveis para descoberta e reuso. Para conjuntos de dados dinâmicos isto implica enriquecimento contínuo ou atualizações para que sigam servindo seus propósitos. Níveis de curadorias mais elevados incluirão links com anotações e outros materiais publicados. (Tradução nossa)*

Ou seja, a curadoria de dados envolve todo o ciclo de vida dos dados (*Data Life Cycle*), desde o planejamento para coleta até sua preservação para acesso e reuso a longo prazo. Pode também ser definida de forma mais específica como as verificações e ações realizadas pelos curadores com o objetivo de garantir que o conjunto de dados estejam estruturado e documentado da forma mais completa possível e seguindo as boas práticas.

No contexto deste guia o termo curadoria será utilizado considerando a segunda definição.

<sup>1</sup> O *CoreTrustSeal* é uma iniciativa do *World Data System* do *International Science Council* (WDS) e do *Data Seal of Approval* (DSA). É uma organização internacional, comunitária, não governamental e sem fins lucrativos que promove infraestruturas de dados confiáveis e sustentáveis.

## 1. Citação do conjunto de dados no artigo vinculado

Para haver conexão entre artigo publicado e conjunto de dados subjacente é recomendado que nas instruções aos autores esteja documentado que os artigos devem conter a seção “Disponibilidade de dados” informando se o conjunto de dados referente à pesquisa está disponível e em caso positivo, onde acessá-lo.

Exemplo de conteúdo para a seção recomendados pelo SciELO:

### **Dados não disponíveis**

*O conjunto de dados que dá suporte aos resultados deste estudo não está disponível publicamente (Não se aplica para artigos com conjuntos de dados no SciELO Data).*

### **Dados disponíveis**

*Todo o conjunto de dados que dá suporte aos resultados deste estudo foi publicado no próprio artigo (Não se aplica para artigos com conjuntos de dados no SciELO Data).*

*Todo o conjunto de dados que dá suporte aos resultados deste estudo foi publicado no artigo e na seção “Materiais suplementares” (Não se aplica para artigos com conjuntos de dados no SciELO Data).*

*Todo o conjunto de dados que dá suporte aos resultados deste estudo foi disponibilizado no SciELO Data e pode ser acessado em [URL ou DOI].*

*Todo o conjunto de dados que dá suporte aos resultados deste estudo foi disponibilizado no SciELO Data com os identificadores [lista dos identificadores].*

*Todo o conjunto de dados anonimizados que dá suporte aos resultados deste estudo foi disponibilizado no SciELO Data e pode ser acessado em [URL ou DOI].*

### **Dados disponíveis mediante solicitação<sup>2</sup>**

*Todo o conjunto de dados que dá suporte aos resultados deste estudo está disponível mediante solicitação ao autor correspondente [nome do autor correspondente]. O conjunto de dados não está publicamente disponível devido a [detalhar motivo da restrição, por exemplo, conter informação que compromete a privacidade dos participantes da pesquisa] (Não se aplica para artigos com conjuntos de dados no SciELO Data).*

*Todo o conjunto de dados que dá suporte aos resultados deste estudo está disponível mediante solicitação à [nome da organização]. O conjunto de dados não está publicamente disponível devido a [detalhar motivo da restrição, por exemplo, conter informação que compromete a privacidade dos participantes da pesquisa] (Não se aplica para artigos com conjuntos de dados no SciELO Data).*

---

<sup>2</sup> No SciELO Data é possível depositar conjuntos de dados, porém mantê-los restritos. Quando alguém quiser acessá-los entrará em contato com o autor depositante do conjunto através da plataforma. Veja mais detalhes no item 5.1 do [Guia de depósito de Dados de Pesquisa](#).

Para informações detalhadas sobre os níveis de aplicação dos critérios de dados, códigos e materiais de pesquisa no [Guia para promoção da abertura, transparência e reprodutibilidade das pesquisas publicadas pelos periódicos SciELO](#).

## 2. Anonimização de dados

A curadoria é também importante para verificar se os dados a serem publicados não ferem as regras dos comitês de ética e pesquisa de suas respectivas áreas, e verificar se há dados que precisam de anonimização (como dados pessoais, sensíveis ou não<sup>3</sup>, informações que ultrapassem o direito de privacidade de pessoas envolvidas, ou as coloca em risco, assim como coordenadas de áreas protegidas, sob ameaça de extinção ou informações que infrinjam acordos comerciais, patentes ou pertençam a terceiros).

As informações abaixo já constam no [Guia de preparação de dados](#), no entanto são repetidas abaixo para a verificação da equipe editorial.

Dados sensíveis (dados que, caso sejam expostos sem autorização ou perdidos, podem resultar em problemas legais), ou pessoais devem ser anonimizados.

Exemplo de anonimização de dados<sup>4</sup>:

Informação sem anonimização	Resposta sem anonimização
Nome	Juan Pérez
País de origem	Argentina
Idade	54
Anos de experiência	25
Modelo de aviões	Boeing 777 Boeing 747
Data do último voo	05/01/2022

Informação anonimizada	Resposta anonimizada
-	-
Continente	América do Sul
Faixa Etária	50-60
Anos de experiência	10-20
Modelo de aviões	Comerciais
Data do último voo	01/2022

<sup>3</sup> Dados pessoais: Podem ser considerados dados pessoais: nome e sobrenome; endereço residencial; endereço de e-mail (se contiver elementos que ajudem a identificar o dono, como nome e sobrenome); gênero; data de nascimento; número de documentos cadastrais, como RG, CPF e carteira de trabalho; dados de geolocalização de um telefone celular; número de telefone pessoal. <https://portal.fiocruz.br/noticia/entenda-melhor-lei-geral-de-protacao-de-dados-pessoais>. Acesso em 21 mar 2023.

Dado pessoal sensível: “dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural”. [https://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/l13709.htm](https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm). Acesso em 30 jan 2023.

<sup>4</sup> Exemplo retirado de: Gestión de Datos de Investigación - Parte I. Disponível em: <https://www.youtube.com/watch?v=BM-lZ2XCCN0>

Em casos em que a prática de anonimização é impossível busque usar pseudônimos ou considere não publicar.

### 3. Níveis de curadoria de dados

Visando a transparência com relação às verificações e ações realizadas pelos curadores nos conjuntos de dados depositados, o SciELO adotará como referência os níveis de curadoria de 1 a 3 usados pelo *CoreTrustSeal*, como requisito na avaliação de repositórios de dados confiáveis:

- A. Curadoria básica: rápida verificação dos metadados / conteúdo, adição de metadados básico ou documentação → **Nível 1**;
- B. Curadoria detalhada: curadoria básica + conversão dos arquivos de dados para novos formatos, aprimoramento da documentação → **Nível 2**;
- C. Curadoria no nível de dados: curadoria básica + curadoria detalhada + edição dos dados depositados para maior precisão → **Nível 3**.

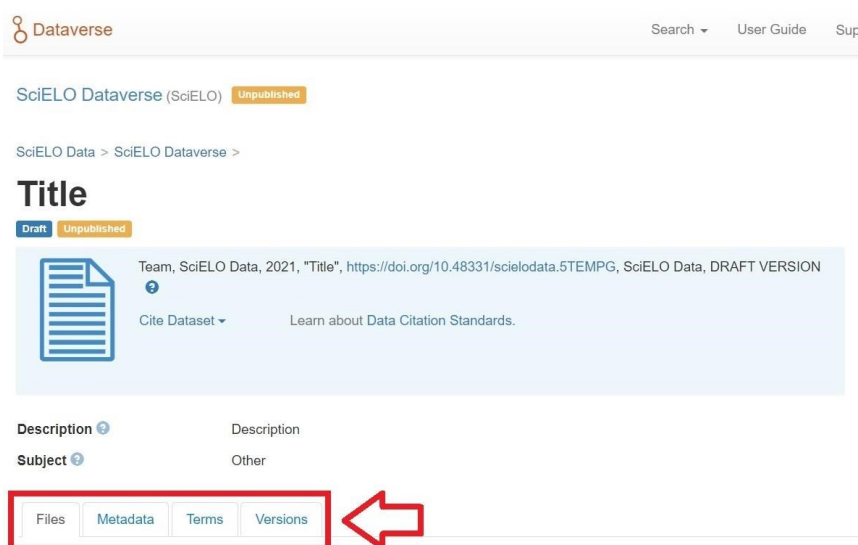
Independente do nível de curadoria, é obrigatório verificar entre os arquivos se há dados pessoais ou potencialmente sensíveis conforme item 2 deste guia.

### 4. Checklists de verificações e ações de curadoria de dados

No repositório SciELO Data há 4 áreas onde os metadados / arquivos de dados precisam passar por curadoria considerando o nível de curadoria adotado e realizado pelo periódico:

- Arquivos (Files)
- Metadados (Metadata)
- Termos (Terms)
- Versões (Versions)

É possível navegar por essas áreas clicando nas abas correspondentes:

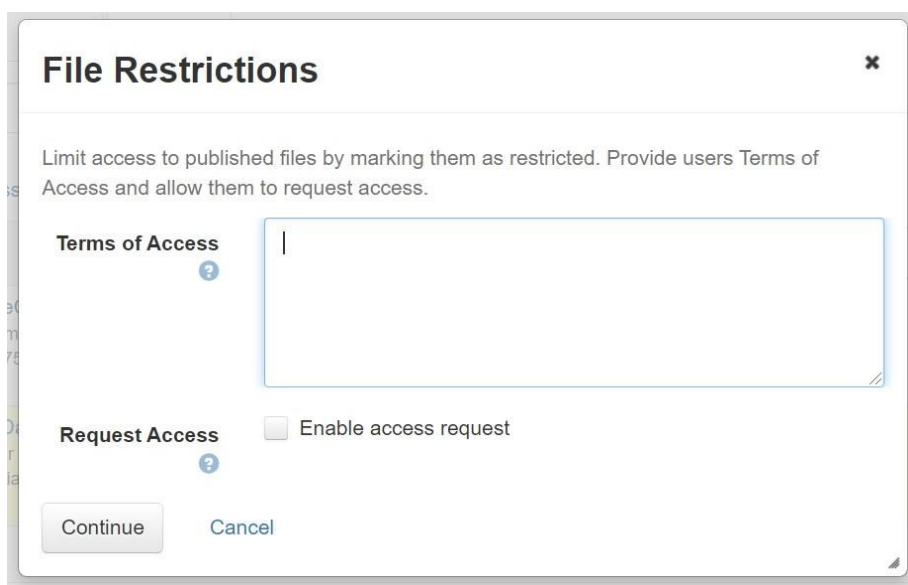


## Nível 1 - Curadoria básica

- Certifique-se que o conjunto de dados está relacionado com algum manuscrito submetido ao periódico. Obs.: Só podem ser publicados conjuntos de dados de artigos aprovados.

### Na aba “Files”:

- Verificar se o conjunto de dados foi documentado em um arquivo nomeado como README. A presença deste arquivo é obrigatória;
- Verificar se a nomeação dos arquivos está adequada (ver tópico 1 do “[Guia de preparação de dados de pesquisa](#)”). Se não estiver recomendar aos autores que façam edição seguindo as recomendações;
- Verificar se os arquivos podem ser abertos (se não estão corrompidos). Se não abrirem, solicitar novo depósito aos autores;
- Verificar se nos arquivos há dados que precisam de anonimização. Se for necessário anonimizar, consulte o item 2 deste guia;
- Verificar se os arquivos estão em formatos recomendados (ver tópico 2 do “[Guia de preparação de dados de pesquisa](#)”). Se não estiver recomendar aos autores que façam edição seguindo as recomendações;
- Verificar se o conjunto de dados possui arquivos com acesso restrito.
  - Se sim, o campo “Terms of Access” deve estar preenchido com informações sobre como os usuários podem obter acesso aos arquivos restritos.



**File Restrictions**

Limit access to published files by marking them as restricted. Provide users Terms of Access and allow them to request access.

**Terms of Access**

**Request Access**  Enable access request

Continue Cancel

### Na aba “Metadata”:

- **Title:** verificar se está preenchido com o título do artigo cujos dados estão relacionados ou com título próprio que seja significativo/descritivo para o conjunto de dados. Se não estiver recomendar aos autores que editem;
- **Author:**

- Verificar se os nomes dos autores foram inseridos na ordem inversa (Sobrenome, Nome);
- Verificar se os autores informaram a afiliação (obrigatório) e o ORCID (recomendado).
- **Subject:** verificar se a área temática selecionada é a mais adequada. Evite o uso de “Other” sempre que possível;
- **Keyword:** Cada palavra-chave foi inserida separadamente (ou seja, cada uma em um campo). Se não estiverem editem na tela de edição de metadados e adicione palavras-chave apertando o sinal “+”;
- **Related Publication:**
  - Se o conjunto de dados estiver relacionado com um manuscrito em processo de avaliação, inserir o **título do manuscrito**;
  - Se o conjunto de dados estiver relacionado com um artigo publicado, inserir a **citação do artigo com DOI**.
- **Funding Information:** Se a pesquisa possui fonte de fomento, clique em “Edit Dataset” e então “Metadata”. Encontre o campo e o preencha.

#### Na aba “Terms”:

- Verificar se a licença de uso escolhida para o conjunto de dados foi a CC BY 4.0. Se desejar usar outra licença, favor entrar em contato com [data@scielo.org](mailto:data@scielo.org).

#### Na aba “Versions”:

- Verificar se trata-se de um novo conjunto de dados ou uma nova versão de um conjunto de dados já publicado.

### Nível 2 - Curadoria detalhada → Realizar curadoria básica e:

#### Na aba “Files”:

- Renomear os arquivos da forma mais adequada (ver tópico 1 do “[Guia de preparação de dados de pesquisa](#)”);
- Avaliar os formatos dos arquivos para determinar se estão em um formato recomendado e se necessário convertê-los (ver tópico 2 do “[Guia de preparação de dados de pesquisa](#)”);
- Avaliar se a documentação fornecida (arquivo README, *codebook* etc.) está completa e compreensível (ver tópico 4 do “[Guia de preparação de dados de pesquisa](#)”). Caso não esteja, solicite as alterações necessárias aos autores.

#### Na aba “Metadata”:

- Avaliar informações fornecidas para determinar se estão completas e compreensíveis. Solicitar correções ou fazer edições se necessário.

#### Na aba “Terms”:

- Avaliar informações fornecidas para determinar se estão completas e compreensíveis. Solicitar correções ou fazer edições se necessário.

**Nível 3 - Curadoria no nível de dados** → Realizar curadoria básica + curadoria detalhada e:

**Na aba “Files”:**

- Fazer download dos arquivos de dados;
- Abrir os arquivos de dados e verificar se necessitam de algum tratamento adicional. Se necessário, solicite correções aos autores ou realize as alterações e os informe;
- Abrir os arquivos de dados e avaliá-los no que se refere a possíveis problemas como: variáveis apropriadas e definições de valor, valores fora do intervalo, descrições de programas usados para arquivos de código e estruturas de dados preferenciais. Se necessário, solicite correções aos autores ou realize as alterações e os informe;
- Executar e solucionar problemas em arquivos de códigos;
- Verificação de consistência (*checksum*<sup>5</sup>) dos arquivos do conjunto de dados para garantir a integridade dos dados no nível de bits;

Se o conjunto de dados não estiver adequadamente estruturado e/ou documentado, o curador pode devolvê-lo para o autor (clique em “Publish Dataset” e depois em “Return to Author”).

Se o conjunto de dados estiver adequado para publicação, envie um e-mail para [data@scielo.org](mailto:data@scielo.org) informando a URL do dataset e solicite verificação. A curadoria SciELO tem caráter temporário. Após retorno da equipe SciELO Data o periódico pode seguir com a publicação (clique em “Publish Dataset” e depois em “Publish”). O autor receberá um e-mail informando que o conjunto de dados foi publicado.

#### **IMPORTANTE:**

Uma vez publicado o conjunto de dados não pode ser deletado. Recomendamos que o conjunto seja editado gerando então uma 2ª versão do conjunto de dados. O DOI será mantido o mesmo.

Para casos excepcionais que exijam exclusão não é possível “sumir” com o conjunto de dados, apenas “desacessá-lo”. O DOI do conjunto de dados levará à mesma página que passará a exibir a citação do conjunto com a informação “DEACCESSIONED VERSION” e o motivo da indisponibilidade.

Se após a publicação algum conjunto de dados precisar de alterações ou perder o acesso, por gentileza entre em contato com [data@scielo.org](mailto:data@scielo.org).

---

<sup>5</sup> *Checksum* ou soma de verificação é uma sequência de números e letras usada para verificar a integridade dos dados, ou seja, se um arquivo é exatamente o mesmo depois de uma transferência, verificar se não foi alterado por terceiros ou se não está corrompido.



O SciELO Data permite compartilhar o conjunto de dados com pareceristas para *peer review* através de uma URL privada<sup>6</sup>.

Para informações sobre curadoria de dados de arquivos em formatos específicos consulte:

Excel (.xlsx)	<ul style="list-style-type: none"><li>• <a href="#">Excel CURATED checklist</a></li></ul>
Google Docs	<ul style="list-style-type: none"><li>• <a href="#">Google Docs CURATED Checklist</a></li></ul>
R (.r, .rmd)	<ul style="list-style-type: none"><li>• <a href="#">Filetype CURATED checklist</a></li></ul>

Para auxiliar na gestão dos conjuntos de dados em rascunho o administrador do Dataverse do periódico ou o curador podem atribuir etiquetas para um conjunto de dados com o objetivo de indicar seu status atual:

- Journal curation: curadoria do periódico pendente / em andamento;
- Author contacted: aguardando correções dos autores;
- Privacy Review: revisão do(s) parecerista(s) pendente / em andamento;
- SciELO curation: curadoria do SciELO pendente / em andamento;
- Awaiting article approval: aguardando aprovação do artigo relacionado para publicação do conjunto de dados.

Para incluir uma etiqueta clique em “Publicar conjunto de dados” → “Alterar status de curadoria” e então escolha o status adequado. As etiquetas serão automaticamente removidas do conjunto de dados quando ele for publicado.

## Referências

Abbott, D. What is Digital Curation? *Digital Curation Centre* [online]. [viewed 20 October 2021]. Available from: <https://www.dcc.ac.uk/guidance/briefing-papers/introduction-curation/what-digital-curation>.

CoreTrustSeal Standards and Certification Board. CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022. *CoreTrustSeal* [online]. [viewed 20 October 2021]. Available from: <https://doi.org/10.5281/zenodo.3638211>.

---

<sup>6</sup> A criação de uma URL privada (“*Private URL*”) permite o compartilhamento (para visualização e download dos arquivos) de um conjunto de dados não publicado com um grupo de indivíduos que podem não ter uma conta de usuário no SciELO Data, ou seja, qualquer pessoa que receba a URL privada poderá acessar o conjunto de dados sem precisar se cadastrar ou fazer o *Log in* no SciELO Data. Para criar uma URL privada vá até o conjunto de dados → clique em “Edit” → “Private URL” → na caixa pop-up escolha “Create Private URL” ou “Create URL for Anonymized Access” (permite a revisão anônima removendo nomes de autores e outras informações potencialmente identificadoras de citações) → copie a URL criada e compartilhe com os pareceristas. Quando quiser desabilitar a URL privado vá até o conjunto de dados → clique em “Editar” → “URL privada” → “Desativar URL privada” → “Sim, desative a URL privada”.

CoreTrustSeal Standards and Certification Board. CoreTrustSeal Trustworthy Data Repositories Requirements: Glossary 2020–2022. *CoreTrustSeal* [online]. [viewed 20 October 2021]. Available from: <https://doi.org/10.5281/zenodo.3632563>.

DataverseNO. Curator Guide. *DataverseNO* [online]. [viewed 05 October 2021]. Available from: <https://site.uit.no/dataverseno/admin-en/curatorguide/>.

Lafferty-Hess, S., et al. Conceptualizing Data Curation Activities Within Two Academic Libraries. *Journal of Librarianship and Scholarly Communication* [online]. 2020, **8**, eP2347 [viewed 20 October 2021]. <https://doi.org/10.7710/2162-3309.2347>.

## Como citar este documento

SciELO. *Guia de curadoria de dados de pesquisa para equipe editoriais* [online]. SciELO, 2023 [cited DD Month YYYY]. Available from: \_\_\_\_\_.

**Este e demais documentos do SciELO Data estão disponíveis em:**

<https://www.scielo.org/pt/sobre-o-scielo/scielo-data-pt/>