



Scientific Electronic Library Online

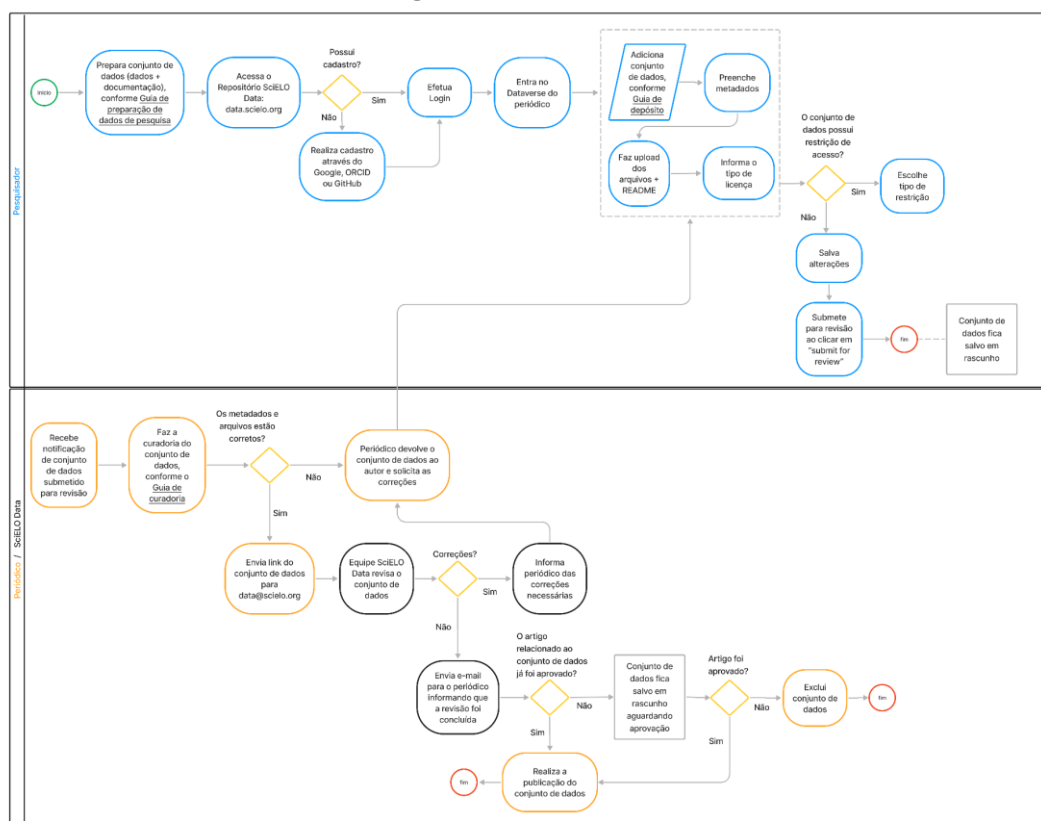
Guia de curadoria de dados de pesquisa para equipes editoriais

Junho de 2022



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*¹ (2019), curadoria de dados é:

The activity of managing and promoting the use of data from their point of creation to ensure that they are fit for contemporary purpose and available for discovery and reuse. For dynamic datasets this may mean continuous enrichment or updating to keep them fit for purpose. Higher levels of curation will also involve links with annotation and with other published materials.

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 esteja 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.

¹ 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 infra estruturas de dados confiáveis e sustentáveis.

1. 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 usados pelo *CoreTrustSeal* como requisito na avaliação de repositórios de dados confiáveis:

- A. Conteúdo publicado como depositado → **Nível 0**;
- B. Curadoria básica: rápida verificação dos metadados / conteúdo, adição de metadados básico ou documentação → **Nível 1**;
- C. Curadoria detalhada: curadoria básica + conversão dos arquivos de dados para novos formatos, aprimoramento da documentação → **Nível 2**;
- D. Curadoria no nível de dados: curadoria básica + curadoria aprimorada + edição dos dados depositados para maior precisão → **Nível 3**.

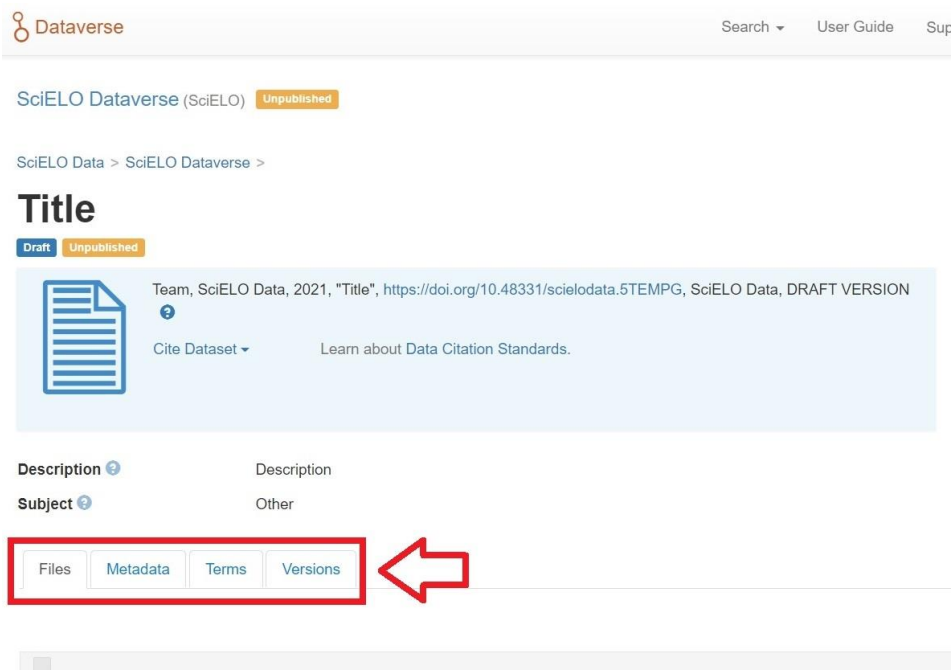
O nível de curadoria adotado e realizado pelo periódico deve ser explicitado na descrição do Dataverse do periódico no SciELO Data e nas instruções aos autores do periódico.

2. 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:



The screenshot shows the SciELO Dataverse interface. At the top, there is a search bar and links for 'User Guide' and 'Sup.'. Below this, the page title is 'SciELO Dataverse (SciELO)' with an 'Unpublished' status tag. The breadcrumb trail shows 'SciELO Data > SciELO Dataverse >'. The main heading is 'Title', followed by 'Draft' and 'Unpublished' tags. A document icon is shown next to the title: 'Team, SciELO Data, 2021, "Title", https://doi.org/10.48331/scielodata.5TEMPG, SciELO Data, DRAFT VERSION'. Below the title, there are links for 'Cite Dataset' and 'Learn about Data Citation Standards.'. At the bottom, there are two tabs: 'Description' and 'Subject'. Below these, there are four navigation tabs: 'Files', 'Metadata', 'Terms', and 'Versions'. A red box highlights these four tabs, and a red arrow points to them from the right.

Nível 1 - Curadoria básica

- Verificar se o conjunto de dados está relacionado com algum manuscrito submetido ao periódico.

Na aba “Versions”:

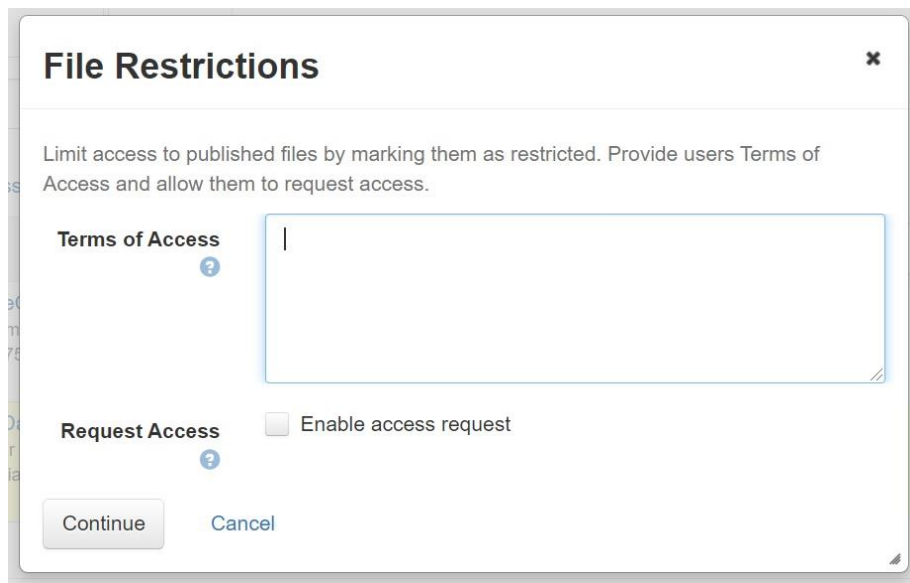
- Verificar se trata-se de um novo conjunto de dados ou uma nova versão de um conjunto de dados já publicado.

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;
- **Author:**
 - Verificar se os nomes dos autores foram inseridos na ordem inversa (Sobrenome, Nome);
 - Verificar se os autores informaram a afiliação e o ORCID (recomendado).
- **Subject:** verificar se a área temática selecionada é a mais adequada;
- **Keyword:** verificar se cada palavra-chave foi inserida separadamente (ou seja, cada uma em um campo);
- **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**.

Na aba “Files”:

- Verificar se o conjunto de dados foi documentado em um arquivo nomeado como README;
- Verificar se a nomeação dos arquivos está adequada (ver tópico 1 do [“Guia de preparação de dados de pesquisa”](#));
- Verificar se os arquivos podem ser abertos (se não estão corrompidos);
- Verificar se os arquivos estão em formatos recomendados (ver tópico 2 do [“Guia de preparação de dados de pesquisa”](#));
- Verificar se o conjunto de dados possui arquivos com acesso restrito.
 - Se sim, verificar se o campo “Terms of Access” foi preenchido com informações sobre e como os usuários podem obter acesso aos arquivos restritos.



Na aba “Terms”:

- Verificar se a licença de uso escolhida para o conjunto de dados foi a CCO - "Public Domain Dedication".
 - Se sim, confirmar com autores se desejam publicar o conjunto de dados sob essa licença;
 - Se não, verificar se o campo “Terms of Use” foi preenchido com alguma licença de uso (ver tópico 7 do [“Guia de depósito de dados de pesquisa”](#)).

Nível 2 - Curadoria detalhada → Antes realizar curadoria básica

Na aba “Metadata”:

- Avaliar informações fornecidas para determinar se estão completas e compreensíveis.

Na aba “Files”:

- Fazer download dos arquivos de dados.
- 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 (ver tópico 2 do [“Guia de preparação de dados de pesquisa”](#));
- Converter os arquivos para os formatos recomendados (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”](#)).

Na aba “Terms”:

- Avaliar informações fornecidas para determinar se estão completas e compreensíveis.

Nível 3 - Curadoria no nível de dados → Antes realizar curadoria básica + curadoria aprimorada

Na aba “Files”:

- Fazer download dos arquivos de dados;
- Abrir os arquivos de dados e verificar se necessitam de algum tratamento adicional;
- Abrir os arquivos de dados e verificar se há dados sensíveis que precisam ser desidentificados/anonimizados (informações de identificação pessoal, informações de saúde protegidas, etc.);
- 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;
- Executar e solucionar problemas em arquivos de códigos;
- Verificação de consistência (*checksum*²) dos arquivos do conjunto de dados para garantir a integridade dos dados no nível de bits;
- Se necessário, compartilhar o conjunto de dados com pareceristas para *peer review* através de uma URL privada³.

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

Excel (.xlsx)	<ul style="list-style-type: none">● Excel CURATED checklist
Google Docs	<ul style="list-style-type: none">● Google Docs CURATED Checklist
R (.r, .rmd)	<ul style="list-style-type: none">● Filetype CURATED checklist

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, o curador pode publicá-lo (clique em “Publish Dataset” e depois em “Publish”). O autor receberá um e-mail informando que o conjunto de dados foi publicado.

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:

² *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.

³ 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”.

- 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>.

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, 2021 [cited DD Month YYYY]. Available from: _____.