



Scientific Electronic Library Online

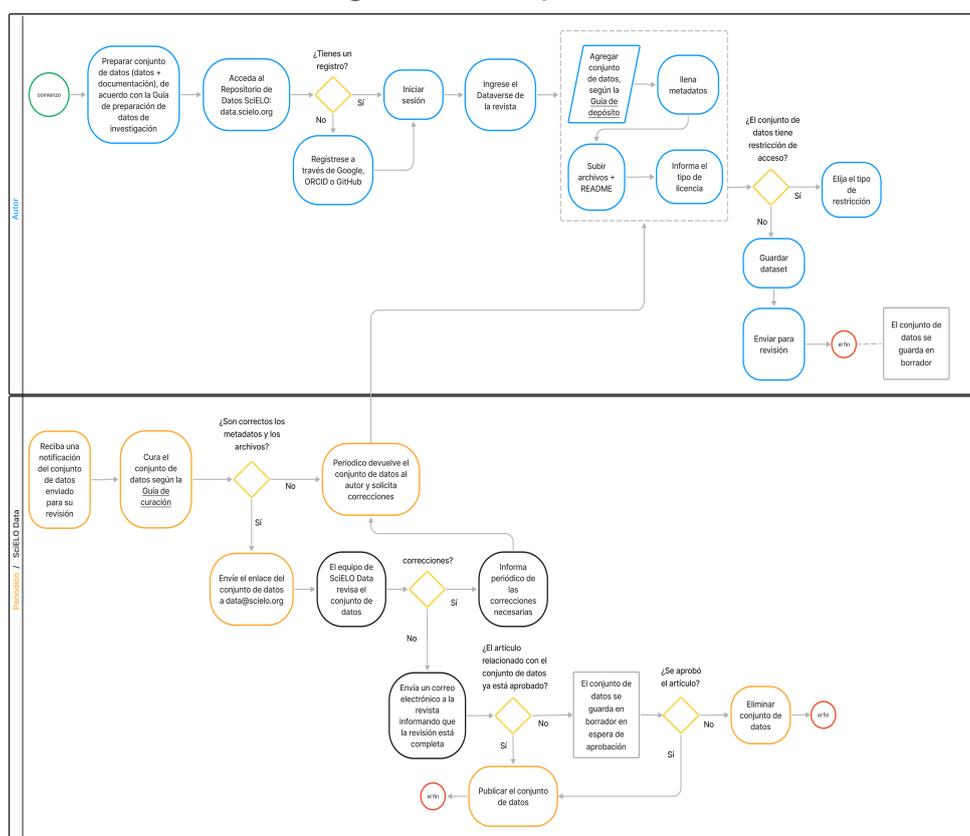
## Guía de curación de datos de investigación para equipos editoriales

Abril 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.

## Diagrama de flujo SciELO Data



De acuerdo con el *CoreTrustSeal*<sup>1</sup> (2019), curación de los datos de investigación es:

*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.*

Es decir, la curación de datos involucra todo el ciclo de vida de los datos (Data Life Cycle), desde la planificación para la recolección hasta la preservación para el acceso y reutilización a largo plazo. La curación de datos también se puede definir más específicamente como las verificaciones y acciones realizadas por los curadores con el fin de asegurar que el conjunto de datos esté estructurado y documentado de la manera más completa posible y de acuerdo con las mejores prácticas.

En el contexto de esta guía, el término curaduría se utilizará teniendo en cuenta esta segunda definición.

<sup>1</sup> CoreTrustSeal es una iniciativa del *World Data System* do *International Science Council* (WDS) e del *Data Seal of Approval* (DSA). Es una organización internacional, comunitaria, no gubernamental y sin fines de lucro que promueve infraestructuras de datos confiables y sostenibles.

## 1. Cita del conjunto de datos en el artículo vinculado

Para que exista una conexión entre el artículo publicado y el conjunto de datos subyacente, se recomienda documentar las instrucciones a los autores de que los artículos deben contener la sección "Disponibilidad de datos" informando si el conjunto de datos que se refiere a la investigación está disponible y, si es así, dónde acceder a él.

Ejemplo de contenido para la sección recomendada por SciELO:

### ***Datos no disponibles***

*El conjunto de datos que apoya los resultados de este estudio no está disponible públicamente (No aplica para artículos con datasets en SciELO Data).*

### ***Datos disponibles***

*Todo el conjunto de datos que apoya los resultados de este estudio fue publicado en el propio artículo (No aplica para artículos con datasets en SciELO Data).*

*Todo el conjunto de datos que apoya los resultados de este estudio fue publicado en el artículo en la sección "Materiales suplementarios" (No aplica para artículos con datasets en SciELO Data).*

*Todo el conjunto de datos que apoya los resultados de este estudio fue publicado en SciELO Data y puede ser accedido en [URL o DOI].*

*Todo el conjunto de datos que apoya los resultados de este estudio fue publicado en SciELO Data con los identificadores [lista de identificadores].*

*Todo el conjunto de datos anónimos que apoya los resultados de este estudio fue publicado en SciELO Data y puede ser accedido en [URL o DOI].*

### ***Datos disponibles a pedido<sup>2</sup>***

*Todo el conjunto de datos que apoya los resultados de este estudio está disponible mediante petición al autor de contacto [nombre del autor de contacto]. El conjunto de datos no está públicamente disponible debido a [detalle motivo de restricción, por ejemplo, contener información que compromete la privacidad de los participantes de la investigación] (No aplica para artículos con datasets en SciELO Data).*

*Todo el conjunto de datos que apoya los resultados de este estudio está disponible mediante solicitud a [nombre de la organización]. El conjunto de datos no está públicamente disponible debido a [detallar el motivo de restricción, por ejemplo, contener información que compromete la privacidad de los participantes de la investigación] (No aplica para artículos con datasets en SciELO Data).*

---

<sup>2</sup> En SciELO Data es posible depositar conjuntos de datos, pero mantenerlos restringidos. Cuando alguien quiera acceder a ellos, se pondrá en contacto con el autor depositante del conjunto a través de la plataforma. Ver más detalles en el ítem 5.1 de la [Guía de depósito de datos de investigación](#).

Para obtener información detallada sobre los niveles de aplicación de los criterios de datos, códigos y materiales de investigación en la [Guía para promover la apertura, transparencia y reproducibilidad de la investigación publicada por las revistas SciELO](#).

## 2. Anonimización de datos

La curaduría también es importante para verificar que los datos a publicar no violen las normas de los comités de ética e investigación de sus respectivas áreas, y para verificar si hay datos que necesitan anonimización (como datos personales, sensibles o no<sup>3</sup>, información que exceda los derechos de privacidad de las personas involucradas, o las ponga en riesgo, así como coordenadas de áreas protegidas, en peligro de extinción o información que infrinja acuerdos comerciales, patentes o pertenecientes a terceros).

La siguiente información ya aparece en la [Guía de preparación de datos de investigación](#), sin embargo, se repite a continuación para que la verifique el equipo editorial.

Los datos sensibles (datos que, si se exponen sin autorización o se pierden, podrían generar problemas legales) o datos personales deben anonimizar.

Ejemplo de anonimización de datos<sup>4</sup>:

Información sin anonimización	Responder sin anonimizar
Nombre	Juan Pérez
País de nacimiento	Argentina
Edad	54
Años de experiencia	25
Modelo de aviones	Boeing 777 Boeing 747
Fecha de último vuelo	05/01/2022

Información anonimizada	Respuesta anonimizada
-	-
Continente	América del Sur
Rango de Edad	50-60
Rango de años de experiencia	10-20
Modelo de aviones	Comerciales
Fecha de último vuelo	01/2022

<sup>3</sup> Datos personales: Pueden ser considerados datos personales: nombre y apellidos; dirección de casa; dirección de correo electrónico (si contiene elementos que ayuden a identificar al titular, como nombre y apellido); género; fecha de nacimiento; número de documentos de registro, como RG, CPF y tarjeta de trabajo; datos de geolocalización desde un teléfono celular; número de teléfono personal. <https://portal.fiocruz.br/noticia/entenda-melhor-lei-geral-de-protecao-de-dados-pessoais>. Portugués. Consultado el 21 de marzo de 2023.

Datos personales sensibles: "datos personales sobre origen racial o étnico, convicción religiosa, opinión política, afiliación u organización sindical de carácter religioso, filosófico o político, datos referentes a la salud o vida sexual, datos genéticos o biométricos, cuando se vinculen a una persona natural". persona". [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). Portugués. Consultado el 30 de enero de 2023.

<sup>4</sup> Ejemplo tomado de: Gestión de Datos de Investigación - Parte I. Disponible en: <https://www.youtube.com/watch?v=BM-lZ2XCCN0>

En los casos en que la práctica de la anonimización sea imposible, trate de usar seudónimos o considere no publicar.

### 3. Niveles de curación de datos

Con el objetivo garantizar transparencia con respecto a las verificaciones y acciones realizadas por los curadores con los conjuntos de datos depositados, SciELO adopta como referencia los niveles de curación, de 1 a 3, utilizados por *CoreTrustSeal* como requisito en la evaluación de repositorios de datos confiables:

- A. Curación básica: verificación rápida de metadatos / contenido, agregando metadatos básicos o documentación → **Nivel 1**;
- B. Curación detallada: curación básica + conversión de los archivos de datos a nuevos formatos, perfeccionar la documentación → **Nivel 2**;
- C. Curación a nivel de datos: curación básica + curación detallada + edición de datos depositados para una mayor precisión → **Nivel 3**.

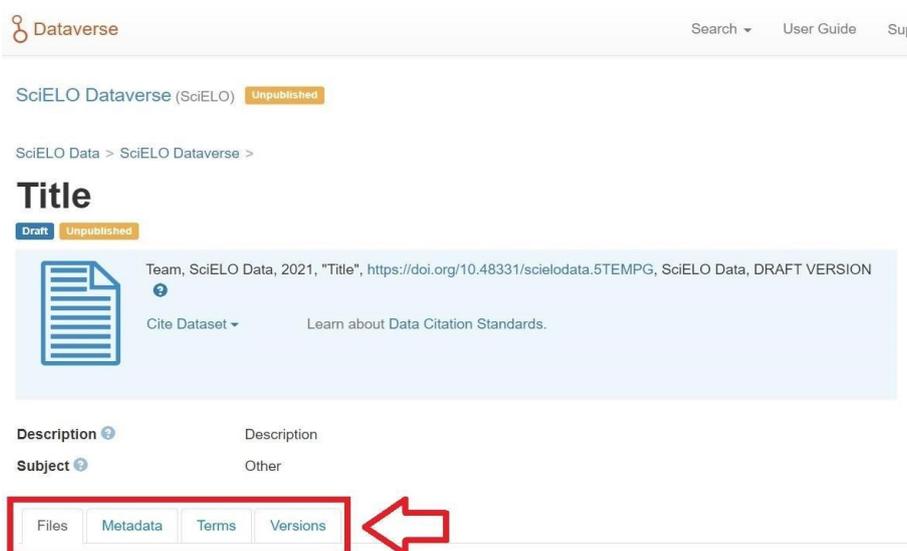
Independientemente del nivel de conservación, es obligatorio verificar los archivos en busca de datos personales o potencialmente confidenciales según el punto 2 de esta guía.

### 4. Listas de verificación y acciones de curación de datos

En el repositorio de SciELO Data existen 4 secciones donde los metadatos / archivos de datos deben ser curados teniendo en cuenta el nivel de curación adoptado y realizado por la revista:

- Ficheros (Files)
- Metadatos (Metadata)
- Condiciones (Terms)
- Versiones (Versions)

Puede navegar por estas secciones haciendo clic en las pestañas correspondientes:

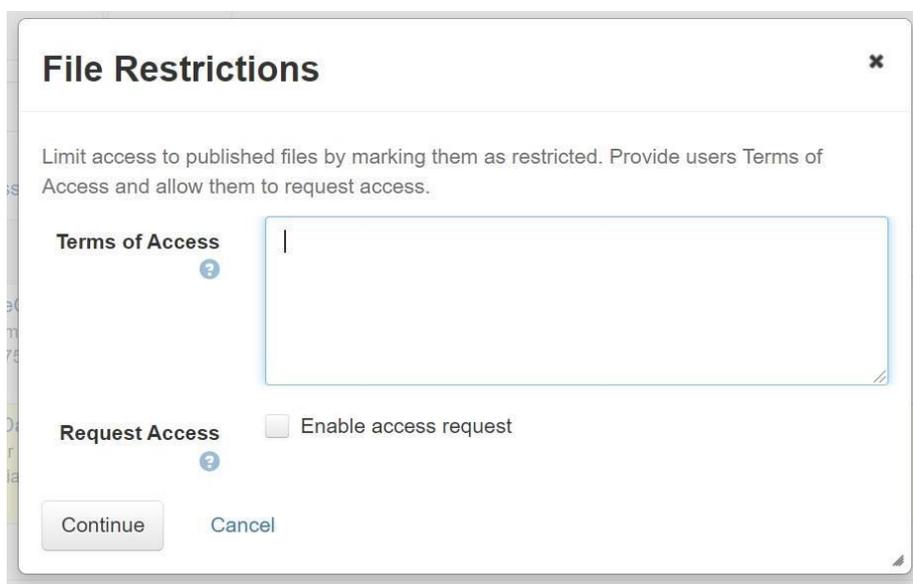


## Nivel 1 - Curación básica

- Asegúrese de que el conjunto de datos está relacionado con algún manuscrito enviado a la revista. Nota: Solo se pueden publicar conjuntos de datos de artículos aprobados.

### En la sección “Files”:

- Verifique si el conjunto de datos se encuentra documentado en un archivo llamado README. La presencia de este archivo es obligatoria;
- Verificar si la denominación de los archivos es adecuada (ver ítem 1 de la “[Guía de preparación de datos de investigación](#)”). Si no está recomiende a los autores que lo editen siguiendo las recomendaciones;
- Verifique si los archivos pueden abrirse (que no estén dañados). Si los archivos no se abren, solicitar un nuevo depósito a los autores;
- Verifique los archivos en busca de datos que necesiten anonimización. Si es necesario anonimizar, consulte el ítem 2 de esta guía;
- Verifique si los archivos están en los formatos recomendados (consulte el tema 2 del [Guía de preparación de datos de investigación](#)). Si no está en los formatos correctos, recomendar a los autores que lo editen siguiendo las recomendaciones;
- Verifique si el conjunto de datos tiene archivos con acceso restringido.
  - En caso afirmativo, verifique si el campo "Terms of Access" esta llenado con información sobre cómo los usuarios pueden obtener acceso a los archivos restringidos.



### En la sección “Metadata”:

- **Title:** verificar si fue llenado el título del artículo cuyos datos pertenecen o con un título propio que sea significativo/descriptivo para el conjunto de datos. Si no recomendar autores para editar;
- **Author:**
  - Verifique si los nombres y apellidos de los autores se ingresaron en orden inverso (Apellido, Nombre);

- Verificar si los autores tienen afiliación informada (obligatorio) y ORCID (recomendado).
- **Subject:** verificar si el área temática seleccionada es la más adecuada. Evite el uso de "Otro" siempre que sea posible;
- **Keywords:** verifique si cada palabra clave fue ingresada por separado (es decir, cada palabra clave un campo). Si no, editarlos en la pantalla de edición de metadatos y agregue palabras clave presionando el signo "+";
- **Related publication:**
  - Si el conjunto de datos está relacionado con un manuscrito en revisión, ingrese el **título del manuscrito**;
  - Si el conjunto de datos se relaciona con un artículo publicado, inserte la **cita del artículo con DOI**.
- **Funding Information:** si la encuesta tiene una fuente de financiación, haga clic en "Editar conjunto de datos" y luego en "Metadatos". Encuentra el campo y complétalo.

#### En la sección "Terms":

- Verifique si la licencia de uso elegida para el conjunto de datos es CC BY 4.0. Si desea utilizar otra licencia, póngase en contacto con [data@scielo.org](mailto:data@scielo.org).

#### En la sección "Versions":

- Verifique si se trata de un nuevo conjunto de datos o de una nueva versión de un conjunto de datos ya publicado.

**Nivel 2 - Curación detallada** → Realizar la curación básica y:

#### En la sección "Files":

- Cambiar el nombre de los archivos según corresponda (consulte el ítem 1 de la "[Guía de preparación de datos de investigación](#)");
- Evaluar los formatos de archivo para determinar si están en un formato recomendado. Si es necesario conviértalos (consulte el ítem 2 de la "[Guía de preparación de datos de investigación](#)");
- Evaluar si la documentación proporcionada (archivo README, *codebook*, etc.) está completa y es comprensible (ver tema 4 de la "[Guía de preparación de datos de investigación](#)"). De no ser así, solicitar los cambios necesarios a los autores.

#### En la sección "Metadata":

- Evaluar la información proporcionada para determinar si esta completa y comprensible. Solicite correcciones o realice ediciones si es necesario.

#### En la sección "Terms":

- Evaluar la información proporcionada para determinar si es completa y comprensible. Solicite correcciones o realice ediciones si es necesario.

**Nivel 3 - Curación a nivel de datos** → Realizar la curación básica + curación detallada y:

**En la sección “Files”:**

- Descargar archivos de datos;
- Abrir los archivos de datos y verificar si necesitan más procesamiento. Si es necesario, solicitar correcciones a los autores o realizar los cambios e informarles;
- Abrir los archivos de datos y evaluarlos en busca de posibles problemas, como: definiciones de valores y variables adecuadas, valores fuera de rango, descripciones de programas utilizados para archivos de código y estructuras de datos preferidas. Si es necesario, solicitar correcciones a los autores o realizar los cambios e informarles;
- Ejecutar y solucionar problemas de archivos de código;
- Comprobación de coherencia (*checksum*<sup>5</sup>) de archivos de conjuntos de datos para garantizar la integridad de los datos a nivel de bit.

Si el conjunto de datos no está debidamente estructurado y/o documentado, el curador puede devolverlo al autor (haga clic en "Publish Dataset" y luego en "Return to Author").

Si el conjunto de datos es apto para publicación, envíe un correo electrónico a [data@scielo.org](mailto:data@scielo.org) informando la URL del conjunto de datos y solicite la verificación. La curaduría SciELO es temporal. Después de que regrese el equipo de SciELO Data, la revista puede continuar con la publicación (haga clic en “Publish Dataset” y luego en “Publish”). El autor recibirá un correo electrónico informándole que el conjunto de datos ha sido publicado.

**IMPORTANTE:**

Una vez publicado, el conjunto de datos no se puede eliminar. Se recomienda editar el conjunto, generando así una segunda versión del conjunto de datos. El DOI seguirá siendo el mismo.

Para casos excepcionales que requieran exclusión, no es posible “desaparecer” con el conjunto de datos, solo “deshabilitarlo”. El DOI del conjunto de datos conducirá a la misma página que mostrará la cita del conjunto con la información "VERSIÓN DESACCESIONADA" y el motivo de la no disponibilidad.

Si, después de la publicación, cualquier conjunto de datos necesita cambios o se pierde el acceso, comuníquese con [data@scielo.org](mailto:data@scielo.org).

SciELO Data permite compartir el conjunto de datos con pares revisores a través de una URL privada<sup>6</sup>.

---

<sup>5</sup> *Checksum* o suma de verificación es una secuencia de números y letras que se utiliza para verificar la integridad de los datos, es decir, si un archivo es exactamente el mismo después de una transferencia, verificar que no ha sido alterado por un tercero o no está dañado.

<sup>6</sup> La creación de una URL privada ("URL privada") permite compartir (para ver y descargar archivos) un conjunto de datos no publicado con un grupo de personas que pueden no tener una cuenta de usuario de SciELO Data, es decir, cualquiera que reciba la URL privada podrá acceder al conjunto de datos sin tener que registrarse o iniciar sesión en SciELO Data.

Para obtener información sobre la conservación de datos de archivos en 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 ayudar en la gestión de borradores de conjuntos de datos, el administrador o curador de Dataverse de la revista puede asignar etiquetas a un conjunto de datos para indicar su estado actual:

- Journal curation: curación de revistas en curso/pendiente;
- Author contacted: en espera de correcciones de los autores;
- Privacy Review: revisión de los revisores pendiente/en curso;
- SciELO curation: Curación SciELO en curso / pendiente;
- Awaiting article approval: en espera de la aprobación del artículo relacionado para la publicación del conjunto de datos.

Para agregar una etiqueta, haga clic en "Publish Dataset" → "Change Curation Status" y luego elija el estado apropiado. Las etiquetas se eliminarán automáticamente del conjunto de datos cuando se publique.

## Referencias

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>.

---

Para crear una URL privada, vaya al conjunto de datos → haga clic en "Editar" → "URL privada" → en el cuadro emergente, seleccione "Crear URL privada" o "Create URL for Anonymized Access" (permite la revisión anónima eliminando los nombres de los autores y otra información que pueda identificar la cita) → copie la URL creada y compártala con los revisores. Cuando desee deshabilitar la URL privada, vaya al conjunto de datos → haga clic en "Editar" → "URL privada" → "Deshabilitar URL privada" → "Sí, deshabilitar URL privada".

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>.

## **Cómo citar este documento**

SciELO. *Guía de curación de datos de investigación para equipos editoriales* [online]. SciELO, 2023 [cited DD Month YYYY]. Available from: \_\_\_\_\_.

**Este y otros documentos del SciELO Data están disponibles en:**

<https://www.scielo.org/es/sobre-el-scielo/scielo-data-es/>