



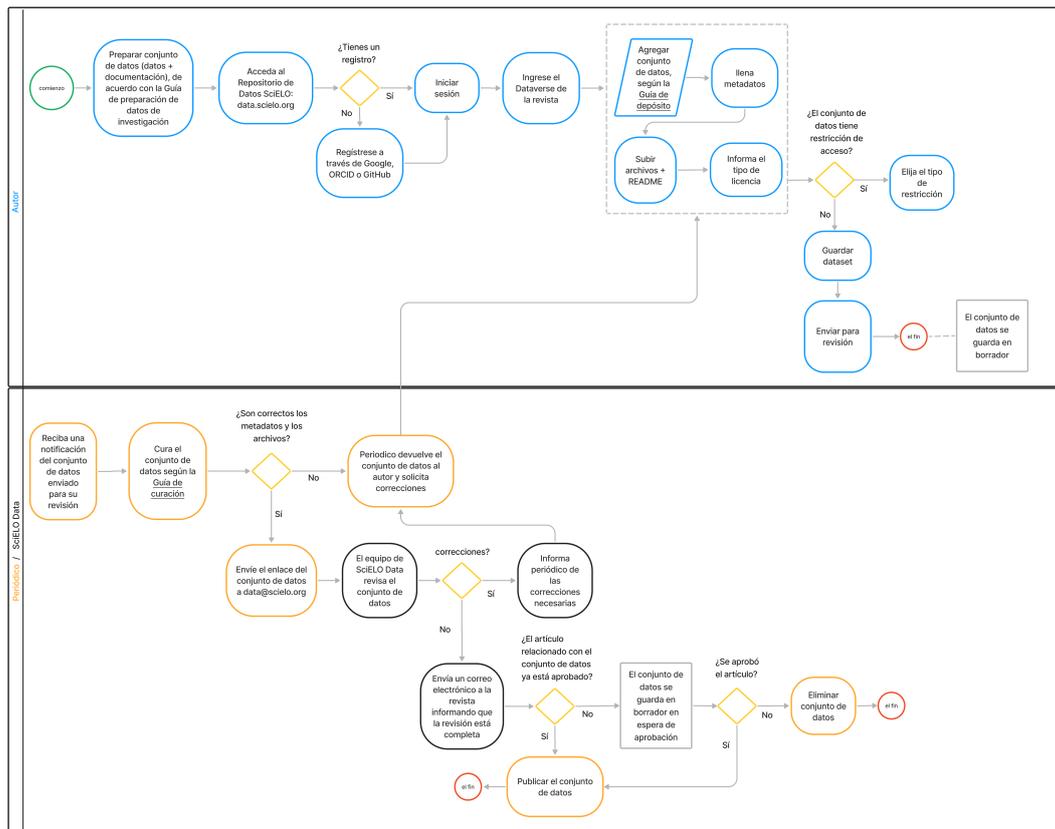
## Guía de preparación de datos de investigación

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.

## Diagrama de flujo SciELO Data



La adopción de buenas prácticas en la preparación de datos para su depósito es un aspecto de fundamental importancia para acceder, compartir y reutilizar los datos de investigación, además de contribuir a que sean lo más FAIR posible.<sup>1</sup>

Para comprender mejor los principios FAIR y cómo hacer que sus datos sean lo más FAIR posible, le recomendamos que utilice la herramienta FAIR-Aware: <https://fairaware.dans.knaw.nl/>.

### 1. Nombre de archivos

La adopción de buenas prácticas de nomenclatura de archivos evita que se sobrescriban y facilita la ubicación y reutilización por parte de otros investigadores.

- Utilice nombres descriptivos / significativos, con un máximo de 100 caracteres;
- No uses acentos;
- Utilice caracteres alfanuméricos, subrayado (my\_data) o guion (my-data);
- Evite los espacios (my.data), puntos (my.data), iniciales en mayúscula (my.data) y caracteres especiales (como \ / ? : \* " > < | : # % " { } | ^ ` ~ @ & ; ° æ Æ ø Ø å Å ä Ä ö Ö);
- Utilice el formato AAAA-MM-DD (my\_data\_2021-01-07) o AAAAMMDD (my\_data\_20210107) para las fechas;

<sup>1</sup> Recomendamos el webinar "¿Cómo hacer los datos sean FAIR? Buenas prácticas para datos (abiertos) de investigación" disponible en: <https://www.youtube.com/watch?v=I14SwZxIRHY>.

- Incluya el número de versión en el nombre cuando corresponda (my-data001.csv, my-data002.csv..., my-data010.csv...);
- Use el mismo nombre para archivos con el mismo contenido, pero en diferentes formatos (my-data.doc y my-data.txt).

## 2. Formato de los archivos

SciELO Data acepta cualquier tipo de archivo, sin embargo, se recomienda depositar formatos:

- No propietarios;
- Abiertos, con padrones internacionales documentados;
- Con codificación de caracteres estándar, preferiblemente Unicode (p. Ej. UTF-8).

Tipo de documento	Formatos recomendados	Formatos no recomendados (pero aceptados)
Análisis estadísticos	R (.r, .rdata) SPSS (.dat/.sps) STATA (.dat/.do)	SPSS Portable (.por) SPSS (.sav)
Archivos compactados	.zip*	.rar
Datos tabulares	Comma Separated Values (.csv) Archivo de texto (.txt)	Excel (.xlsx)
Datos textuales	Archivo de texto (.txt) OpenOffice (.odt, .ods ou .odp) PDF (.pdf)	Word (.doc ou .docx)

\* Los archivos compactados con la extensión .zip serán descompactados después del upload de los archivos.

## 3. Tamaño de los archivos

El tamaño límite de los archivos individuales es de 2 GB. Para agregar archivos por encima de este límite, comuníquese con [data@scielo.org](mailto:data@scielo.org).

## 4. Descripción de los datos

Para ayudar a garantizar que los datos depositados puedan ser interpretados y reutilizados correctamente, tanto por usted en una fecha posterior como por otros investigadores, es esencial que se describan de la manera más detallada y comprensible posible.

Esta descripción detallada debe proporcionarse completando los campos durante el depósito y de un archivo README, que actúa como una guía para el usuario, y debe ser depositado con los archivos de datos.

El archivo README debe escribirse como texto sin formato con codificación de caracteres Unicode UTF-8 (.txt) o como PDF si necesita ilustrar o formatear la descripción de los datos.

El archivo README debe contener al menos la siguiente información:

- Título del conjunto de datos;
- Información de contacto (nombre, institución y correo electrónico) del investigador principal / corresponsal o responsable de la recopilación de datos;
- Fecha de recopilación de datos (fecha única o intervalo de tiempo);
- Resumen de datos y archivos (breve descripción de los datos que contiene cada archivo, fecha de creación de cada archivo y cómo se relacionan entre sí, etc.);
- Descripción de los métodos de recopilación o generación de datos;
- Descripción de los métodos utilizados para procesar los datos;
- Información de datos específicos (lista de variables, unidades de medida, definiciones de códigos o símbolos, calibración de equipos, etc.).

Para ver ejemplos de otra información que se puede agregar al archivo README, consulte:

- [Guide to writing "readme" style metadata](#)

Ejemplos de plantillas para el archivo README:

<b>General</b>	<ul style="list-style-type: none"><li>• <a href="https://drive.google.com/file/d/167cJdaRy4sxQWA5qEEp2cgfWqKV-smZV/view">https://drive.google.com/file/d/167cJdaRy4sxQWA5qEEp2cgfWqKV-smZV/view</a></li><li>• <a href="https://cornell.app.box.com/v/ReadmeTemplate">https://cornell.app.box.com/v/ReadmeTemplate</a></li></ul>
<b>Ciencias Sociales</b>	<ul style="list-style-type: none"><li>• <a href="https://social-science-data-editors.github.io/template_README/template-README.html">https://social-science-data-editors.github.io/template_README/template-README.html</a></li></ul>
<b>Código de software</b>	<ul style="list-style-type: none"><li>• <a href="https://drive.google.com/file/d/1VIDF489DDr044Uta8z1G7EuLj-Wm_UtG/view">https://drive.google.com/file/d/1VIDF489DDr044Uta8z1G7EuLj-Wm_UtG/view</a></li></ul>

Para obtener más información sobre la preparación de datos, consulte también:

<b>Códigos y datos de computadores</b>	<ul style="list-style-type: none"><li>• <a href="#">Experiences on reproducibility of paper experiments</a></li><li>• <a href="#">Research Code</a></li></ul>
<b>Datos de Ciencias Sociales</b>	<ul style="list-style-type: none"><li>• <a href="#">Guide to Social Science Data Preparation and Archiving</a></li></ul>
<b>Datos tabulares</b>	<ul style="list-style-type: none"><li>• <a href="#">Preparing tabular data for description and archiving</a></li></ul>

## 5. Anonimización de datos

Deberán anonimizar: Los datos personales, sean o no sensibles<sup>2</sup>, la información que exceda el derecho a la privacidad de las personas involucradas, o las ponga en riesgo, así como las coordenadas de áreas protegidas, en peligro de extinción o la información que viole los derechos comerciales. acuerdos, patentes o pertenece a otros.

Ejemplo de anonimización de datos<sup>3</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

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

## Referencias

Cornell University. Guide to writing "readme" style metadata. *Cornell University* [online]. [viewed 12 February 2021]. Available from: <https://data.research.cornell.edu/content/readme>.

DataverseNO. Prepare your data. *DataverseNO* [online]. [viewed 12 February 2021]. Available from: <https://site.uit.no/dataverseno/deposit/prepare/>.

<sup>2</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>. Nuestra traducción. 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). Nuestra traducción. Consultado el 30 de enero de 2023.

<sup>3</sup> Ejemplo tomado de: Gestión de Datos de Investigación - Parte I. Disponible en: <https://www.youtube.com/watch?v=BM-lZ2XCCNO>

Nanyang Technological University. DR-NTU (Data) User Guides and Policies. *Nanyang Technological University* [online]. [viewed 12 February 2021]. Available from: <https://libguides.ntu.edu.sg/drntudataguidespolicies/depositor#s-lg-box-21651979>.

UC Santa Barbara Library. The Dos and Don'ts of file naming. *UC Santa Barbara* [online]. [viewed 30 March 2021]. Available from: <https://www.library.ucsb.edu/sites/default/files/dls-n01-2021-filenaming.pdf>.

University of Illinois at Urbana-Champaign. Dataset Documentation. *Illinois Data Bank* [online]. [viewed 12 February 2021]. Available from: [https://databank.illinois.edu/help#dataset\\_documentation](https://databank.illinois.edu/help#dataset_documentation).

## Cómo citar este documento

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