

## Cómo depositar sus datos de investigación en e-cienciaDatos

<p><b>¿Quién puede depositar?</b></p>	<p>Personal investigador de las universidades miembro del Consorcio Madroño.</p>
<p><b>¿Cómo?</b></p>	<p>El depósito se realiza mediante archivo delegado: el personal investigador/grupo de investigación solicita a la biblioteca de su universidad la creación de un dataset.</p> <p><a href="#">Direcciones de contacto</a></p>  <p><a href="https://goo.gl/sX2pGk">https://goo.gl/sX2pGk</a></p>
<p><b>¿Qué datos se pueden depositar?</b></p>	<p>Conjuntos de datos de cualquier disciplina. Datos finales o datos completos, preparados para su distribución pública y su reutilización. Deberán contener la pertinente documentación anexa (<a href="#">Readme.txt</a>)</p>
<p><b>¿Cuándo depositar?</b></p>	<p>Si están financiados por fondos públicos, cuando lo determine la entidad financiadora. <a href="#">Open Research Data in Horizon2020</a>.</p> <p>Los datos que validen resultados de investigación, lo antes posible.</p> <p>Los otros datos susceptibles de ser compartidos que figuren en el Plan de Gestión de Datos una vez finalizado el proyecto.</p>
<p><b>Licencias de depósito y uso</b></p>	<p>En la <a href="#">licencia de depósito</a> se establece que los datos que se depositan son originales y no infringen ningún derecho de otros autores. Hay disponibles diversas <a href="#">licencias libres</a> de uso para los datasets siendo las más habituales las licencias del tipo <b>CC0</b> y <b>CC-BY</b>. Si un investigador necesita otra licencia o una combinación de las mismas, deberá ponerse en contacto con los administradores de <a href="#">e-cienciaDatos</a>. Los datos estarán en acceso abierto.</p>
<p><b>¿Qué formatos de ficheros se admiten?</b></p>	<p>Formatos abiertos no propietarios para garantizar la accesibilidad y preservación y formatos propietarios que tienen un uso extendido entre las comunidades científicas que tienen interés en los datos.</p>
<p><b>Datos FAIR</b></p>	<p>Los datasets deben ser localizables, accesibles, interoperables y reutilizables, para lo que tienen metadatos descriptivos estandarizados, licencias de uso, <a href="#">documentación explicativa</a> (readme.txt) y DOI.</p>
<p><b>Valores añadidos</b></p>	<p>e-cienciaDatos se basa en el asesoramiento por parte de las bibliotecas, cumple con los requisitos de OpenAire y ofrece estadísticas y <a href="#">geolocalización</a>. Permite descargar los metadatos en varios formatos compatibles con los principales gestores bibliográficos para citarlos correctamente.</p>

## Buenas prácticas en la gestión de datos de investigación

### Plan de Gestión de Datos

Un Plan de Gestión de Datos (PGD) o *Data Management Plan* (DMP) es un documento formal que debe realizarse en los inicios de la investigación en el que se describe cómo se planifica la gestión de datos de investigación durante todo el proceso. Es un documento vivo.

Siempre es una buena práctica. Además, es necesario elaborarlo siempre que el proyecto esté financiado por Horizonte 2020 (Piloto de Datos de Investigación en Abierto).

En el caso de que el proyecto esté financiado por Horizonte 2020 pero no genere datos, o éstos puedan suponer un conflicto con la obligación de proteger los resultados, o con las obligaciones de confidencialidad, seguridad nacional o con las leyes de protección de datos personales, el investigador puede optar por no realizar un PGD. Más información en las [FAQs](#).

El Consorcio Madroño ha adaptado la herramienta [PGD Online](#) para facilitar la elaboración de los Planes de Gestión de Datos.

### Documentación de los Datos

Es necesario documentar el ciclo de vida de los datos para que sean accesibles, reutilizables e interoperables con otros conjuntos de datos. Existen diversos estándares o esquemas para ello. El Consorcio Madroño ha preparado un fichero "[Readme.txt](#)" con la información requerida. La información proporcionada en esta plantilla se incluirá junto al resto de ficheros del dataset.

### Características de los ficheros

Utilizar siempre que sea posible formatos abiertos, no propietarios.

Elegir formatos estándar para que la mayoría de softwares sean capaces de interpretarlos.

El formato utilizado ha de permitir la indización del contenido para su potencial recuperación.

Evitar formatos desconocidos y poco documentados.

Es recomendable optar por formatos que tengan un uso extendido para tener una mayor probabilidad de apoyo en el caso de que sea necesario.

Un formato de compresión de datos utiliza menos espacio de almacenamiento.

Algunos ejemplos de formatos más recomendados: Imagen (JPEG, JPG-2000, PNG, TIFF); Texto (HTML, XML, PDF/A); Audio (AIFF, WAVE); Almacenamiento (TAR, GZIP, ZIP).

Más información en [UK Data Service guidance on recommended formats](#)

### Organización y nombre de los ficheros

No existe una normativa al respecto, pero se pueden seguir unas recomendaciones básicas:

Evitar los nombres muy largos y estructura complejas de carpetas. Los nombres de los ficheros deberían ser ricos en información con los suficientes elementos para una fácil recuperación.

Usar el guión bajo (\_) como delimitador de los diferentes elementos en el nombre de los ficheros. No utilizar espacios, tildes ni caracteres especiales.

Usar guión medio (-) para separar palabras dentro del mismo elemento o usar mayúsculas.

Los elementos deberían ordenarse de lo general a lo específico siempre que sea posible.

Las fechas deberían ordenarse: AÑO, MES, DÍA (ej. YYYYMMDD) Los tiempos deberían seguir el formato: HORA, MINUTOS, SEGUNDOS (HHMMSS).

Los nombres personales deberían comenzar por el apellido.

Abreviar el contenido de los elementos siempre que sea posible.

El control de las versiones debería incluirse en el nombre de los ficheros, ejemplo: MCIM\_Proposal\_V09.

Existen software específicos para renombrar ficheros y carpetas: [https://www.bulkrenameutility.co.uk/Main\\_Intro.php](https://www.bulkrenameutility.co.uk/Main_Intro.php)

## Buenas prácticas en la gestión de datos de investigación

### Datos tabulares

Recomendaciones generales para preparar datos tabulares para su descripción y archivo:

No se deben incluir gráficos ni análisis en un fichero de datos.

Los repositorios individuales pueden tener requisitos específicos con respecto a los formatos de archivo. Si un repositorio no tiene requisitos de formato de archivo, se recomienda utilizar texto delimitado por tabulaciones o comas (\*.txt o \*.csv).

Es conveniente agregar los datos en pocos ficheros de mayor tamaño en lugar de utilizar muchos más pequeños.

Se debe usar solo la primera línea para identificar el encabezado de una columna.

Los encabezados deben ser significativos, pero no demasiados largos.

Utilizar un método estándar para identificar los datos que faltan. Está comúnmente aceptado el uso de -999 o -9999.

Más información en: <https://data.research.cornell.edu/content/tabular-data>

Existe una herramienta gratuita que permite limpiar y transformar ficheros de datos de gran tamaño: <http://openrefine.org/>

### Ética, seguridad y preservación de los datos

La estrategia de almacenamiento de datos debe contemplar la obsolescencia del hardware y del software, por lo que conviene copiar los datos en diferentes tipos de soporte físico. Lo más aconsejable es almacenar los ficheros en un servidor institucional seguro gestionado por el servicio de informática, lo que permitirá el acceso en red y la sincronización con el trabajo de otros investigadores que participan en la gestión de los datos.

Se debe programar de forma automática la realización de copias de seguridad y se debe constatar la integridad de los ficheros, el tamaño y la fecha.

Si fuera necesario anonimizar los datasets, porque aparecen datos de tipo personal, OpenAire ofrece la herramienta Amnesia: <https://amnesia.openaire.eu/>

### Cómo citar los datos

Es importante citar correctamente los datos para favorecer su reutilización en nuevas investigaciones, en actividades académicas y para otros fines de la sociedad en general.

Cómo citar correctamente los datos de investigación: <http://www.rebiun.org/acceso-abierto/cita-tus-datos-de-investigacion>