



# GESTIÓN DE DATOS DE INVESTIGACIÓN

Apoyo a la Investigación  
Ciencia Abierta

# CIENCIA ABIERTA Y DATOS DE INVESTIGACIÓN

Los datos de investigación son una información muy valiosa. La buena gestión de éstos, permite que sea accesibles y reutilizables por otros investigadores.

En el contexto de la [Ciencia Abierta](#), los datos de investigación deben estar disponibles en abierto con el fin de conseguir:

- Transparencia.
- Transferencia del conocimiento más efectiva.
- Replicabilidad (disminución del fraude).
- Mayor productividad.
- Mejora del impacto social de la investigación.



# Ciencia Abierta: La investigación y los datos científicos accesibles y abiertos a todos los ciudadanos

## Open Science



**Open Repositories**  
Repositorios Abiertos



**Open Access**  
Acceso Abierto

Acceso sin trabas económicas, tecnológicas o jurídicas a las publicaciones científicas

**Open Access Journals**  
Revistas de Acceso Abierto



**Open Peer Review**  
Revisión por Pares Abierta



**Open Science Evaluation**  
Evaluación de la Ciencia en Abierto

Evaluación abierta de los resultados de investigación, ampliando la revisión tradicional con la contribución de la comunidad

**Open Metrics and Impact**  
Impacto y Métricas Abiertas



**Open Reproducible Research**  
Investigación Reproducible en Abierto

Acceso libre a los elementos experimentales para la reproducción de la investigación



**Open Research Data**  
Datos de Investigación Abiertos

**Open Source in Open Science**  
Código Abierto para la Ciencia Abierta



**Open Data**  
Datos Abiertos

Datos que están disponibles en línea de forma gratuita y que se pueden usar, reutilizar y distribuir



**Open Big Data**  
Datos Masivos Abiertos

**Open Government Data**  
Datos Gubernamentales Abiertos



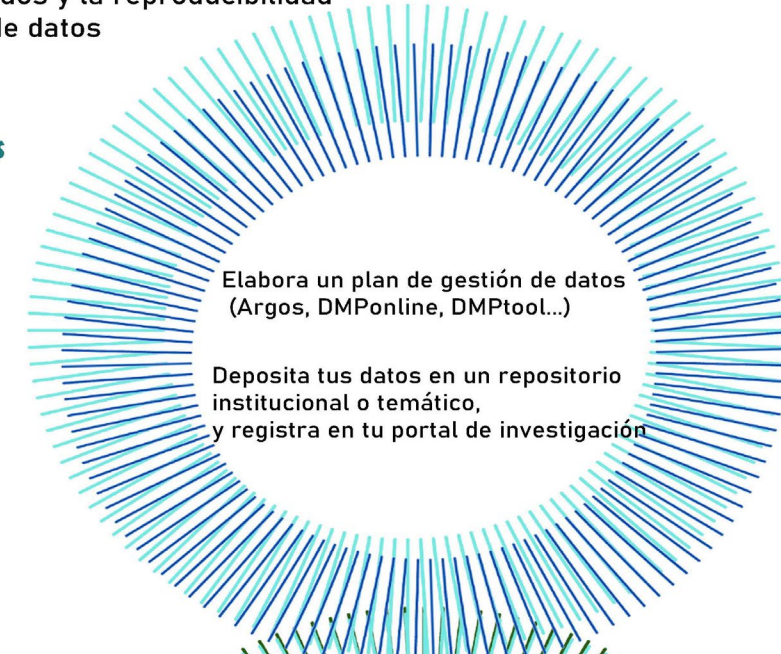
# LOS DATOS DE INVESTIGACIÓN IMPORTAN: GESTIONA, COMPARTE, PUBLICA Y REUTILIZA

Los datos de investigación son aquellos materiales generados o recolectados durante el transcurso de una investigación. En general es todo aquello que un investigador necesita para validar los resultados obtenidos en esa investigación

## Compartir tus datos

Promueve la innovación  
Facilita la colaboración  
Maximiza la transparencia y la fiabilidad  
Permite la verificación de los resultados y la reproducibilidad  
Reduce costes al evitar duplicación de datos

## Planifica la gestión de tus datos



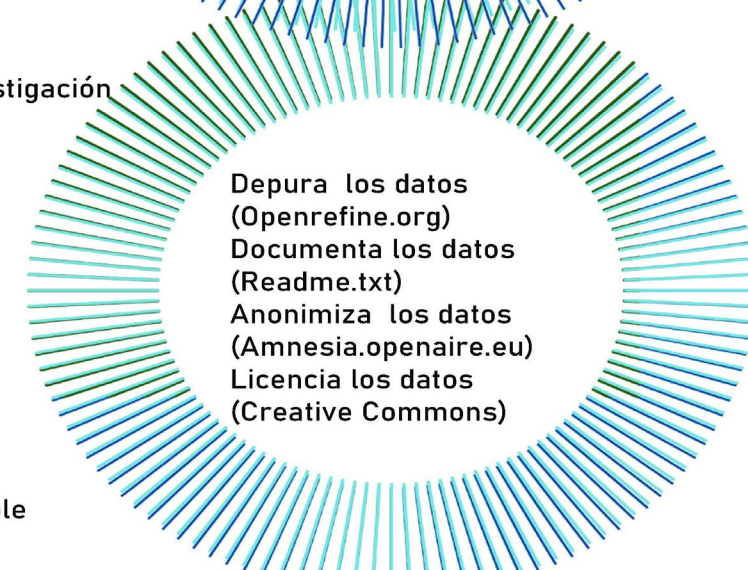
Políticas institucionales y legislativas, agencias de financiación, requieren abrir y compartir los datos de investigación

## Haz tus datos

F<sub>indable</sub> A<sub>ccessible</sub> I<sub>nteroperable</sub> R<sub>eusable</sub>



Localizable Accesible Interoperable Reutilizable



## DATOS DE INVESTIGACIÓN

- Según su naturaleza: cualitativos o cuantitativos.
- Según su nivel de procesamiento: En estado bruto (datos primarios), procesados o analizados.
- Según la fuente de la que provienen pueden ser experimentales (ej. cromatografías), observacionales (ej. encuestas) y computacionales (obtenidos mediante simulación).
- Según su formato: Textuales (Word, PDF, RTF, etc.), Numéricos (Excel, CSV, etc.), Multimedia (JPEG, MPEG, WAV, etc.), Estructurados (XML, MySQL, etc.), Código de software (Java, C, etc.), Específicos de un software (Mesh, 3D CAD, modelo estadístico, etc.), específicos de una disciplina o instrumento.
- Según su apertura: abiertos (aquellos a los que cualquier persona puede acceder, usar y compartir) o restringidos (datos confidenciales, datos personales o datos comerciales confidenciales).

# GESTIÓN DE DATOS DE INVESTIGACIÓN

- Por gestión de datos de investigación (RDM, Research Data Management) entendemos el conjunto de tareas que permiten la recopilación, organización, documentación, almacenamiento y preservación de los datos utilizados o generados durante un proyecto de investigación. Gestionar los datos de investigación es importante para asegurar la reproducibilidad de tu investigación, así como para demostrar la veracidad de la información y la metodología de la misma.

# GESTIÓN DE DATOS DE INVESTIGACIÓN

- Algunos de los motivos de gestionar los datos de investigación son:
  1. Cumplir con los requerimientos de los organismos financiadores.
  2. Mayor transparencia para la validación de los resultados de la investigación.
  3. Mejorar la protección de datos y minimizar el riesgo de pérdida de los mismos.
  4. Asegurar que los datos sean localizables, accesibles, interoperables y reutilizables: datos FAIR (Findable, Accessible, Interoperable and Reusable).
  5. Ahorrar tiempo evitando duplicaciones y haciendo un uso eficiente de los recursos disponibles.
  6. Mejorar el perfil del investigador, el impacto y la visibilidad de los proyectos

# PRINCIPIOS FAIR

- Son un conjunto de principios rectores para hacer que los datos de investigación sean fáciles de encontrar, accesibles, interoperables y reutilizables (Wilkinson et al., 2016).
- En el contexto de la Ciencia Abierta, es importante el cumplimiento de estos principios bajo el lema:  
**TAN ABIERTOS COMO SEA POSIBLE, TAN CERRADOS COMO SEA NECESARIO.**
- Puedes consultar estos recursos para profundizar en los datos FAIR:
- [¿FAIR o no FAIR? He ahí la gestión de datos](#). REBIUN.
- [Principios FAIR: Buenas prácticas para la gestión y administración de datos científicos](#). Gobierno de España.
- Existen diversas **herramientas gratuitas** online que evalúan el grado de cumplimiento con los Principios FAIR, como:
- [OpenAire](#) mantiene una sección con recursos de apoyo sobre los datos FAIR.



## PRINCIPIOS FAIR FINDABLE-LOCALIZABLES

- Cómo se cumple:
- Identificador persistente: DOI, Handle...
- Descritos con metadatos enriquecidos.
- Los metadatos están en línea en un recurso de búsqueda, están recolectados por motores de búsqueda.
- El registro de metadatos especifica el identificador persistente.

## PRINCIPIOS FAIR

### ACCESIBLE - ACCESIBLES

- Cómo se cumple:
- **El ID persistente enlaza con los datos o metadatos asociados para que estos puedan ser recuperados.**
- El protocolo para recuperar los datos sigue estándares reconocidos, por ejemplo <http://>.
- Acceso con autenticación y autorización en caso de que sea necesario.
- **Los metadatos serán accesibles, siempre que sea posible, aunque los datos no sean.**

## PRINCIPIOS FAIR INTEROPERABLE - INTEROPERABLES

- Cómo se cumple:
- Formatos abiertos
- Metadatos siguen estándares reconocidos.
- Vocabularios controlados, palabras clave, tesauros u ontologías.
- Referencias y enlaces a otros datos relacionados.

## PRINCIPIOS FAIR REUSABLE - REUTILIZABLES

- Cómo se cumple:
- Los datos son precisos y están bien descritos con todos los atributos relevantes.
- **Los datos tienen una licencia de uso de datos clara y accesible.**
- Queda claro cómo, por qué y quién ha creado y procesado los datos.
- Los datos y sus metadatos cumplen con los estándares que usa la comunidad de esa disciplina.

## GESTIÓN DE DATOS EN 10 PASOS

- 1. Formatos
- 2. Nombre de los archivos
- 3. Organización de los datos
- 4. Control de versiones
- 5. Almacenamiento
- 6. Seguridad
- 7. Documentación de los datos
- 8. Depósito
- 9. Licencias
- 10. Citas

## 1. USA FORMATOS ABIERTOS

- Formatos estándar y abiertos para asegurar el acceso a largo plazo.
- Guardar los archivos en formatos abiertos y propietarios (por ejemplo, data.xlsx y data.csv) para mejorar la usabilidad.
- Mantener el mismo nombre de archivo para el mismo archivo en diferentes formatos (por ejemplo, data.doc y data.txt).
- Comprimir los datos para facilitar el intercambio y la descarga de archivos grandes.
- Incluir la documentación de los datos, las transformaciones y las directrices del software para acceder al formato propietario.

## 2. CUIDA EL NOMBRAMIENTO DE LOS ARCHIVOS

- **La nomenclatura ha de ser significativa, coherente y descriptiva, de forma que facilite la gestión y**
- **localización de archivos:**
- Nombrar los archivos con nombres cortos y relevantes.
- No utilizar caracteres especiales:
- ~ ¡ ! @ # \$ % ^ & \* ( ) ` ; < > ¿ ? , [ ] { } ' " |
- Usar el guion bajo, mejor que el espacio en blanco.
- Identificar la actividad o proyecto en el nombre del archivo.
- Ser consistente con la nomenclatura que se escoge, mayúsculas, minúsculas, forma de las fechas, AAAA-MM-DD o AAAA-MM (ISO 8601 aplicada a la gestión de datos de
- investigación).
- Conservar las extensiones de 3 letras específicas a códigos de la aplicación (p. ej., .doc, .xls, .mov, .tif).

### 3. ORGANIZA LOS ARCHIVOS DEL PROYECTO

- El nombre, la estructura de las carpetas y el control de versiones de los ficheros deben facilitar la búsqueda, localización y comprensión de los datos.
- Pensar la mejor jerarquía para los archivos: profunda o superficial.
- Organización sistemática de las carpetas y ficheros.
- Restringir el nivel de carpetas a tres o cuatro.
- Separar los trabajos finalizados de los que los que están en curso.



## 4. CONTROL DE VERSIONES

- Si hay varias versiones nombrar por número (por ejemplo v01, v02, etc.).
- Para la versión final se puede nombrar como FINAL.
- Decidir cuántas versiones de un archivo y qué versiones se conservarán y durante cuánto tiempo.
- Registrar los cambios realizados en un archivo cuando se crea una nueva versión.

## 5. ALMACENAMIENTO

- Qué sistemas o herramientas se utilizarán para hacer que los datos sean accesibles, qué personas accederán, quien controlará el acceso a los datos y cómo se controlará que el acceso sea seguro.

## 6. SEGURIDAD

- Utilizar contraseñas robustas y cambiarlas cada cierto tiempo para resistir ataques informáticos y encriptación de datos altamente sensibles.
- No enviar datos por correo electrónico y utilizar sistemas de encriptado

## 7. DOCUMENTACIÓN DE LOS DATOS

- Ayuda a interpretar los datos permitiendo usarlos de forma rápida y ágil.
- Creación del fichero Readme.txt para adjuntar a los datasets.
- Fichero en que se describe la información necesaria para que los datos sean comprensibles y reutilizables.
- Se debe de publicar en inglés.
- Para depositar el conjunto de datos que se generan o recolectan durante la investigación es necesario además de los datos propiamente dichos, adjuntar un archivo Readme.txt en el repositorio.

## 8. DEPÓSITO DE LOS DATOS: REPOSITORIO

Es necesario depositar los datos en un repositorio temático o institucional:

- Que cumpla con los criterios FAIR.
- Que permita almacenar datos abiertos, con embargo, restringidos o cerrados,
- Que sirva para materiales diversos de investigación (distintos tipos, formatos...
- Que proporcione visibilidad, para facilitar la citación y aumentar el impacto.
- Que se integre con otras infraestructuras.
- Que sea transparente y fiable.


## 8. DEPÓSITO DE LOS DATOS: REPOSITORIO

- Identificador persistente (DOI/handle).
- Resumen del contenido (puede ser en varios idiomas).
- Palabras clave (pueden estar en más de un idioma).
- Licencia de reutilización (Creative Commons).
- Indicar el código del proyecto (en caso de tener una ayuda para la investigación).
- Indicar la obra relacionada (artículo, tesis, libro...).
- 2 Gb gratuitos para cada dataset (si necesitas más espacio contacta con la Biblioteca).


## 9. UTILIZA LICENCIAS DE USO

- Las licencias sirven para especificar restricciones y permisos para que otras personas puedan hacer uso de un trabajo.
- Las licencias dependerán del tipo de datos, de si están sujetos a legislación, así como de los usos permitidos o restricciones aplicables.
- Aunque existen diferentes [tipos de licencias](#), las más conocidas y frecuentes son las Creative Commons.
- Para saber más sobre Creative Commons [aquí](#)

# 10. CITA DE LOS DATOS



## Cita tus datos de investigación



### Por qué es importante citar los datos:

- Los conjuntos de datos también son resultados de investigación como los artículos, monografías, etc.
- Facilita la identificación y el acceso a los datos y de esta forma su localización, validación y reutilización.
- Permite reconocer la autoría de sus creadores.
- Facilita la métrica e impacto de los datos.
- Favorece la transparencia de la investigación científica.

### Buenas prácticas para citar datos:

- Se debe facilitar la identificación, localización y el acceso a los datos mediante un identificador único y persistente (DOI, Handle, etc.).
- Cada conjunto y subconjunto de datos (dataset) debe citarse de forma independiente.
- Las citas de los datos utilizados han de aparecer en la sección de referencias bibliográficas de la publicación resultante.
- Se recomienda incluir un identificador único de autor (ORCID, etc.).

### Elaboración de la cita

- Existen elementos mínimos obligatorios (O) y otros recomendados (R) que se combinan para elaborar la cita en cualquier estilo estándar (APA, MLA, Chicago, etc.) o los propuestos por los principales repositorios de datos (Dataverse, Dryad, etc.).

Autor(es) (O)

Identificador autor (R)

Fecha (O)

Título (O)

Identificador único persistente (O)

Tipo de recurso (O)

Versión y/o Edición (O)

Repositorio de datos (R)

Publicación (R)

Productor (R)

Ámbito geográfico (R)


Ámbito temporal (R)

### Ejemplo de cita estilo APA


Autoría
Título

Remesar Betlloch, X., Antelo, A., Llivina, C., Albà, E., Berdié, L., Agnelli, S.,... Alemany, M. (2015). *Influence of a hyperlipidic diet on the composition of the non-membrane lipid 6 pool of red blood cells of male and female rats*. [Dataset]. Versión de 22 de junio de 2015. Dipòsit digital de documents de la UAB. <http://hdl.handle.net/2445/66010>.


Fecha
Repositorio
Tipo de recurso
Identificador único y persistente
Versión



**Enlaza los datos con los documentos resultado de investigación y viceversa, y crea las referencias bibliográficas de los mismos.**



El personal de tu Biblioteca te puede asesorar





## EL PLAN DE GESTIÓN DE DATOS

- Un Plan de Gestión de Datos - (Data Management Plan (DMP))- es un documento que describe como se gestionarán los datos generados y procesados durante la investigación y al finalizar el proyecto de investigación.
- Es obligatorio para los proyectos financiados con fondos públicos.
- En los proyectos subvencionados hay entregar un plan a mitad del proyecto y un plan definitivo al final.



# 10 pasos para elaborar un Plan de Gestión de Datos

Un **Plan de Gestión de Datos** (PGD) o Data Management Plan (DMP) es un **documento formal, que debe presentarse al inicio de la investigación, en el que se describe qué**

**vas a hacer con tus datos durante y después de finalizar tu investigación** y que puede modificarse si se producen cambios en el proceso de la misma.

## ¿Por qué?

Es una **buena práctica**, es un **elemento clave de Open Science** y es **obligatorio** en los nuevos proyectos H2020.

## Herramientas gratuitas para elaborar un PGD



PGDonline  
(Consortio Madroño)  
<http://dmp.consortiomadrono.es/>



DMPonline (Digital Curation  
Centre, UK)  
<https://dmponline.dcc.ac.uk/>



Revisa los **requerimientos** de la entidad financiadora (H2020).



**Identifica los datos:** tipología, procedencia, volumen, formatos y ficheros.



**Define cómo se organizarán y gestionarán los datos:** nombre de los ficheros, control de versiones, software necesario...



**Explica cómo se documentarán los datos:** identifica la información a procesar, consulta si hay estándares o esquemas de metadatos, identifica herramientas que permitan gestionarlos.



Describe los procesos que aseguran una **buena calidad de los datos**.



**Prepara una estrategia de almacenamiento** (durante el proceso) y de preservación de datos (repositorio).



**Define las políticas de datos del proyecto:** cuestiones sobre propiedad intelectual y cómo se tratarán los datos sensibles y personales.



**Describe cómo se difundirán los datos:** dónde, cuáles, cuándo se van a difundir. Si publicarás los datos en un repositorio, como información suplementaria del artículo o como un "data paper".



**Asigna roles y responsabilidades** para las personas y organizaciones participantes en el proyecto.



**Prepara un presupuesto realista:** la gestión de datos cuesta tiempo y dinero en términos de software, hardware, servicios y personal.



crue

Universidades  
Españolas

Red de Bibliotecas  
REBIUN

Adaptado de Michener, WK. Ten simple rules for creating a good data management plan. PLOS Comp Biol. 2015; 11(10)



El personal de tu  
Biblioteca  
te puede asesorar



# EL PLAN DE GESTIÓN DE DATOS:

Plantillas de PGD:

- [Plantilla Europa](#) (en español)
- Horizon 2020 Template for the Data Management Plan

## EL PLAN DE GESTIÓN DE DATOS:

- **Herramientas para elaborar un PGD**
- Dmp online del Digital Curation Centre del Reino Unido
- [PaGoDa](#) traducción al español
- [Argos](#) de la Comisión Europea  
[Guía de uso de Argos](#)

Ejemplos de Plan Gestión Datos: <https://www.dcc.ac.uk/resources/data-management-plans/guidance-examples>

## GUÍAS PARA HACER UN PLAN DE GESTIÓN DE DATOS

- Directrices de la Comisión Europea dirigidas a los solicitantes y beneficiarios de proyectos de investigación:
- [Directrices para la Gestión de Datos en H2020](#). Traducción al español consorcio Madroño. Versión 1.0. 11 de diciembre de 2013.
- [Guidelines on Fair Data Management in Horizon 2020](#). Versión 3.0. 26 de julio 2016.



MÁS INFORMACIÓN:

[biblioteca.rediumh@umh.es](mailto:biblioteca.rediumh@umh.es)



**RediUMH**  
Universidad Miguel Hernández