

## Cómo gestionar los datos de investigación

Entendemos por datos de investigación a todo el material que ha sido registrado y sirve para demostrar los resultados alcanzados. Estos datos pueden ser numéricos, descriptivos o visuales, en estado bruto o analizado, experimentales u observacionales.

### A) Medios para difundir los datos de investigación:

1. **Como material unido al artículo:** las editoriales indican cómo se ha de adjuntar ese material suplementario. Según la tipología de los datos los formatos cambiarán.  
Así pues, los protocolos, secuencias, estructuras, se albergarán en un repositorio de datos según la materia, como Genbank; y el artículo resultante debe contener el nº de identificación asignado en dicho repositorio.
2. **Como un artículo de datos en una revista de datos:** éstas aparecen recogidas en [Data Journals](#) son de acceso abierto con el fin de que puedan ser reutilizados y citados.
3. **Buscar un repositorio de datos para depositarlos:** éstos pueden ser institucionales o temáticos, puede ser útil consultar en [re3data](#); también la revista [Scientific Data](#) tiene una guía de repositorios



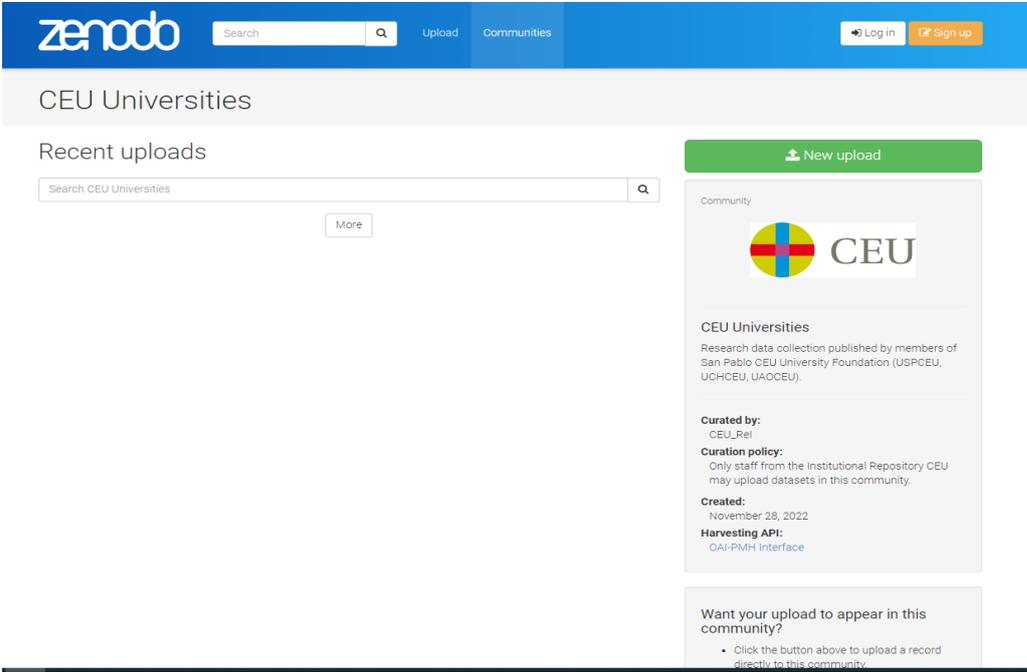
Fuente: [re3data](#) permite buscar el repositorio de datos por materias.

**Zenodo** es un repositorio multidisciplinar creado por OpenAIRE y el CERN con el apoyo de la Comisión Europea:

- Admite datos de investigación multidisciplinarios de cualquier investigador, comunidad científica o institución de I+D.
- Acepta todo tipo de datos y formatos
- Permite un gran almacenamiento (1 GB)
- Fomenta el depósito de datos en abierto (aunque ofrece todo tipo de licencias)
- Asigna un identificador digital (un DOI) que demuestra su publicación

Hemos creado una comunidad en **Zenodo** para albergar los datos de investigación de nuestras universidades: **CEU Universities**

[https://zenodo.org/communities/ceu\\_rei/?page=1&size=20](https://zenodo.org/communities/ceu_rei/?page=1&size=20)



The screenshot shows the Zenodo interface for the CEU Universities community. At the top, there is a blue navigation bar with the Zenodo logo, a search bar, and buttons for 'Upload', 'Communities', 'Log in', and 'Sign up'. Below the navigation bar, the page title 'CEU Universities' is displayed. A 'Recent uploads' section contains a search bar for 'Search CEU Universities' and a 'More' button. To the right, there is a 'New upload' button and a community profile card for 'CEU Universities'. The profile card includes the CEU logo, a description of the community as a research data collection, and details about its curation policy, creation date (November 28, 2022), and harvesting API (OAI-PMH Interface). At the bottom of the profile card, there is a prompt: 'Want your upload to appear in this community?' with a bullet point: 'Click the button above to upload a record directly to this community.'

Se pueden hacer envíos a esta Comunidad y se irán formando Colecciones; siempre que se suban los dataset el administrador de CEU Rel los revisará para que se adecuen a los requerimientos de pertenencia. Por otro lado, se invita a todos los que tengan ya creada su Colección en Zenodo, puedan pasarla a CEU Universities.

Existen buscadores de datasets:

- **Google Dataset Search:** buscador de conjunto de datos para investigadores de Google
- **DataCiteSearch:** buscador generalista de datasets abiertos con DOI.

**B)** Se aconsejan una serie de **buenas prácticas** para que todo el proceso sea eficaz:

**Tratamiento:** los ficheros de datos han de estar ordenados y con una estructura cohesionada. Se aconseja:

- Nombrar los ficheros con el mismo criterio: que sean significativos y breves (sin caracteres especiales ni tildes). Indicar la versión al nombrar el fichero.
- Hacer copias de seguridad y procurar hacer una versión definitiva.

**Datos públicos y privados:** seleccionar aquellos datos sensibles que no pueden aparecer en abierto relativos a la propiedad intelectual [Ley 23/2006](#), a la [protección de datos personales](#) y a la propiedad industrial.

Los datos obtenidos a través de entrevistas, cuestionarios, etc. debe tenerse en cuenta: solicitar el consentimiento de las personas implicadas mediante un formulario y por otra parte aplicar un proceso de anonimización en aquellos ficheros que contienen datos sensibles o confidenciales.

**OpenAire** recomienda la herramienta [Amnesia](#) que transforma los datos relacionales y transaccionales en conjuntos de datos anónimos donde se mantienen las garantías formales de privacidad

**Formatos de archivos:** abiertos y estándar para que puedan ser leídos en el futuro.

Existen unas [recomendaciones internacionales](#) :

- **Contenedores:** TAR, GZIP, ZIP
- **Bases de datos:** XML, CSV, JSON
- **Geoespacial:** SHP, DBF, GeoTIFF, NetCDF
- **Video:** MPEG, AVI, MXF, MKV
- **Sonido:** WAVE, AIFF, MP3, MXF, FLAC
- **Estadísticas:** DTA, POR, SAS, SAV
- **Imágenes:** TIFF, JPEG 2000, PDF, DNG, GIF, BMP, SVG
- **Datos tabulares:** CSV, TXT
- **Texto:** XML, PDF / A, HTML, JSON, TXT, RTF
- **Archivo web:** WARC

**Licencias de uso y reutilización:** debe aparecer adjunto a los datos “readme file” la licencia que hayamos elegido, las más comunes son: [Creative Commons](#) y [Open Data Commons](#)

Resulta necesario disponer de un **Identificador, Único y Persistente** para localizar y preservar los datos de la investigación en el Repositorio donde se ha elegido almacenarlos. Existen varios: **URI** (identificador persistente de recursos), **DOI** (identificador único y permanente para las publicaciones electrónicas), etc.

**Citar** de forma independiente cada conjunto de datos, bien sea un artículo de datos publicado en una revista de datos o un conjunto de datos archivados en

un repositorio. Se recomienda incluir un identificador persistente de autor (Ej. ORCID).

**C) Plan de Gestión de Datos** (PGD, en inglés DMP Data Management Plan) es el informe que engloba todas las fases: creación o recopilación, procesamiento y análisis, preservación, publicación y reutilización.

Recoge desde el inicio cómo se van a gestionar los datos, señala los datos que se compartirán, cómo se podrán utilizar y cómo se preservarán, antes de empezar a crearlos o a recogerlos. No es un documento definitivo, sino que se va modificando durante el proyecto.

Horizonte 2020 ha dado paso al nuevo programa [Horizonte Europa 2021-2027](#) que es el Programa Marco de Investigación e Innovación de la Unión Europea.

Para diseñar el **PGD** nos puede servir la plantilla diseñada por [Digital Curation Center](#) para cumplimentar los campos que se solicitan:

- **Información administrativa:** identificador del centro, financiador, número de identificación de la financiación, nombre del proyecto, descripción, IP, identificador [Orcid](#) del IP, fecha de la primera versión, etc;
- **Descripción de los datos:** tipología, volumen, formatos, estructura de los archivos y gestión de las versiones;
- **Documentación y metadatos** que acompañarán a los datos;
- **Requisitos legales y éticos:** consentimiento de las personas implicadas, protección de la identidad, tratamiento de los datos sensibles y gestión de derechos de propiedad intelectual;
- **Almacenamiento y copias de seguridad:** control de acceso a los colaboradores y gestión de las copias de seguridad;
- **Selección y preservación:** qué datos y cómo serán conservados, dónde se preservarán y restricciones en el intercambio de datos;
- **Compartir los datos:** bajo qué condiciones se compartirán, cómo encontrarán los usuarios los datos, cuándo estarán disponibles, obtendrán un identificador persistente, etc.;
- **Responsabilidades y recursos:** asignación de responsabilidades, costes asociados al acceso abierto de los datos, ya que son elegibles para su reembolso durante el periodo del proyecto.

Existen varias **herramientas** para orientarnos en el desarrollo del **PGD**:

- [PAGODA](#) que es el repositorio de datos exclusivo para el Consorcio Madroño
- [ARGOS](#) promovido por la Comisión Europea, permite crear DMP (Plan de Gestión de Datos, en inglés) que pueden intercambiarse libremente entre infraestructuras.

Es necesario seguir los principios [FAIR](#) (localizables, accesibles, interoperables y reutilizables) para poder compartir y reutilizar los datos de investigación de forma legal en investigaciones futuras.

1. **Localizables**: describir los metadatos lo más completos posibles junto con el identificador permanente asignado, otro de los metadatos es el DOI o la URI que lo otorga la revista de datos y el repositorio de datos respectivamente.
2. **Accesibles**: utilizar los protocolos estándar que facilite su comunicación abierta y gratuita.
3. **Interoperables**: los metadatos deben utilizar formatos, lenguajes y vocabularios admitidos por la comunidad científica.
4. **Reutilizable**: deben utilizar una licencia abierta y legible por ordenador, aportar información sobre la procedencia.

El servicio de **CEU Repositorio Institucional de la Universidad San Pablo-CEU** le atenderá a través del correo [repositorioinstitucional@ceu.es](mailto:repositorioinstitucional@ceu.es)

## 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 (Consorcio Madroño) <http://dmp.consorcioadroño.es/>
- DMPonline (Digital Curation Centre, UK) <https://dmponline.dcc.ac.uk/>

- 01 **Revisa los requerimientos** de la entidad financiadora (H2020).
- 02 **Identifica los datos:** tipología, procedencia, volumen, formatos y ficheros.
- 03 **Define cómo se organizarán y gestionarán los datos:** nombre de los ficheros, control de versiones, software necesario...
- 04 **Explica cómo se documentarán los datos:** identifica la información a procesar, consulta si hay estándares o esquemas de metadatos, identifica herramientas que permitan gestionarlos.
- 05 **Describe los procesos que aseguran una buena calidad de los datos.**
- 06 **Prepara una estrategia de almacenamiento** (durante el proceso) y de preservación de datos (repositorio).
- 07 **Define las políticas de datos del proyecto:** cuestiones sobre propiedad intelectual y cómo se tratarán los datos sensibles y personales.
- 08 **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".
- 09 **Asigna roles y responsabilidades** para las personas y organizaciones participantes en el proyecto.
- 10 **Prepara un presupuesto realista:** la gestión de datos cuesta tiempo y dinero en términos de software, hardware, servicios y personal.

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

El personal de tu Biblioteca te puede asesorar

crue Universidades Españolas Red de Bibliotecas REBIUN

CC BY