



CEU | *Universidad  
San Pablo*

# Guía de apoyo para la creación de un Plan de Gestión de Datos (PGD)

Marzo de 2025

**Biblioteca | Unidad de Apoyo a la Investigación**



## Índice

0. Objeto de esta Guía | p. 3
1. Introducción | p. 3
2. ¿Qué son los datos de investigación? | p. 4
3. Ciencia abierta y difusión de datos de investigación | p. 5
4. Recomendaciones | p. 7
5. Procedimiento | p. 11
6. Estructura y aspectos clave en la elaboración de un Plan de Gestión de Datos | p. 14
7. Plantillas y guías recomendadas para la creación de un Plan de Gestión de Datos | p. 16
8. Herramientas en línea para la creación de un Plan de Gestión de Datos | p. 16
9. Evaluación de un Plan de Gestión de Datos | p. 17

## 0. Objeto de esta Guía

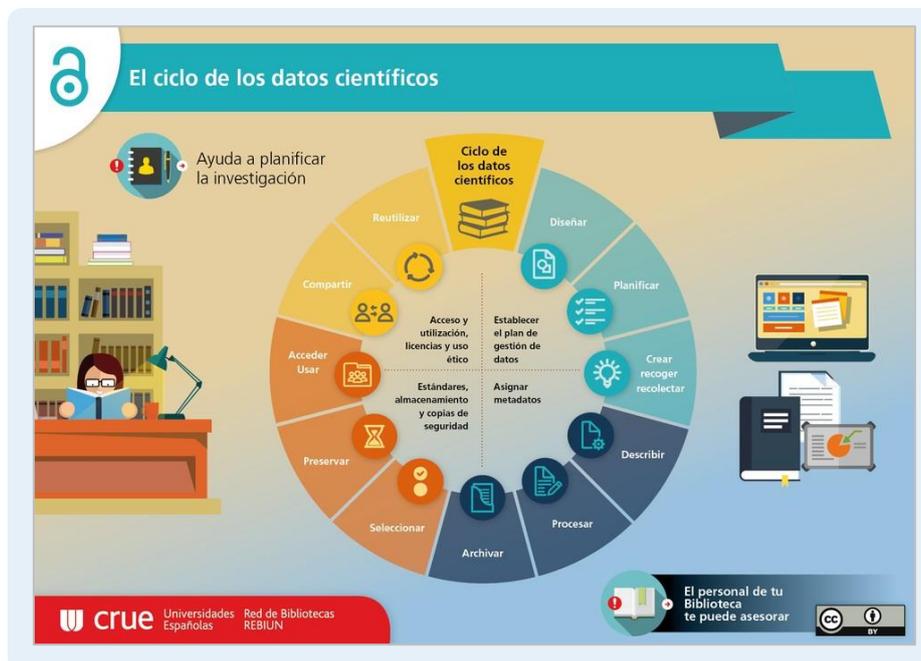
El presente documento tiene como finalidad proporcionar una guía de soporte destinada al Personal Docente Investigador (PDI) de la comunidad académica, científica e investigadora de la Universidad San Pablo-CEU para la elaboración de un Plan de Gestión de Datos (PGD) asociado al proyecto de investigación llevado a cabo o que se vaya a iniciar, de acuerdo con la normativa vigente y obligaciones legales, y con los requisitos y estándares de organismos nacionales e internacionales en materia de ciencia, innovación, transferencia del conocimiento y acreditaciones.

Cabe señalar que esta guía se complementa con el procedimiento de Repositorio Institucional – [CEU Rel](#)

## 1. Introducción

De acuerdo con los requisitos de programas de financiación pública como Horizonte 2020, Horizonte Europa, así como de los Proyectos Competitivos, y con la normativas nacionales e internacionales, los investigadores participantes en proyectos financiados deben elaborar un PGD que garantice el acceso abierto a los datos de investigación, salvo excepciones justificadas.

Un PGD debe contener información precisa y concreta sobre cómo se van a gestionar los datos de investigación desde el inicio del proyecto: recogida de datos, almacenamiento, protección, seguridad, recuperación, accesibilidad, reutilización, aspectos éticos y legales, recursos y responsables... hasta su fin: archivo/depósito a lo largo del ciclo de vida del proyecto de investigación.





La Unidad de Apoyo a la Investigación de la Biblioteca de la Universidad San Pablo-CEU ofrece formación, asesoramiento y consultoría destinados al PDI para la creación de este plan.

## 2. ¿Qué son los datos de investigación?

Los datos de investigación son “todo aquel material que ha sido generado, recopilado, observado o registrado durante el ciclo de vida de un proyecto de investigación y que se utiliza como evidencia de un proceso de investigación, está reconocido por la comunidad científica y sirve para validar los resultados de la investigación y garantizar su reproducibilidad”.

Información tomada de la [Estrategia Nacional de Ciencia Abierta \(ENCA\) 2023-2027](#). Ministerio de Ciencia e Innovación. Edita: Secretaría General Técnica del Ministerio de Ciencia e Innovación, 2023. CC BY-NC-SA.

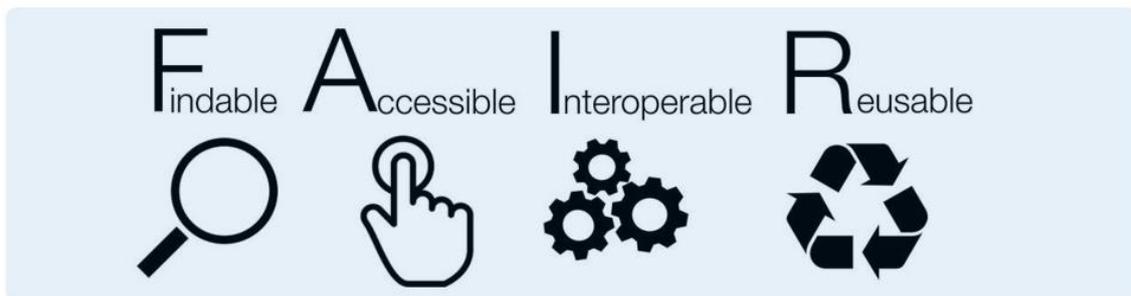
A continuación, se incluyen distintas tipologías y ejemplos de datos de investigación (dataset) y formatos más recomendados:

Tipo de datos	Formatos recomendados	Formatos aceptables
<b>Datos tabulares complejos</b>	<ul style="list-style-type: none"> <li>• SPSS portable format (.por), R (.r)</li> <li>• Texto estructurado o xml etiquetado, e.j. DDI XML, json...</li> </ul>	<ul style="list-style-type: none"> <li>• Formatos propietarios de paquetes estadísticos: SPSS (.sav), Stata (.dta). Ficheros de bases de datos en acceso abierto exportadas, con las instrucciones para recuperarlos (como bases de datos mysql, postgresql o sqlite)</li> </ul>
<b>Datos tabulares simples</b>	<ul style="list-style-type: none"> <li>• comma-separated values (.csv)</li> <li>• tab-delimited file (.tab, .tsv)</li> </ul>	<ul style="list-style-type: none"> <li>• Formatos muy conocidos y extendidos: MS Excel (.xls, .xlsx), OpenDocument Spreadsheet (.ods)</li> </ul>
<b>Imágenes</b>	<ul style="list-style-type: none"> <li>• TIFF 6.0 uncompressed (.tif, .tiff)</li> </ul>	<ul style="list-style-type: none"> <li>• PNG (.png)</li> <li>• SVG (.svg)</li> <li>• JPEG (.jpeg, .jpg, .jp2)</li> <li>• GIF (.gif) • TIFF other versions (.tif, .tiff)</li> <li>• RAW image format (.raw)</li> <li>• Photoshop files (.psd)</li> <li>• BMP (.bmp)</li> </ul>
<b>Audio</b>	<ul style="list-style-type: none"> <li>• Free Lossless Audio Codec (FLAC) (.flac)</li> </ul>	<ul style="list-style-type: none"> <li>• MPEG-1 Audio Layer 3 (.mp3)</li> <li>• Audio Interchange File Format (.aif)</li> <li>• Waveform Audio Format (.wav)</li> </ul>
<b>Video</b>	<ul style="list-style-type: none"> <li>• OGG video (.ogv, .ogg)</li> <li>• WEBM (.webm)</li> <li>• Matroska (.mkv) con codecs de vídeo con licencia libre</li> </ul>	<ul style="list-style-type: none"> <li>• AVCHD video (.avchd, .mts, .m2ts)</li> <li>• MPEG-4 (.mp4)</li> <li>• motion JPEG 2000 (.mj2)</li> </ul>

Fuente: Consorcio Madroño

### 3. Ciencia abierta y difusión de datos de investigación

Los datos de investigación deben depositarse en abierto siguiendo los principios FAIR (acrónimo de Findable, Accessible, Interoperable y Reusable) y la política de ciencia abierta que describen cómo deben organizarse los resultados de la investigación para que sean encontrables, accesibles, interoperables y reutilizables:



Publicar los datos siguiendo los principios FAIR permite cumplir con el lema de ciencia abierta «as open as possible, as closed as necessary».

- > **Findable (“Econtrable”) a través de:**
  - > Identificador global único y persistente (handle/DOI).
  - > Metadatos enriquecidos.
  - > Metadatos indexados en un sistema que disponga de motor de búsqueda.
  
- > **Accesible (“Accesible”) siguiendo o cumpliendo con:**
  - > Protocolo de interoperabilidad.
  - > Protocolo abierto, gratuito y universal.
  - > Accesibilidad de los metadatos (aun no estando disponibles los datos).
  
- > **Interoperable (“Interoperable”) utilizando:**
  - > Lenguaje formal, accesible y compartido.
  - > Vocabularios controlados.
  - > Referencias óptimas a otros metadatos.
  
- > **Reusable (“Reutilizable”) mediante:**
  - > Metadatos descritos con pluralidad de atributos precisos: licencia de uso, procedencia detallada y cumplimiento de normativa.



Adicionalmente, se recomienda que los datos de investigación sean inteligibles, legibles y comprensibles tanto para humanos como para máquinas (en un formato que una computadora/ordenador pueda procesar fácilmente).

Según el Plan Estatal de Investigación Científica, Técnica y de Innovación 2021-2023, “los datos de investigación se deberán depositar en repositorios institucionales, nacionales y/o internacionales antes de que transcurran dos años desde la finalización del proyecto, con el fin de impulsar el acceso a datos de investigación de las ayudas financiadas”.

Una vez se ha obtenido la aprobación de la financiación del proyecto y este ha comenzado, se debe enviar la primera versión del PGD dentro de los primeros 6 meses del proyecto.

La mayoría de las convocatorias y proyectos pertenecientes al Plan de Investigación Científica, Técnica y de Innovación de 2024-2027 indican que:



Quando se opte por diseminar los resultados de investigación en publicaciones científicas, se deberá depositar una copia de la versión final aceptada para publicación y los datos asociados a las mismas en repositorios institucionales o temáticos de acceso abierto, de forma simultánea a la fecha de publicación.

A tales efectos, los autores de trabajos científicos que hayan sido aceptados para su publicación en publicaciones seriadas o periódicas podrán optar por publicar en revistas de acceso abierto o autoarchivar en repositorios institucionales o temáticos de acceso abierto, recogidos en la plataforma RECOLECTA, de la Fundación Española de Ciencia y Tecnología (FECYT), o en otros repositorios promovidos por las propias instituciones.

Los beneficiarios de estas ayudas se asegurarán de que conservan los derechos de propiedad intelectual necesarios para dar cumplimiento a los requisitos de acceso abierto.

Los datos generados por la investigación se deberán depositar en repositorios institucionales, nacionales y/o internacionales tan pronto como sea posible, y siempre antes de que transcurran dos años desde la finalización de la actuación, con el fin de facilitar e impulsar el libre acceso y gestión de los datos de investigación de las ayudas financiadas, siguiendo los principios internacionales FAIR (del inglés – *Findable, Accessible, Interoperable, and Reusable*).

No obstante, los requisitos, recomendaciones y tiempos concretos dependen de la convocatoria a la que se aplique o se haya aplicado, por lo que se recomienda leer y analizar al detalle y con detenimiento el texto que corresponda a cada proyecto de investigación.



Respondiendo a las necesidades detectadas, la institución ha desarrollado un repositorio institucional ([CEU Rel](#))  destinado a reunir, conservar y difundir la producción científica y académica generada por los investigadores de la universidad. Esta herramienta promueve el acceso abierto a los resultados de las investigaciones, aumentando la visibilidad de la producción y el impacto de los trabajos académicos. [CEU Rel](#) está sustentado por el software de código abierto DSpace 8, que no solo proporciona una mayor flexibilidad gráfica, sino que permite la incorporación de los identificadores de autor para cada uno de los autores con filiación CEU: ORCID, Scopus ID, perfil de Dialnet y de Google Académico. Ello permite al autor controlar la información disponible en CEU Rel de cada uno de sus trabajos y conocer el impacto y difusión de su producción científica.

Además de garantizar el cumplimiento de la normativa vigente, la Unidad de Apoyo a la Investigación ofrece asesoría para que los investigadores se ajusten a la legislación y las normativas relacionadas con la publicación y difusión de sus trabajos.

Complementariamente, el servicio de Biblioteca ha creado una comunidad en Zenodo (repositorio multidisciplinar de acceso abierto gestionado por CERN, desarrollado por OpenAIRE y financiado por la Unión Europea) denominada “[CEU Universities](#)” , para el depósito de datos de investigación generados por sus investigadores, poniendo así a su disposición, un espacio propio dentro de este repositorio. De esta forma los investigadores podrán depositar y difundir sus datos de investigación de forma abierta o cumplir, cuando se necesite, con los requerimientos de editores o financiadores.

Opcionalmente, cabe señalar que a decisión y bajo el criterio del investigador responsable del proyecto, los resultados de la producción científica y los datos de investigación asociados podrán depositarse en un repositorio temático de su elección.

## 4. Recomendaciones

### A) TRATAMIENTO: ¿CÓMO HAN DE TRATARSE LOS DATOS DE INVESTIGACIÓN?

- > Deben utilizarse criterios **homogéneos** para denominar ficheros.
- > Emplear nombres **significativos y breves**.
- > Indicar el **número de versión** si hay varias revisiones (control de versiones).
- > Realizar **copias de seguridad**.



## B) SELECCIÓN: ¿QUÉ DATOS SELECCIONO Y CÓMO?

- > Se deben seleccionar **aquellos datos que vayan a ser públicos y accesibles** (propiedad intelectual y protección de datos).
- > Se garantizará la firma de un **consentimiento informado / acuerdo de participación** si hay datos personales o confidenciales.

## C) ANONIMIZACIÓN Y SEUDONIMIZACIÓN\*

- > Se deben **eliminar identificadores directos**: apellidos, nombres, DNI...
- > **Reducir precisión de información**: reemplazar edad o fecha de nacimiento por rangos de edad.
- > **Anonimizar datos geo-referenciados**: reemplazando las coordenadas exactas por regiones o zonas genéricas; incluyendo símbolos y caracteres especiales en lugar de códigos postales, etc.
- > Utilizar **pseudónimos**.

\* Véase más información sobre los conceptos de anonimización y seudonimización en la web de la [Agencia Española de Protección de Datos \(AEPD\)](#). 

Existen diversas herramientas y software que facilitan la labor de eliminar los identificadores personales o que dificultan la atribución a una persona física de los datos que se hayan recabado y que sean de naturaleza sensible, satisfaciendo los requisitos de privacidad. Por ejemplo, "[Amnesia Anonymization Tool](#)"  un software gratuito y de acceso libre, con versión de prueba.

No obstante, se recomienda acudir al [Comité de Ética de la Investigación de la Universidad San Pablo-CEU](#) , cumpliendo con el procedimiento establecido y la normativa más vigente en asuntos éticos, de integridad y de dignidad del ser humano, del buen trato a los animales, y de experimentación.

## D) FORMATOS ACEPTADOS

- > Datos textuales\*: txt, xml, rtf. (Se aceptan HTML, doc, docx).
- > Datos tabulares con metadatos ampliados: SPSS por, SPSS, Stata, SAS, DDI xml. (Se permiten sav, dta, mdb, accdb).
- > Datos tabulares con metadatos mínimos: csv, tab. (Se permiten txt, xls,xlsx, mdb, accdb, dbf, ods).

- > Datos geoespaciales: shp, shx, dbf, prj, sbx, sbn, tif, tfw, dwg, gml. (Se aceptan mdb, mif, kml, ai, dxf, svg).
- > Imagen digital: tif. (Se aceptan jpeg, jpg, jp2, gif, tiff, psd, bmp, png, pdf, pdf/a).
- > Audio: flac. (Se aceptan mp3, aif, wav).
- > Audio digital: mp4, ogv, ogg, mj2. (Se aceptan avchd).

\* Para la documentación y el plan de gestión de datos: rtf, pdf, xhtml, htm, odt (Se aceptan txt, doc, docx, xls, xlsx, xml).

### E) LICENCIAS

- > Definir las [licencias de uso y reutilización de datos](#).
- > Licencias más comunes de uso: [Creative Commons](#) y [Open Data Commons](#).



## Licencias de Creative Commons

			¿Usarlas de forma comercial?	¿Modificarlas?
Atribución	BY			
Compartir bajo la Misma Licencia	BY-SA			Solo bajo una licencia similar
Sin Obras Derivadas	BY-ND			
No comercial	BY-NC			Solo con propósitos no comerciales
No comercial bajo la Misma Licencia	BY-NC-SA			Solo con propósitos no comerciales
No comercial Sin Obras Derivadas	BY-NC-ND			

En todas es necesario darle el crédito al autor

norfipc.com



## F) IDENTIFICADORES

- Asignar identificador único y persistente (DOI, handle...).

## G) FICHERO "README": archivo complementario en formato .txt titulado "readme file"

```
Readme Datos de Investigación (1): Bloc de notas
Archivo Edición Formato Ver Ayuda

Nombre del dataset:

Nombre o iniciales del responsable:

Número de versión:

Proyecto asociado al dataset:

|

Tipología de los datos albergados (breve descripción):
```

Se recomienda la creación y el depósito de un fichero "README" junto a los datos de investigación que se hayan publicado y sean accesibles. Dicho fichero debe resultar de utilidad para aquellos investigadores externos al proyecto que deseen reutilizar ese dataset, siendo este fichero "README" una guía o ayuda complementaria para la lectura o interpretación del dataset sin necesidad de haber participado en el proyecto de investigación, posibilitando así su reutilización.

Continuando con la pauta de accesibilidad de los datos, se recomienda igualmente que dicho fichero esté disponible tanto en lengua inglesa como en lengua española.

## H) REDACCIÓN ELABORADA Y CON DETALLE DEL PGD

- Redactar frases completas (evitando respuestas simples: Sí/NO), aportando ejemplos y detalles concretos, y evitando respuestas generales.



- > Seleccionar el tiempo verbal adecuado (H2020: pasado vs HEuropa: futuro) según la concepción de PGD (documento finalizado o vivo).

## 5. Procedimiento

Para aquellos investigadores que así lo soliciten, ofrecemos asesoramiento y herramientas que aseguren la correcta gestión de los datos generados a lo largo del proceso de investigación, así como soporte para la elaboración de un Plan General de Datos. Esto incluye el almacenamiento seguro, la preservación a largo plazo y el cumplimiento de las normativas nacionales y de la propia institución en cuanto a la gestión de datos.

Las solicitudes para el asesoramiento de un PGD deben ser enviadas mediante un [formulario](#) al correo electrónico: [repositorioinstitucional@ceu.es](mailto:repositorioinstitucional@ceu.es)



**CEU**  
Biblioteca

**UNIVERSIDAD SAN PABLO-CEU**  
Biblioteca  
Unidad de Apoyo y Soporte a la Investigación

Una vez cumplimentado el formulario, por favor, remítalo a: [repositorioinstitucional@ceu.es](mailto:repositorioinstitucional@ceu.es)

**SOLICITUD DE INFORMACIÓN PARA PGD**

- Fecha de solicitud:
- Código del proyecto:
- Responsable del PGD:
- Fecha inicio / concesión de la ayuda o financiación:
- Título del proyecto:

Nombre completo	Identificador ORCID
<input type="text"/>	<input type="text"/>

- ¿Tienes la publicación resultante del proyecto de investigación depositada en un repositorio de confianza (temático, institucional: CEU Rel,...)? Si es así, especifica en cuál.
- Tipos de datos o resultados de investigación:
  - A) Describir el tipo de datos que espera o se ha recabado en el desarrollo del proyecto de investigación. Ejemplo: experimentales, observacionales, imágenes, texto, numéricos, otro,...
  - 
  - B) Describir el formato del dataset. Ejemplo: txt, csv, xlsx, software, bdd, otros (indicar cuáles):
  - 
  - C) Señalar el tamaño estimado del dataset (por fichero o archivo):
  -



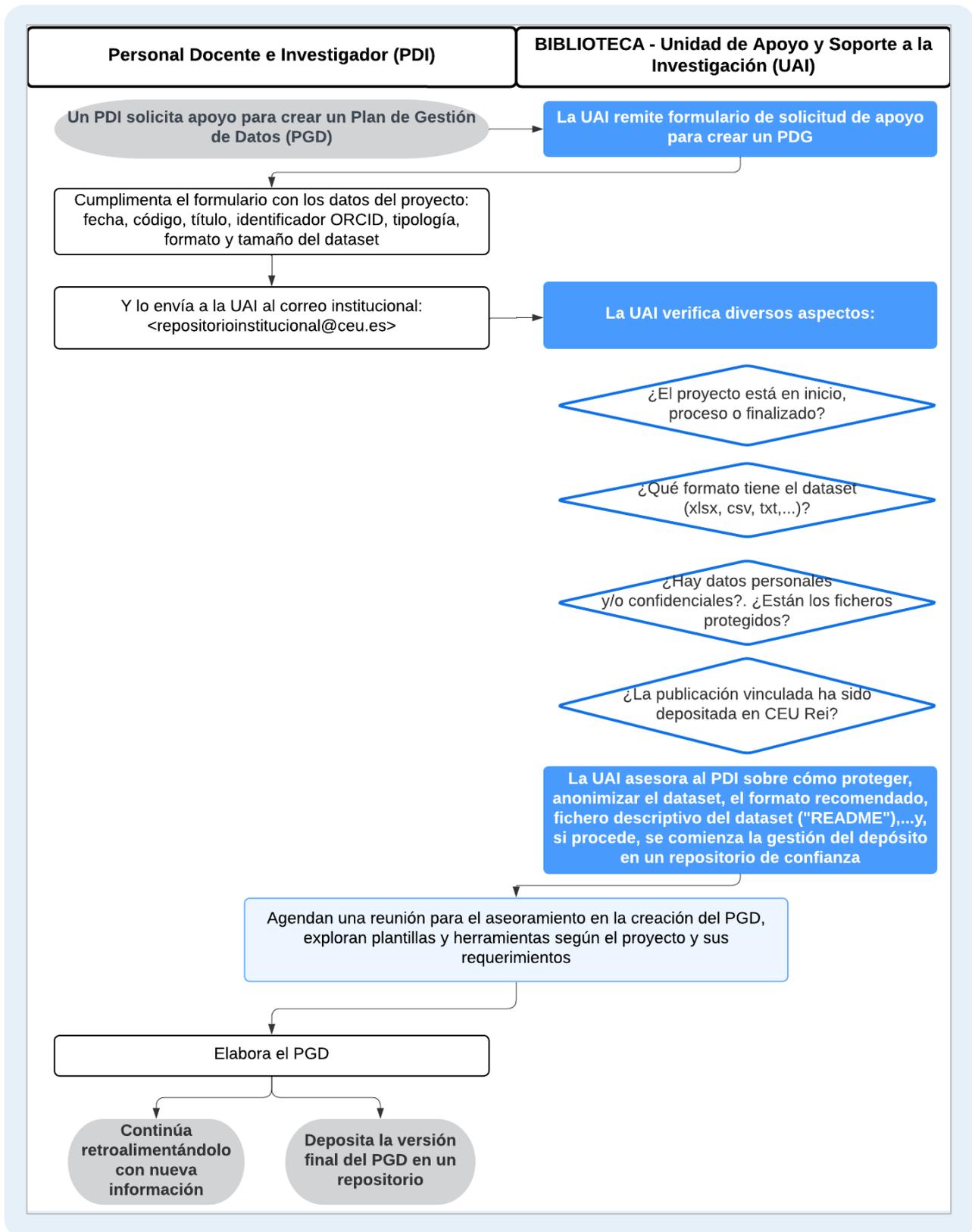
En dicho formulario el PDI interesado deberá incluir información sobre el código y el título del proyecto, fecha de inicio y fin de la concesión de la ayuda, identificador de ORCID del investigador responsable, tipo de dataset, formato y tamaño...

Desde la Unidad de Apoyo a la Investigación se proporcionarán herramientas y plantillas recomendadas, cumpliendo con los principios FAIR, para asegurar que los resultados de la investigación sean fácilmente localizables, accesibles, interoperables y reutilizables y crear el Plan de Gestión de Datos. Además, se ofrece orientación sobre cómo proteger y anonimizar datos de carácter personal o confidencial, cumpliendo con las normativas de protección y seguridad de datos. Así mismo, se asesorará al investigador sobre cómo citar los datos de investigación, garantizando su identificación, localización y reconocimiento de autoría.

Adicionalmente, la Unidad organiza y facilita diferentes sesiones de formación dirigidas a investigadores y personal académico, en diversas áreas como el uso de recursos bibliográficos, la gestión de datos de investigación y las mejores prácticas en la publicación científica.



> Flujograma del procedimiento de asesoramiento





## 6. Estructura y aspectos clave en la elaboración de un Plan de Gestión de Datos

### ESTRUCTURA

1. HACER UN RESUMEN DE LOS DATOS DE INVESTIGACIÓN
2. INDICAR RESPONSABILIDADES
3. ESPECIFICAR CUMPLIMIENTO DE DATOS FAIR
4. ASIGNAR RECURSOS
5. DESCRIBIR PROTOCOLOS DE SEGURIDAD Y PROTECCIÓN DE DATOS
6. INCLUIR ASPECTOS ÉTICOS Y LEGALES
7. OTROS

- 1) *Resumen de los datos de investigación*: Incluir el propósito de la recopilación de datos; describir la relación con los objetivos del proyecto de investigación; y especificar los tipos, el formato, el origen, el tamaño y la utilidad de los datos.
- 2) *Responsabilidades*: Indicar la/s persona/s responsable/s de la salvaguarda, la protección y la seguridad de los datos de investigación; así como los roles desarrollados por la/s misma/s.
- 3) *Cumplimiento de datos FAIR*: Describir cómo se procederá para que los datos de investigación cumplan con los principios FAIR:
  - > Datos encontrables: ¿Los datos serán recuperables mediante un identificador persistente y único (handle/DOI...)? ¿Qué metadatos, estándares o modelos de esquemas de metadatos se seguirán para la gestión de los datos de investigación? ¿Se utilizarán palabras clave para facilitar la recuperación de los datos y su reutilización?



- > Datos accesibles: ¿Se depositarán los datos de investigación en un repositorio de confianza? ¿Se podrá acceder a los datos a través de un protocolo de acceso libre y estandarizado? Si hay restricciones de acceso, ¿cómo se accederá a los datos de investigación (acceso restringido por embargo con una editorial, acceso a metadatos, período de tiempo de restricción...)?
  - > Datos interoperables: ¿Qué vocabularios, ontologías, estándares se utilizarán para asegurar que los datos de investigación puedan ser compartidos o para favorecer el intercambio de datos con la posibilidad de su reutilización? En el caso de que los datos no sean originales, indicar las referencias apropiadas de los metadatos asociados.
  - > Datos reutilizables: ¿El dataset favorece la reutilización de los datos mediante un fichero “README”, suponiendo este una ayuda complementaria para asegurar la interpretación y reutilización de los datos por parte de otros investigadores?
- 4) **Asignar recursos:** Estimar qué costes directos e indirectos conllevan estos protocolos y actuaciones para asegurar que los datos sean FAIR, y determinar el valor potencial de la preservación de los datos de investigación. Estos costes podrían incluir, por ejemplo: costes de almacenamiento, hardware, tiempo del personal, costes de preparación de datos para depósito y cargos de depósito. En caso de duda, existen herramientas en línea que posibilitan la asignación o estimación de los costes asociados a la planificación, gestión y mantenimiento de un PGD, por ejemplo:
- > [Data Management Costing Tool and Checklist](#) 
- 5) **Describir protocolos de seguridad y protección de datos:** Indicar si se utilizarán datos personales y/o información confidencial y si es necesaria la firma de un contrato con consentimiento informado de los participantes; cómo se protegerán esos datos: anonimización y seudonimización...
- 6) **Incluir aspectos éticos y legales:** Especificar si el proyecto será o ha sido evaluado previamente por el Comité de Ética y de Investigación, garantizando el respeto a la dignidad, integridad e identidad del ser humano y el buen trato a los animales en los campos de la experimentación biológica, genética y de la investigación científica (incluir comisiones *ad hoc*, si procede). Aspectos legales relacionados con los derechos de autor, propiedad intelectual (copyright), licencias de uso (Creative Commons...) y permisos o cumplimiento con la normativa nacional en materia de protección de datos (consentimiento informado si van a intervenir terceros y si se van a utilizar datos personales y/o confidenciales) han de contemplarse igualmente.



- 7) *Otros*: Si procede, referenciar el cumplimiento de requisitos adicionales de organismos o entidades financiadoras, departamentales, etc.

## 7. Plantillas y guías recomendadas para la creación de un Plan de Gestión de Datos

- > **Horizonte 2020** (Plantilla en inglés para proyectos concedidos en 2020) 
- > **Horizonte Europa** (Plantilla en inglés para proyectos concedidos entre 2021-2027) 
- > **Guía de la UCM** (Guía en español basada en plantilla Horizonte 2020) 
- > **Guía de la Universidad de Cantabria** (Guía en español basada en plantilla Horizonte 2020) 
- > **Plantilla de la Universidad de Huelva** (Guía en español según Plantilla Horizonte Europa con ejemplos de redacción para cada uno de los puntos y subapartados) 
- > **Plantilla de la Universidad de Cádiz** (Guía en español según Plantilla Horizonte Europa con ejemplos de redacción para cada uno de los puntos y subapartados) 

## 8. Herramientas en línea para la creación de un Plan de Gestión de Datos



Herramienta en inglés que permite con una cuenta de Google Chrome (Gmail) crear PGDs siguiendo una plantilla, visualizar, exportar o incluso clonar PGDs públicos de otras instituciones o investigadores.



Herramienta en inglés que permite con una cuenta institucional crear PGDs siguiendo una plantilla, visualizar o exportar PGDs públicos de otras instituciones o investigadores.



Herramienta en español, basada en ARGOS, que permite con una cuenta Google Chrome (Gmail) crear PGDs siguiendo una plantilla, visualizar, exportar o incluso clonar PGDs públicos\* de otras instituciones o investigadores.

\* Actualmente solo hay un PGD público en español en PGD Madroño (Universidad de Jaén).

## 9. Evaluación de un Plan de Gestión de Datos

Aunque no existe actualmente ninguna herramienta facilitada por parte de ningún organismo de financiación que permita evaluar y comprobar que el PGD elaborado cumple con los requisitos, tenemos a nuestra disposición diversas plantillas en línea que posibilitan la evaluación punto a punto del nivel de detalle incluido en el documento, y que por tanto permiten conocer si se está dando respuesta a los interrogantes o aspectos a contemplar en un PGD:



## Plantilla de Evaluación de PGD (Science\_Europe)

### Evaluation Rubric for Data Management Plans

DMP Question	DMP Guidance	Performance Level	
Guidance for Researchers		Sufficiently Addressed The DMP...	Insufficiently Addressed The DMP...
Administrative information	<ul style="list-style-type: none"> <li>Provide information such as name of applicant, project number, funding programme, version of DMP.</li> </ul>	<ul style="list-style-type: none"> <li>contains the minimal information required to identify the applicant and the references of the project.</li> </ul>	<ul style="list-style-type: none"> <li>provides no or limited information, which makes it hard to identify who is responsible for the project.</li> </ul>
<b>1 DATA DESCRIPTION AND COLLECTION OR RE-USE OF EXISTING DATA</b>			
Guidance for Researchers		Sufficiently Addressed The DMP...	Insufficiently Addressed The DMP...
<b>1a</b> How will new data be collected or produced and/or how will existing data be re-used?	<ul style="list-style-type: none"> <li>Explain which methodologies or software will be used if new data are collected or produced.</li> <li>State any constraints on re-use of existing data if there are any.</li> <li>Explain how data provenance will be documented.</li> <li>Briefly state the reasons if the re-use of any existing data sources has been considered but discarded.</li> </ul>	<ul style="list-style-type: none"> <li>gives clear details of where the existing data come from and how new data will be collected or produced. It clearly explains methods and software used.</li> <li>explains, if existing data are re-used, how these data will be accessed and any constraints on their re-use.</li> <li>explains clearly, if applicable, why new data must be collected, rather than re-using existing data.</li> </ul>	<ul style="list-style-type: none"> <li>provides little or no details on where the data come from and what data will be collected or re-used.</li> <li>does not, if applicable, provide sufficient rationale for generating new data.</li> </ul>
<b>1b</b> What data (for example the kind, formats, and volumes) will be collected or produced?	<ul style="list-style-type: none"> <li>Give details on the kind of data: for example, numeric (databases, spreadsheets), textual (documents), image, audio, video, and/or mixed media.</li> <li>Give details on the data format: the way in which the data is encoded for storage, often reflected by the filename extension (for example</li> </ul>	<ul style="list-style-type: none"> <li>clearly describes or lists what data types will be generated (for example numeric, textual, audio, or video) and their associated data formats, including, if needed, data conversion strategies.</li> <li>explains why certain formats have been chosen and indicates if they are in open and standard format. If a proprietary</li> </ul>	<ul style="list-style-type: none"> <li>provides no or little details on what data types will be generated and does not provide a valid reason for this omission (for example a statement that no data will be produced or generated).</li> <li>only lists/describes the kinds of data</li> </ul>

### Rúbrica de Evaluación para Planes de Manejo de Datos

Pregunta DMP	Guía de DMP	Nivel de desempeño	
Orientación para investigadores		Suficientemente Abordado El DPM...	Abordado insuficientemente El DPM...
Información administrativa	<ul style="list-style-type: none"> <li>Proporcione información como el nombre del solicitante, número de proyecto, programa de financiación, versión de DMP.</li> </ul>	<ul style="list-style-type: none"> <li>contiene la información mínima requerida para identificar al solicitante y las referencias del proyecto.</li> </ul>	<ul style="list-style-type: none"> <li>proporciona información limitada o nula, lo que dificulta identificar quién es el responsable del proyecto.</li> </ul>
<b>1 DESCRIPCIÓN DE DATOS Y RECOPIACIÓN O REUTILIZACIÓN DE DATOS EXISTENTES</b>			
Orientación para investigadores		Suficientemente Abordado El DPM...	Abordado insuficientemente El DPM...
<b>1a</b> ¿Cómo se recopilarán o producirán nuevos datos y/o cómo se reutilizarán los datos existentes?	<ul style="list-style-type: none"> <li>Explique qué metodologías o software se utilizarán si se recopilan o producen nuevos datos.</li> <li>Indique cualquier restricción sobre la reutilización de los datos existentes, si las hay.</li> <li>Explique cómo se documentará la procedencia de los datos.</li> <li>Indique brevemente los motivos por los que se ha considerado la reutilización de cualquier fuente de datos existente pero se ha descartado.</li> </ul>	<ul style="list-style-type: none"> <li>brinda detalles claros de dónde provienen los datos existentes y cómo se recopilarán o producirán nuevos datos. Explica claramente los métodos y el software utilizado.</li> <li>explica, si se reutilizan los datos existentes, cómo se accederá a estos datos y cualquier restricción sobre su reutilización.</li> <li>explica claramente, si corresponde, por qué se deben recopilar nuevos datos, en lugar de reutilizar los datos existentes.</li> </ul>	<ul style="list-style-type: none"> <li>proporciona pocos o ningún detalle sobre el origen de los datos y qué datos se recopilarán o reutilizarán.</li> <li>no proporciona, si corresponde, justificación suficiente para generar nuevos datos.</li> </ul>
<b>1b</b> ¿Qué datos (por ejemplo, el tipo, los formatos y los volúmenes) se recopilarán o producirán?	<ul style="list-style-type: none"> <li>Proporcione detalles sobre el tipo de datos: por ejemplo, numéricos (bases de datos, hojas de cálculo), textuales (documentos), imágenes, audio, video y/o medios mixtos.</li> <li>Proporcione detalles sobre el formato de los datos: la forma en que se codifican los datos</li> </ul>	<ul style="list-style-type: none"> <li>claramente qué tipos de datos se generarán (por ejemplo, numéricos, textuales, de audio o video) y sus formatos de datos asociados, incluidas, si es necesario, las estrategias de conversión de datos.</li> </ul>	<ul style="list-style-type: none"> <li>proporciona pocos o ningún detalle sobre qué tipos de datos se generarán y no proporciona una razón válida para esta omisión (por ejemplo, una declaración de que no se producirán ni generarán datos).</li> </ul>



Plantilla de Evaluación de PGD (Belmont Forum DMP) 

Criteria	Complete Response (score = 2)	Incomplete Response (score = 1)	No Response (score = 0)	SCORE
<b>1. What types of datasets and other digital outputs of long-term value do you expect the project will produce or reuse?</b>				
<p>1.1 Plan lists the <u>types</u> of data and other digital outputs of long-term value.</p> <p><i>(e.g. text, databases, images, 3D models, software, audio files, code, video files, reports, surveys, patient records, samples, and so forth)</i></p>	<p>Datasets and other digital outputs of long-term value are identified, including data type and encoding.</p> <p><i>"Environmental data will be delivered as NetCDF (Network Common Data Format) files. Raster files will use the raster2pgsql PostGIS module. Maps and other geographic data will use shapefiles."</i></p> <p><i>"Transcripts and coding will be provided in text files. Audio recording will be MP3 format."</i></p>	<p>Datasets and other digital outputs of long-term value are identified, but lack detail for users beyond the project to understand.</p> <p><i>"A combination of geo-referenced data at various spatial, temporal, and taxonomic scales (e.g., populations, regions, nations, circumpolar, biomes, habitats) will comprise our data of long-term value."</i></p> <p><i>"Long-term value data include data from anthropological field studies: transcripts of interviews and discussion workshops, associated metadata."</i></p>	No information about data types is included.	X
<p>1.2 Plan describes how the data and other digital outputs will be <u>collected, captured, or created</u>.</p> <p><i>(e.g., new observations, results from models, reuse of other data, or other)</i></p>	<p>Clearly defines how data will be collected, captured or created, including methods, instruments, software, or infrastructure where relevant.</p> <p><i>"The MIP compatible model runs to the ISI-MIP servers, where existing Fish-MIP runs are already stored. Scenario outputs will be archived."</i></p> <p><i>"Socio-economic data will include household food security, nutrition, and demographic data. Spatial data produced will include ground-truthed land use/land cover data ~3 km from 50 farms, land use scenario maps for 12 villages and 4 regions. All ecological/social data will be recorded on physical datasheets and entered directly into Excel or STATA."</i></p>	<p>Missing some details regarding how some of the data will be produced; makes assumptions about reviewer knowledge of methods or practices.</p> <p><i>"Models will produce a broad range of output simulation data."</i></p> <p><i>"Data collection includes gathering in-the-field various phytoplankton, zooplankton, fish and flooded forest biodiversity."</i></p>	No information about data collection, capture or creation.	X
1.3	Datasets and other digital outputs volume estimated.	Datasets and other digital outputs amount is vaguely estimated or	Amount of expected data is	X

