Plan Estatal

(Versión 2, mayo 2025)

(Doc.CO25/05) (4 RDM\Plans Gestio Dades, 26.05.25)

Este documento tiene como objetivo servir de guía para los investigadores en la elaboración de sus Planes de Gestión de Datos (*Data Management Plan,* DMP), especialmente en el contexto de proyectos financiados en el marco del Plan Estatal de Investigación Científica, Técnica y de Innovación.

A continuación, se muestra:

* Con un número, los diferentes apartados de la plantilla.
* Con una letra mayúscula, los elementos que se deberían tener en cuenta en el momento de rellenar el apartado.
* Con una letra minúscula, las descripciones y una muestra de ejemplos reales.

Este documento ha sido elaborado por el Grupo de Trabajo de Apoyo a la Investigación del CSUC que está formado por representantes de las siguientes universidades: Universitat de Barcelona, Universitat Autònoma de Barcelona, Universitat Politècnica de Catalunya, Universitat Pompeu Fabra, Universitat de Girona, Universitat de Lleida, Universitat Rovira i Virgili, Universitat Oberta de Catalunya, Universitat de Vic-Universitat Central de Catalunya, Universitat Ramon Llull y Universitat de les Illes Balears.

Los ejemplos son una muestra1 de diferentes DMPs disponibles en la red.

Este documento está sujeto a la licencia de Reconocimiento de Creative Commons ([http://creativecommons.org/licenses/by/4.0/](https://guidelines.openaire.eu/en/latest)).

Versión digital: [http://hdl.handle.net/2072/370510](http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm)

**Información preliminar**

El entregable también suele incluir otra información preliminar como: el logotipo del proyecto, el nivel de diseminación, las revisiones históricas, la tabla de contenido y una lista de acrónimos utilizados.

**1. Resumen de los datos**

**1.A Especificar el origen de los datos, su tipo, formato y tamaño. Describir su utilidad desde el punto de vista de su reutilización, y si se podrán compartir**

1.A a) Descripción

Especificar de manera breve si los datos se generarán dentro del proyecto o se recopilarán de otras fuentes, así como el estándar, la metodología o el software utilizado para su generación o recopilación. Si se reutilizan datos ya existentes (ya sea datos abiertos, comprados, etc.) especificar la fuente de datos de donde se ha extraído (DOI, handle, URL, etc.) y explicar cómo se han tratado temas como los derechos de autor y la LPI.

Describir las tipologías de datos, formatos (texto, numérico, imagen, etc.) y tamaños (o estimación aproximada) de los datos utilizados en el proyecto. Preferiblemente utilizar estándares abiertos para que los datos puedan ser leídos por múltiples programas, lo que facilita su conservación y su uso compartido con otros usuarios. Si no los utiliza, explique por qué y las estrategias para convertirlos.

Especificar también para cada conjunto de datos las posibilidades de compartición en acceso abierto (acceso abierto total, restringido o embargado) en repositorios, licencias abiertas.

Indicar para qué colectivo y con qué finalidad pueden ser de interés los datos del proyecto.

1.A b) Ejemplo real

**Ej.** Los datos existentes provendrán de varias fuentes que incluirán: datos preexistentes de *[http://doi.org/xxxx]*, datos de la literatura científica, datos de mediciones del mundo real y datos de experimentos de simulación. También se crearán datos recopilados por el propio proyecto mediante encuestas *[XXXXX explicar]*.

Los datos tabulares estarán guardados en formatos abiertos como *[.csv (variables separadas por comas)],* que pueden importarse a archivos de texto enriquecido para procesamiento de textos o a hojas de cálculo. Si se utiliza software especializado, se proporcionará información sobre lectores gratuitos. Los ficheros de texto se guardarán en formato *[.txt y .pdf]*. Actualmente no se conoce el tamaño esperado de los datos, pero es probable que sea *[<10 GB]* y que los archivos individuales sean *[≤1 MB]*.

**Ej. RDR.** Los datos generados por el proyecto se publicarán el repositorio consorciado [*Repositori de Dades de Recerca (RDR)*](https://dataverse.csuc.cat/) una vez finalizado el proyecto, preferiblemente bajo la licencia <https://creativecommons.org/publicdomain/zero/1.0/> . Los datos serán útiles para su uso por otros grupos de investigación que trabajen en los siguientes temas: *[indicar temas]*.

# 2. Datos FAIR

**2.1 Datos encontrables (incluyendo los metadatos)**

**2.1.A Especificar el estándar de metadatos utilizado**

2.1.A a) Descripción

Esquema de metadatos que utiliza el repositorio de datos en abierto donde se ha depositado el conjunto de datos.

2.1.A b) Ejemplo real

**Ej. RDR.** Los datos se han depositado en el  [*Repositori de Dades de Recerca (RDR)*](https://dataverse.csuc.cat/), que está basado en [*Dataverse*](https://dataverse.org/) y que utiliza el estándar de metadatos [*DataCite*](https://schema.datacite.org/).

**2.1.B Identificar el tipo de identificador persistente asignado**

2.1.B a) Descripción

Qué tipo de identificador único y permanente otorga el repositorio de datos.

2.1.B b) Ejemplo real

**Ej. RDR.** Dicho repositorio utiliza como identificador único y permanente el DOI, con el formato siguiente: *[doi.org/10.34810/dataXXX]*

**2.1.C Indicar si se aplica control de versiones**

2.1.C a) Descripción

Especificar si el repositorio en el que se han depositado los datos realiza control de versiones.

2.1.C b) Ejemplo real

**Ej. RDR.** El RDR permite el control de versiones ‘minors’ (V1.1 -> V1.2) y ‘majors’ (V.1.0 -> V2.0).

**2.1.D Indicar la localización institucional o temática en el repositorio**

2.1.D a) Descripción

Mencionar, si se ha creado un Dataverse/Instancia específico para alojar los datos de la institución o del proyecto de investigación, la dirección del Dataverse/Instancia.

2.1.D b) Ejemplo real

**Ej.** Todos los conjuntos de datos de investigación vinculados con este proyecto se pueden consultar des del enlace *[especificar el enlace específico para la institución o el proyecto]*

**2.2 Datos accesibles**

**2.2.A Durante el periodo de ejecución de la investigación**

2.2.A a) Descripción

Dónde se han almacenado los datos. Quién ha podido acceder, con qué permisos.

2.2.A b) Ejemplo real

**Ej.** Durante el periodo de ejecución de este proyecto los datos se han almacenado en el ECM (Enterprise Content Management) de la institución, *[SharePoint de Microsoft]*, que permite compartir el acceso al conjunto de datos con diferentes permisos en función del usuario.

Para esta investigación el IP del proyecto ha sido el administrador de todos los conjuntos de datos generados; por otro lado, cada investigador ha creado su propia estructura de carpetas y ficheros, dentro del espacio definido para el proyecto. Los permisos hacia otros espacios en los que no se es propietario han sido solo de consulta.

**2.2.B Al finalizar la investigación (en la que se pueden dar varios casos)**

2.2.B.1 a) Descripción

Dónde se han almacenado los datos. ¿Quién ha podido acceder, con que permisos?

2.2.B.1 b) Ejemplo real

**Ej. 1** Los datos están depositados en acceso abierto sin ningún tipo de restricción para su acceso.

**Ej. RDR A.** Los datos están depositados en el RDR, con el siguiente DOI: *[doi.org/10.34810/dataXXX]*. Los metadatos están disponibles en acceso abierto, los ficheros tienen interpuesto un periodo de embargo que finaliza *[dd/mm/aaaa]*, y a partir de ese momento estarán accesibles sin ningún tipo de restricción.

**Ej. RDR B.** Los datos de investigación están depositados en el RDR con el siguiente DOI: *[doi.org/10.34810/dataXXX]*. Los metadatos están disponibles en acceso abierto, los ficheros tienen interpuesto un acceso restringido dado que para poder ser compartidos se debe generar un convenio de colaboración para su reutilización.

**Ej. RDR C.** Los datos de investigación contienen datos de carácter personal por lo que no se han depositado en cumplimiento con *[....]* / Dado que el reglamento del RDR especifica que no se pueden depositar este tipo de datos.

2.2.B.2 a) Descripción

Se debe mencionar los métodos y programas necesarios para acceder a los datos. La recomendación es realizar el depósito en formatos abiertos. En los casos de que la disciplina utilice formatos propietarios como estándares del ámbito, se debe indicar el programa y la versión específica que se ha utilizado para crear el conjunto de datos de investigación.

2.2.B.2 b) Ejemplo real

**Ej. 1** Los datos se han depositado en formato *[(.csv, .tiff, .pdf, .tab....)]* que es abierto y no se precisa de ningún programa específico para poder acceder a ellos.

**Ej. 2** Los datos se han generado en *[formato .raw]*, que es el formato estándar y propietario del *[software XXX versión XXX]* utilizado en espectrometría de masas. Este formato requiere el uso del programa mencionado para la apertura, visualización y análisis de los datos depositados en *[.raw]*. *[Software XXX]* no permite exportar los datos en su totalidad a un formato abierto sin perder parte de la información estructural del experimento (metadatos de adquisición, parámetros del instrumento, etc.). Por ello, el acceso completo a los datos originales está condicionado a disponer de dicho software.

**2.3 Datos interoperables**

**2.3.A Especificar la interoperabilidad de los datos. Explicar qué datos y qué vocabularios de metadatos, estándares o metodologías se seguirán para facilitar la interoperabilidad**

2.2.A a) Descripción

Explicar qué vocabularios, estándares o metodologías de datos y metadatos se seguirán para facilitar la interoperabilidad. La interoperabilidad de datos del proyecto permite el intercambio y reutilización de datos entre investigadores, instituciones, organizaciones, países, etc. Adherirse a los estándares de formatos que sean, en la medida de lo posible, compatibles con programas y aplicaciones abiertas.

2.2.A b) Ejemplo real

**Ej. 1** Los datos generados en el proyecto serán interoperables, ya que los conjuntos de datos se ajustarán a formatos estandarizados: *[ASCII, TXT, CSV, XML y TIFF]*. Si no se puede utilizar *[MS Office]*, un visor de PDF o un visor de imágenes, se proporcionará un archivo de texto (ASCII) junto con el conjunto de datos que explica dónde se puede obtener un lector gratuito.

**Ej. 2** No existen estándares de metadatos consensuados por la comunidad para esta disciplina. Para garantizar la interoperabilidad, los datos se documentarán adecuadamente (archivos README, etc.) para permitir su reutilización y operatividad entre disciplinas. Además, los datos subyacentes a las publicaciones estarán disponibles a través de *[nombre del repositorio]* (que utiliza *[esquema de metadatos DataCite y Dublin Core]*).

**Ej.3** Para garantizar la interoperabilidad entre los conjuntos de datos disponibles a través del portal de *[asociación]* y otros datos espaciales, se adoptan los protocolos estándar *[OGC WMS y WFS]*. Los metadatos se publican mediante el protocolo estándar *[OGC CSW]*. Además, la disponibilidad de los conjuntos de datos del [portal] como archivos descargables en formatos habituales *[(archivos GeoJSON, archivos ESRI shapefiles, tablas de MapInfo, KML)]* facilita a los usuarios que lo necesiten la combinación de los conjuntos de datos del *[portal]* con otros datos geográficamente referenciados en *[SIG de escritorio]*.

**2.4 Datos reutilizables**

**2.4.A Especificar cuando los datos estarán disponibles para su reutilización y en qué condiciones (licencias). Si procede, especificar porqué y durante qué período será necesario embargar los datos**

2.4.A a) Descripción

Si los datos se ponen a disposición de otros investigadores y del público en general, especificar el grado de reutilización permitido a través de las licencias. La CE propone las licencias *Creative Commons CC-BY* o *CC0*. Existen licencias específicas para software/código. Para más información, consultar la guía [Derechos y licencias para datos de investigación](https://confluence.csuc.cat/x/vJDLBQ).

Si el conjunto de datos se publica en abierto, indicar cuándo los datos estarán disponibles para su consulta y reutilización (durante el proyecto, al final del proyecto, etc.). Si el conjunto de datos tiene alguna restricción de acceso, especificar el motivo y durante qué período será necesario embargar o restringir.

Indicar la documentación que acompañará a los datos publicados (archivo readme, diccionarios de datos, libros de códigos, etc.) necesaria para comprenderlos y reutilizarlos. Esta documentación debe incluir información sobre cómo se recopilaron los datos, información analítica y procedimental, definiciones de variables, unidades de medida, etc.

Especificar durante cuánto tiempo los datos estarán disponibles para su reutilización (diferentes estándares internacionales recomiendan un mínimo de 10 años).

2.4.A b) Ejemplo real

**Ej. RDR.** Los datos del proyecto se publicarán una vez finalizado el proyecto en el [*Repositori de Dades de Recerca (RDR)*](https://dataverse.csuc.cat/) bajo la licencia *[CC-BY]*. Los datos publicados no tendrán restricciones de acceso ni embargo. Toda la documentación necesaria para validar el análisis de datos y facilitar su reutilización acompañará a los datos a través de un archivo README creado de acuerdo con la plantilla y las pautas RDR. El repositorio RDR garantiza la preservación y facilita el acceso a los conjuntos de datos depositados durante un mínimo de 10 años.

# 3. Responsabilidades

**3.A Indicar las personas responsables de las actividades de la gestión de los datos y especificar cuestiones relacionadas con la titularidad de los derechos de propiedad intelectual**

3.A a) Descripción

Indicar las personas responsables de la creación, seguimiento y actualización del PGD. Especificar quién es el responsable de la gestión de los datos y quién determina el acceso a los mismos, así como su publicación, preservación y conservación. En el caso de que hubiera derechos de propiedad intelectual, determinar quién es el titular. En el caso de proyectos colaborativos, determinar las responsabilidades de cada equipo y la responsabilidad de coordinación, así como se gestiona la propiedad intelectual entre los diferentes equipos. Señalar cualquier normativa institucional aplicable.

3.A b) Ejemplo real

**Ej. 1** La/el IP del proyecto, será el/la responsable de la gestión de datos durante todo el periodo de ejecución de la investigación. Sus responsabilidades han sido las de crear y supervisar el espacio compartido que contendrá toda la estructura de archivos y carpetas generadas como resultado del proceso de investigación. Antes de publicar los conjuntos de datos en el repositorio, se realizará una curación de los datasets por parte del servicio de apoyo a la investigación de la Biblioteca.

**Ej. 2** La/el IP del proyecto, supervisará la aplicación y el progreso del Plan de Gestión de Datos, asegurándose de que se cumplan las metas. El IP también supervisará los esfuerzos de gestión de datos, con la participación de todos los miembros del equipo, en la manipulación, preservación y conservación de los datos. Los investigadores serán responsables de garantizar, a lo largo de todo el proyecto, la integridad de los datos, su correcto etiquetado y archivo. El acceso a los datos será supervisado por el IP y el gestor de datos.

# 4. Seguridad de los datos

**4.A Especificar las medidas de seguridad para la gestión y conservación de los datos**

4.A a) Descripción

Se debe especificar como se realiza el tratamiento de la recuperación de los datos, así como el almacenamiento seguro y el traspaso de datos sensibles y las medidas adoptadas para la seguridad de los datos. Este punto incluye el almacenamiento y las copias de seguridad, la preservación a largo plazo y los procedimientos para mantener los datos seguros y protegidos

4.A b) Ejemplo real

**Ej. RDR.** Durante la recolección o elaboración de los datos de investigación se seguirán las directrices de la *[institución X]*. Y al final del proyecto se depositarán en el RDR. La siguiente lista describe su configuración de seguridad:

* **Versiones:** Los archivos de datos están versionados. Los archivos no están versionados. Los datos cargados se archivan como paquete de información de envío. Se generan derivados de los archivos de datos, pero el contenido original nunca se modifica. Los registros pueden ser retirados de la vista pública; sin embargo, los archivos de datos y los registros se conservan.
* **Réplicas:** Todos los archivos de datos se almacenan en el Centro del CSUC, principalmente en Barcelona, con réplicas en el Consorcio Madroño de Madrid. Los archivos de datos se mantienen en múltiples réplicas en un sistema de archivos distribuido, del que se realizan copias de seguridad en cinta cada noche.
* **Periodo de conservación:** Los elementos se conservarán durante la vida útil del repositorio. El RDR ha definido una vida útil para el repositorio de los próximos 10 años como mínimo.
* **Conservación funcional:** El RDR no hace promesas sobre la usabilidad y comprensibilidad de los objetos depositados a lo largo del tiempo.
* **Conservación de archivos:** Se realizan copias de seguridad nocturnas de los archivos de datos y metadatos, que se replican en múltiples copias en el sistema en línea.
* **Fijeza y autenticidad:** Todos los archivos de datos se almacenan junto con una suma de comprobación MD5 del contenido del archivo y el archivo tabular se almacena con la huella digital numérica universal (UNF). Los archivos se comprueban periódicamente con sus sumas de comprobación para garantizar que su contenido permanece constante.
* **Planes de sucesión:** En caso de cierre del repositorio, desde RDR se ha garantizado la migración. Se puede consultar el plan de contingencia y especificaciones técnicas del RDR en el siguiente enlace: [Policies/Plans](https://confluence.csuc.cat/x/4ADxCQ).

**4.B Describir la cobertura ética en el manejo de los datos**

4.B a) Descripción

Cobertura ética, se debe describir, si fuera de aplicación, los problemas éticos en la recogida, almacenamiento, procesamiento y archivo de los datos, así como en los procedimientos de aprobación ética relacionados en el proyecto.

4.B b) Ejemplo real

**Ej. 1** Esta investigación tiene la aprobación del comité ético de la *[institución]*, con un informe favorable emitido por *[órgano competente en cada institución]* el día *[dd/mm/aaaa]*, tal y como se adjunta en la justificación de este proyecto.

**Ej. 2** Esta investigación no necesita aprobación del comité ético, ya que no se manejan datos de carácter personal, sensibles o relacionados con los supuestos que la ley establece para el registro en el comité ético.