

BIBLIOTECA DE LA
UNIVERSIDAD DE MURCIA
NOVIEMBRE, 2022

**GUÍA PARA LA
ELABORACION
DE UN PLAN DE
GESTIÓN DE
DATOS DE
INVESTIGACIÓN**

MARÍA ÁNGELES MOLINA MICOL
SUSANA SÁNCHEZ ALMARCHA

1. INTRODUCCIÓN
2. PARTES QUE CONSTITUYEN UN PLAN DE GESTIÓN DE DATOS
 - 2.1. DATOS ADMINISTRATIVOS
 - 2.2. DESCRIPCIÓN DE LOS DATOS
 - 2.3. DATOS FAIR.
 - 2.4. RESPONSABILIDADES Y RECURSOS
 - 2.5. SEGURIDAD DE LOS DATOS
 - 2.6. ASPECTOS ÉTICOS Y LEGALES
 - 2.7. OTROS
3. EJEMPLOS DE PLANES DE DATOS
4. GENERADORES DE PLANES DE GESTIÓN DE DATOS
5. BIBLIOGRAFÍA

1. Introducción

La Biblioteca universitaria ha redactado esta guía para la elaboración del Plan de Gestión de Datos (PGD), cuyo objetivo es ayudar a los investigadores de la Universidad de Murcia (UM) a realizar una buena gestión de los datos de investigación y que cumplan con los principios FAIR.

Es un complemento para facilitar el depósito de los datos con una serie de orientaciones y recomendaciones tanto generales, como propias de la UM. También se incluyen enlaces a recursos para ampliar información y ejemplos de otros planes de gestión de datos.

Tanto la guía como la plantilla, están principalmente basadas en la herramienta **DMPonline** de **The Digital Curation Centre (DCC)**, el uso de la misma es general y válida para otras herramientas, como por ejemplo **PGDOnline**, **Argos**, etc. En esta guía se señalan con un * los campos obligatorios para la herramienta Argos.

Las agencias de financiación solicitan generalmente a los investigadores un plan de gestión de datos, al inicio, durante y al acabar el proyecto. En la directorio **Sherpa Juliet** puedes consultar las políticas de agencias que financian proyectos de investigación con respecto a los datos de investigación bajo el paraguas de la ciencia abierta, tanto públicas como privadas.

En el actual **Programa Horizonte Europa (2021-2027)**, es obligatorio elaborar un plan de gestión de datos para todos los proyectos que generen o reutilicen datos, dentro de los 6 primeros meses del proyecto, con el fin de:

- Realizar una gestión responsable de los datos de la investigación de acuerdo con los principios FAIR de “localización”, “accesibilidad”, “interoperabilidad” y “reutilización”.
- Garantizar el acceso abierto a los datos de investigación, según el principio “tan abierto como sea posible, tan cerrado como sea necesario”, depositando los datos en un repositorio seguro, bajo una licencia Creative Commons (CC BY), de dominio público (CC0) o equivalente.

Más información en el portal Web de **Horizonte Europa**

En la web de la Biblioteca puedes encontrar información actualizada respecto a cómo crear un plan de gestión de datos

 <https://www.um.es/web/biblioteca/investigar-publicar/datos-investigacion/plan-gestion>

La Biblioteca Universitaria ofrece también a los investigadores, un servicio de asesoramiento para la realización de los planes de gestión de datos, a través del Grupo de apoyo a la investigación (enlazar web), bib.gestiondedatos@um.es.

2. Partes que constituyen un plan de gestión de datos



2.1. Datos administrativos

En este primer apartado, el investigador debe especificar todos los datos administrativos relacionados con el proyecto de investigación:

Datos propios del proyecto:

- Nombre del proyecto, tal y como se ha solicitado en la subvención.
- Resumen del proyecto de la investigación, por el cual se obtienen o recopilan los datos, su procedencia, naturaleza y objetivos generales.
- Código identificativo del proyecto de investigación asignado por la agencia de financiación o la institución correspondiente.
- Fecha de comienzo y finalización del proyecto de investigación.
- *Organización o institución que financia el proyecto de investigación.
- *Idioma en el que está redactado el proyecto de investigación.

Datos propios del investigador o investigadores:

- Especificar el investigador responsable y los investigadores principales del proyecto.
- Identificador personal del investigador responsable del proyecto, puede ser un ORCID, Author ID, IraLIS, ...
- *Nombre de la persona de contacto, correo electrónico y teléfono para comunicarse con los investigadores.

2.2. Descripción de los datos

In the second section, the researcher should describe the information about the data obtained or collected in the research project:

- Summary of the data, the nature, content and purpose of the data.

Special attention should be paid to this field so as not to confuse it with the summary specified in administrative data, as this refers to the summary of the research project and its objectives. And in this section the summary is the summary of the data that will be or has already been collected during the research project.

- Specify keywords to facilitate collection and retrieval by databases, repositories and Internet search engines.

- The types of data can be numerical, textual, images, audio, video, etc., the format in which the data are encoded for storage, e.g., PDF, DOCX, XLSX, CSV, PNG, etc.

- Indicate the volume of data expressed in KB/MB/TB...

It is recommended that open source and non-proprietary formats be used whenever possible at the time of data dissemination to facilitate reuse, usability and long-term preservation.

.There are disciplines that recommend specific data formats. Several tools allow the identification of the most commonly used standards, for example:

Sustainability of Digital Formats: Planning for Library of Congress Collections

 http://www.digitalpreservation.gov/formats/fdd/browse_list.shtml

 https://www.loc.gov/preservation/digital/formats/content/gis_intro.shtml

 <https://fairsharing.org/>

Por su parte, la Biblioteca del Congreso de los Estados Unidos también mantiene **un listado actualizado de formatos recomendados para objetos analógicos y digitales** con fines de preservación.

- Metodología empleada en la generación o recolección de los datos de investigación.
- Especificar si los datos se pueden reutilizar para otras investigaciones futuras, para qué y para quién podrían ser útiles. En el caso de reutilizar datos existentes, indicar las posibles restricciones que existan para su reutilización.

Si finalmente se descarta la reutilización de datos, indicar brevemente los motivos: como, por ejemplo, por seguridad ciudadana o medioambiental, restricciones de propiedad intelectual y/o industrial, restricciones relacionadas con la protección de datos personales, etc.

- Si los ficheros de datos se modifican a lo largo de la investigación, hay que especificar las diferentes versiones de los mismos.
- Citar correctamente los datos de investigación, es importante para que puedan ser reutilizados en otras investigaciones, en actividades académicas y para otros fines de la sociedad en general.

Se pueden citar en cualquier estilo bibliográfico: APA, Chicago, ISO 690, MLA, Vancouver, etc., Existe una herramienta online que permite la generación de citas en numerosos formatos, tan solo con facilitar el DOI del fichero de datos en **DOI Citation Formatter**

Más información sobre cómo elaborar la cita en “**Cita tus datos de investigación**” .

2.3. FAIR Data

En este tercer punto, se trata de describir como se tratan los datos para garantizar que los datos sean **FAIR** (Findable, Accesible, Interoperable, Reusable).

Los datos deben seguir los principios FAIR, conjunto de directrices para hacer que los datos de investigación sean localizables, accesibles, interoperables y reutilizables (**Wilkinson et al., 2016**). Estos principios sirven de base para la gestión de datos científicos y se dirigen directamente a los productores de datos y a los editores de datos para promover el acceso y la reutilización de los datos de investigación.

Ser localizable (findable):

- F1. Asignar un identificador único y persistente a los datos y metadatos.
- F2. Describir los datos con metadatos enriquecidos (es decir, agregando declaraciones semánticas, anotaciones, etc. sobre los datos que describen, que mejoran la calidad de los metadatos).
- F3. Registrar e indizar datos y metadatos en un sistema (repositorio, servicio, etc.) que disponga de motor de búsqueda.
- F4. Incluir un elemento de metadatos específico para el identificador persistente de los datos

Se incluyen los identificadores, palabras clave, normas de metadatos que los describen para optimizar su localización.

Ser accesible (accessible):

- A1. Utilizar protocolos estandarizados para poder recuperar datos y metadatos mediante su identificador.
 - A1.1 Los protocolos han de ser abiertos, gratuitos y universalmente aplicables.
 - A1.2 Los protocolos deben permitir, además, un procedimiento de autenticación y autorización, si fuera necesario.
- A2. Garantizar la disponibilidad de accesibilidad de los metadatos, aunque los datos ya no estén disponibles

Se recomienda que los datos se depositen en el repositorio institucional de la Universidad de Murcia: **DIGITUM**,

 <https://digitum.um.es/digitum/handle/10201/37686>.

Y además, si el investigador lo desea en un repositorio de datos específico para la disciplina científica. Para localizar estos repositorios se pueden utilizar directorios como **r3data.org** y **FAIRsharing**.

Ser interoperable (interoperable):

- 11. Utilizar lenguajes formales (accesibles, compartidos y normalizados) para representar datos y metadatos.
- 12. Describir los datos y metadatos con vocabularios (esquemas, ontologías, etc.) que también sigan los principios FAIR.
- 13. Incluir referencias cruzadas y enlaces entre datos y metadatos.

Explicar qué datos y qué vocabularios de metadatos, estándares o metodologías se siguen para facilitar la interoperabilidad. Se podrán usar esquemas generales (por ejemplo, Dublin Core o DataCite), aunque se recomienda la utilización de esquemas de metadatos específicos ampliamente utilizados por la comunidad científica de una disciplina concreta.

La interoperabilidad es la capacidad de los sistemas de información y de los procedimientos a los que éstos dan soporte, de compartir datos y posibilitar el intercambio de información y conocimiento entre ellos.

En el caso de depositar el conjunto de datos en la Universidad de Murcia, DIGITUM utiliza metadatos Dublin Core Metadata, protocolo OAI-PMH, y tiene como requisito que el formato utilizado para la creación de los datos sea estándar y el software sea abierto. Si se deposita en otro repositorio, esta información la proporcionara el repositorio elegido.

Ser reutilizable (reusable):

- R1. Asegurarse de que los datos y metadatos utilizados tienen una variedad de elementos o atributos precisos y relevantes.
 - R1.1. Publicar datos y metadatos con una licencia de reutilización clara y accesible.
 - R1.2. Utilizar criterios de provenance (creación, atribución e historial de versiones) para asociar metadatos a los datos durante su ciclo de vida.
 - R1.3. Asegurarse de que los estándares de datos y metadatos utilizados cumplen con los estándares comunes del área de conocimiento a la que se refieren los datos.

Documentar la procedencia de los datos y proporcionar la información necesaria para validar los datos y facilitar su reutilización.

La Biblioteca ha elaborado una plantilla **README.txt** que acompañará al resto de ficheros con los datos de investigación, donde se indicará la metodología empleada, variables usadas, etc. para facilitar la reutilización de los datos.

Otras plantillas:

- o Plantilla de la **Universidad de Cornell**
- o Plantilla de la **Universidad Autónoma de Barcelona**
- o Plantilla del **Consortio Madroño**
- o Plantilla de la **Universidad Politécnica de Madrid**

- Si no entiendes muy bien a qué estándares, metadatos o identificadores se refieren estos principios, pregunta a tu biblioteca universitaria.

Una herramienta muy útil para comprobar si tus datos son FAIR es **F-UJI**. F-UJI es el resultado del **proyecto FAIRsFAIR** "Fostering FAIR Data Practices In Europe" de la convocatoria del proyecto Horizon 2020 de la Unión Europea.

También puedes consultar **GO-FAIR** y OpenAire mantiene una **sección** con recursos de apoyo sobre los datos FAIR.



2.4. Responsabilidades y recursos

En este punto, hay que describir la asignación de los recursos al proyecto y los responsables.

·Especificar el responsable del PGD, el cual se encargará de su redacción, implementación, revisión y actualización si fuera necesario.

Identificar los roles y responsabilidades de las diferentes actividades relacionadas con la gestión de los datos: captura de los datos, introducción de metadatos, calidad de los datos, almacenamiento y copias de seguridad, depósito de los datos, difusión, etc.

Si en el proyecto participan varias organizaciones, explicar la gestión de los datos entre los diferentes participantes en el proyecto.

·Costes en tiempo y recursos empleados, considerar y justificar los costes de almacenamiento, personal, hardware, costes para la preparación de los datos previos a su difusión, costes de depósito en el repositorio, etc.

En el caso de que se necesiten nuevos recursos especificar cómo se cubrirán los costes. Se recuerda que en ocasiones las agencias de financiación cubren estos costes, por ejemplo, en el actual programa marco europeo, Horizonte Europa.

Si recomienda consultar la guía desarrollada por la **Universidad de Utrecht**, para el cálculo de los costes derivados de la gestión de datos.

2.5. Seguridad de los datos

En el caso de recabar datos personales, confidenciales, o por si están sujetos a algún tipo de confidencialidad, los datos deben estar en un lugar seguro y con acceso controlado mediante los sistemas de seguridad adecuados.

Otra cuestión relevante se refiere a la **anonimización** de los datos, si fuera necesario por cuestiones de privacidad.

Existen herramientas diferentes para anonimizar los datos de investigación que contengan información personal confidencial, por ejemplo **AMNESIA** de OpenAire.

Explicar cómo y quién gestiona y controla el acceso para trabajar con los datos o la transferencia de éstos, manteniendo la seguridad, integridad y confidencialidad (si es el caso) de los mismos, así como su almacenamiento.

Se recomienda realizar copias de seguridad de forma segura y con vistas a la conservación a largo plazo. La Universidad de Murcia dispone de un disco personal en red llamado **UMUBOX**. Puedes consultar la **política de seguridad** de la Universidad de Murcia, así como sus normativas, las cuales cumplen con el **Esquema Nacional de Seguridad (ENS)**.

2.6. Aspectos éticos y legales

En este bloque, se mencionarán todos los aspectos éticos y legales del proyecto de investigación.

·Para garantizar la protección de los datos personales, se debe tener en cuenta la normativa europea (**Reglamento (UE) 2016/679**) y la legislación nacional (**Ley Orgánica 3/2018**) sobre protección de datos personales.

En la Universidad de Murcia se recomienda consultar a la Comisión de Ética y Buenas prácticas en la Investigación y al Delegado de Protección de datos para recibir asesoramiento sobre los posibles riesgos en el tratamiento de los datos y las acciones para la gestión de estos riesgos:

 Comisión de Ética y Buenas prácticas en la Investigación

 https://www.um.es/web/hrs4r/ethics/comision_etica

 Contacto con el delegado de protección de datos: 868 88 9330 dpd@um

Es importante aclarar estos aspectos, especialmente en los casos en los que los propietarios de los datos pertenecen a diferentes organizaciones.

·Se recomienda la utilización de licencias **Creative Commons** para definir las condiciones de uso y reutilización de los datos. Las licencias que se recomiendan son la CC By o la CC0 1.0 Universal (CC0 1.0), Dedicación de Dominio Público.

Para depositar los Datos de investigación en el repositorio institucional de la Universidad de Murcia, Digitum, además de asignarle una licencia Creative Commons, hay que aceptar la licencia de **Digitum**.



2.7. Otros temas

Por último, referenciar los proyectos o procedimientos empleados durante el proyecto de investigación.

·Si durante el proyecto de investigación se utilizan otros procedimientos nacionales, sectoriales o departamentales para la gestión de datos, hay que especificarlos en este apartado.

3 .Ejemplos de planes de datos

¿Dónde encontrar planes de gestión de datos reales?

Si bien aún no es una práctica muy extendida cada vez hay más grupos de investigación que publican sus planes de gestión de datos como un resultado más de sus proyectos. Los siguientes sitios ofrecen un catálogo de planes reales:

- **LIBER: Data Management Plan Catalogue**
- **Zenodo: data management plan**
- **Digital Curation Centre: DMP online**
- **Digital Curation Centre: example DMPs and guidance**
- **DMP Tool**

4. Generadores de Planes de Gestión de Datos

Varias herramientas gratuitas online permiten generar planes de gestión de datos de manera colaborativa entre los miembros del equipo de investigación. Estas herramientas integran explicaciones, recomendaciones y sugerencias para cada sección del plan y están cada vez más estandarizadas. Entre ellas destacamos:

- **DMPOnline:** es la herramienta recomendada por la Comisión Europea y contiene las plantillas para generar planes de gestión de datos de diversas agencias financiadoras de investigación.
- **DSW Wizard:** es una herramienta con funcionalidades específicas para investigadores e instituciones y su cuestionario se alinea con los Principios FAIR.
- **RDMO:** requiere instalación previa y permite planificar y llevar a cabo la gestión de datos de un proyecto de investigación.
- **Argos:** es una herramienta en línea, iniciativa de OpenAIRE y EUDAT. Ofrece una plataforma para la Planificación de Gestión de Datos y adopta los principios FAIR para datos de investigación.
- Data management Plan generators

5. Bibliografía

Argos[consulta: 16 de octubre de 2022]

ConSORCI de Serveis Universitaris de Catalunya. 2019. **Data Management Plans**. Versió 4, June 2019. Adaptación de las H2020 Guidelines... V.4.0.

Costs of data management. **Universidad de Utrecht** [consulta: 16 de octubre de 2022]

DCC. 2013. **Checklist for a Data Management Plan**. V.4.0. Edimburgh: Digital Curation Centre.

F-UJI (Servicio de Evaluación de Objetos de Datos de Investigación FAIRsFAIR) [consulta: 16 de octubre de 2022]

Pagoda [consulta: 16 de octubre de 2022]

Principios FAIR: Buenas prácticas para la gestión y administración de datos científicos [consulta: 11 de noviembre de 2022]

Recomendaciones para la gestión de datos de investigación Dirigidas a investigadores [consulta: 11 de noviembre de 2022]

The Digital Curation Centre (DCC)[consulta: 16 de octubre de 2022]

Universidad de Cantabria. Biblioteca. Datos de investigación.

<https://web.unican.es/buc/investigacion/datos-de-la-investigacion> [consulta: 16 de octubre de 2022]

Universidad de Málaga. Orientaciones para elaborar un Plan de Gestión de Datos (PGD) en la Universidad de Málaga. **https://biblioguias.uma.es/RIUMA/Datasets_PGD_UMA**, [consulta: 16 de octubre de 2022]

Universidad de Murcia. Biblioteca. Apoyo a la investigación.

<https://www.um.es/web/biblioteca/investigar-publicar/apoyo-investigacion> [consulta: 16 de octubre de 2022]

<https://www.um.es/web/biblioteca/investigar-publicar/datos-investigacion>[consulta: 16 de octubre de 2022]

Universidad de Sevilla. Biblioteca.

https://bib.us.es/estudia_e_investiga/investigacion/estrategias/gdi/Elaborar_un_Plan_de_Gestion_de_Datos [consulta: 16 de octubre de 2022]

Universidad de Valencia, Plan de Gestión de Datos de investigación UPV: guía para la elaboración de un Plan de Gestión de Datos de investigación en el contexto de la Universitat Politècnica de València (UPV).

<https://riunet.upv.es/handle/10251/143184> [consulta: 16 de octubre de 2022].