

# Guía para elaborar un plan de gestión de datos para doctorandos

(Doc.CO22/26) (4 RDM\DMP\EinaDMP\Plantilles\Doctorands\DMP\_Doctorandos\_ES.docx, 24.11.22)

Este documento pretende ser una guía de apoyo a los doctorandos a la hora de crear sus Planes de Gestión de Datos (Data Management Plan, DMP)

Este documento está elaborado por el Grupo de Trabajo de Soporte a la Investigación del CSUC que está formado por representantes de las siguientes universidades: Universidad de Barcelona, Universidad Autónoma de Barcelona, Universidad Politécnica de Catalunya, Universidad Pompeu Fabra, Universidad de Girona, Universidad de Lleida, Universidad Rovira i Virgili, Universidad Oberta de Catalunya, Universidad de Vic, Universidad Central de Cataluña, Universidad Ramon Llull y Universidad de las Illes Balears.

Este documento sujeto a la licencia de Reconocimiento de Creative Commons (<http://creativecommons.org/licenses/by/4.0/>).

Versión digital: <http://hdl.handle.net/2072/370509>

NOTA: El punto 6. Datos sensibles/personales ha sido elaborado por la Universitat de Lleida y es necesario para el tratamiento de datos de las tesis doctorales de la UdL.

## Sobre tu investigación

---

|                                |  |
|--------------------------------|--|
| Nombre y correo electrónico    |  |
| Director/es de tesis           |  |
| Título provisional de la tesis |  |
| Describe tu investigación      | <i>Aproximadamente 50 palabras</i>                                   |
| Duración de la investigación   | Inicio: DD-MM-AAAA<br>Fin: DD-MM-AAAA                                |
| Proyecto vinculado             | <i>¿Se trata de una tesis vinculada a un proyecto? ¿Cuál?</i>        |
| Financiamiento                 | <i>¿Has recibido financiación para realizar la tesis? ¿De quién?</i> |

## Sobre este plan de gestión de datos

---

|                      |   |
|----------------------|---|
| Fecha de creación    | <i>DD-MM-AAAA</i>   |
| Última actualización | <i>DD-MM-AAAA</i>   |
| Versión y fecha      | <i>Haz una nueva versión cada vez que haya cambios significativos (nuevos datasets, cambios en tu investigación u otros factores)</i> |

|                            |  |
|----------------------------|--|
| Datos sensibles/personales | <p><i>Si trabajas <u>con datos de carácter personal o sensible</u>, tienes la obligación legal de tratarlos de acuerdo con la normativa vigente. Los datos personales son cualquier información que permite identificar a una persona (nombre, dirección, localización, etc.).</i></p> <p><input type="checkbox"/> No voy a trabajar con datos personales</p> <p><input type="checkbox"/> Trabajaré con datos personales [véase punto 2 y 6]</p> |
|----------------------------|--|

## 1. Captura de datos

*Describe los datos que vas a crear/recoger*

### 1.1 ¿Durante tu búsqueda utilizarás datos ya existentes? Si no es así, indica la procedencia de los datos que vas a utilizar

- No
- Datos propios o datos del grupo de investigación en los que participas
- Colaboradores académicos
- Colaboradores de carácter comercial
- Bases de datos/archivos disponibles de forma pública
- Proveedores comerciales de datos
- Otros (indica cuáles):

### 1.2 Descripción de los datos

*Describe los datos que vas a crear o los datos de terceros que vas a reutilizar e indica:*

- *si vas a usar protocolos o estándares usados en tu área de investigación*
- *qué herramientas, instrumentos, equipamiento, hardware o software utilizarás*

*Si reutilizas datos de terceros, asegúrate de que dispones de los permisos adecuados y de conocer las condiciones de uso de estos datos.*

### 1.3 Tipo y formatos de los datos

*Ten en cuenta:*

- *la tipología de datos: por ejemplo, si trabajarás con medidas, simulaciones, observaciones, texto (texto, MS Word), imágenes, audiovisuales o muestras, estadísticas (hojas de cálculo), con modelos computacionales, con datos de una encuesta cualitativa (cuestionarios), grabaciones (audio, vídeo), software (código), etc.*
- *la longevidad de los formatos de archivo: utiliza preferentemente estándares abiertos para que los datos puedan ser leídos por varios programas facilitando la preservación y la compartición con otros usuarios.*

#### 1.4 Indica el volumen de los datos

- < 10 GB
- 10-30 GB
- 30-50 GB
- 50 GB-250 GB
- 250 GB-500 GB
- 500 GB-2 TB
- 2 TB

## 2. Almacenamiento de datos y seguridad

---

*Asegúrate de que todos tus datos de investigación se almacenan de forma segura y que se realizan copias de seguridad periódicamente*

### 2.1 Indica las restricciones (comerciales, éticas o de confidencialidad) que pueden afectar a tus datos

- Obligaciones contractuales
- Obligaciones legales: protección de datos personales ([LOPDGDD](#), [RGPD](#)...) [ver 4.1]
- Obligaciones legales: copyright, propiedad intelectual [ver 4.1]
- Restricciones éticas [ver 4.1]
- Aspectos comerciales (ej. patentabilidad)
- Estándares formales de seguridad
- Sin obligaciones
- Otros, indica:

*Explica brevemente las restricciones*

### 2.2 Riesgos principales en la seguridad de los datos

*Indica los principales riesgos, por ejemplo: borrado accidental de datos, pérdida o robo de los datos. Describe las consecuencias de una posible pérdida de datos*

### 2.3 Medidas a tomar para mitigar los riesgos de una posible pérdida de datos

- Restricciones de acceso
- Encriptación
- Procesamiento de datos
- Pseudoanonimización
- Anonimización
- Copias de seguridad periódicas
- Otros, indica:

*Indica también los procedimientos que vas a utilizar para asegurar la confidencialidad de los datos personales*

## **2.4 Dónde almacenar los datos?**

*En caso de **tratar y almacenar datos personales**, contesta el punto 6.6. ¿Dónde vas a almacenar los datos en caso de tratar datos personales?*

- En la red de tu departamento o grupo de investigación
- En la red de la universidad
- Almacenamiento físico (p. ej. USB, disco duro externo)
- Servicio en la nube (p. ej. Dropbox)
- Otros, indica:

*Describe brevemente las condiciones de almacenamiento y copias*

### 3. Documentación de los datos

---

*Documenta los datos para facilitar su comprensión y reutilización*

#### 3.1 Nombre y estructura de los archivos y carpetas

*Describe cómo organizarás y nombrarás las carpetas y archivos de datos*

#### 3.2 Control de versiones

*Describe brevemente cómo controlarás las versiones. Indica también qué harás en caso de eliminar datos*

- Sin control de versiones (p. ej. los archivos originales se sobrescriben)
- Software con control de versiones, indícalo:
  - Software con opción de seguimiento de cambios
  - Número de versión y fecha en el nombre del archivo o carpeta
  - Haciendo una copia del script con el que se procesan los datos
  - Otros, indica:

#### 3.3 ¿Qué estándares de metadatos vas a utilizar?

*Ver [“Disciplinary metadata standards”](#) del DCC o [“Metadata standards”](#) en Wikipedia.*

- No usaré ningún estándar (especifica los metadatos necesarios para entender los datos)
- Esquema de metadatos genérico (p. ej. Dublin Core)
- Esquema de metadatos automático de Windows (p.ej. de Word, Excel)
- Esquema de metadatos temático especializado, indica:
- Otro esquema de metadatos, indica:

*Indica cómo se crearán (en un fichero “readme”, una hoja de cálculo, incrustadas en los datos) y qué documentación elaborarás para que los datos sean comprensibles por otros*

## 4. Acceder, compartir y reutilizar los datos

---

**4.1 ¿Tienes restricciones para compartir los datos en relación con el reglamento vigente ([General Data Protection Regulation](#)) u otros (éticas, comerciales, de seguridad, propiedad intelectual o copyright)?**

*Indica cuáles.*

**4.2 ¿Cuáles son los usuarios potenciales de tus datos y cómo los van a descubrir?**

*Describe brevemente quién podría estar interesado en tu investigación y qué difusión harás (p.ej. datos en el repositorio, página web, publicaciones en congresos, etc.)*

**4.3 Especifica qué licencias vas a aplicar a los datos para permitir su reutilización al máximo**

*Se recomienda el uso de licencias Creative Commons (CC - BY o CC Zero) o GNU*



## 5. Depósito y preservación de los datos

---

*Ten en cuenta, también, que todas las tesis doctorales evaluadas favorablemente deben ser publicadas en los respectivos repositorios institucionales de acceso abierto.*

### 5.1 ¿Qué criterios utilizarás para seleccionar los datos a preservar a largo plazo?

- Tipo de datos (raw, procesados) y facilidad en la generación
- Relevancia del contenido por otros
- Facilidad de reutilización del formado por otros
- Datos vinculados a una publicación
- Verificación de la investigación
- Tiempo disponibles
- Recursos económicos disponibles
- Otros, indica:

### 5.2 ¿Cuánto tiempo preservarás los datos?

*Diferentes estándares internacionales recomiendan un mínimo de 10 años.*

### 5.3 ¿En qué repositorio depositarás tus datos?

*Ten en cuenta los [requerimientos](#) específicos en cuanto a formato, metadatos, tamaño, coste, etc. que pueden tener para depositar datos*

*El repositorio para depositar datos en la UdL es [CORA-Repositorio de Datos de Investigación](#). Es un repositorio de datos federado y multidisciplinar que permite a las universidades catalanas, centros de investigación CERCA y otras entidades que investigan la publicación de conjuntos de datos de investigación en modo FAIR y siguiendo las directrices del EOSC.*

- Repositorio institucional
- Repositorio temático (internacional), indica:
- Repositorio multidisciplinar (p. ex. Zenodo, Figshare, Dryad)
- Otros, indica:

## 6. Datos sensibles/personales

---

### 6.1 Trabajarás con datos personales?

*Un dato personal es el relativo a una persona física identificada o identificable. En este sentido, se pueden considerar datos como el nombre o DNI de una persona que permitirían identificarlo de forma directa, así como un conjunto de datos de una persona que, a pesar de no saber su nombre, permiten identificarla.*

*Por ejemplo, una dirección IP o una dirección de correo electrónico son consideradas datos personales o saber la fecha de nacimiento, etc.*

*En caso de trabajar con datos personales específicas con qué datos.*

- No
- Sí, con datos personales de menores de edad de menos de 14 años
- Sí, con datos personales de menores de edad entre 14 y 17 años
- Sí, con datos personales de mayores de edad

### 6.2 ¿Trabajarás con datos especialmente protegidos, como datos de salud, biométricas o genéticas, datos de origen racial o étnico, convicciones filosóficas o religiosas, ideas políticas u otros datos con riesgo elevado para las personas afectadas?

*En caso afirmativo, indica con qué datos.*

### 6.3 ¿En mi investigación es imprescindible identificar personas concretas?

- No, porque los datos personales son anónimos (véase apartado 6.4)
- Sí, (véase apartado 6.5)

#### **6.4 ¿Por qué los datos personales son anónimos?**

- Porque no se captan ni se guardan ningún identificador de las personas afectadas (direcciones de correo, números de IP de dispositivos conectados a internet, número de teléfonos móviles, nombres y apellidos o DNI, etc.
- Porque la participación en el proyecto sólo genera un código para cada participante, lo que hace imposible al equipo investigador de la UdL identificar su identidad.
- Porque en ningún caso se realizan grabaciones de la voz ni de la imagen de las personas afectadas.
- Porque de forma razonable y sin esfuerzos desproporcionados no es posible deducir, de los resultados de la ejecución de los Proyectos (contenido de las encuestas y entrevistas, contenido de las analíticas médicas, etc.), la identidad de las personas.
- Porque, aunque se capten o se guardan identificadores de las personas afectadas, se han implementado las medidas técnicas indispensables para evitar la reidentificación (“anonimización” a través del cifrado de los datos o encriptación).

#### **6.5 ¿Por qué los datos personales identificando personas concretas?**

- Porque se captan o guardan identificadores de las personas afectadas y no se han implementado medidas técnicas que eviten la reidentificación.
- Porque se realizan grabaciones de la voz o de la imagen de las personas afectadas.
- Porque de manera razonable y sin esfuerzos desproporcionados es posible deducir, de los resultados de la ejecución del Proyecto, la identidad de las personas.
- Porque se utilizan sistemas de pseudonimización de los datos personales.

#### **6.6 Confirma qué herramientas utilizarás para la gestión de los datos personales.**

*En caso de seleccionar la segunda opción es necesario contestar los puntos:*

*6.7 ¿Qué herramientas ajenas a la UdL utilizarás para la gestión de datos personales? ¿Por qué son imprescindibles?*

*6.8 Confirmando que a pesar de utilizar herramientas ajenas a la UdL para la gestión de datos personales, estos datos también los almacenaré en un entorno de la UdL.*

- Confirmando mi obligación de utilizar sólo herramientas de la UdL, o de las que la UdL ha adquirido la licencia, bajo el control de mi director de tesis.
- Confirmando que, a pesar de que conozco la obligación de usar sólo herramientas de la UdL, o de las que la UdL ha adquirido la licencia, es imprescindible para mi investigación utilizar otras adicionales y cuento con el visto bueno de mi director de tesis.

**6.7 ¿Qué herramientas ajenas a la UdL utilizarás para la gestión de datos personal? ¿Por qué son imprescindibles?**

*Por ejemplo, plataformas específicas para encuestas, como Google forms, EUSurvey, etc.*

**6.8 Confirmando que, a pesar de utilizar herramientas ajenas a la UdL para la gestión de datos personales, estos datos también los almacenaré en un entorno de la UdL.**

Confirmando que, en cualquier caso, los datos permanecerán almacenados en un entorno (preferentemente electrónico) de la UdL.