

EJEMPLO DE PREVISIÓN DEL PLAN DE GESTIÓN DE DATOS (PGD) PARA LA MEMORIA CIENTÍFICO-TÉCNICA DE LA PROPUESTA DE UN PROYECTO DE INVESTIGACIÓN

Se consideran datos de investigación a todo aquel material que ha sido generado, recolectado, observado o registrado durante el ciclo de vida de un proyecto de investigación y que se utilizan como evidencia de un proceso de investigación, están reconocidos por la comunidad científica y sirven para validar los resultados de la investigación y garantizar su reproducibilidad. Los datos de investigación pueden ser datos experimentales, observacionales, operativos, datos de terceros, del sector público, datos de seguimiento, datos brutos, procesados o datos reutilizados.

En la **memoria científico-técnica de la propuesta de un proyecto solamente se deberá incorporar una descripción inicial** que contenga qué datos se van a recoger o generar en el marco del proyecto (tipologías y formatos), cómo será el acceso a los mismos (quién, cómo y cuándo se podrá acceder a ellos), de quién son los datos y en qué repositorio está previsto su depósito, difusión y preservación. Asimismo, se recogerán, en su caso, las condiciones éticas o legales específicas que los regulen (ej. privacidad de los datos y su reglamentación; datos protegidos o protegibles por propiedad intelectual o industrial, etc.) que condicionen su disponibilidad, uso y/o reutilización. Esta declaración inicial en el momento de la solicitud no condicionará la creación de un Plan de Gestión de Datos formal en el caso de que el proyecto resulte finalmente financiado. El PGD se deberá presentar, si así es requerido, junto con los informes de seguimiento intermedio y final del proyecto financiado. Asimismo se recomienda la publicación en acceso abierto del PGD junto a los datasets utilizados, y en formato legible por máquina.

Este documento es un ejemplo que pretende servir de apoyo al personal investigador en el momento de crear la propuesta de previsión del Plan de Gestión de Datos-PGD (Data Management Plan, DMP) enmarcado en la memoria científico-técnica de la propuesta de un proyecto. Esta previsión hace referencia a una **descripción inicial que no suele superar los dos folios**, extensión que podría variar según la convocatoria en cuestión. Puede ajustarse a cualquier proyecto de investigación.



1. Descripción de los datos:

Los datos recogidos son de naturaleza biomédica. Serán cuantitativos, y se registrarán en hojas de datos tabulares (csv, xlsx) tras ser procesados por software (se debe especificar qué software se usará). Serán de interés para la comunidad científica internacional, y para los ciudadanos en general, por lo que los datos serán compartidos tras la finalización del proyecto.

2. Documentación de los datos:

Los datos serán documentados y descritos incluyendo la metodología en un fichero readme.txt que acompaña al dataset. Así, los datos generados durante el proyecto serán acompañados por metadatos estandarizados y estructurados explicando el propósito, origen, creador(es), condiciones de acceso y términos de uso de los datos.

Los metadatos producidos cumplirán con el estándar Dublin Core (para su almacenamiento en Zenodo y en RUO, y en el estándar que requiera el repositorio temático elegido) y se exportarán en formato .xml. Se publicarán bajo licencia CC Zero (Public Domain Dedication).

3. Almacenamiento, acceso y depósito de los datos:

Durante el proyecto, solo los miembros del equipo de investigación tendrán acceso a los datos. Dado que la recogida de datos se realizará durante un período de tiempo extenso (de dos años), y varios investigadores deberán tener acceso simultáneo y en tiempo real a dichos datos, los mismos se almacenarán en la nube de la Universidad de Oviedo en una carpeta encriptada. Se realizarán backups locales diarios, y se almacenará la información en dos discos duros externos de forma semanal. Se llevará a cabo el



pertinente control de calidad para garantizar la fiabilidad de los datos a lo largo de la investigación (calibración, repetición de mediciones, validación de entrada de datos, captura estandarizada, vocabularios controlados, etc.).

Una vez terminado el proyecto, los datos se anonimizarán antes de su publicación, de forma que no exista ninguna correlación potencial de los pacientes que han participado en el estudio clínico con los datos publicados. Para garantizar el cumplimiento de la Ley Orgánica de Protección de Datos, los ficheros con los contenidos sin anonimizar serán permanentemente eliminados, conservando solo los ficheros anonimizados.

Para garantizar la *preservación a largo plazo* de los datos generados durante la investigación, el dataset se almacenará en 3 repositorios: Zenodo (cuya persistencia está garantizada por el CERN), un repositorio temático de confianza (TDR, Trusted Digital Repository), y el repositorio institucional, RUO. El dataset recibirá un DOI, lo que garantiza su identificación persistente. Cualquiera de los repositorios elegidos asignará dicho DOI.

Para facilitar la localización del dataset, se asignarán palabras clave tomadas del tesauro MESH cuando se deposite en los repositorios.

Para garantizar la máxima legibilidad de los datos, se depositarán en formato CSV, soportado por la mayoría de software de análisis numérico y de datos.

El intercambio y reutilización quedan garantizados gracias a la adherencia a los estándares Dublin Core, al uso de vocabularios controlados (MESH) y a la utilización de estándares abiertos.

4. Procedimiento previsto para garantizar el cumplimiento de los requisitos éticos y/o legales de aplicación:

Las personas participantes en el estudio han firmado un consentimiento informado, al tratarse de un estudio que va más allá de lo meramente observacional, requiriéndose intervención para la recogida de datos clínicos.

El proyecto ha sido analizado por el [Comité de Ética de la Investigación de la Universidad de Oviedo](#), recibiendo el correspondiente informe favorable.

Los datasets se publicarán sin ninguna restricción de uso, bajo licencia CC-BY.

5. Titularidad y/o responsabilidad sobre los datos:

El responsable de la gestión de los datos será el IP del proyecto, quien garantizará que los datos cumplan con los principios FAIR. El responsable entregará un DMP en los primeros 6 meses del proyecto, y se asegurará de que sea revisado cada vez que sea necesario.

FUENTES:

Biblioteca de la Universidad de Cantabria. (2023). *Modelo básico de Plan de Gestión de Datos*.

https://web.unican.es/buc/Documents/Formacion/Guia_modelo_basico_PGD.pdf

Iglesias Gutiérrez, E., Amado Rodríguez, L., Fernández Fernández, A., Pinín Osorio, C., Fernández Pérez, R., & Fernández García, G. (2020). *Consejos y ejemplos para la elaboración de un Plan de Gestión de Datos (PGD) en el marco de la Convocatoria 2021 del Programa Estatal de Promoción del talento y su empleabilidad en I+D+i, Subprograma Estatal de Generación de Conocimiento, Proyectos de Investigación en Salud*. <https://www.ispasturias.es/plan-gestion-datos/>

Instituto de Salud Carlos III. (2023). *Preguntas Frecuentes AES 2023. Proyectos de desarrollo tecnológico en Salud (DTS23)*.

https://www.isciii.es/QueHacemos/Financiacion/Documents/DTS_2023/DTS23_PREGUNTAS_FRECUENTES.pdf

Milanés Guisado, Y. (2022). *Previsión del Plan de Gestión de Datos en la propuesta del proyecto. Plantilla*. Universidad Pablo de Olavide.

<http://hdl.handle.net/10433/15272>

Sánchez Almarcha, S., & Molina Micol, M. A. (2023). *Plantilla Readme.Txt*.

<http://hdl.handle.net/10201/128767>

