



GUÍA PARA ELABORAR UN PLAN DE GESTIÓN DE DATOS DE INVESTIGACIÓN

SERVICIO DE APOYO A LA INVESTIGACIÓN

VERSIÓN 01 - DICIEMBRE 2021

INTRODUCCIÓN

Esta guía ha sido elaborada por el Grupo de trabajo de gestión de datos de Investigación de la Biblioteca de la Universidad de Sevilla con el objetivo de servir de orientación a los investigadores que tengan que realizar sus Planes de gestión de datos de investigación.

La gestión de datos de investigación se enmarca en las estrategias de la ciencia abierta para facilitar la conservación y tratamiento de los datos en investigaciones científicas. Asimismo, tiene que ver con la obligación que establece la normativa de los proyectos de investigación que son financiados por el Marco Europeo, **Horizonte Europa** de la Comisión Europea; aquellos proyectos financiados por el Ministerio de Ciencia y Tecnología o cualquier otro proyecto de investigación cuya agencia financiadora así lo requiera.

La Unión Europea establece en la **Guidelines on Data Management in H2020 (2016)** que los datos de investigación deben cumplir los principios FAIR, es decir que los datos tienen que poder ser encontrados, deben estar accesibles, interoperables y reutilizables, se solicita también el depósito de los datos en un repositorio y su preservación.

Para elaborar el Plan de gestión de datos de investigación (Data Management Plan- DMP), se debe tener en cuenta el ciclo de vida de los datos que están vinculados a los proyectos de investigación, tanto durante el proyecto, como cuando este haya finalizado. Las actividades relacionadas con la gestión de los datos de investigación se resumen en las siguientes etapas: planificación, documentación, formato, almacenamiento, anonimización, control de acceso. En cada etapa de este ciclo de vida se requieren una serie de actividades y tareas que permitan preservar los datos y poder acceder a los mismos, así como su reutilización.

El Plan de gestión de datos es de un documento vivo que debe crearse al inicio del proyecto y actualizarse a medida que se va desarrollando el mismo o siempre que se produzcan modificaciones importantes, debidas a la inclusión de nuevos datos de investigación o a factores externos.

Este documento sirve para cualquier proyecto y está orientado a cualquier disciplina, aunque como hemos mencionado con anterioridad, los datos que se incluyen son de carácter orientativo.

Se ha consultado el **Código de Buenas Prácticas en Investigación de la Universidad de Sevilla**, documento que fue aprobado en 2017 y en el que se establecen las pautas de actuación en la realización de las actividades de investigación. Así como los documentos editados por la Comisión Europea, Horizonte 2020, Horizonte Europa y los que otras universidades han realizado al respecto y que se relacionan en la bibliografía final.

Los apartados de esta guía siguen la estructura de la **plantilla del PGD** incluida en el **Programa Guía de Horizon Europe (HORIZON)**

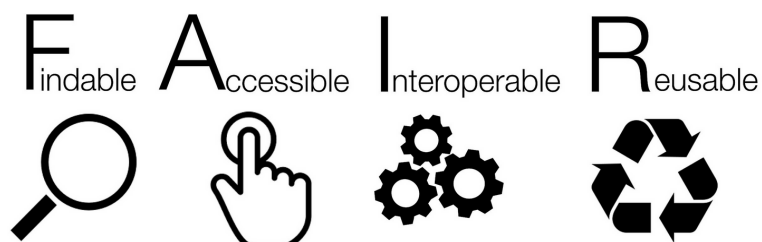
En la portada se deben incluir los datos administrativos del proyecto: número de referencia, acrónimo, título, fecha y versión del plan.

1. RESUMEN DE DATOS

En este apartado se incluirá la información relativa a los datos de investigación, procedencia, reutilización de datos existentes, propósito y relación con los objetivos del proyecto, tipos y formatos, tamaño y utilidad de estos fuera del proyecto.

- Descripción de la naturaleza, contenido y alcance de los datos de investigación.
- Finalidad, describir la relación con los objetivos del proyecto de investigación.
- Procedencia de los datos, tanto de los nuevos como de los reutilizados.
- Si se trata de reutilización de datos existentes, se debe especificar la fuente de la que se han extraído, explicar, en su caso, cómo se han tratado temas como los derechos de autor y la ley de propiedad intelectual, indicar posibles restricciones para su reutilización.
- Motivos por los que han descartado datos.
- Tipología de los datos generados: datos numéricos, textuales, imágenes, vídeo, audio, etc.
- Formatos utilizados: se recomienda utilizar en los posible formatos abiertos y no propietarios; datos codificados en ASCII o UTF-8; mejor CSV que Excel, mejor txt o PDF/A que Word o PPT, mejor TIFF que .psd, etc. Más información sobre formatos.
- Estructura de los datos: organización, jerarquización, nombres, etc.
- Control de las versiones: describir cómo se organizará el control de las versiones de los ficheros.
- Volumen del conjunto de datos (datasets) en las unidades de medida de los mismos Kb, Mb, TB, posible crecimiento gradual a lo largo del proyecto y el número de ficheros utilizados.
- Metodología o software que se va a emplear en la creación de los datos de investigación, si se trata de datos nuevos.
- Reutilización de los datos: usos previsibles de los datos para investigaciones futuras, para qué y para quien podrían ser útiles.

2. DATOS FAIR



2.1. Hacer que los datos se pueden encontrar, incluidas las disposiciones para metadatos

Información relativa a la facilidad para encontrar los datos, asignación de un identificador persistente, Handle, DOI, etc. Metadatos que se crearán, si son estándares en la disciplina o generales, palabras claves para facilitar la búsqueda.

- **Identificador:** asignación de un identificador estándar y permanente, DOI, Handle.
- **Estándares de metadatos:** describir los estándares que se usarán para describir los datos y como se utilizarán. Pueden ser esquemas generales (Dublin Core, Datacite) o esquemas de metadatos específicos para cada disciplina.
- **Palabras claves:** creación de palabras clave para optimizar la recuperación y reutilización de los datos.

2.2. Hacer que los datos sean accesibles

En este apartado se recoge la información relativa a la accesibilidad de los datos en cuanto al repositorio donde se van a depositar (institucional, temático, general) y el acceso a los propios datasets (acceso abierto, acceso restringido, embargo, comité de acceso a los datos), software necesario para acceder, tiempo en que permanecerán disponibles.

Depósito:

Indicar el repositorio donde se van al depositar los datos, asegurándose que dicho repositorio asigna un identificador persistente y que se han explorado adecuadamente cuáles son los requisitos del repositorio elegido.

Datos:

- Describir correctamente las condiciones de acceso definidas por el repositorio elegido (indicando el URL de donde proviene la información).
- Accesibilidad a los datos: describir si los datos estarán disponibles en acceso abierto.
- Embargos temporales: si los datos estarán embargados durante un tiempo, por futuras publicaciones, solicitud de patentes, etc., se especificarán los motivos y la duración del embargo.

- Restricciones sobre datos: indicar las razones, en el caso de que haya restricciones a terceros para la reutilización de datos ajenos generados durante el proyecto de investigación, siempre que los datos no estén en abierto.
- Si los datos no pueden ser abiertos totalmente o hay restricciones para la reutilización de algunos datos por razones éticas, de propiedad intelectual y/o industrial, intereses comerciales, etc. Debe justificarse el porqué.
- Preservación y compartición de los datos: qué datos deben ser retenidos o destruidos por razones contractuales o legales y en virtud de otros criterios basados en su reutilización para validar los resultados de la investigación, realizar nuevos estudios, aprovecharlos para la enseñanza, etc.
- Especificar si es necesario un comité de acceso a los datos, en caso de datos personales/sensibles.

Metadatos

- Plazos de conservación de datos: tiempo previsto para la preservación de los datos. Qué datos se mantendrán a largo plazo. Cuanto tiempo serán los datos reutilizables una vez finalizado el proyecto. Indicar si los metadatos permanecerán disponibles después de que los datos ya no lo estén.
- Métodos o softwares necesarios para acceder a los datos: requerimientos técnicos para acceder a los datos y reutilizarlos.
- Incluir la documentación del software que se necesita para acceder a los datos. En caso de que sea un software específico, incluirlo si es posible. Por ejemplo, el código fuente.

2.3. Hacer que los datos sean interoperables

Explicar qué formatos de datos y qué vocabularios de metadatos, estándares o metodologías se siguen para facilitar la interoperabilidad de los datos y permitir el intercambio y la reutilización de datos dentro y entre disciplinas.

2.4. Aumentar la reutilización de los datos

Este apartado recoge la información que facilita la reutilización de los datos, adjuntar archivos readme, asignación de licencias de reutilización estándar, validez de los datos para terceros, calidad de los datos y otros aspectos distintos de los datos como los relacionados con la asignación de recursos, la seguridad de los datos y los aspectos éticos.

- Crear un archivo Readme.txt que acompañará al resto de ficheros con los datos de investigación donde se incluirá una breve descripción del dataset, información de los autores, la metodología empleada, variables usadas, fechas de recogida de creación del dataset, etc.

- Asignación de licencias de reutilización bajo las cuáles se va a permitir la reutilización de los datos. Se recomienda la utilización de licencias ampliamente reconocidas para definir las condiciones de uso de los datos que deberán cumplir los investigadores que quieran reutilizar los datos. Por ejemplo, Licencias Creative Commons. La Comisión Europea propone la asignación de licencias CC-BY o CC0.

- Indicar la utilidad de los datos para otras investigaciones, en particular cuando termine el proyecto.

- Control de calidad de los datos: descripción de los procesos para asegurar la fiabilidad de los datos a lo largo de la investigación. Especificar cómo y cuándo se llevarán a cabo las evaluaciones internas

3. OTROS RESULTADOS DE LA INVESTIGACIÓN

Además de la gestión de los datos, los beneficiarios también deben considerar y planificar la gestión de otros productos de investigación que puedan generarse o reutilizarse a lo largo de sus proyectos.

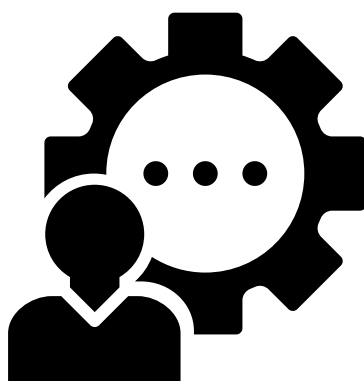
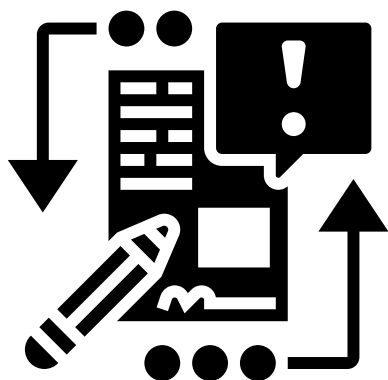
Dichas salidas pueden ser digitales (por ejemplo, software, flujos de trabajo, protocolos, modelos, etc.) o físicas (por ejemplo, nuevos materiales, anticuerpos, reactivos, muestras, etc.).

Los beneficiarios deben considerar cuál de las preguntas relativas a los datos FAIR anteriores puede aplicarse a la gestión de otros resultados de la investigación, y deben esforzarse por proporcionar suficientes detalles sobre cómo se gestionarán y compartirán sus resultados de investigación, o se pondrán a disposición para su reutilización, de conformidad con los principios FAIR

4. ASIGNACIÓN DE RECURSOS

Estimación del coste para hacer que los datos sean FAIR (Findable, Accesible, Interoperable, Reusable) en el proyecto.

- Funciones y responsables: descripción de los diferentes roles y responsabilidades de las diferentes actividades relacionadas con la gestión de los datos: captura de datos, introducción de metadatos, etc. En el caso de proyectos en los que participen varias organizaciones, explicar las responsabilidades entre los diferentes participantes en el proyecto. Determinar quién es el responsable del PGD.
- Recursos temporales y humanos: estimar la previsión de personal adicional especial para la gestión de los datos, así como los costes en tiempo necesarios.
- Recursos tecnológicos: previsión de software, hardware adicional y especial para la gestión de los datos. Programas de acceso: Indicar si se necesitarán herramientas específicas para el acceso y/o reutilización de los datos. Se tendrá en cuenta que el software pueda ser utilizado en el futuro y no esté obsoleto, por lo que se recomienda la utilización de formatos abiertos.
- Recursos económicos: previsión de costes adicionales especiales para la gestión de los datos. Previsión de costes del repositorio y de las tareas de preparación de los datos para su conservación a largo plazo. Indicar si serán necesarios nuevos recursos y, en caso afirmativo, indicar qué se necesitará y cómo se cubrirán los costes.



5. SEGURIDAD DE LOS DATOS

Medidas adoptadas respecto a la seguridad de los datos. Incluye la recuperación de los datos, la transferencia de los datos confidenciales, el almacenaje y las copias de seguridad, preservación a largo plazo y mantenimiento de datos de forma que queden protegidos y sean seguros.

Como se recoge en el art. 11 del Código de Buenas Prácticas en Investigación de la Universidad de Sevilla "Todos los datos, debidamente identificados, deberán conservarse durante un período mínimo de diez años, como recomiendan las directivas europeas (excepto en los casos en que se haya acordado o convenga un período superior que no contradiga las normas nacionales y europeas). En el caso de los datos almacenados en soporte electrónico, deberá disponerse de un sistema de copias de seguridad".

- Almacenamiento: procedimientos para almacenar los datos y metadatos durante el proceso de investigación. Capacidad, infraestructura y medios disponibles. Se recomienda la utilización de los servicios de almacenamiento de la US (disco virtual y Onedrive). Se desaconseja el almacenamiento en dispositivos locales.
- Copias de seguridad: realización y mantenimiento de estas. Número de copias, frecuencia, soportes, etc.
- Responsabilidad: personas responsables de la gestión de los datos.
- Gestión de riesgos: medidas previstas para la protección y gestión de riesgos potenciales. Protección de datos sensibles-
- Restauración de datos: recuperación de los datos en caso de pérdida accidental.
- Acceso y control de acceso a los datos: control y gestión del acceso y/o la transferencia de los datos, para mantener su seguridad y confidencialidad.



6. ÉTICA

Problemas éticos o legales que puedan surgir durante la recogida, almacenaje, procesamiento y archivo de datos. Necesidad de incluir el consentimiento informado para el intercambio de datos y la conservación a largo plazo de los cuestionarios que tratan con datos personales.

Según consta en el art. 7 del Código de Buenas Prácticas en Investigación de la Universidad de Sevilla "Si la investigación implica directamente a personas, material o datos de origen humano o animales de experimentación, el documento se someterá a la evaluación previa del Comité de Ética de la Investigación, o al Comité Ético de Experimentación de la US (CE) respectivamente, o al órgano que asuma dichas funciones".

Hay que tener en cuenta las instrucciones para el **Tratamiento de Datos Personales en Actividades de Investigación** de la Universidad de Sevilla.

Uso de datos personales: acciones sobre el cumplimiento de la normativa europea (**Reglamento UE 2016/679**) y nacional (**Ley Orgánica 3/2018**) sobre protección de datos personales.

- Accesos autorizados: procedimiento para el acceso a los datos personales por parte de usuarios autorizados. Servicio protección datos US.

- Propiedad intelectual de los datos: titulares de la propiedad intelectual de los datos. Directrices institucionales al respecto. Indicar si afecta el derecho sui generis recogido en la Ley de Propiedad Intelectual

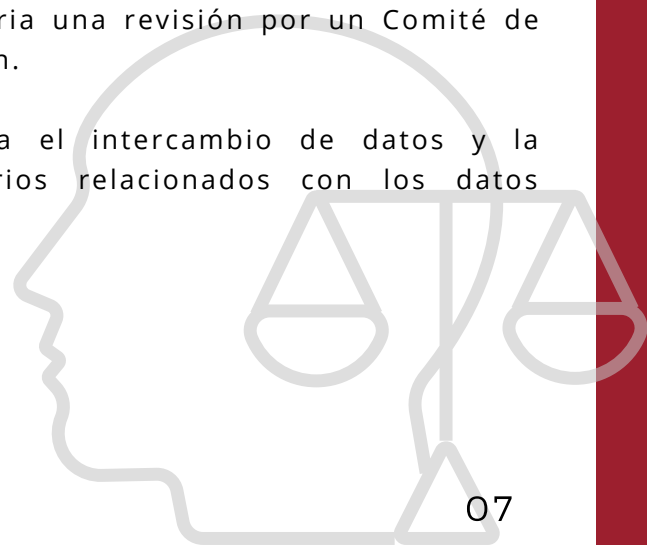
- Titularidad de los datos: quién tendrá derecho sobre el control de acceso a los datos.

- Códigos éticos: códigos de conducta que se deben seguir internacionales, nacionales y el de la Universidad de Sevilla.

- Preservación y compartición de datos: procedimientos para conservar y compartir datos. Anonización y seguridad en el almacenamiento y transferencia de la información.

- Reutilización potencial: comprobar si es necesaria una revisión por un Comité de Ética para utilizar los datos durante la investigación.

- Se incluye el consentimiento informado para el intercambio de datos y la conservación a largo plazo de los cuestionarios relacionados con los datos personales.



7. OTRAS CUESTIONES

Indicar si se harán uso de otros procedimientos nacionales/ financiadores/sectoriales/departamentales para la gestión de datos.

8. BIBLIOGRAFÍA

Comisión Europea. Programa H2020 (2013). Directrices para la gestión de datos en Horizonte 2020. Recuperado el día 20 de junio de 2021.

Comisión Europea. H2020. (2016). Guidelines on FAIR Data Management in Horizon 2020. Recuperado el día 21 de junio de 2021.

Comisión Europea. Horizon Europe (HORIZON) Programme Guide. Version 1.2. 04 October 2021 https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf

Consorci de Serveis Universitaris de Catalunya. (2019). Data Management Planz Recuperado de <https://www.csuc.cat/es/servicios/gestion-de-datos-de-investigacion>

CSIC. Red de Bibliotecas sobre datos de investigación en abierto. (2018) Recomendaciones para la gestión de datos de investigación en abierto dirigida a investigadores, recuperado de <https://digital.csic.es/bitstream/10261/173801/1/Maredata-recomendaciones-ESP.pdf>

REBIUN. Difusión de datos de investigación, Recuperado el día 2 de julio de 2021 de <https://www.rebiun.org/difusion-datos-investigacion>

Universidad de Cantabria. (2019) Modelo básico de Plan de Gestión de datos, recuperado el día 23 de junio de 2021 de https://web.unican.es/buc/Documents/Formacion/Guia_modelo_basico_PGD.pdf

Universidad de Granada. (2020). Breve guía de elaboración de un plan de gestión de datos, recuperado el día 22 de junio de 2021

Universidad Politécnica de Valencia. (2021). Recuperado el 24 de junio de 2021 de <https://riunet.upv.es/handle/10251/143184>

Universidad de Utrech. (2021). Recuperado el día 14 de junio de 2021 <https://www.uu.nl/en/research/research-data-management/guides>

MÁS INFORMACIÓN

SERVICIO DE APOYO A LA INVESTIGACIÓN



BIBINVESTIGACION@US.ES



954551130