Biblioteca de la Universidad de Sevilla
INVESTIGA
monográficos sobre investigación

Nº 36 (2022) Cómo abordar un plan de gestión de datos

En este número:

- Introducción
- El marco de la Ciencia Abierta
- Los Datos y las agencias financiadoras
- Los Principios *FAIR*
- Qué es un plan de gestión de datos y cómo elaborarlo.

Introducción

El último monográfico de investigación (Investiga, nº 35) ya lo dedicamos a exponer el panorama científico actual, fruto del movimiento de la Ciencia Abierta y, más concretamente, a las recomendaciones (y en ocasiones, requerimientos) que derivan de ella en cuanto a la disponibilidad de los datos de investigación. Por ello, hablamos de los tipos de datos, de sus formatos, de las herramientas para su localización, etc., y de la posibilidad de depositarlos en nuestro repositorio institucional idUS.

Ahora, en este monográfico, mostraremos algunas pautas o recomendaciones necesarias para la elaboración de un Plan de Gestión de Datos (PGD), así como la organización de su contenido, los beneficios que podrán obtenerse para la comunidad investigadora en su conjunto, etc.

No obstante, para conocer de manera pormenorizada todo lo concerniente a los datos desde que nace un proyecto de investigación hasta que éste llega a su culminación, la Biblioteca ha

creado un espacio en su web con la información necesaria para ofrecer a la comunidad investigadora soporte a la hora de abordar un Plan de Gestión de Datos que acompañe a un proyecto de investigación.

Ciclo de vida de los Datos de Investigación (Universidad de Tasmania, adapatado deUK Data Service Model 2017).

El marco de la Ciencia Abierta

El concepto de **Ciencia Abierta** proviene de un movimiento surgido en las comunidades científicas a nivel internacional. Se caracteriza por la apertura, no solo de las publicaciones (Open Access), sino también de los datos de investigación, las metodologías, los procesos y la implicación de la ciudadanía en un entorno de investigación e innovación responsables.

Componentes de la Ciencia Abierta según la UNESCO.

En definitiva, **se trata de hacer libremente accesible todas las facetas del ciclo de vida de la investigación financiada con fondos públicos** y, no sólo para la comunidad científica que los produce sino también para la sociedad que los financia, potenciando la reproducibilidad de la Ciencia y la reutilización de los resultados.

Instituciones, editores y agencias financiadoras van adaptando sus estructuras, servicios y requerimientos a este nuevo paradigma, y comienza a intensificarse el debate sobre la necesidad de redefinir un nuevo modelo de evaluación de la ciencia.

En este sentido empezamos a ver los primeros movimientos en **ANECA** plasmados en sus Principios y Directrices publicados en octubre de 2021; la valoración del depósito de las publicaciones en repositorios institucionales y la apertura de los datos vinculados a las publicaciones.

DIRECTRICES DE INVESTIGACIÓN – ESTÁNDARES E INDICADORES PARA LA EVALUACIÓN CUALITATIVA DE LA PRODUCCIÓN CIENTÍFICA

DI13. Se recomiendo que tanto las aportaciones presentadas para obtener el sexenio de investigación como las 4 aportaciones relevantes que deben identificarse en ACADEMIA estén depositadas en un repositorio con *Green Open Access* como los repositorios institucionales de los que disponen las universidades. En el futuro será obligatorio, salvo impedimento legal o imposibilidad técnica, este depósito para todas las aportaciones mencionadas.

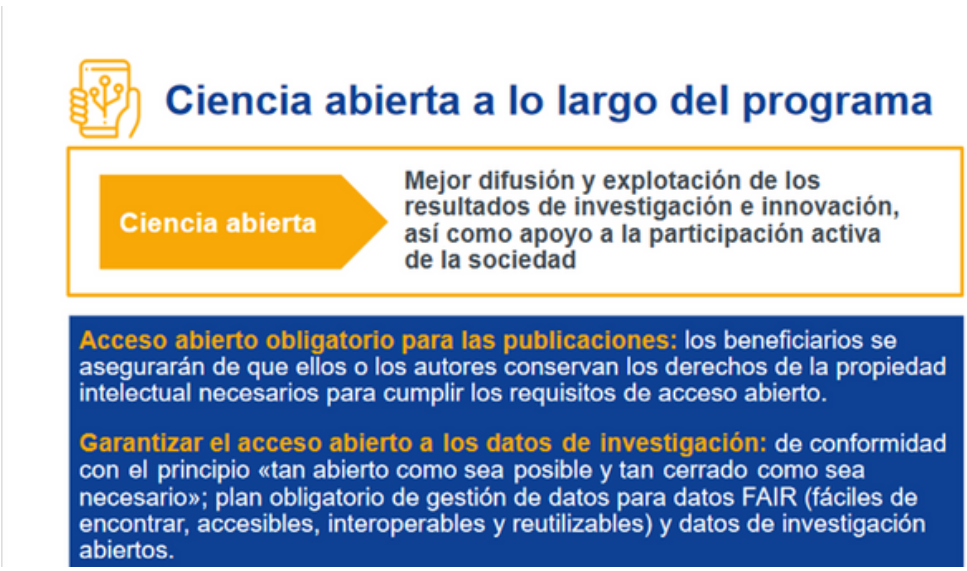
DI20. Se podrá valorar, dependiendo del campo, la apertura de datos vinculados a una publicación dentro de la revista.

Los datos y las agencias financiadoras

En este nuevo ecosistema, las agencias financiadoras toman un papel relevante en el impulso de la Ciencia Abierta y, por tanto, de la apertura de las fases de la investigación no sólo obligando a compartir resultados y datos, sino creando infraestructuras como la Nube Europea de Ciencia Abierta (EOSC) y exigiendo una adecuada planificación de la gestión de los datos de investigación que facilite su posterior accesibilidad.

- **Horizonte Europa** es el programa marco de investigación e innovación (I+I) de la Unión Europea (UE) para el período 2021 -2027. Como su predecesor Horizonte 2020, será el instrumento fundamental para llevar a cabo las políticas de I+D+I de la UE y su financiación. En este programa se promueve y fomenta la Ciencia Abierta exigiendo a los proyectos financiados con los fondos del programa la consecución de las siguientes acciones:

- El Acceso Abierto obligatorio para las publicaciones.
- Garantizar del mismo modo el acceso abierto a los datos obtenidos en la investigación de conformidad con el principio “tan abierto como sea posible, tan cerrado como sea necesario” a través de la elaboración de un Plan de Gestión de Datos que facilite el cumplimiento de los principios FAIR (fáciles de encontrar, accesibles, interoperables y reutilizables) de los datos de investigación que contenga.



Estos requerimientos también se incluyen en las convocatorias de la Agencia Estatal de Investigación para la financiación de proyectos de la investigación.

→ En las solicitudes de financiación del **PEICTI (Plan Estatal de Investigación Científica, Técnica y de Innovación, 2021-2023)** se exige entregar con la Memoria del proyecto un Plan de Previsión de los datos que contenga la información siguiente:

- 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.

Es importante aclarar que el **Plan de Gestión de Datos** es un documento exigido por las agencias financiadoras sólo **en el caso de aprobar el proyecto de investigación**. Sin embargo, sí es obligatorio presentar un **Plan de Previsión de los Datos** a la hora de solicitar las ayudas a estas entidades.

Los Principios FAIR

Como ya hemos señalado, Horizonte Europa establece una serie de principios que regulen las características de los datos de investigación para que éstos puedan depositarse en abierto siguiendo la premisa **“tan abierto como sea posible, tan cerrado como sea necesario”.**

Estos principios, denominados FAIR (acrónimo de Findable, Accessible, Interoperable y Reusable) describen cómo deben organizarse los resultados de la investigación para que se sean:

- **Localizables:** se incluyen los identificadores, palabras clave, normas de metadatos que los describen para optimizar su localización.
- **Accesibles:** se indica el repositorio dónde se depositan los datos y los metadatos en acceso abierto y se justifica si hay datos cerrados.
- **Interoperables:** se exponen los vocabularios de datos, las normas, formatos o metodologías que se utilizarán para permitir el intercambio y la interoperabilidad de los datos.
- **Reutilizables:** Publicar datos y metadatos con una licencia clara y accesible sobre su uso y reutilización; asociarlos con información sobre su procedencia y cumplir con los estándares relevantes que usa la comunidad de ese dominio concreto.

Para comprobar si los datos empleados y derivados del proyecto de investigación cumplen con estos principios FAIR, existen herramientas de ayuda para su verificación.

Qué es un Plan de Gestión de Datos y cómo elaborarlo

El Plan de Gestión de Datos (PGD) –Data Management Plan (DMP) es un documento formal que establece desde el inicio cómo se van a gestionar los datos en un proyecto de investigación. En él se describe el ciclo de vida de los datos recopilados, generados y procesados durante todo el proyecto.

El PGD define la metodología de extracción de la información, la unidad de medida, los procesos de creación, limpieza y análisis, el formato, el software de lectura y procesamiento, las licencias y permisos de acceso y (re)utilización, su sistema de control de versiones, su autoría, su difusión, y su lugar de almacenamiento y preservación a largo plazo.

Se creará al principio del proyecto de investigación financiado y dependiendo de la naturaleza y evolución del proyecto, posiblemente variará, se completará y/o modificará a lo largo del ciclo de vida de la investigación. No es un documento definitivo, sino que se va modificando durante el proyecto.

❑ Qué beneficios se obtienen con la Gestión de Datos

Es importante ser conscientes de los beneficios que emanan de una óptima gestión de los datos, tanto para el grupo de investigación que los crea y deposita, como para aquellos que tienen la oportunidad de reutilizarlos.

En líneas generales, los principales beneficios obtenidos por la gestión de los datos son las siguientes:

- **Ahorrar tiempo y hacer un uso eficiente de los recursos disponibles:** encontrar, comprender y utilizar los datos en el momento que se necesite.
- **Facilitar la reutilización de los datos** que se han generado o recopilado, la correcta gestión y documentación de los datos permitirá que estos se mantengan exactos, completos y fiables.
- **Cumplir con las exigencias de agencias financiadoras:** cada vez son más las agencias que exigen la presentación de planes de gestión de datos y/o el depósito de datos en repositorios.
- **Asegurar que los datos sean FAIR:** localizables, accesibles, interoperables y reutilizables: datos FAIR (findable, accessible, interoperable and reusable).
- **Proteger y preservar los datos:** al gestionar y depositar los datos en repositorios apropiados se preservan de forma segura a lo largo del tiempo.

❑ Qué contiene un Plan de Gestión de Datos

La elaboración de un Plan de Gestión de Datos consiste en una recopilación y organización de los datos empleados y generados a lo largo del desarrollo de un proyecto de investigación, de una manera sistemática, para que su almacenamiento y su reutilización en posteriores investigaciones sea óptima.

Para ello, es imprescindible que en la elaboración del Plan se lleven a cabo las siguientes acciones:

- **Organizar y describir los datos:** el PGD debe describir qué tipos de datos, formatos, normas y métodos de captura se utilizarán.
- **Gestionar los requisitos legales y éticos:** cómo se protegerán los datos sensibles y si se restringirá su acceso y uso.
- **Almacenar los datos:** cómo se almacenarán los datos a corto y largo plazo.
- **Compartir los datos:** cómo se compartirán los datos y qué posibilidades de reutilización existen.
- **Gestionar los recursos de personal y la formación:** quién será el responsable de la gestión de los datos, qué recursos necesita y si necesita formación.

Fuente:
https://library.leeds.ac.uk/info/14062/research_data_management/62/data_management_planning.

A la hora de abordar la elaboración del Plan de Gestión de Datos hay que tener en cuenta que hay que hacer una primera versión de éste (Plan de Previsión de Datos que ya hemos mencionado anteriormente) dentro de los primeros seis meses del desarrollo de la investigación y, en virtud de la evolución del proyecto, debe actualizarse al menos dos veces: a mitad del proyecto y a la finalización del mismo.

La elaboración del PDG implica la especificación de los siguientes aspectos:

1. **Resumen de los datos:** se debe describir de forma general los datos que se utilizarán en el proyecto. Incluye el tipo (textuales, numéricos...) y el formato (cvs, PDF...), la finalidad, el tamaño y el origen de los datos.
2. **Comprobar que los datos cumplan con los Principios FAIR**
3. **Otros resultados de la investigación:** se consideran qué cuestiones relativas a los datos FAIR pueden aplicarse a la gestión de otros productos de la investigación generados o reutilizados en el proyecto como, por ejemplo: softwares, flujos de trabajo, protocolos, nuevos materiales, muestras, etc.
4. **Asignación de recursos:** se indican cuáles serán los costes para hacer que los datos sean FAIR (por ejemplo, costes directos e indirectos relacionados con el almacenamiento, el archivo, la reutilización, la seguridad, etc.) y quién será responsable de la gestión de los datos.
5. **Seguridad de los datos:** se garantiza que los datos se almacenan de forma segura en repositorios de confianza para su preservación y conservación a largo plazo.
6. **Ética:** se presentan todas las cuestiones éticas o legales que puedan tener un impacto en el intercambio de datos. Además, cuando la investigación utilice datos personales, deberá hacerse referencia a aspectos como el consentimiento informado y la conservación a largo plazo.
7. **Otras cuestiones:** se describe si se van a utilizar, otros procedimientos (nacionales, de proveedores, sectoriales, departamentales) para la gestión de datos.

En la Guía para elaborar un Plan de Gestión de Datos de Investigación realizada por la Biblioteca puedes obtener más información sobre estas cuestiones.

❑ Cómo elaborar el PGD

Una vez que se han tenido en consideración los aspectos descritos hasta ahora, debemos almacenar los datos que se están utilizando en el proyecto de investigación de una forma adecuada y segura teniendo en cuenta los aspectos siguientes:

→ Facilidad de **acceso** para los miembros autorizados, al tiempo que se restringe a los no autorizados.
→ Sistema de copias de seguridad compatible con el sistema de almacenaje.
→ La **seguridad** de los datos, sobre todo si se está trabajando con datos personales y/o confidenciales.
→ Problemas de seguridad relacionados con el envío de datos a través de correo electrónico y la compartición de copias.
→ Las especificaciones, si existen, de la agencia financiadora.
→ **Almacenar** los datos en: PC/ordenador portátil, almacenamiento óptico (CDs, DVDs), unidades externas (USB, discos duros), sistemas de almacenamiento institucional (Disco Virtual y One Drive) y otros servicios en la nube (Google Drive).
→ Utilizar **contraseñas** robustas y cambiarlas de manera periódica para proteger los archivos.

Existe aplicaciones informáticas que facilitan la elaboración de un PGD, entre ellas cabe destacar: Pagoda, DMP online, Pla de Gestió de Dades de Recerca y Argos.

Para hacer un uso correcto de ellas y elaborar el PGD de manera óptima también existen guías o directrices entre las que cabe destacar la elaborada en 2021 por el programa **Horizonte Europa** (en español) que puedes descargar desde aquí.

❑ Beneficios de la gestión y apertura de datos

Por último, animamos desde aquí a la comunidad investigadora a conocer y tomar conciencia sobre los beneficios que derivan de una correcta planificación de la gestión de los datos, tanto para facilitar el desarrollo de la investigación como para su posterior apertura siguiendo las recomendaciones de la Ciencia Abierta:

- Validación de los resultados obtenidos.
- Localización y comprensión de los datos.
- Reducción de la duplicidad de recogida de datos y los costes que ello comporta.
- Cumplimiento de los requisitos de las convocatorias de investigación.
- Fomento del debate científico.
- Fomento de la innovación y los nuevos usos potenciales de los datos.
- Fomento de la colaboración entre los usuarios de datos y sus creadores.
- Incremento del impacto y la visibilidad de la investigación.
- Aumento de la propia reputación cuando los demás citen la obra propia.

Por todo ello, instamos a la comunidad investigadora a tomar conciencia de la importancia de la gestión de los datos, de su almacenamiento y difusión; ofreciendo un asesoramiento desde cada una de las bibliotecas de cada centro y también poniendo a disposición de todos el material informativo que la Biblioteca ha elaborado para tal fin.

Novedades y recursos-e del mes

La Biblioteca de la Universidad de Sevilla obtiene el Sello de Excelencia Europea EFQM 600

El Club Excelencia en Gestión reconoce así el progreso y los avances positivos de la BUS

PRISMA integra, actualiza, mide tu investigación

Grandes ventajas para el PDI de la US y la gestión de la investigación.

Campaña ANECA 2022

Información, formación y asesoramiento desde la Biblioteca.

Actualizado SJR 2021 en PRISMA

El índice de impacto y posición (cuartil, tercil y decil) de SJR 2021 se aplica a las publicaciones de ese año y posteriores.

Formación en línea de CAS SciFinder-n

Conoce todas las funcionalidades de esta plataforma con estas sesiones de formación.

PRISMA integra, actualiza, mide tu investigación

Campaña de presentación de PRISMA en los diferentes campus y facultades, para mostrar sus ventajas y recoger sugerencias

Acuerdos transformativos en la Universidad de Sevilla: informe 2021

Más de 270 artículos de la US en abierto financiados con estos acuerdos.

La Biblioteca presenta su nueva Carta de Servicios de Apoyo a la Investigación

Recoge los servicios que se prestan, así como los compromisos que adquirimos para su cumplimiento.