

BREVE GUÍA PARA LA ELABORACIÓN DE PLANES DE GESTIÓN DE DATOS DE INVESTIGACIÓN



UNIVERSIDAD
DE GRANADA



www.ibsgranada.es



	REALIZADO	REVISADO	APROBADO
FECHA	Septiembre 2024	Octubre 2024	Noviembre 2024
NOMBRE	Francisco Luis Andújar Vera	Equipo de Direccion Comité Científico Interno	Equipo de Direccion
CARGO	Responsable Plataforma Bioinformática		

CONTROL DE EDICIONES		
VERSION	FECHA	TIPO DE REVISION
01	Septiembre 2024	Edición inicial



Contenido

1. INTRODUCCIÓN	4
2. CIENCIA ABIERTA.....	4
3. MANDATOS Y POLÍTICAS.....	5
4. GESTIÓN DE LOS DATOS.....	6
5. PRINCIPIOS FAIR	7
6. PLANES DE GESTIÓN DE DATOS (PGD)	8
7. GENERADORES Y PLANTILLAS	9
8. SECCIONES DEL PLAN DE GESTIÓN DE DATOS	10
8.1. Aspectos administrativos: Descripción del proyecto y principales protagonistas.....	10
8.2. Descripción de los datos y recogida o reutilización de los datos existentes.	10
8.3. Documentación y calidad de los datos.....	11
8.4. Almacenamiento y copia de seguridad durante el proceso de investigación.....	12
8.5. Requisitos legales y éticos, códigos de conducta.....	13
8.6. Intercambio de datos y conservación a largo plazo	14
8.7. Responsabilidades y recursos para la gestión de los datos	16
9. DOCUMENTACIÓN RELACIONADA	16



1. INTRODUCCIÓN

En el contexto actual de la investigación científica, la gestión adecuada de los datos es fundamental para asegurar la integridad, la reproducibilidad y la accesibilidad del conocimiento. Conscientes de esta necesidad, el Instituto de Investigación Biosanitaria ibs.GRANADA ha redactado esta guía breve destinada a la elaboración de los Planes de Gestión de Datos de investigación (PGD). Este documento tiene como objetivo principal orientar a nuestros investigadores, y a todo aquel que quiera consultarla, en el desarrollo de estrategias efectivas para manejar sus datos a lo largo de todo el ciclo de investigación, desde la planificación y recopilación hasta la publicación y preservación.

La importancia de esta breve guía radica en su capacidad para fomentar buenas prácticas en la gestión de datos, lo cual no solo beneficia a los propios investigadores, sino que también contribuye al avance colectivo de la ciencia. Al proporcionar un marco claro y accesible, ibs.GRANADA reafirma su compromiso con la excelencia investigadora y la transparencia, promoviendo un entorno en el que los datos se gestionen de manera responsable y ética.

Entre los aspectos que los investigadores encontrarán en esta guía se incluyen directrices sobre la organización de los datos, almacenamiento, seguridad y preservación de los mismos de acuerdo con los **principios FAIR** (*Findable, Accesible, Interoperable y Reusable*) y garantizado el acceso abierto a los datos de investigación siguiendo el lema: “*tan abierto como sea posible, tan cerrado como sea necesario*”.

Confiamos en que esta guía sirva como una herramienta de consulta para contribuir a la consolidación de ecosistemas de investigación más eficientes y colaborativos.

2. CIENCIA ABIERTA

Para situar el presente documento en el contexto adecuado, es importante hacer una reseña sobre un concepto que se irá repitiendo a lo largo del mismo y que es base fundamental para entender la necesidad de elaborar un PGD y depositar los datos en abierto de manera correcta, se trata de la *Ciencia Abierta*.

Aunque el concepto en sí no es reciente y probablemente surgió en el momento en que se reconoció que el acceso limitado a los resultados de la investigación obstaculizaba el progreso científico, no fue hasta la implementación del *Acceso Abierto* y, más recientemente, con el *Plan Piloto de Datos de Investigación en Abierto de la Comisión Europea* en 2016, cuando se formalizó el concepto de **Ciencia Abierta**. Este concepto, no solo abarca el acceso libre a las publicaciones, sino también a los datos, metodologías y otros recursos, reconociendo así la importancia de los datos como fuente de nuevo conocimiento.

Basándonos en la definición que la UNESCO proporciona en el documento sobre *Recomendaciones de la UNESCO sobre la Ciencia Abierta*, podemos entenderla como un enfoque inclusivo que integra diversos movimientos y prácticas para que el conocimiento científico sea accesible y reutilizable para todos. Promueve la colaboración científica y el intercambio de información, beneficiando tanto a la ciencia como a la sociedad. Este enfoque aboga por la apertura de los procesos de creación, evaluación y comunicación del conocimiento científico a



actores sociales más allá de la comunidad científica tradicional. Abarca todas las disciplinas científicas y aspectos de las prácticas académicas, incluyendo ciencias básicas y aplicadas, ciencias naturales y sociales, y humanidades. Se fundamenta en pilares clave como el conocimiento científico abierto, infraestructuras de ciencia abierta, comunicación científica, participación abierta de actores sociales y diálogo con otros sistemas de conocimiento.



<https://www.rebiun.org/>

3. MANDATOS Y POLÍTICAS

En respuesta a la necesidad de garantizar la integridad, reproducibilidad y accesibilidad del conocimiento, dentro del contexto de la gestión adecuada de los datos, numerosos mandatos y políticas se han ido implementando sucesivamente a nivel global, abogando por la importancia de documentar la gestión de los datos de investigación y la promoción de la ciencia abierta.

Estas políticas surgen como un reconocimiento del valor de los datos y su beneficio a la hora de ser compartidos entre la comunidad científica y la sociedad en su conjunto. Instituciones como la Comisión Europea, a través de programas como Horizonte 2020 y Horizonte Europa, han establecido requisitos claros que obligan a los investigadores a elaborar PGD, entendiéndolos como documentos estratégicos que definen el manejo, compartición y preservación de los datos generados durante un proyecto de investigación y una vez finalizado éste.

Organizaciones como OCDE, ONU o UNESCO también han promovido principios y directrices para la ciencia abierta, enfatizando la importancia de la transparencia, la colaboración y el acceso a los datos como pilares fundamentales para el avance científico. Estas iniciativas buscan



fomentar un entorno en el que los datos sean considerados recursos valiosos, propiciando su intercambio y reutilización entre investigadores, instituciones y el público.

A nivel nacional, se han establecido diversas iniciativas que promueven la ciencia abierta a través de un marco normativo y estratégico que reflejan el compromiso de España al entorno colaborativo y accesible de la ciencia. La Ley de la Ciencia, la Tecnología y la Innovación (Ley 17/2022) garantiza el acceso abierto a publicaciones y datos de investigación financiados con fondos públicos, fomentando la transparencia y la colaboración científica. El Plan Estatal de Investigación Científica, Técnica y de Innovación (PEICTI) 2021-2022 respalda la gestión de datos mediante la recomendación de elaborar un PGD y el cumplimiento de los principios FAIR. Además, la Estrategia Española de Ciencia, Tecnología e Innovación 2021-2023 establece objetivos para mejorar infraestructuras digitales y promover la colaboración internacional, incluyendo la obligatoriedad de depositar datos en acceso abierto y presentar PGD para proyectos financiados con fondos públicos.

En este contexto, la implementación de los PGD no solo responde a exigencias regulatorias, sino que también refleja un compromiso con las buenas prácticas en la investigación. Al establecer directrices claras y accesibles, se contribuye a una cultura de responsabilidad y ética en el manejo de los datos, que a su vez fortalece la confianza pública en la ciencia y sus resultados. Así, los mandatos y políticas que promueven la ciencia abierta y la gestión de datos se presentan como herramientas esenciales para impulsar un futuro de investigación más inclusivo, transparente y colaborativo.

4. GESTIÓN DE LOS DATOS

Los datos constituyen la base de la que se deriva el conocimiento, por lo que, para la producción de unos resultados de investigación originales, es necesario analizar unos datos que han sido recolectados, observados o creados durante el desarrollo de una investigación.

A partir del momento en el que los datos se incluyen en una investigación, tiene que haber una **gestión de los datos de investigación** (GDI), es decir, el conjunto de buenas prácticas en relación a la recopilación, organización, almacenamiento, seguridad, documentación, preservación e intercambio de los datos. Todos estos aspectos estarán encuadrados en los diferentes apartados que forman el ciclo de vida de los datos y que ocurren antes del proyecto, con una planificación inicial, así como durante y una vez finalizado el proyecto. Además, durante el propio ciclo de vida de los datos, se implicarán a una serie de elementos necesarios para llevar a cabo esta gestión, tales como servicios, herramientas o infraestructuras. Y toda esta gestión que incluye la descripción del ciclo de vida de nuestros datos, así como las decisiones tomadas o que se tomarán en cada apartado, es lo que quedará recogido en el documento formal llamado Plan de Gestión de Datos.



<https://www.rebiun.org/>

5. PRINCIPIOS FAIR

El primer aspecto que hay que tener en cuenta para la elaboración del PGD es que toda la **información y materiales deben estar contextualizados bajo principios FAIR** (*Findable, Accessible, Interoperable, Reusable* en su significado en inglés), los cuales son una serie de principios rectores que promueven la gestión eficaz de los datos con el objetivo de mejorar el valor de los recursos digitales y su potencial reutilización, ya que cada vez con más frecuencia se realizan descubrimientos utilizando datos ajenos.

Estos principios están organizados en cuatro conocimientos:

- **Localizable:** Los datos se describen con metadatos legibles por máquina. Los (meta)datos tendrán asignados un identificador único y persistente; y ambos estarán registrados o indexados en un sistema que disponga de motor de búsqueda.
- **Accesible:** que los datos sean alcanzables y accesibles para humanos y máquinas utilizando formatos y protocolos estándar de comunicaciones en abierto. Estos protocolos pueden necesitar de un procedimiento de autenticación o autorización cuando sea necesaria. Los metadatos siempre deben de estar disponibles aun cuando los datos no estén disponibles (por ejemplo, por ser sensibles, confidenciales o personales).
- **Interoperable:** que los (meta)datos sean legibles por máquina y se anoten utilizando lenguajes controlados y en formatos reconocibles y de uso extendido. Indicar referencias a otros datos o soportes o si los propios datos pertenecen a una publicación.



- **Reutilizable:** que los (meta)datos están suficientemente bien descritos para permitir la integración automatizada con otras fuentes de datos compatibles, licencias de uso para compartir y reutilizar, y usar estándares comunes en la comunidad investigadora.

Para entender cómo se enriquecen unos datos bajo los principios FAIR, podríamos imaginarnos una esfera cuyo núcleo serían los propios datos, a los que se van añadiendo determinadas capas, una primera capa que correspondería a añadir un identificador único y persistente, una segunda capa estandarizando la información y una tercera capa en la que se incluiría documentación contextual, es decir Metadatos y, si fuera necesario, información adicional tipo Readme.txt o diccionarios de datos.

La organización FAIRSharing (<https://fairsharing.org/>) pone a disposición un recurso curado, informativo y educativo sobre estándares de datos y metadatos, interrelacionados con bases de datos y políticas de datos.

6. PLANES DE GESTIÓN DE DATOS (PGD)

Un PGD es un documento complementario al proyecto de investigación, elaborado por el investigador, que detalla todas las tareas relacionadas con la gestión de los datos a lo largo de su ciclo de vida, siguiendo los principios FAIR. Este plan identifica las acciones y estrategias clave para asegurar que los datos de investigación sean de alta calidad, seguros, accesibles y reutilizables.

El PGD abarca desde la planificación inicial del proyecto hasta su finalización, garantizando la correcta custodia de los datos para evitar su pérdida, identificando posibles debilidades y asegurando su preservación para futuras validaciones o reutilización de resultados. Es un documento dinámico que se actualiza continuamente con nuevas decisiones y estrategias a lo largo del proyecto, por tanto, ganando precisión en cada versión.

Generalmente, junto con la solicitud del proyecto, se puede requerir una propuesta inicial del PGD, que es un documento breve con una descripción preliminar de los datos a utilizar y su gestión prevista. Durante el desarrollo del proyecto, se deben presentar al menos tres versiones del PGD: una en los primeros seis meses, otra en etapas intermedias y una final al concluir el proyecto. Según las necesidades, pueden entregarse versiones adicionales.

La elaboración de un plan de gestión de datos puede resultar una tarea ardua y demandante, requiriendo de esfuerzo y dedicación. Sin embargo, cuenta con una serie de beneficios directos para el investigador y su grupo:

- Permite la visualización del proyecto en general anticipándose a posibles imprevistos o riesgos, aplicando soluciones en fases tempranas
- Permite una gestión más fácil y es requisito de organizaciones y financiadores
- Clarifica el presupuesto necesario para llevar a cabo la gestión de los datos
- Favorece que los datos cumplan con los criterios FAIR
- Remarca una seria responsabilidad sobre la investigación



- Demuestra el interés del investigador por favorecer el intercambio, reutilización y conservación de los datos derivados del proyecto

7. GENERADORES Y PLANTILLAS

Actualmente, existen multitud de plantillas para la elaboración de PGD, con diferentes formatos y maneras de introducir la información, pero prácticamente en todas ellas, la información mínima que se incluye es:

- Información general sobre el proyecto.
- Descripción de los datos, tanto si son generados como reutilizados.
- Documentación y calidad de los datos.
- Almacenamiento y seguridad de los datos y la estrategia de conservación durante y después del proyecto.
- Requisitos legales y éticos, como la privacidad, propiedad intelectual y licencias.
- Intercambio y conservación que se va a dar.
- Responsabilidades y recursos económicos que conlleva la propia gestión de los datos.

Las plantillas suelen estar disponibles en páginas de organizaciones e instituciones, las cuales pueden ser descargadas y modificadas para nuestros proyectos. Suelen tener licencia abierta. Algunos ejemplos pueden ser la realizada por la Universidad de Granada y depositada en su repositorio Digibug o las de la Universidades de Huelva, Sevilla, Cantabria, entre otras. Estas plantillas suelen estar adaptadas a la institución, ya que utilizan sus propios recursos tales como el repositorio o algún otro servicio o herramienta.

Existen también herramientas gratuitas generadoras de PGD, que suelen estar bastante estandarizadas e integran explicaciones, recomendaciones o sugerencias para cada apartado del plan. Las más comunes son:

- DMPonline (<https://dmponline.dcc.ac.uk/>)
- PGDonline del consorcio Madroño y que es una traducción del anterior (<https://pgd.consorciomadrono.es/home>)
- DSW (<https://ds-wizard.org/>),
- ARGOS (integrada en OpenAIRE, <https://argos.openaire.eu/splash/index.html>)
- RDMO (*Research Data Management Organiser*, <https://rdmorganiser.github.io/en/>)

Por otro lado, también se puede contar con acceso a múltiples de PGD reales que pueden servir como referencia que pueden encontrarse en:

- LIBER: Data Management Plan Catalogue (<https://libereurope.eu/working-group/research-data-management/plans/>).
- Utilizando el buscador del repositorio Zenodo (<https://zenodo.org/>).



- En propia herramienta DMPOnline (<https://dmponline.dcc.ac.uk/>).
- Digital Curation Centre: example DMPs and guidance (<https://www.dcc.ac.uk/resources/data-management-plans/guidance-examples>).
- DMP Tool (<https://dmptool.org/>).

A todo esto hay que añadir servicios más allá de la propia redacción del PGD, como son los que Europa pone a disposición a través de OpenAIRE-Nexus (<https://www.openaire.eu/openaire-nexus-project>) que fomentan la Ciencia Abierta y donde se incluyen desde ARGOS (generador de GMP) y Zenodo (repositorio), hasta herramientas para anonimizar datos de investigación confidenciales (Amnesia) entre otros servicios, interoperables entre ellos.

8. SECCIONES DEL PLAN DE GESTIÓN DE DATOS

En general, el PGD podría dividirse en las siguientes secciones:

- 1) Aspectos administrativos: Descripción del proyecto y principales protagonistas
- 2) Descripción de los datos y recogida o reutilización de los datos existentes
- 3) Documentación y calidad de los datos
- 4) Almacenamiento y copia de seguridad durante el proceso de investigación
- 5) Requisitos legales y éticos, códigos de conducta
- 6) Intercambio de datos y conservación a largo plazo
- 7) Responsabilidades y recursos para la gestión de los datos

8.1. Aspectos administrativos: Descripción del proyecto y principales protagonistas

Incluir datos generales del proyecto como el título del proyecto y su código de identificación, objetivos del proyecto, organización o institución, agencia de financiación, nombre e identificador del investigador principal o investigadores principales, persona de contacto y modo de contactar, fechas y versiones del PGD.

8.2. Descripción de los datos y recogida o reutilización de los datos existentes.

Se debe describir toda la información sobre el proceso de obtención de datos, tanto si son datos nuevos como si son datos reutilizados. Además, hay que incluir un resumen del propio proyecto de investigación y los objetivos que pretenden alcanzarse con los datos. También, debe aportarse todo tipo de información respecto a la tipología y formato de los datos, así como una previsión del volumen necesario para almacenarlos durante el proyecto y de su estructuración y gestión de versiones.



Para desarrollar este punto hay que tener en cuenta que la propia organización de los datos en una investigación es crucial para asegurar que los datos se almacenen de manera útil y accesible, permitiendo una investigación más transparente. Además, un buen sistema de organización de los datos aporta seguridad, protegiendo a los datos frente a pérdidas y accesos no autorizados. Y cumple con las exigencias de financiadores e instituciones en cuanto a normativas y política de datos, que serán desarrolladas en otras secciones del PGD.

Formatos

La recomendación general es utilizar formatos que faciliten el intercambio, la reutilización y la preservación a largo plazo. Por lo tanto, se aconseja utilizar formatos con características de interoperabilidad, es decir que estén ampliamente aceptados y sean compatibles con diferentes sistemas y software (no propietarios, basados en estándares abiertos y comunes a la disciplina). Confirmar que, en caso de ser necesaria una conversión entre formatos, no se altere o se produzca pérdida de información. Finalmente, se aconseja tener en cuenta los requisitos del repositorio donde se depositarán los datos. Para saber más y entender la priorización sobre determinados formatos puede consultarse, entre otros, el servicio de datos de UK: <https://ukdataservice.ac.uk/learning-hub/research-data-management/format-your-data/recommended-formats/> donde, por ejemplo, se prioriza el formato .csv / .tab frente a formatos .xls/.xlsx de MS Excel, que aun siendo propietario, es aceptado por ser ampliamente extendido.

Estructura de Carpetas y Nombres Archivos

Una estructuración clara y consistente de carpetas y archivos facilita la localización y el acceso a los datos, ahorrando tiempo y esfuerzo. Utilizar un máximo de 3 niveles de carpetas y subcarpetas organizados jerárquicamente y organizados por contexto o tipología es un buen sistema de estructura de archivos. También se aconseja utilizar nombres descriptivos, únicos y estables, evitando caracteres especiales o nombres completos de autor. Un ejemplo podría ser de este tipo: *01_AnalisisPCRgenRUNX_v3_20231201_ACR.txt*. Se recomienda eliminar archivos intermedios o que no aporten valor.

Versionado

Debido a que la mayoría de los archivos suelen ser dinámicos es necesario llevar un control total de las diferentes versiones, incluyendo su fecha de modificación y documentar los cambios realizados en los datos o realizar un historial de versiones. Mantener siempre la versión original por seguridad.

8.3. Documentación y calidad de los datos

Los datos tienen que ir acompañados de unos archivos complementarios y metadatos que permitan contextualizar los datos y darles un significado, permitiendo su completo



entendimiento, tanto para los miembros del equipo que pueden acceder a ellos, como para su reutilización en el futuro.

Documentación extra: Es la información mínima necesaria para poder reutilizar los datos:

- *Readme file:* archivo/s con la información para interpretar los datos
- *Diccionario de datos:* lista descriptiva que incluye tanto los nombres de variables y etiquetas de categorías como las definiciones que se han utilizado en la gestión de los datos
- *Data paper:* descripción de los conjuntos de datos, suelen ser revistas que recogen la documentación de los datos

Metadatos: Son los datos de los datos y poseen una estructura que permite ser procesado por computadoras, permitiendo su transmisión e interoperabilidad. Existen varios tipos: PREMIS para preservación, DDI (*Data Documentation Initiative*) para encuestas, JSON Schema, pero uno de los más recomendados es el Dublin Core (<https://www.dublincore.org/>) que debería de contener información sobre:

- | | |
|---------------|-------------------|
| ● Título | ● Contribuciones |
| ● Creador | ● Fuentes |
| ● Fecha | ● Cobertura |
| ● Descripción | ● Materia |
| ● Derechos | ● Identificadores |
| ● Tipo | ● Formato |
| ● Idioma | ● Publicaciones |

Actualmente, los propios repositorios o portales de datos utilizan estos tipos de estándares de metadatos, cuya información será proporcionada por el repositorio elegido. Por ejemplo, el repositorio Digibug de la Universidad de Granada utiliza estándares Dublin Core para la descripción de los materiales alojados en él; o el repositorio Zenodo, incluido en OpenAire que utiliza un sistema basado en esquema JSON que ofrece exportación a Dublin Core.

8.4. Almacenamiento y copia de seguridad durante el proceso de investigación

Esta sección está dirigida a describir el lugar de almacenamiento de los datos durante el proceso de investigación y el respaldo frente a situaciones de pérdida o alteración de los datos, es decir qué datos serán copiados, dónde y cuándo se realizarán copias de seguridad.

Durante el desarrollo de la investigación hay que establecer una política de copias de seguridad en ubicaciones seguras. Se puede utilizar la regla del 3-2-1 dirigida a copias de seguridad y recuperación de desastres. Esta regla contempla la existencia de 3 copias de los datos e información relevante, dos copias en soportes distintos (servidor interno, nube) y una tercera copia en una localización diferente al recinto donde se almacenan las demás.



Por otro lado, para garantizar la seguridad y el acceso a los datos o a parte de ellos, se recomienda el uso de contraseñas seguras, encriptación de datos sensibles si fuera necesario, directrices para la eliminación segura y la utilización de antivirus. Debería de indicarse el responsable de toda esta gestión de acceso y mantenimiento del almacenamiento de los datos.

8.5. Requisitos legales y éticos, códigos de conducta

En general, para este apartado hay que describir como se garantiza el cumplimiento de la legislación sobre datos personales, seguridad de los datos, derechos de propiedad intelectual y otras posibles cuestiones éticas que haya que tener en cuenta.

Privacidad y Confidencialidad

Los datos personales son aquellos relacionados con una persona viva, lo que permite que ésta sea identificada. Los datos confidenciales o sensibles son datos personales sobre aspectos como origen racial o étnico, opinión política, religión, salud física y mental, vida sexual, delitos y procedimientos judiciales. Si se va a trabajar con estos tipos de datos, hay que tener en cuenta los siguientes aspectos:

- Estrategias para proteger la privacidad de los sujetos involucrados
- Determinar el tipo de acceso
- Siempre debe de haber un consentimiento informado
- Garantizar los derechos de las personas (RGPD y la Ley 3/2018 sobre datos personales)

Anonimización y seudonimización

En el caso de necesitar evitar la identificación de personas y/o de los datos sensibles o personales habría que realizar una encriptación y anonimización, pudiendo entonces compartir, ceder los datos o publicarlos abiertamente. Existen herramientas para esta anonimización como Amnesia de OpenAIRE (<https://amnesia.openaire.eu/>), Text anonymization helper tool de UK Data Archive o ARX Data anonymization Tools (<https://arx.deidentifier.org/>).

Licencias

Por otro lado, para que la compartición se haga de manera correcta, hay que tener una gestión de las licencias y de la propiedad intelectual. El titular de los derechos de explotación es el que tiene que determinar el acceso y las condiciones de explotación. Normalmente, los derechos de propiedad intelectual son del autor, pero los derechos de explotación hay que determinarlos mediante el acceso y condiciones para utilizarlos.

En general, disponemos de dos tipos de licencias: *Creative Commons* (CC, <https://creativecommons.org/>), que es la más conocida y utilizada para multitud de casos; y *Open Data Commons* (<https://opendatacommons.org/>), que es específica para los datos.



Las primeras tienen cuatro condiciones posibles como son: atribución, no comercial, sin obra derivada y compartir igual; estas 4 condiciones, combinadas entre ellas, dan lugar a las 6 licencias de CC. La que se recomienda principalmente para el contexto de ciencia abierta y datos FAIR es la de atribución CC-BY 4.0, es decir abierto y con reconocimiento al autor. Un caso especial de las licencias CC es la de dominio público (CC0), que libera la obra para que pueda ser utilizada bajo cualquier condición sin necesidad de reconocimiento de autoría. Esta es la licencia que recomienda la Unión Europea para el caso de los metadatos, los cuales deberían estar disponibles sin ningún tipo de restricción.

8.6. Intercambio de datos y conservación a largo plazo

Aunque en puntos anteriores se trata el tema del almacenamiento durante el desarrollo de la investigación, en esta sección hay que mencionar la manera cómo se preservarán los datos a largo plazo, es decir, una vez finalizado el proyecto. Para ello, habrá que indicar qué datos serán los que vamos a conservar a largo plazo, y en el caso de ser necesario, el método o herramienta informática para acceder y utilizar los datos, así como si existe algún tipo de intervalo de tiempo para su acceso, debido a embargo o similar, y si se dispone de identificador único y persistente para los archivos.

Por otro lado, hay que describir ciertas características del repositorio donde se depositarán, como puede ser el método de acceso a los archivos, si existe algún sistema de identificación, el tiempo de preservación, si permite la captura de identificadores únicos (DOI o Handle), el sistema de seguridad y de visibilidad.

Compartición y Preservación a largo plazo

La importancia de compartir los datos radica en la generación de nuevo conocimiento promoviendo innovación y aportando transparencia. Con esta dinámica de compartir los datos procedentes de la investigación se benefician tanto el propietario de los datos, como la sociedad y la comunidad investigadora en general, principalmente debido al aumento en el impacto y en la visibilidad de los resultados de los trabajos realizados. Pero hay que tener en cuenta que, aunque existe la obligación como requisito de depósito en abierto de los datos cuando la investigación proviene principalmente de fondos públicos, hay casos en los que no se pueden compartir, o al menos de compartir de manera completa, como puede ser cuando se incluya información sensible, cuando no son propiedad de quienes los hayan recopilado o cuando, por ejemplo, tiene un valor financiero asociado a la propiedad intelectual.

En cualquier caso, cuando sí se puede llevar a cabo el depósito en abierto de los datos, hay que tener en cuenta que éste se realiza a largo plazo, por lo que es fundamental contar con un repositorio, ya sea institucional, temático o multidisciplinar, que sea de calidad y seguro, y que cumpla los requisitos mínimos recomendados:

- I. Cumplir los criterios FAIR
- II. Almacenamiento de datos abiertos, con embargo, restringidos o cerrados
- III. Que admita diferentes tipos de materiales (formatos, tipos...)



- IV. Proporcionar visibilidad, transparencia y fiabilidad
- V. Conexión con otras infraestructuras que permita intercambio de información

A la hora de elegir el repositorio, se aconseja seguir la dirección: institucional - temático – generalista, aunque no es incompatible la utilización de más de uno, siempre y cuando conserven el mismo identificador único. Existen buscadores de repositorios como Repository Finder o re3data (Resitry of Research Data Repositories). Para nuestro caso, los repositorios más utilizados y recomendados que cumplen con los requisitos necesarios son **RISalud, repiSalud, DIGIBUG, Zenodo, Figshare**.

Difusión y Reutilización

El acceso a los datos puede variar en diferentes niveles, que van desde completamente abierto hasta cerrado, pasando por situaciones de embargo o acceso restringido. Tal como se mencionó previamente, la política general establece que los datos deben estar disponibles de manera abierta para todos. No obstante, existen casos en los que, debido a colaboraciones con empresas u organizaciones privadas que financian la investigación, se requiere la imposición de un embargo. Además, puede ser necesario restringir el acceso a los datos cuando se trata de información sensible o de carácter personal.

Independientemente del tipo de acceso, es imprescindible cumplir con ciertos requisitos para su publicación y reutilización. Estos incluyen la asignación de metadatos, siempre de manera abierta, el depósito de los datos en repositorios adecuados, el uso de licencias que regulen su distribución y la asignación de identificadores persistentes que garanticen su trazabilidad.

Los datos deberán ser publicados en un repositorio o en una revista específica de datos (por ejemplo, Scientific Data o Data in Brief) que serían artículos sobre datos donde se incluye toda la información necesaria y cuenta con revisión por pares, DOI, etc. La última opción es publicarse junto al artículo en repositorios propios de la revista, aunque para estos casos es más recomendable, por preservación, depositar este archivo en repositorio externo a la revista y citarlo en el artículo, lo cual añade más disponibilidad y publicidad a los autores y por tanto, más impacto al trabajo.

Citar correctamente los datos es importante porque:

- Identifica inequívocamente tanto a los datos como a los creadores de estos, sean nuestros o de terceros, generando reconocimiento y recompensa al productor.
- Permite la reutilización de datos y su verificación, gracias a garantizarse el acceso a los datos, metadatos, documentación o código.
- Se puede hacer seguimiento de su impacto, métricas específicas (Data Citation Index (Clarivate) o Dimensions).



8.7. Responsabilidades y recursos para la gestión de los datos

En esta sección será la referente a la asignación de los roles y responsabilidades de personas y organizaciones que participan en el proyecto, principalmente sobre la gestión de los datos y sobre la que decide qué datos serán los que finalmente sean compartidos.

Por otro lado, hay que detallar un presupuesto realista que indique el coste económico y humano que conlleva la preparación, almacenamiento y difusión de los datos de investigación siguiendo los principios FAIR. Actualmente, algunas agencias de financiación permiten contabilizar dichos costes dentro del proyecto, por ejemplo, Horizonte 2020 y Europa, podrán cubrir este tipo de costes, de hecho ponen a disposición de investigadores herramientas que ayudan a estimarlos (<https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs>).

9. DOCUMENTACIÓN RELACIONADA

- Horizon 2020 online manual
https://ec.europa.eu/research/participants/docs/h2020funding-guide/cross-cutting-issues/open-access-data-management/datamanagement_en.htm
- Guidelines on FAIR Data Management in Horizon 2020.
https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016).
<https://doi.org/10.1038/sdata.2016.18>
- GO FAIR Initiative. <https://www.go-fair.org/go-fair-initiative/>
- Instrucciones para la elaboración del Plan de Gestión de Datos (PGD). Agencia Estatal de Investigación. https://www.aei.gob.es/sites/default/files/convocatory_info/2023-07/Plan%20de%20Gesti%C3%B3n%20de%20Datos%202020.pdf
- Infografía: 10 pasos para elaborar un plan de investigación.
<http://hdl.handle.net/20.500.12105/10220>