

GESTIÓN DE DATOS DE INVESTIGACIÓN

GUÍA DE APOYO PARA LA DIFUSIÓN Y GESTIÓN DE LOS DATOS DE INVESTIGACIÓN

ÍNDICE:

- ¿Qué son los datos de investigación?
- ¿Por qué publicar los datos en acceso abierto? Beneficios de compartir los datos en abierto.
- Requisitos para la publicación en abierto de los datos de investigación.
- ¿Cómo gestionar los datos? Recomendaciones para la gestión de datos de investigación.
- Creación del Plan de Gestión de Datos:
 - Finalidad.
 - Plantillas.
 - Herramientas: ARGOS y Pla de Gestió de Dades de Recerca.
- Guías para elaborar un Plan de Gestión de Datos.
- Cómo citar los datos de investigación.
- Cómo difundir los datos – REPOSITARIOS:
 - RediUMH.
 - ZENODO/Universidad Miguel Hernández Research Data.
- Buscar datos de investigación:
 - Directorios o buscadores de repositorios.
 - Repositorios de conjuntos de datos (datasets) generalistas y especializados.
- ENLACES DE INTERÉS.

GESTIÓN DE DATOS DE INVESTIGACIÓN

GUÍA DE APOYO PARA LA DIFUSIÓN Y GESTIÓN DE LOS DATOS DE INVESTIGACIÓN

- ¿Qué son los datos de investigación?

Entendemos como datos de investigación todo aquel material generado o recolectado durante el desarrollo de una investigación y que sirve como evidencia de la misma. La tipología y variedad de los datos es muy amplia.

Los datos pueden ser numéricos, descriptivos o visuales. Pueden ser datos sin tratar o datos tratados y analizados. Se presentan en cualquier rama del conocimiento, como resultado de muy diferentes técnicas o metodologías, configurando conjuntos o colecciones heterogéneas de datos, o datasets, en cualquier formato y soporte.

Como datos de investigación se incluyen: datos primarios (en soporte papel o informático), cuestionarios, materiales audiovisuales, cuadernos de campo, cuadernos de laboratorio, desarrollo de modelos, pruebas, etc.

- ¿Por qué publicar los datos en acceso abierto? Beneficios de compartir los datos en abierto.

Los datos abiertos de la investigación (**Research Open Data**) son datos que, tanto legal como técnicamente, se pueden consultar, analizar, revisar, reutilizar, redistribuir y redirigir a otros fines.

Se caracterizan por:

1. **Aportar transparencia** a los procesos de investigación al facilitar la reproducción y validación de los experimentos e indagaciones realizadas.
2. **Permitir a cualquier otro investigador que use los datos** en nuevos estudios y trabajos, y por tanto también a empresas, profesionales o personas interesadas.

Uno de los elementos claves de la Ciencia Abierta es la difusión de los datos de investigación. Desde diferentes instituciones y foros, se insiste cada vez más en el valor de los datos y los **beneficios de compartir los datos en abierto**:

- Acceso libre y universal a los datos, sin barreras legales ni económicas.
- Aumenta la visibilidad, repercusión, impacto y prestigio de los investigadores.
- Favorece los procesos de revisión y reproducibilidad de la investigación.
- Promueve la ciencia abierta, impulsando el conocimiento y la innovación.
- Mejoran la gestión y conservación de los datos, minimizando el riesgo de pérdida.
- Incrementan la eficiencia y la colaboración en la investigación, optimizando esfuerzos.
- Estimulan la innovación y la productividad científica e industrial.

Muchos organismos financiadores incluyen como uno de sus requisitos la publicación en abierto de los datos de investigación:

- [Horizon Europe](#). Obligaciones para los beneficiarios:
 - Gestionar los datos de investigación de forma responsable y que los datos cumplan con los principios FAIR (Findable, Accesible, Interoperable, Reusable).
 - **Elaborar un Plan de gestión de datos** (Data Management Plan).
 - Depositar los datos en un repositorio de confianza.
 - Asegurar el acceso abierto a los datos vía repositorio con una licencia CC-BY o CC-0 o equivalente.
 - Los metadatos de los datos de investigación tienen que ser abiertos y con una licencia CC-0 o equivalente.
- [Plan Estatal de Investigación Científica y Técnica y de Innovación](#) (PEICTI) 2021-2023. Entre los principios básicos está:
 - Acceso abierto a los resultados de las actividades subvencionadas, tanto las publicaciones científicas como los datos de investigación.
 - Los datos de investigación tienen que cumplir los principios FAIR, y siempre que sea posible, se tienen que difundir en abierto.

- [H2020](#). A partir de 2017, todos los proyectos beneficiarios de este programa tienen la obligación de:
 - Elaborar un Plan de gestión de datos (Data Management Plan).
 - Depositar los datos, respetando siempre la normativa relativa a los datos de carácter personal y otras posibles excepciones, en un repositorio de acceso abierto.
 - Los datos de investigación tienen que cumplir los principios FAIR.

https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

https://ec.europa.eu/info/research-and-innovation/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_es



<https://www.rebiun.org/acceso-abierto/ciencia-abierta-la-investigacion-y-los-datos-cientificos-accesibles-y-abiertos-todos>



Infografía Cómo cumplir con los mandatos sobre gestión y publicación de datos en Horizonte 2020; Realizada por REBIUN Línea 2 (3er. P.E.) Grupo de Acceso Abierto (2016): <http://hdl.handle.net/20.500.11967/68>

- **Requisitos para la publicación en abierto de los datos de investigación:**
 - Gestionar los datos de investigación de forma responsable y que los datos cumplan con los **principios FAIR** (Findable, Accesible, Interoperable, Reusable).
 - **Elaborar y mantener** un Plan de Gestión de Datos.
 - **Depositar** al final del proyecto los datos en un repositorio público.
 - **Permitir el acceso** y explotación de los datos, preferiblemente mediante licencias CC.
 - **Documentar y describir** los datos mediante metadatos, etc. Los metadatos de los datos de investigación tienen que ser abiertos y con una licencia CC-0 o equivalente
- **¿Cómo gestionar los datos? Recomendaciones para la gestión de datos de investigación (por la red Maredata):**

La **gestión** de los datos de la investigación requiere **elaborar y mantener un plan de gestión de datos** desde la propuesta de financiación o proyecto de investigación, **que prevea los retos y soluciones al manejo de los datos**.

También es conveniente seguir algunas buenas prácticas recomendadas, sobre denominación y formatos de archivos, almacenamiento, acceso y seguridad durante el proyecto, descripción y metadatos, licencias de reutilización, etc.:

- Promover la innovación y la reutilización de los datos que potencialmente puedan tener nuevos usos.
- Facilitar la colaboración entre usuarios, creadores y reutilizadores de datos.
- Maximizar la transparencia y la fiabilidad de los datos.
- Favorecer la reproducibilidad de los ensayos experimentales.
- Permitir la verificación de los resultados de investigación.
- Reducir costes al evitar la duplicación de datos.
- Aumentar el impacto y la visibilidad de la investigación.
- Promover los proyectos de investigación de los que provienen los datos y sus publicaciones.
- Generar un reconocimiento directo de los investigadores productores de datos, como ocurre con cualquier otro resultado de investigación.



Infografía El ciclo de los datos científicos; Realizada por REBIUN Línea 2 (3er. P.E.) Grupo de Acceso Abierto (2016) <http://hdl.handle.net/20.500.11967/69>

Documentos generales e introductorios sobre el tema:

- MELERO, R. 2019. **Recomendaciones para la gestión de datos de investigación dirigidas a investigadores.** En: *Digital CSIC*. <https://biblioteca.umh.es/files/2019/03/Maredata-recomendaciones-ESP.pdf>
- **Gestión de datos de investigación en las universidades españolas y CSIC:** Memoria de buenas prácticas de los servicios ofrecidos <http://hdl.handle.net/20.500.11967/244>

- **Principios FAIR:** Buenas prácticas para la gestión y administración de datos científicos [Fuente: datos.gob.es].
- Más información en: "Formatos de archivos para almacenar datos a largo plazo" del Portal de datos iuFOR, y en: "Cómo elegir el formato correcto para los datos abiertos" Curso European Data Portal.
- **Infoguía sobre buenas prácticas en la gestión de los datos de investigación** (elaborada por el Servei de Biblioteques, Publicacions i Arxius de la Universitat Politècnica de Catalunya y traducida en la Sección de Información y Referencia del Servicio de Biblioteca de la Universidad de A Coruña).



- **Creación del Plan de Gestión de Datos:**

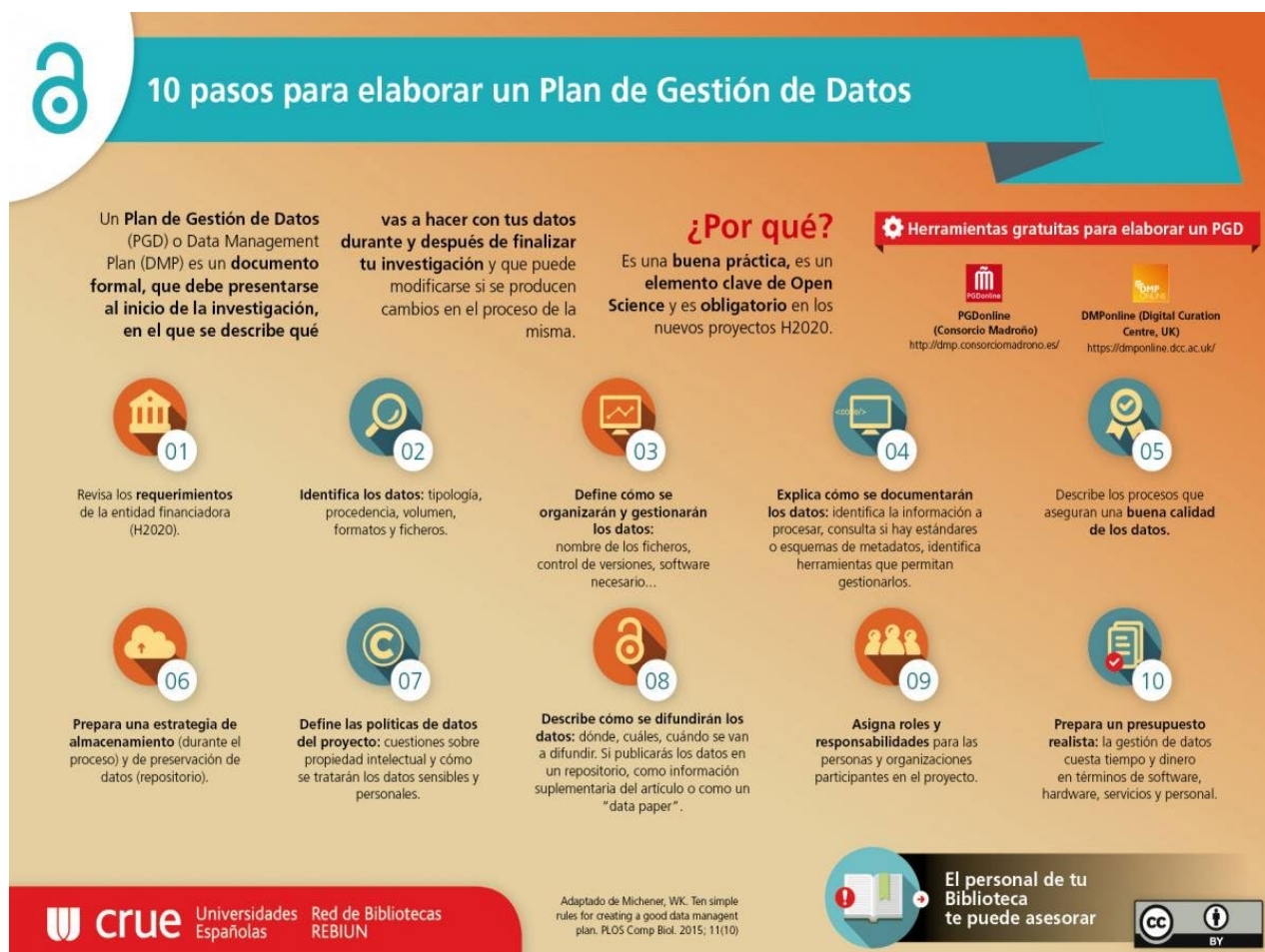
- **Finalidad:** El Plan de Gestión de datos (Data Management Plan) tiene como finalidad describir el ciclo de vida de todos los datos que se generarán, recogerán o procesarán a lo largo de la investigación. También hay que indicar la metodología y estándares utilizados para su descripción, así como las acciones que se llevarán a cabo para la difusión, conservación y preservación de estos datos. Incluye cuestiones como nombre y descripción del proyecto, metodología y estándares empleados, cómo se preservarán y compartirán los datos, nombre y tipo de repositorio donde se depositarán, etc.

Se trata de un documento abierto, que se presenta al inicio del proyecto y se va actualizando a medida que avanza el proyecto de investigación. Es un requisito en muchas convocatorias de ayudas.

Los Planes de Gestión de Datos (PGD) son elementos imprescindibles para asegurar una correcta gestión de los datos de investigación. Además, también se está convirtiendo en un requisito para optar a financiación de proyectos de investigación, por ejemplo: El Horizonte Europa (2021-2027) obliga a redactar y mantener un PGD, siguiendo con lo establecido en el anterior plan, el H2020 que ya obligaba a realizar un PGD actualizable a lo largo del proyecto. Se debe redactar un PGD en los Proyectos I+D+i correspondientes a la convocatoria del año 2021 de ayudas del Plan Estatal de Investigación Científica, Técnica y de Innovación 2021-2023 siguiendo lo marcado en el anterior Plan Estatal de Investigación Científica y Técnica y de Innovación 2017-2020.

Los planes de gestión de datos persiguen por un triple objetivo:

- ✓ Garantizar que no se pierden los datos, sobre todo al finalizar un proyecto de investigación.
- ✓ Facilitar su correcta custodia desde su producción.
- ✓ Permitir su preservación mediante documentación de todo su entorno tecnológico.



Infografía 10 pasos para elaborar un Plan de Gestión de Datos; Realizada por REBIUN Línea 2 (3er. P.E.) Grupo de Acceso Abierto (2016)
<http://hdl.handle.net/20.500.11967/71>

Un Plan de Gestión de Datos describe el ciclo de vida de la gestión de los datos en la utilización, procesamiento y generación para conseguir que los datos de investigación sean localizables, accesibles, interoperables y reutilizables. Es decir, que sean FAIR (Findable, Accessible, Interoperable and Reusable). En resumen, indica qué se va a hacer con los datos durante y después de finalizar la investigación.

Un Plan de Gestión de Datos debe incluir información sobre:

- Determinar los requisitos de la entidad financiadora.
- Identificar los datos que se utilizarán: tipología, procedencia, formatos, etc.

- Definir cómo se organizarán y gestionarán los datos: control de versiones, nombre de ficheros, etc.
 - Explicar cómo se documentarán los datos.
 - Describir cómo se asegura la calidad de los datos.
 - Preparar una estrategia de almacenamiento y preservación de los datos.
 - Definir las políticas de datos del proyecto: propiedad intelectual, tratamiento de datos sensibles y personales, etc.
 - Describir cómo se difundirán los datos: dónde, cuáles y cuándo se van a difundir, etc.
 - Asignar roles y responsabilidades.
 - Preparar un presupuesto realista para la gestión de los datos.
- **Plantillas:** Horizonte Europa. Plantilla de plan de gestión de datos.
- **Herramientas:**



Argos es una herramienta online desarrollada por Open Aire para la creación, gestión, difusión y enlace de un PGD, que hace énfasis en la aplicación de los principios FAIR y en las mejores prácticas para fomentar la accesibilidad a los datos de investigación. Éstas son algunas de sus características:

- Asistente para la creación del PGD en función del financiador.
- Asistente para la descripción de los datos de investigación, de manera que permite reutilizarlos en diferentes PGD's.
- Asignación de un DOI para el PGD si se publica, por lo que se facilita la visibilidad y citación del propio PGD.
- Exportación del PGD en formatos comprensibles por máquinas (xml, json) y por personas.
- Servicio incluido en el European Open Science Cloud (EOSC), iniciativa promovida por la Comisión Europea.

Se puede encontrar más información sobre Argos consultando el manual de usuario o las FAQs <https://argos.openaire.eu/home>



El [Pla de Gestió de Dades de Recerca](#) Es una herramienta en línea, desarrollada por el Consorcio de Servicios de las Universidades Catalanas, basada en el modelo creado por el DCC (Digital Curation Centre, UK). Incluye ejemplos reales en el cuestionario para ayudar en la elaboración del Plan de Gestión de Datos.

- **Guías para elaborar un Plan de Gestión de Datos:**

- Directrices de la Comisión Europea dirigidas a los solicitantes y beneficiarios de proyectos de investigación.
- Guías para investigadores: Cómo crear un plan de gestión de datos para proyectos H2020
<https://www.openaire.eu/how-to-create-a-data-management-plan>
- Manuales de ciencia abierta: introducción a las buenas prácticas
<https://www.openaire.eu/os-primers>
- Directrices para la Gestión de Datos en H2020. Traducción al español consorcio Madroño. Versión 1.0. 11 de diciembre de 2013.
https://www.consorciomadrono.es/wpcontent/uploads/2017/05/directrices_gestion_datos_horizon_2020_es.pdf
- Un manual de gestión de datos de investigación: Una introducción a la gestión de sus datos de investigación <https://www.openaire.eu/rdm-handbook>

• Cómo citar los datos de investigación:

Los datos deben citarse correctamente, siguiendo un formato de citas específico y deben aparecer con el resto de referencias bibliográficas de la publicación resultante. [DOI Citation Formatter](#) es un servicio ofrecido por [DataCite](#) que construye automáticamente las citas conforme al estilo que seleccionemos a partir del DOI asignado a los datos. Citar los datos de investigación te permitirá:

- Identificarlos fácilmente y atribuirlos correctamente a su creador.
- Reutilizar los datos fácilmente y, además, verificarlos.
- Hacer un seguimiento del impacto que puedan tener los datos.
- Crear una estructura académica que reconozca y recompense a los productores de datos.



Cita tus datos de investigación

Por qué es importante citar los datos:

- Los conjuntos de datos también son resultados de investigación como los artículos, monografías, etc.
- Facilita la identificación y el acceso a los datos y de esta forma su localización, validación y reutilización.
- Permite reconocer la autoría de sus creadores.
- Facilita la métrica e impacto de los datos.
- Favorece la transparencia de la investigación científica.

Buenas prácticas para citar datos:

- Se debe facilitar la identificación, localización y el acceso a los datos mediante un identificador único y persistente (DOI, Handle, etc.).
- Cada conjunto y subconjunto de datos (dataset) debe citarse de forma independiente.
- Las citas de los datos utilizados han de aparecer en la sección de referencias bibliográficas de la publicación resultante.
- Se recomienda incluir un identificador único de autor (ORCID, etc.).

 **Enlaza los datos con los documentos resultado de investigación y viceversa, y crea las referencias bibliográficas de los mismos.**

Elaboración de la cita

- Existen elementos mínimos obligatorios (O) y otros recomendados (R) que se combinan para elaborar la cita en cualquier estilo estándar (APA, MLA, Chicago, etc.) o los propuestos por los principales repositorios de datos (Dataverse, Dryad, etc.).

Autor(es) (O)

Identificador autor (R)

Fecha (O)

Título (O)

Identificador único persistente (O)

Tipo de recurso (O)

Versión y/o Edición (O)

Repositorio de datos (R)

Publicación (R)

Productor (R)

Ámbito geográfico (R)

Ámbito temporal (R)

Ejemplo de cita estilo APA

Autoría

Título

Remesar Betloch, X., Antelo, A., Llivina, C., Albà, E., Berdié, L., Agnelli, S.,... Alemany, M. (2015). *Influence of a hyperlipidic diet on the composition of the non-membrane lipid 6 pool of red blood cells of male and female rats*. [Dataset]. Versión de 22 de junio de 2015. Dipòsit digital de documents de la UAB: <http://hdl.handle.net/2445/66010>

Fecha

Repositorio

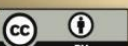
Tipo de recurso

Identificador único y persistente

Versión



El personal de tu Biblioteca te puede asesorar



- **Cómo difundir los datos – Repositorios:**

Muchas agencias de financiación, instituciones y revistas académicas tienen mandatos y políticas sobre la publicación en acceso abierto de los datos de investigación. El cumplimiento de estos mandatos y políticas se suele realizar a través del depósito de los datos de investigación en un repositorio de datos. A la hora de elegir un repositorio debemos tener en cuenta algunos aspectos:

- Área temática.
- Capacidad de almacenamiento.
- Facilidad de recuperación de datos.
- Asignación de un identificador único y persistente para cada conjunto de datos (DOI).
- Establecimiento de un periodo de embargo para los datos.
- Selección de la licencia de uso de los datos.
- Preservación a largo plazo de los datos.
- Cumplimiento con la certificación Core Trust Seal.

Según el OpenAIRE Research Data Management Briefing Paper, los datos deben depositarse en un repositorio de datos según el siguiente orden de preferencia:

1. Repositorio temático de datos consolidado para esa disciplina.
2. Repositorio institucional de datos.
3. Repositorio multidisciplinar de datos.
4. Otros repositorios de datos.

El destino definitivo de los datos, el repositorio donde vayan a depositarse, es objeto de una decisión importante, pues puede condicionar las opciones sobre formatos de datos, metadatos, licencias de uso, etc. Para la difusión de los datos de investigación en abierto, los investigadores disponen de diferentes opciones, puesto que la diversa tipología, el volumen o la finalidad de los datos hacen que no sea posible actualmente la existencia de un único repositorio donde depositar. Según las necesidades, puede ser más adecuada para la difusión un repositorio temático, el repositorio institucional o un repositorio general.

En el caso de la Universidad Miguel Hernández disponemos:

- **Repositorio UMH/REDIUMH/Investigación:** Para depositar y difundir los datos vinculados a una publicación o cualquier proyecto de investigación, hasta 500 MB. A cada dataset se le asigna un DOI, además del handle.

<http://dspace.umh.es/>

- **ZENODO/Universidad Miguel Hernández Research Data:** Para depositar y difundir conjunto de datos (datasets), códigos de software, etc. con un tamaño superior a 500 MB hasta 50 GB.

<https://zenodo.org/communities/1021134rediumh/>

- **Buscar datos de investigación:**

Reseñamos a continuación una serie de Buscadores o Directorios de repositorios de datos de investigación y de enlaces a repositorios concretos, tanto generales como especializados.

- **Directorios o buscadores de repositorios:**

- [Re3data](#)
- [OAD Data Repositories](#)
- [DataMED](#)
- [Repository Finder](#)

- **Repositorios de conjuntos de datos (datasets) generalistas y especializados:**
 - [DataCite Search](#): buscador generalista de datasets abiertos con DOI.
 - [DataMED](#): buscador de datasets y repositorios en el ámbito biomédico.
 - [DataONE](#): buscador de datasets en repositorios sobre la vida en la tierra y medio ambiente.
 - [DataSearch](#): buscador de datasets de Elsevier, en su mayoría de artículos de Elsevier.
 - [Dryad](#): buscador de datasets asociados a publicaciones de ciencias y medicina.
 - [e-cienciaDatos](#): buscador de datasets del Consorcio Madroño (Universidades de Madrid y UNED).
 - [Eudat B2FIND](#): buscador generalista de datasets mantenido por la organización EUDAT Collaborative Data Infrastructure.
 - [Figshare](#): buscador internacional en el que se pueden recuperar datos por categorías temáticas.
 - [Google Dataset Search](#): buscador de datasets en repositorios de datos científicos, bases de datos de gobiernos locales y nacionales, sitios web de editores y autores y otras fuentes.
 - [Mendedey Data](#): buscador de datasets de Elsevier, en su mayoría de artículos contenidos en sus publicaciones.
 - [Zenodo](#): buscador de datos de investigación de la Unión Europea. Desarrollado por CERN Data Center e Invenio.

- **ENLACES DE INTERÉS:**

- Gestión de datos de investigación en las universidades españolas y CSIC: memoria de buenas prácticas de los servicios ofrecidos
<https://repositoriorebiun.org/handle/20.500.11967/244>
- Recomendaciones para la gestión de datos dirigidas a investigadores. Maredata.
<https://datos.gob.es/es/noticia/recomendaciones-para-la-gestion-de-datos-dirigidas-investigadores>
<https://digital.csic.es/handle/10261/173801>
- Datasets: plantilla normalizada para la descripción de registros en DIGITAL.CSIC
Oficina Técnica de DIGITAL.CSIC
- FAIR x FAIR: recomendaciones para repositorios de datos de investigación FAIR
<https://www.ub.edu/blokdebid/es/content/fair-x-fair-recomendacionespara-repositorios-de-datos-de-investigacion-fair>
- OpenAIRE. Guides for Researchers. How to select a data repository?
<https://www.openaire.eu/opendatapilot-repository-guide>
- [Infografías de REBIUN](#)

Para finalizar, debemos tener en cuenta varias normas básicas en materia de protección de datos:

- [Reglamento \(UE\)2016/679](#) del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (GDPR).
- [Ley Orgánica 3/2018](#), de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.
- [Guía de buenas prácticas en materia de Transparencia y Protección de Datos](#). Editado por la CRUE.