



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Biblioteca y Documentación Científica

Depósito de datos de investigación en RiuNet.

28 de julio de 2024





ÍNDICE:

- ¿Qué son datos de investigación?
- ¿Qué se incluye dentro de dato de investigación?
- ¿Qué son datos abiertos de investigación?
- ¿Qué datos se depositan?
- ¿Quién puede depositar?
- Condiciones de aceptación
- Proceso de depósito
- ¿Cuándo se depositan los datos?
- ¿Cómo hacemos los datos FAIR?
- ¿Qué es la Data Availability Statement o declaración de disponibilidad de datos?
- ¿Cómo ingreso el DOI de un dataset en la publicación?
- ¿Qué licencias usar?
- ¿Qué formatos se recomiendan?
- ¿Qué documentación debo adjuntar?
- Buenas prácticas de trabajo con datos
- ¿Cómo procedo si hay datos personales en el dataset?
- ¿Debo depositar el Plan de Gestión de Datos?
- Enlazar publicaciones y datos



¿Qué son datos de investigación?

Se definen los datos de investigación como material factual registrado, aceptado en la comunidad científica y necesario para documentar, respaldar y validar la información (National Institutes of Health <https://www.nih.gov/>).

La Directiva (UE) 2019/1024 del Parlamento europeo y del Consejo de 20 de junio de 2019 define «datos de investigación» como “documentos en formato digital, distintos de las publicaciones científicas, recopilados o elaborados en el transcurso de actividades de investigación científica y utilizados como prueba en el proceso de investigación, o comúnmente aceptados en la comunidad investigadora como necesarios para validar las conclusiones y los resultados de la investigación”.

De forma sencilla se puede decir que los datos de investigación son hechos, observaciones o experiencias creados durante el proceso de una investigación.

¿Qué se incluye dentro de datos de investigación?

Se incluye una amplia gama de resultados de investigación: cuestionarios, cintas de audio, vídeos, desarrollo de modelos, fotografías, diseños y muestras, software, mediciones tomadas por máquinas, colecciones documentales, entrevistas, encuestas, observaciones de campo, videgrabaciones, simulaciones computacionales ..., tanto investigación cuantitativa como cualitativa.

No se aceptan notas de laboratorio, conjuntos parciales de datos, análisis preliminares, borradores, planes, reportes, objetos o informes en pdf.

¿Qué son datos abiertos de investigación?

Se definen los datos abiertos de investigación como “datos de acceso gratuito que pueden ser reutilizados, remezclados y redistribuidos para la investigación académica y la docencia, entre otros usos. Idealmente, los datos abiertos no tienen restricciones para su reutilización y redistribución y cuentan con licencias acordes a ello”: https://open-science-training-handbook.github.io/Open-Science-Training-Handbook_ES/02OpenScienceBasics/02OpenResearchDataAndMaterials.html



La definición de apertura no se refiere únicamente a la idea de libre disponibilidad y acceso sin trabas. Respecto a datos debe incluir, además:

- Reutilización y redistribución: los datos deben ser suministrados de forma que permitan ser reutilizados y redistribuidos, mediante el cumplimiento de los principios FAIR.
- Interoperabilidad: que permitan la integración con otros conjuntos de datos, de forma que puedan combinarse para construir sistemas más compuestos.

¿Qué datos se depositan?

Es importante proporcionar ambos tipos de datos, los brutos y los procesados. Los datos brutos permiten que otros investigadores verifiquen y reproduzcan los resultados, ofrecen la posibilidad para que otros investigadores analicen de forma diferente o identifiquen posibles errores. Los datos procesados facilitan la comprensión inmediata de los hallazgos y permiten que otros vean cómo se llegó a las conclusiones. Al proporcionar datos brutos se ofrece información de forma más completa, permitiendo a otros investigadores realizar nuevas investigaciones. No obstante, los mandatos de los financiadores hablan de depositar únicamente los datos que verifican los resultados de la publicación.

¿Quién puede depositar?

Para poder iniciar el autoarchivo de un ítem:

- Tiene que identificarse pinchando en el enlace *Mi cuenta > Identificarse* y una vez identificado entrar en *Mi cuenta > Nuevo envío*.
- Seleccionar la colección en la que desea incorporar el nuevo documento mediante el desplegable, y pulsar "Siguiete" o bien desde la propia colección Datasets a través de la opción "*Enviar un ítem a esta colección*".

Para depositar un documento en una colección, debe estar previamente autorizado.

Además, para que un documento pueda ser difundido a través de RiuNet, es necesario que su autor haya firmado el [Acuerdo Marco](#) de cesión de derechos de autor de obras digitales en RiuNet, que autoriza a la Universidad Politécnica de Valencia a difundir su obra digital. Sólo es necesario firmarlo la primera vez que se deposite un documento en RiuNet, a partir de entonces basta con conceder la licencia de distribución que aparece al final del proceso de "Envío de ítems".



Condiciones de aceptación

- Los datos deben haber sido producidos:
 - Dentro de proyectos en los que participen investigadores de la UPV.
 - Entidades y/o servicios de la UPV.
- Los autores de los datos deben estar en condiciones de conceder los derechos necesarios a la UPV para asegurar la correcta distribución y preservación de los datos a través de RiuNet.
- Si el conjunto de datos contiene datos personales, se debe tener en cuenta lo expresado en la normativa nacional e internacional sobre protección de datos personales.
- Se debe indicar el tipo de versión de los datos que se quiere depositar: datos en bruto, datos procesados, versión final, etc.
- Los datos deben estar debidamente organizados para facilitar su comprensión y reutilización. Para ello, se seguirán las siguientes recomendaciones:
 - Nombre de los ficheros.
 - Formatos de los ficheros de datos.
 - Documentación de los ficheros de datos.
- Aunque se pueden depositar varios ficheros con datos de investigación, cada uno de los ficheros no debe superar los 2 GB de tamaño.

Proceso de depósito

- Los investigadores de la UPV podrán realizar el depósito de los datos de investigación en la colección [Datasets](#) que se encuentra en RiuNet.
- Si los datos se han obtenido/procesado dentro de un proyecto de investigación con financiación (H2020, MINECO, GVA, HORIZONTE EUROPA, etc.) deberá reflejarse en RiuNet durante el depósito, indicando, por una parte el agente financiador y por otra, el código del proyecto. De esta forma se facilita la justificación del cumplimiento de la política o mandato del agente financiador.
- La fecha de publicación de los datos de investigación deberá introducirse de forma completa, es decir, año/mes/día.
- Los datos depositados deberán cumplir con las condiciones descritas anteriormente. Especialmente, además de los ficheros con los datos, se deberá depositar un [Fichero README.txt](#) donde se describan los datos de investigación.



Biblioteca y Documentación Científica

- La publicación de los datos en RiuNet no es inmediata, ya que la biblioteca debe realizar un proceso de validación. Durante este proceso se comprobará el cumplimiento de las condiciones para la aceptación, así como el nivel de descripción de los datos.
- [Contacta con la Biblioteca](#) para cualquier duda sobre el depósito de los datos de investigación en RiuNet.

¿Cuándo se depositan los datos?

La Ley 17/2022 por la que se modifica la Ley 14/2011 de la Ciencia dicta, en su artículo 37.2 que “el personal de investigación del sector público o cuya actividad investigadora esté *financiada mayoritariamente con fondos públicos y que opte por diseminar sus resultados de investigación en publicaciones científicas, deberá depositar una copia de la versión final aceptada para publicación y los datos asociados a las mismas en repositorios institucionales o temáticos de acceso abierto, de forma simultánea a la fecha de publicación*”.

La Ley Orgánica 2/2023 del Sistema Universitario (LOSU) recoge el mismo mandato en su artículo 12.2: “El personal docente e investigador deberá depositar una copia de la versión final aceptada para publicación y los datos asociados a la misma en repositorios institucionales o temáticos de acceso abierto, de forma simultánea a la fecha de publicación.”

¿Cómo hacemos los datos FAIR?

Los mandatos de los financiadores como [Horizonte Europa](#) así como la reciente legislación aprobada como la [Ley Orgánica 2/2023, de 22 de marzo, del Sistema Universitario](#) o la [Estrategia Nacional de Ciencia Abierta \(2023-2027\)](#) piden respetar los principios FAIR en el depósito de los datos:

FINDABLE:

- Datos con identificadores persistentes. Debe aparecer en el registro de metadatos.
- Metadatos en línea en recursos de búsqueda (los motores de búsqueda localizarán los datos).
- Elegir un repositorio adecuado (si los datos no se pueden compartir compartiremos los metadatos).
- Incluir el identificador persistente en la publicación asociada.



ACCESSIBLE:

- Elegir un repositorio adecuado. Metadatos siempre disponibles, aunque los datos no estén abiertos.
- Elegir formatos abiertos, convertirlos si es necesario.
- Documentación adecuada: README, metadatos en XML, vocabularios y nomenclaturas según disciplinas...
- Documentar los procesos de autenticación y autorización si los hay.

INTEROPERABLE:

- Elegir formatos abiertos o estándar reconocibles, convertir si es necesario.
- Documentación adecuada: README, metadatos en XML, vocabularios y nomenclaturas según disciplinas...
- Lenguaje formal y accesible. Vocabulario controlado.
- Referencias y enlaces a otros datos.

REUSABLE:

- Elegir formatos abiertos o estándar, convertir si es necesario.
- Documentación adecuada: README, metadatos en XML, vocabularios y nomenclaturas según disciplinas...
- Usar licencias abiertas.

RiuNet se adhiere a los principios FAIR a través del cumplimiento de las siguientes características:

- Protocolos de interoperabilidad abiertos y gratuitos (OAI-PMH).
- Esquema de metadatos estándar (Dublin Core).
- Plantillas de descripción.
- Asignación de PIDs (Handle, DOIs).
- Indexación en agregadores.
- Presencia en www.re3data.org

¿Qué es la Data Availability Statement o declaración de disponibilidad de los datos?

La declaración de disponibilidad de datos es la indicación que le damos al lector de nuestra publicación acerca de la disponibilidad y ubicación de los datos asociados así como la indicación de qué puede hacer con ellos, los usos permitidos. Para ello indicaremos el depósito que aloja los



datos, la licencia que hemos elegido y el identificador persistente, a ser posible mediante un metadato enlazable.

La declaración de disponibilidad de los datos se ofrece como una sección al final del artículo.

¿Cómo ingreso el DOI de un dataset en la publicación?

El DOI de los datasets se obtiene de forma inmediata al realizar el depósito con lo que podemos obtener el DOI antes de publicar la investigación.

¿Qué licencias usar?

Para indicar el tipo de uso permitido, se pueden utilizar licencias Creative Commons 4.0, como las CC0 y CC-BY. También es posible utilizar licencias específicas para datos Open Data Commons:

- [Attribution License \(ODC-By\)](#) — “Attribution for data/databases” permite a terceros copiar, distribuir y usar la base de datos, así como utilizarla para crear nuevos contenidos, bases de datos o colecciones de bases de datos (siempre y cuando se cite la base de datos original).
- [Open Database License \(ODC-ODbL\)](#) — “Attribution Share-Alike for data/databases” permite a terceros copiar, distribuir y usar la base de datos, así como utilizarla para crear nuevos contenidos, bases de datos o colecciones de bases de datos, siempre y cuando que a las bases de datos derivadas se les otorgue la misma licencia que a la base de datos original.
- [Public Domain Dedication and License \(ODC-PDDL\)](#). No impone restricción alguna, permite copiar, distribuir y utilizar la base de datos, así como crear nuevos productos y adaptaciones.

Sugerencia de redacción:

This {DATA(BASE)-NAME} is made available under the Open Data Commons Attribution License: <http://opendatacommons.org/licenses/by/{version}>.

This {DATA(BASE)-NAME} is made available under the Open Database License: <http://opendatacommons.org/licenses/odbl/1.0/>.

Any rights in individual contents of the database are licensed under the Database Contents License: <http://opendatacommons.org/licenses/dbcl/1.0/>.



¿Qué formatos se recomiendan?

Para cumplir con el requisito FAIR de interoperabilidad se requiere utilizar formatos abiertos. De no ser posible, se utilizarán formatos propietarios reconocibles y ampliamente utilizados. Es recomendable convertir esos datos a formatos abiertos mediante programas de conversión como [OpenRefine](#).

Se recomienda consultar los siguientes listados de formatos recomendados y aceptados:

- [UK Data Service](#)
- [Library of Congress](#)

¿Qué documentación debo adjuntar?

La [plantilla README](#) es obligatoria para realizar el depósito. Se trata de un .txt en el que se incluye la información necesaria para que personas ajenas a la investigación puedan comprender y utilizar los datos.

- Información general: título del dataset, equipo investigador (contacto, afiliaciones ...), fechas, financiación, descripción, palabras clave.
- Información para compartir y acceder a los datos: licencias, enlaces a las publicaciones relacionadas, propuesta de cita para los datos, etc..
- Información sobre la organización de archivos: listado de archivos y sus contenidos, versiones, tamaño, etc.
- Información metodológica: métodos utilizados para la recopilación, generación y procesamiento de los datos y software necesario para interpretar los datos.
- Información específica sobre los datos: contenido exacto, número de variables, listas de códigos, abreviaturas utilizadas, etc.

Asimismo, se recomienda aportar toda la documentación que permita reutilizar los datos, como:

- Codebook: conjunto de códigos, definiciones y ejemplos que se utilizan a menudo como guía para proporcionar contexto y ayudar a analizar los datos, esenciales para analizar la investigación cualitativa (por ejemplo, encuestas).
- Data dictionary: colección de nombres, definiciones y atributos sobre elementos de datos que se utilizan o captura, en una base de datos, sistema de información o parte de un proyecto de investigación.



En este [enlace](#) puedes encontrar un tutorial.

Buenas prácticas de trabajo con datos

Carpetas:

- Seguir procedimientos internos.
- Utilizar siempre las mismas estructuras de carpetas incluso entre proyectos.
- Denominar por áreas nunca por investigadores.
- Establecer un máximo de niveles de subcarpetas.
- Separar trabajos en curso de los ya finalizados.
- Evaluar periódicamente los materiales y eliminar archivos intermedios innecesarios.

Archivos:

- No utilizar nombres de ficheros demasiado largos.
- Evitar caracteres especiales ~ ¡ ! @ # \$ % ^ & * () ` ; < > ¿ ? , [] { } ' " |
- Guion bajo _ en lugar de espacio en blanco
- Indicar al final del nombre el número de versión, v01, v02, etc.
- Utilizar nombres únicos y numerar de forma continua: 001, 002, 003, etc.
- Comenzar con el nombre más común, el más específico detrás.
- Si hay que poner al autor especificar por iniciales.
- Evitar interfaz y programas intermedios para operaciones masivas, mejor con comandos.

Formatos:

- Bases de datos: XML, CSV.
- Texto: RFT, TXT, XML.
- Estadísticas: ASCII, DTA, POR, SAS, SAV.
- Datos tabulados: CSV, TSV.
- Geoespaciales: SHP, DBF, GeoTIFF, NetCDF.
- Vídeo: OGG, MP4.
- Sonido: FLAC, WAV, AIFF, MP3.
- Imágenes: TIFF, BMP.



- No se recomienda el uso de ficheros comprimidos. Hay formatos que pierden información en la compresión.

¿Cómo procedo si hay datos personales en el dataset?

Se requiere una atención especial al trabajo con datos personales y sensibles. Este tipo de datos solo se pueden depositar si se han anonimizado previamente. Se recomienda la consulta del documento: [Orientaciones y garantías en los procedimientos de anonimización de datos personales.](#)

Para cualquier aclaración aconsejamos la consulta personalizada en dos Servicios de apoyo disponibles:

- [Delegación de Protección de Datos \(DPD UPV\)](#): supervisión del cumplimiento del Reglamento General de Protección de Datos y de la Ley Orgánica de Protección de Datos Personales y garantía de los derechos digitales en el ámbito de la Universitat Politècnica de València.
- [Comité de Ética en Investigación de la UPV](#): cualquier actividad llevada a cabo en la UPV cuya experimentación esté afectada por alguna normativa, por incluir aspectos con implicaciones éticas o de bioseguridad, debe ser previamente autorizada por el Comité de Ética en Investigación de la UPV. El investigador principal debe seguir el procedimiento y cumplimentar el [formulario](#) correspondiente para solicitar la evaluación por parte del comité.

¿Debo depositar el Plan de Gestión de datos?

Es obligatoria su presentación en el [X PEICTI Plan Estatal de Investigación 2024-2027](#) y en Horizonte Europa, así como para todos los proyectos con financiación pública ([ENCA 2023-2027](#), Eje B). No lo es el depósito, no obstante, se considera una buena práctica de ciencia abierta.



Enlazar publicaciones y datos

La [Estrategia Nacional de Ciencia Abierta 2023-2027](#), en su Eje B “Gestión de datos de investigación siguiendo los principios FAIR”, Actuación 2 “Mejora de la comunicación” presta atención a la vinculación entre datos y publicaciones.

Para realizar esta vinculación cumplimentaremos metadatos enlazables tanto en la publicación (a los datos) como en los datos (a las publicaciones) permitiendo así la recolección y seguimiento de los datos y su alineación bajo los principios FAIR. La “Declaración de disponibilidad de los datos” en las publicaciones científicas debe contener la siguiente información:

- Repositorio donde se han depositado los datos
- DOI de los datos
- Licencias de uso de los datos

Los datos deberán llevar un metadato enlazable del tipo “Items relacionados” o similar con el DOI de la publicación. Para poder realizar esta práctica adecuadamente se deben depositar los datos previamente, para obtener el DOI e insertarlo en la publicación. Posteriormente se puede suministrar el DOI de la publicación al responsable del repositorio para la creación del metadato enlazable