

# Gestión de datos de investigación y Planes de Gestión de Datos

Virginia Jiménez Planet – Silvia Hernández Villegas  
Biblioteca Nacional de Ciencias de la Salud

Ciclo de Seminarios del Centro Nacional de Epidemiología  
29 de junio de 2023



*\* Para la preparación de este seminario se han utilizado los siguientes materiales docentes:*

Melero, Remedios. Gestión de datos de investigación y Planes de Gestión de Datos (PGD) para documentalistas. Formación online ISCIII del 5-22 de junio de 2023.

<https://repisalud.isciii.es/handle/20.500.12105/16205>

Melero, Remedios. Open Science. Gestión de Datos de Investigación. Curso de formación interna ISCIII, 16-23 de mayo 2023.

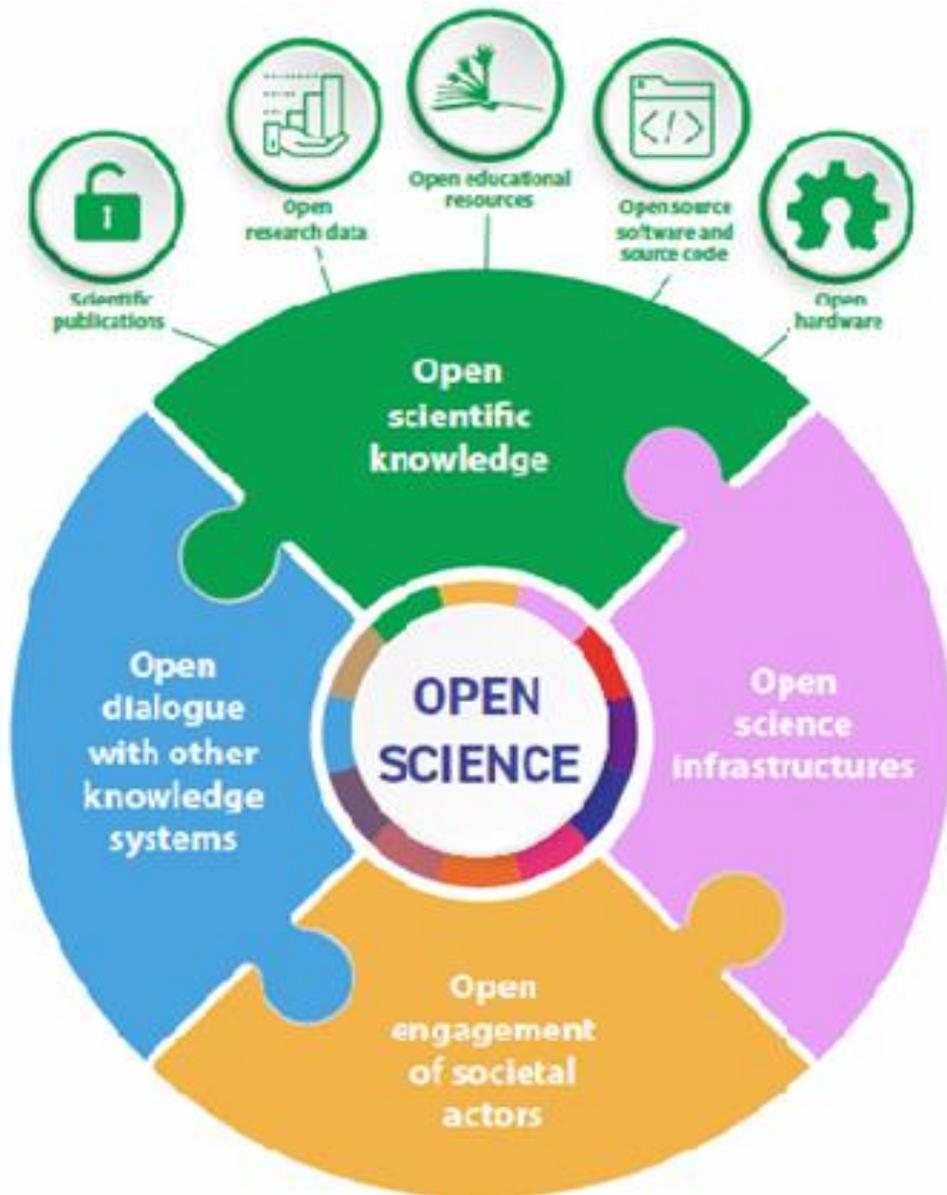
<https://repisalud.isciii.es/handle/20.500.12105/16204>



Remedios Melero. IATA-CSIC, Valencia, España



# Ciencia abierta



La **ciencia abierta** se define como un: Constructo inclusivo que combina varios movimientos y prácticas para hacer que el conocimiento científico multilingüe:

- Esté disponible **de forma abierta, accesible y reutilizable para todos**, aumentar las **colaboraciones** científicas y el **intercambio** de información en **beneficio** de la ciencia y la sociedad, y **abrir los procesos de creación, evaluación y comunicación** del conocimiento científico a los actores de la sociedad más allá de la comunidad científica tradicional.

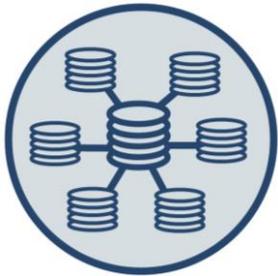
Comprende todas las disciplinas científicas y se basa en los siguientes pilares clave:

- conocimiento científico abierto
- infraestructuras científicas abiertas
- compromiso abierto de los actores sociales
- diálogo abierto diálogo con otros sistemas de conocimiento

# Aclarando conceptos



**Datos:** Información cuantitativa o cualitativa recogida por los investigadores en el curso de su trabajo. Facilitan la información necesaria para apoyar o validar los resultados o conclusiones de la investigación



**Datasets:** conjunto de los datos tabulados en cualquier sistema de almacenamiento de datos estructurado



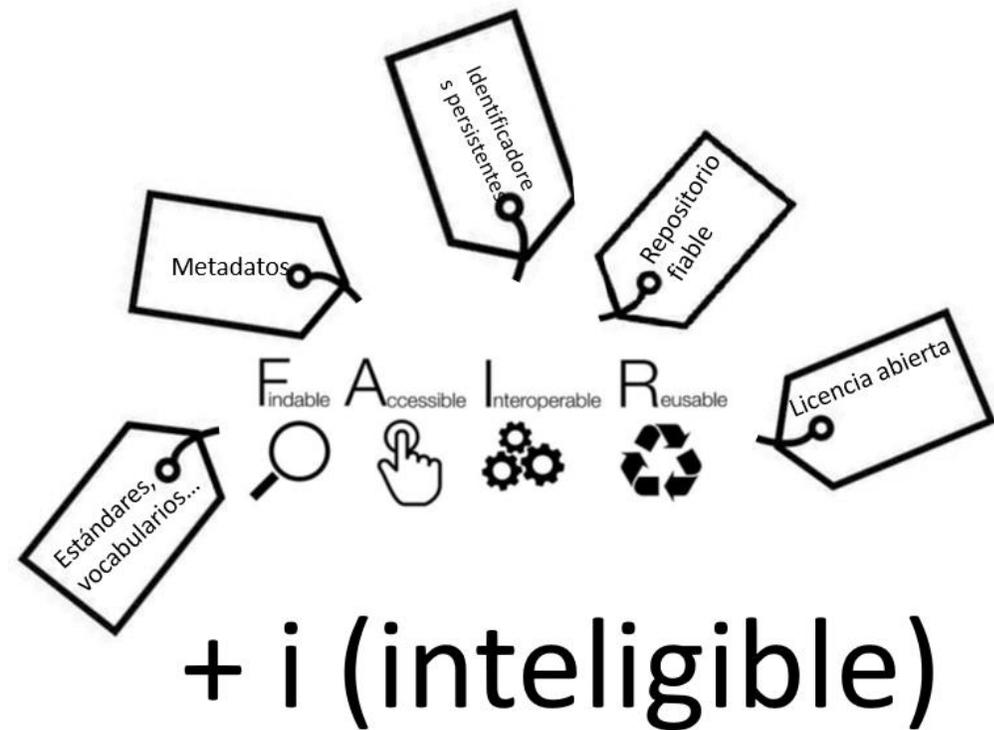
**Plan de gestión de datos (PGD):** documento/informe/guía que sirve para planificar, organizar y documentar cómo se van a obtener o se han obtenido los datos en el marco de un proyecto de investigación. Es un documento vivo que va modificándose o actualizándose en función del desarrollo de una investigación. Es recomendable hacer un plan de gestión de datos al inicio del proyecto e ir creando versiones nuevas en función de los posibles cambios que se deriven durante su ejecución.

# Gestión de datos de investigación

**Gestión de datos de investigación:** tareas realizadas durante y después de la investigación para manejar los datos:

- Recopilación (qué)
- Organización (cómo)
- Almacenamiento (dónde)
- Documentación (cómo)
- Preservación (cómo)
- Puesta en Circulación (cómo, dónde)

Una buena gestión de los datos ayuda a garantizar que los investigadores compartan sus datos de forma **FAIR** (localizables, accesibles, interoperables y reutilizables).



# Ciclo de vida de los datos



## Beneficios de compartir datos

- Promueve la innovación y potenciales nuevos usos.
- Promueve la colaboración entre los usuarios y los creadores de datos.
- Maximiza la transparencia y la fiabilidad.
- Permite la verificación de los resultados de investigación.
- Reduce costes al evitar duplicación de datos.
- Aumenta el impacto y la visibilidad de la investigación.
- Promueve la investigación de dónde salieron los datos y sus publicaciones.
- Puede generar un reconocimiento directo a los investigadores como cualquier otro resultado de la investigación.
- Genera nuevos datos a partir de los originales.

# Políticas de acceso a los datos de investigación

**Ley 17/2022, de 5 de septiembre, por la que se modifica la Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación**  
<https://www.boe.es/eli/es/l/2022/09/05/17/con>

«Artículo 37. *Ciencia abierta.*

1. ...
2. 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 su publicación y los datos asociados a las mismas en repositorios institucionales o temáticos de acceso abierto, de forma simultánea a la fecha de publicación.**
3. Los beneficiarios de ayudas y subvenciones públicas se asegurarán de que **conservan los derechos de propiedad intelectual necesarios para dar cumplimiento a los requisitos de acceso abierto.**

## Depósito

<b>¿Qué?</b>	Versión aceptada del artículo + datos subyacentes
<b>¿Cuándo?</b>	En el momento de su publicación
<b>¿Dónde?</b>	Repositorios institucionales o temáticos de acceso abierto

**Artículo 36 bis.** Aplicación del Derecho privado a las transmisiones a terceros de derechos sobre los resultados de la actividad investigadora por Organismos Públicos de Investigación, universidades públicas y entidades dependientes de la Administración General del Estado.

### Punto 5, letra c

*Reserva por la entidad titular de una **licencia no exclusiva**, intransferible y gratuita de uso limitada a actividades docentes, sanitarias y de investigación, siempre que la actividad carezca de ánimo de lucro.*

# Políticas de acceso a los datos de investigación

## EJES ESTRATÉGICOS

La *Estrategia* se implementará a través de medidas concretas que garanticen la implicación efectiva, la sensibilización, la formación y la capacitación del personal que habrá de ser artífice de este cambio cultural: personal investigador, de apoyo, de gestión y ciudadanía. Estas medidas deberán estar adecuadamente coordinadas entre los agentes de financiación, ejecución y evaluación de la actividad investigadora y respaldadas por el marco legal vigente y la acción del Gobierno de España en su conjunto.

Los ejes estratégicos sobre los que se estructura la ENCA son los siguientes:

A

Infraestructuras digitales para la ciencia abierta.

B

Gestión de datos de investigación siguiendo los principios FAIR.

C

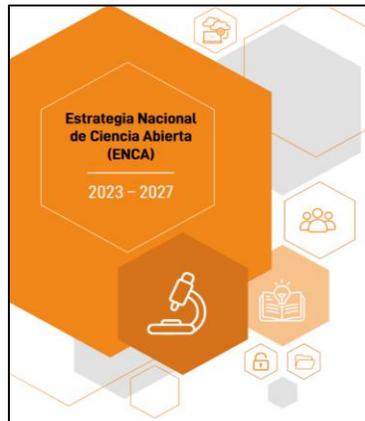
Acceso abierto a publicaciones científicas.

D

Incentivos, reconocimientos y formación.

D2. Consideración de las prácticas de ciencia abierta para la financiación pública de la I+D

**CÓMO:** Las convocatorias de ayudas y subvenciones para la realización de proyectos de investigación establecerán requerimientos de ciencia abierta para las instituciones beneficiarias cuyo adecuado cumplimiento será objeto de seguimiento. Además, en el proceso de evaluación de los proyectos de investigación que soliciten financiación pública, se considerará, como un **criterio puntuable**, la **incorporación de prácticas de ciencia abierta en el diseño y la ejecución del proyecto**. Para la concesión de otras subvenciones públicas, incluidos los programas de ayudas para recursos humanos y fortalecimiento institucional, se incluirá **la evaluación de prácticas de ciencia abierta en la evaluación de méritos científicos y curriculares individuales y/o institucionales**.



# Políticas de acceso a los datos de investigación



## Ciencia abierta a lo largo del programa

Ciencia  
abierta

Integración de las prácticas de ciencia abierta para mejorar la calidad y la eficiencia de la I+i y el compromiso activo de la sociedad.

**Acceso abierto obligatorio e inmediato a las publicaciones:** los beneficiarios deben conservar los derechos de la propiedad intelectual necesarios para cumplir los requisitos de acceso abierto.  
**Intercambio de datos según el principio «tan abierto como sea posible y tan cerrado como sea necesario»:** plan obligatorio de gestión de datos de investigación FAIR (fáciles de encontrar, accesibles, interoperables y reutilizables).

- Los programas de trabajo pueden fomentar o imponer la adhesión a las **prácticas de ciencia abierta**, como la participación de los ciudadanos, o el uso de la **Nube Europea de la Ciencia Abierta**.
- Evaluación de las prácticas de ciencia abierta mediante **criterios de adjudicación** para la evaluación de propuestas.
- Apoyo específico a **medidas relativas a la política de ciencia abierta**.
- Plataforma de publicación **Open Research Europe**.

# Políticas de acceso a los datos de investigación

Resolución legislativa del Parlamento Europeo, de 4 de abril de 2019, sobre la Directiva del Parlamento Europeo y del Consejo relativa a la reutilización de la información del sector público (versión refundida) (COM(2018)0234 – C8-0169/2018 – 2018/0111(COD))

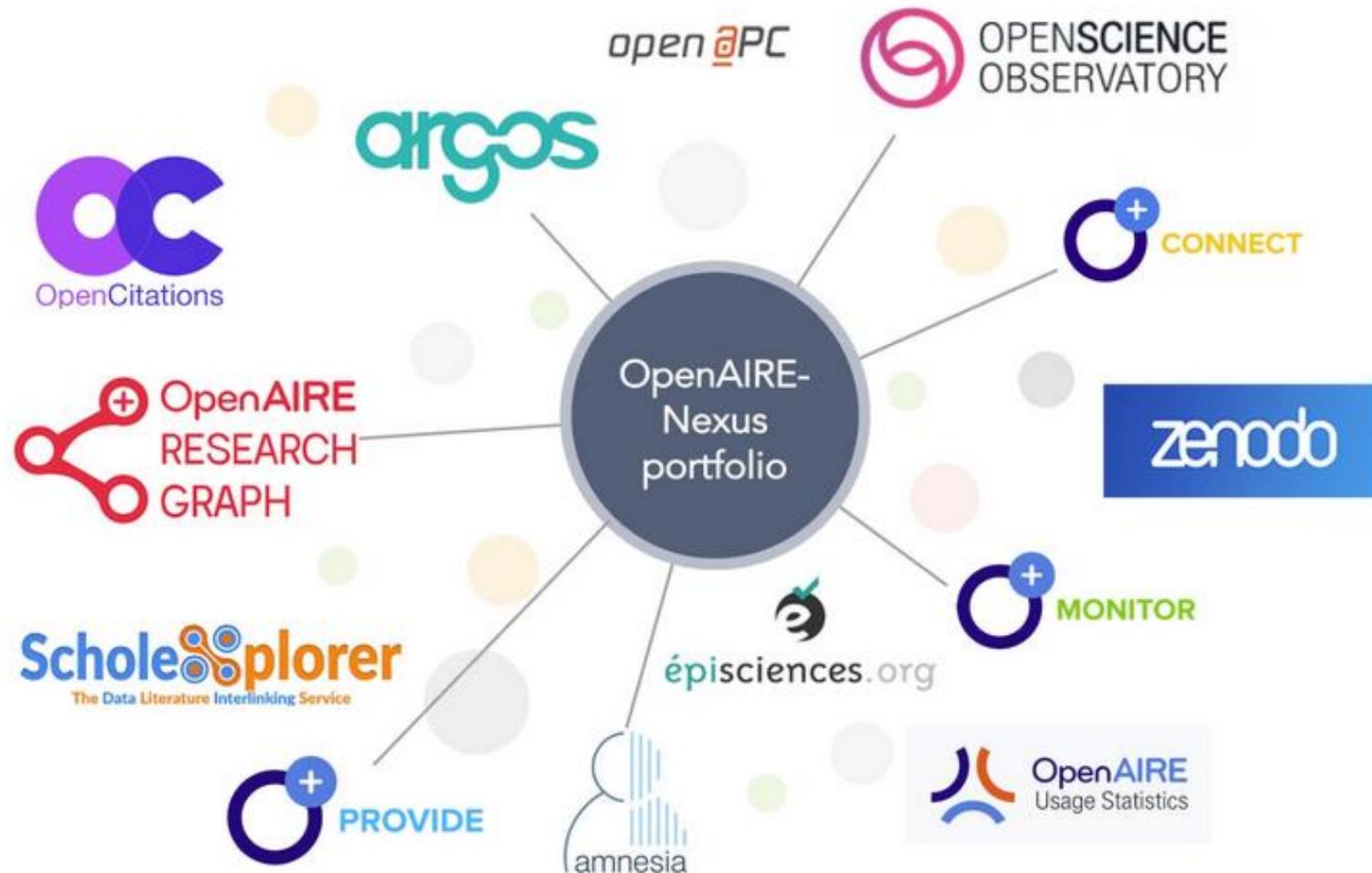
*“Con arreglo a las políticas nacionales de acceso abierto, los datos de la investigación financiada públicamente serán abiertos por defecto. Sin embargo, en este contexto, deben tenerse debidamente en cuenta las inquietudes relacionadas con la privacidad, la protección de datos personales, la confidencialidad, la seguridad nacional, los intereses comerciales legítimos, como los secretos comerciales, y los derechos de propiedad intelectual de terceros, conforme al principio «tan abiertos como sea posible, tan cerrados como sea necesario». Por otra parte, los datos de investigación a los que no puede accederse por motivos de seguridad nacional, defensa o seguridad pública no deben estar cubiertos por la presente Directiva.”*

.....

*“Los Estados miembros pueden hacer extensiva la aplicación de la presente Directiva a los datos de investigación puestos a disposición del público a través de **infraestructuras de datos distintas de los registros, a través de publicaciones de acceso abierto o en forma de fichero adjunto a un artículo, a un artículo de datos o a un artículo en una revista especializada en datos.** Los documentos distintos de los datos de investigación deben seguir estando excluidos del ámbito de la presente Directiva.”*

# Servicios disponibles para el cumplimiento de la Ciencia Abierta

<https://www.openaire.eu/openaire-nexus-public-launch-event>



# Guiding you in Open Science



## Researchers

### Horizon Europe

How to comply with Horizon Europe mandate for Research Data Management **New!**

How to comply with Horizon Europe mandate for publications **New!**

Open Science in Horizon Europe proposal **New!**

RDM in Horizon Europe Proposals **New!**

### Research Data Management

How to comply with H2020 mandates - research data

Research Data Management costs in H2020 projects

RDM service development checklist

How to make your data FAIR

How to find a trustworthy repository for your data

England, Jonathan, & Malaguarnera, Giulia. (2022, November 22). Horizon Europe Open Science requirements in practice - OpenAIRE webinar. Zenodo. <https://doi.org/10.5281/zenodo.7324364>

- PGD: idea básica en la propuesta, 1ª versión 6 meses, se actualizará a mitad y al final del proyecto.
- Depositar los (meta)datos lo antes posible tras su producción/generación o tras su procesamiento y controles de calidad.
- Depositar los datos en un repositorio de confianza y abrirlos lo antes posible (plazos PGD).
- Datos cerrados si es necesario, pero los metadatos deben ser FAIR y estar bajo licencia CC-BY o CC0
- Justificación válida para no abrir los datos: valor comercial, dificultar la protección de la PI de los resultados, normas de protección de datos/privacidad de datos sensibles y/o personales.
- PGD: No hay respuestas “correctas”, sea claro, específico y detallado... y justifique las decisiones. Costes subvencionables (almacenamiento, procesamiento, preservación) en términos de tiempo y recursos necesarios.
- El plan de gestión de datos sirve para demostrar al financiador que el investigador se ha tomado el tiempo necesario para reflexionar sobre lo que debe hacer, que ha reflexionado y que su enfoque parece razonable. Y que los datos son "tan abiertos como sea posible, tan cerrados como sea necesario" (principios FAIR).

## Costs of Research Data Management

Online RDM-costing tool <https://www.openaire.eu/estimating-costs-rdm-tool>

O'Connor, Ryan, Delipalta, Alexandra, & Jones, Sarah. (2020). What will it cost to manage and share my data?. Zenodo. <https://doi.org/10.5281/zenodo.4548344>

## Toolkit for researchers on legal issues

How do I know if my research data is protected? <https://www.openaire.eu/how-do-i-know-if-my-research-data-is-protected>

How do I license my research data? <https://www.openaire.eu/how-do-i-license-my-research-data>

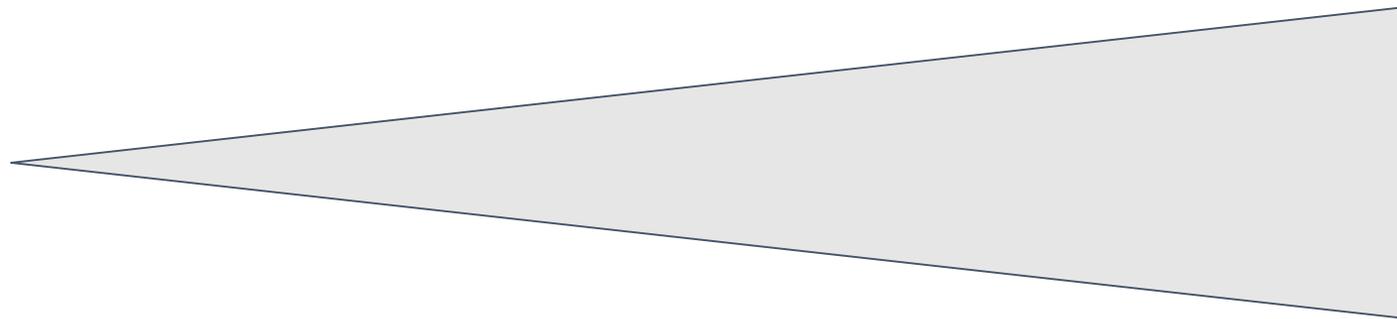
Can I reuse someone else's research data? <https://www.openaire.eu/can-i-reuse-someone-else-research-data>

# Cómo nombrar de forma normalizada nuestros datos

- Incluir los datos en formato estandarizado( p.e. fecha, aaaa-mm-dd)
- Utilizar abreviaturas significativas
- Indicar tipo de fichero
- Usar identificadores de personas/grupo
- Indicar versión

**Fecha**                      **Experimento**                      **ID**  
20130825\_DOEProject\_Ex1Test1\_Data\_Gonzalez\_v3-03.xlsx  
**Proyecto**                      **Tipo**                      **Versión**

General



Específico

## Recomendaciones respecto a los formatos de los ficheros

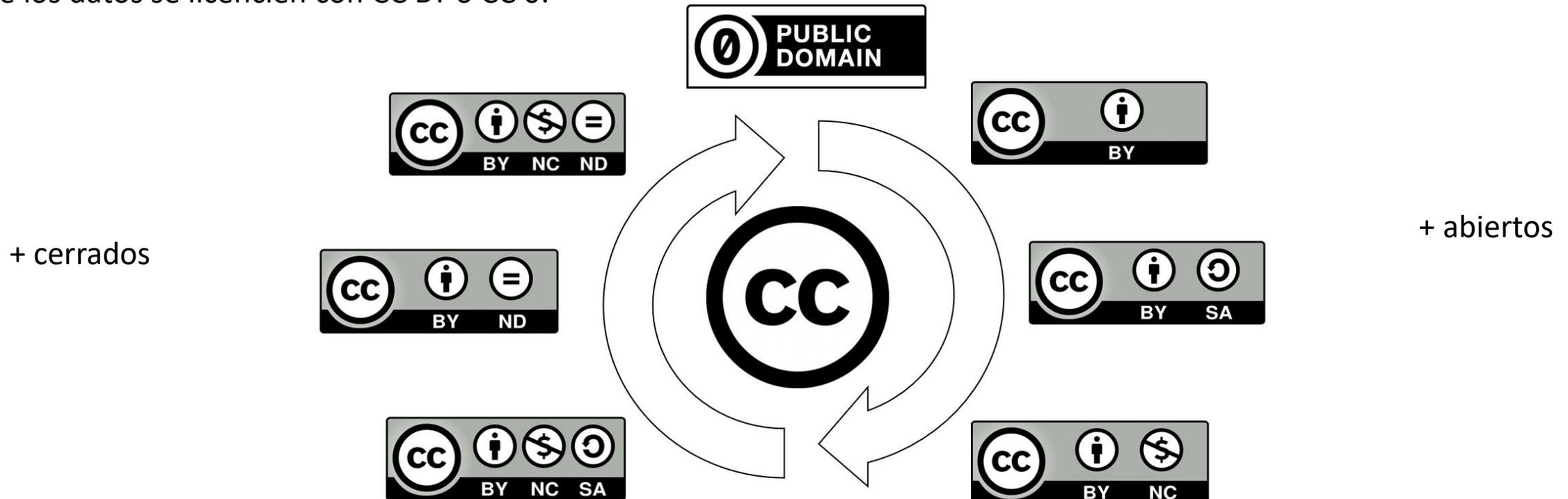
- Publicar los datos en formatos abiertos, no propietarios y legibles por máquina, si es posible.
- Guarde los archivos en formatos abiertos (por ejemplo, data.csv) para mejorar la usabilidad, si procede.
- Comprima los datos para facilitar el intercambio y la descarga de archivos grandes. Compruebe los formatos de archivos comprimidos (zip) en la investigación.
- Mantenga el mismo nombre de archivo para el mismo archivo en diferentes formatos (por ejemplo, data.doc y data.txt).
- Incluya la documentación de los datos, las transformaciones y las directrices del software para acceder al formato propietario.
- Considere la posibilidad de compartir el código de software alternativo a los ejecutables propietarios.
- Compruebe si hay errores u omisiones si convierte el archivo a un formato diferente.

# Licencie sus datos

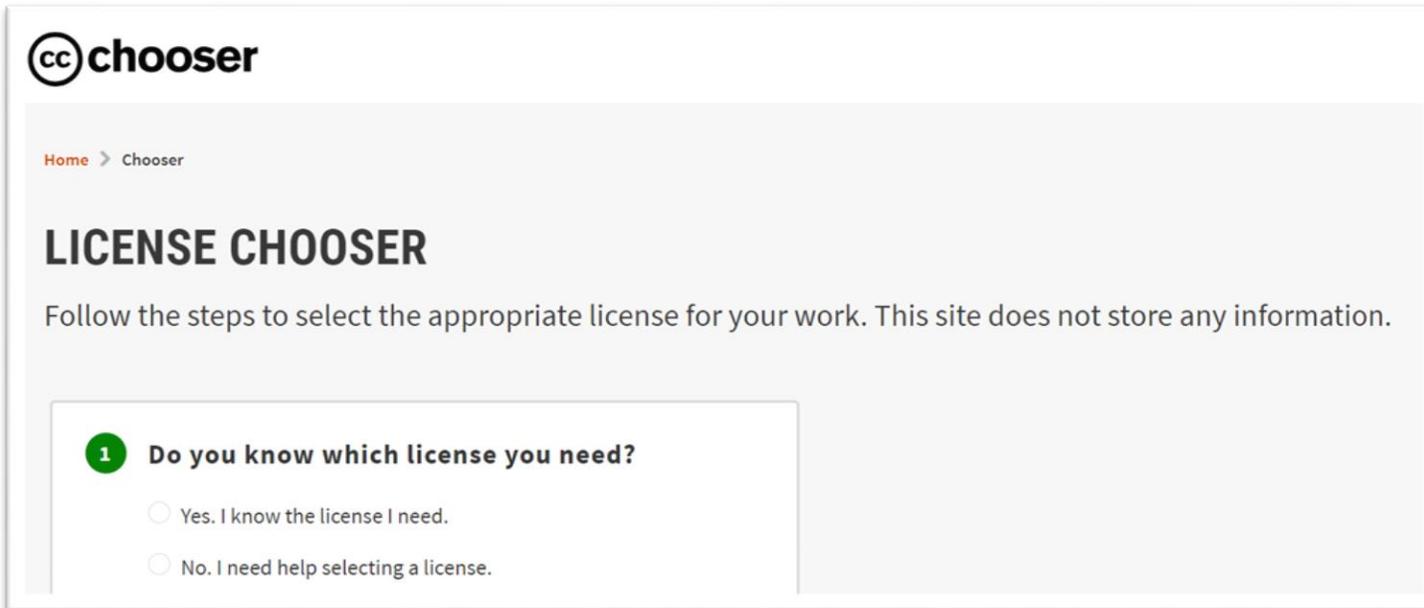
Licencia: Documento por el cual el licenciador (de origen el autor) establece los términos de la explotación/uso de la obra (reproducción, distribución, comunicación pública, transformación).

Las licencias más usadas en ciencia abierta son las Licencias Creative Commons

**Siguiendo el lema de la legislación europea “tan abiertos como sea posible, tan cerrados como sea necesario”** se recomienda que los datos se licencien con CC BY o CC 0.

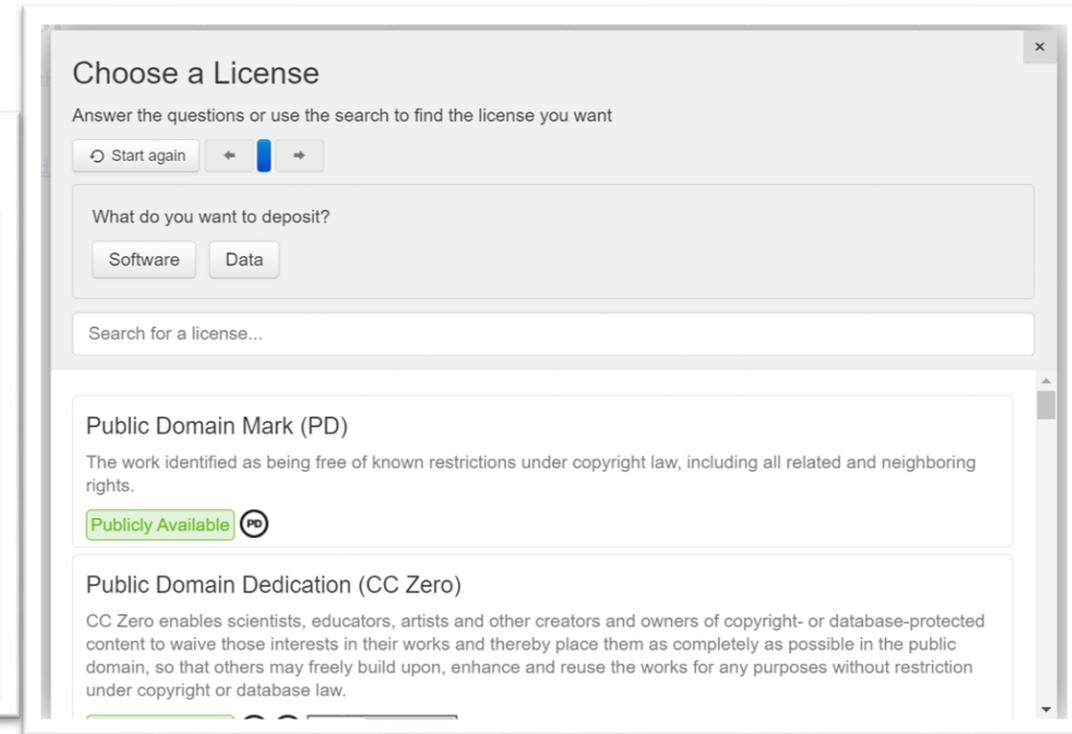


# Herramienta de ayudar para asignar el tipo de licencia:



The screenshot shows the main page of the CC Chooser tool. At the top left is the 'CC chooser' logo. Below it, a breadcrumb trail reads 'Home > Chooser'. The main heading is 'LICENSE CHOOSER' in large, bold, black letters. Underneath, a paragraph states: 'Follow the steps to select the appropriate license for your work. This site does not store any information.' A green circle with the number '1' is followed by the question 'Do you know which license you need?'. Below this question are two radio button options: 'Yes. I know the license I need.' and 'No. I need help selecting a license.'

<https://chooser-beta.creativecommons.org/>



The screenshot shows a 'Choose a License' dialog box. At the top, it says 'Choose a License' and 'Answer the questions or use the search to find the license you want'. Below this are navigation buttons: 'Start again', a left arrow, a blue vertical bar, and a right arrow. A section titled 'What do you want to deposit?' contains two buttons: 'Software' and 'Data'. Below that is a search input field with the placeholder text 'Search for a license...'. The dialog lists two license options: 'Public Domain Mark (PD)' and 'Public Domain Dedication (CC Zero)'. The 'Public Domain Mark (PD)' option is highlighted with a green border and includes a 'Publicly Available' label and a 'PD' icon. The 'Public Domain Dedication (CC Zero)' option is also visible with its description.

<https://ufal.github.io/public-license-selector/>



# 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:

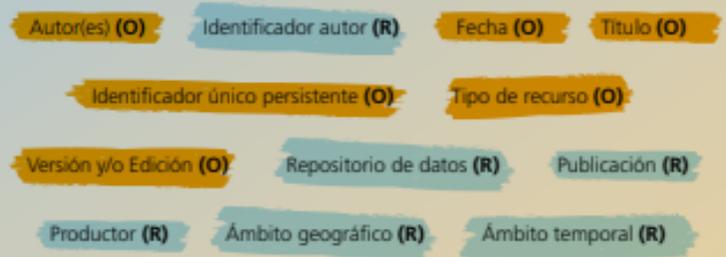
- Se debe facilitar la identificación, localización y el acceso a los datos mediante un identificador único y persistente (DOI, Handle, etc.).
- Cada conjunto y subconjunto de datos (dataset) debe citarse de forma independiente.
- Las citas de los datos utilizados han de aparecer en la sección de referencias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



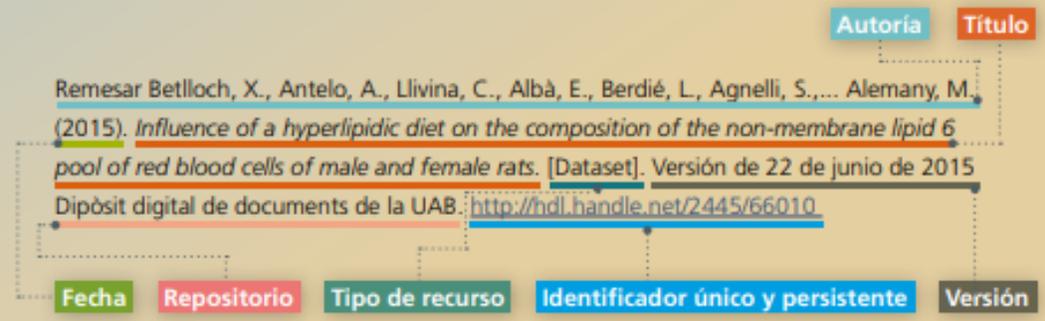
Enlaza los datos con los documentos resultado de investigación y viceversa, y crea las referencias bibliográficas de los mismos.

## 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.).



## Ejemplo de cita estilo APA





## Plantilla readme.txt

Readme.txt ([acceso](#) [descarga](#)) es una plantilla para que rellenen los investigadores que quieran depositar un dataset en e-cienciaDatos.

Todos los conjuntos de datos han de estar documentados para ser comprensibles y reutilizables: autoría, título, descripción, metodología, proyectos financiadores, cobertura temporal y geográfica, derechos de uso y privacidad, etc.

Los gestores del repositorio e-cienciaDatos utilizarán la información proporcionada en la plantilla readme.txt para:

- cumplimentar los metadatos estandarizados que describen el dataset y que son recolectados por diversos servidores científicos.
- elaborar un fichero readme.txt que se incluirá junto al resto de ficheros del dataset.

Formulario "readme" en e-cienciaDatos

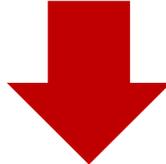
```
-----  
INFORMACIÓN GENERAL  
-----  
1. Título del dataset  
  
2. Contacto  
  
   Investigador/a de contacto  
   Nombre:  
   Filiación:  
   Correo electrónico:  
   ORCID:  
  
3. Descripción del proyecto  
  
4. Descripción del dataset  
  
5. Notas  
  
6. Fecha de depósito de los ficheros  
  
7. Fecha de creación de los ficheros  
  
8. Idioma
```

Depurar los datos  
Documentar los datos  
Anonimizar (si corresponde)  
Licenciar  
p.e.  
Enriquecer  
estandarizados

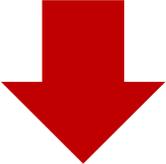
<https://openrefine.org/>  
Readme.txt  
<https://amnesia.openaire.eu/>  
Licencias Creative Commons,  
Vocabularios/tesaurus



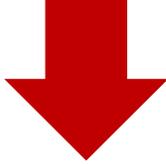
**Elige repositorio**



**Deposita + Añade metadatos**



**Distribuye en abierto**



**Cita los datos**



## LISTAR

### Todo Repisalud

- ▶ Comunidades y Colecciones
- ▶ Por fecha de publicación
- ▶ Autores
- ▶ Títulos
- ▶ Tipo de documento
- ▶ Institución
- ▶ Palabras clave
- ▶ MeSH
- ▶ Agencias financiadoras

### Esta comunidad

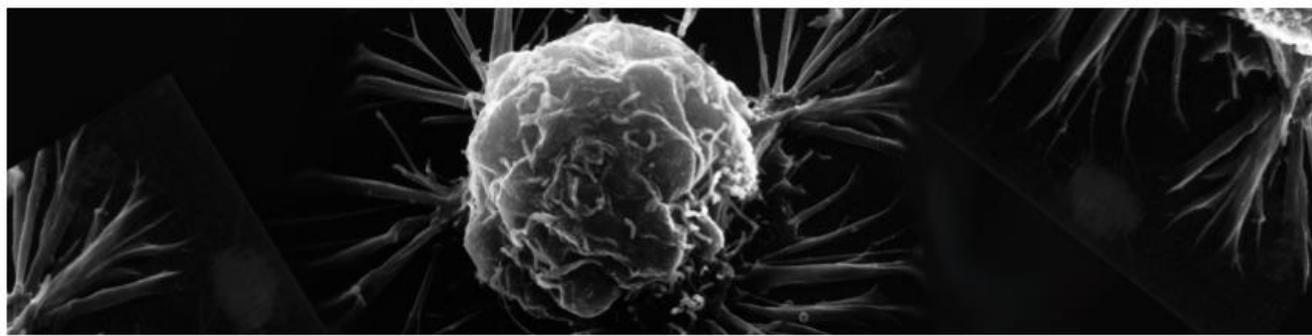
- ▶ Por fecha de publicación
- ▶ Autores
- ▶ Títulos
- ▶ Tipo de documento

Inicio | Sobre Repisalud | Info autores | FAQs | Login | Contacto/Sugerencias español

Búsquedas

Repisalud Principal / Investigación

## Investigación



Artículos, contribuciones a congresos, datos de investigación, libros, capítulos de libro, patentes, tesis y trabajos de investigación.

### Subcomunidades en esta comunidad

- ISCIII** [3528]  
Instituto de Salud Carlos III
- CNIC** [1337]  
Centro Nacional de Investigaciones Cardiovasculares
- CNIO** [465]  
Centro Nacional de Investigaciones Oncológicas
- IIS** [595]  
Institutos de Investigación Sanitaria

<b>Título</b>	Social engagement within the facility increased life expectancy in nursing home residents: a follow-up study [Dataset]
<b>Autor(es)</b>	Pastor-Barriuso, Roberto <a href="#">ISCIII</a>   Padron-Monedero, Alicia <a href="#">ISCIII</a>   Parra-Ramírez, Lina M. <a href="#">ISCIII</a>   García Lopez, Fernando Jose <a href="#">ISCIII</a>   Damian, Javier <a href="#">ISCIII</a>
<b>Fecha de publicación</b>	2020
<b>Cita</b>	Pastor-Barriuso, R, Padron-Monedero, A, Parra-Ramírez, LM, García Lopez, FJ, Damian, J. (2020). Social engagement within the facility increased life expectancy in nursing home residents: a follow-up study [Dataset]. <a href="http://hdl.handle.net/20.500.12105/11202">http://hdl.handle.net/20.500.12105/11202</a>
<b>Idioma</b>	Inglés
<b>Resumen</b>	Social engagement (SE) has been consistently shown to improve survival among community-dwelling older people, but the evidence in nursing home residents is inconclusive and prone to short-term reverse causation and confounding by major health determinants. A representative cohort of 382 nursing home residents in Madrid without severe physical and cognitive impairments at baseline was followed up for 10-year all-cause mortality. Standardized
<b>Tabla de contenidos</b>	"Pastor-Barriuso_et_al_2020_BMC_Geriatrics_Social_engagement_and_mortality_Data.csv"; "Pastor-Barriuso_et_al_2020_BMC_Geriatrics_Social_engagement_and_mortality_Stata_dofile.txt"; "Pastor-Barriuso_et_al_2020_BMC_Geriatrics_Social_engagement_and_mortality_R_script.txt"; "Pastor-Barriuso_et_al_2020_BMC_Geriatrics_Social_engagement_and_mortality_README.txt"
<b>Palabras clave</b>	<a href="#">Cohort study</a>   <a href="#">Inverse probability weighting</a>   <a href="#">Mortality</a>   <a href="#">Nursing homes</a>   <a href="#">Social engagement</a>
<b>Descripción</b>	Study data, STATA do-file, and R script to study the potential causal effect of social engagement on long-term, all-cause mortality in care home residents <a href="https://repisalud.isciii.es/handle/20.500.12105/11382">https://repisalud.isciii.es/handle/20.500.12105/11382</a> [Artículo] <a href="https://repisalud.isciii.es/handle/20.500.12105/9378">https://repisalud.isciii.es/handle/20.500.12105/9378</a> [Preprint]
<b>DOI</b>	<a href="https://doi.org/10.4321/repisalud.11202">10.4321/repisalud.11202</a>
<b>Ubicación geográfica</b>	Madrid
<b>Aparece en las colecciones</b>	<a href="#">Investigación &gt; ISCIII &gt; Centro Nacional de Epidemiología (CNE) &gt; ISCIII - Datos de investigación</a>

**Ficheros en el ítem**

	<b>Nombre:</b> Pastor-Barriuso_et_al_2020_BMC ...	<a href="#">Visualizar/Abrir</a>
	<b>Tamaño:</b> 26.79Kb	
	<b>Formato:</b> Fichero CSV	
	<b>Nombre:</b> Pastor-Barriuso_et_al_2020_BMC ...	<a href="#">Visualizar/Abrir</a>
	<b>Tamaño:</b> 14.53Kb	
	<b>Formato:</b> Fichero de texto	
	<b>Nombre:</b> Pastor-Barriuso_et_al_2020_BMC ...	<a href="#">Visualizar/Abrir</a>
	<b>Tamaño:</b> 16.40Kb	
	<b>Formato:</b> Fichero de texto	
	<b>Nombre:</b> Pastor-Barriuso_et_al_2020_BMC ...	<a href="#">Visualizar/Abrir</a>
	<b>Tamaño:</b> 3.178Kb	
	<b>Formato:</b> Fichero de texto	



Este ítem está sujeto a una licencia Creative Commons: Atribución-NoComercial-Compartidguial 4.0 Internacional

[Mostrar el registro completo del ítem](#)

**Ficheros en el ítem**



<b>Título</b>	Effects of 700 and 3500 MHz 5G radiofrequency exposure on developing zebrafish embryos [Dataset]
<b>Autor(es)</b>	Torres-Ruiz, Mónica <a href="#">ISCIII</a>   Suárez, Oscar J.   Lopez-Alonso, Victoria <a href="#">ISCIII</a>   Marina-Boillos, Pablo <a href="#">ISCIII</a>   Sanchis Otero, Aranzazu <a href="#">ISCIII</a>   Liste-Noya, Isabel <a href="#">ISCIII</a>   De Alba-González, Mercedes <a href="#">ISCIII</a>   Ramos-Gonzalez, María Victoria <a href="#">ISCIII</a>
<b>Fecha de publicación</b>	2023-05
<b>Idioma</b>	Inglés
<b>Resumen</b>	En este estudio, hemos utilizado una configuración disponible comercialmente validada que produce un campo uniforme para exponer embriones de pez cebra (ZFe) a frecuencias no moduladas de 700 y 3500 MHz. Hemos combinado una batería de ensayos de toxicidad, desarrollo y comportamiento para explorar más a fondo los posibles efectos de las radiaciones de radiofrecuencia (RFR). Nuestros perfiles neuroconductuales incluyen un ensayo de
<b>Palabras clave</b>	<a href="#">5G</a>   <a href="#">Telecomunicaciones</a>   <a href="#">Radiación no ionizante</a>   <a href="#">Desarrollo modelo animal</a>   <a href="#">Toxicidad</a>   <a href="#">Comportamiento</a>   <a href="#">Embriones</a>   <a href="#">Pez cebra (ZFe)</a>   <a href="#">Radiaciones de radiofrecuencia (RFR)</a>
<b>DOI</b>	<a href="https://doi.org/10.4321/repisalud.16125">10.4321/repisalud.16125</a>
<b>Fecha inicio</b>	2021-10
<b>Aparece en las colecciones</b>	<a href="#">Investigación &gt; ISCIII &gt; Unidad de Investigación en Salud Digital (UITeS) &gt; ISCIII - Datos de investigación</a>

[Mostrar el registro completo del ítem](#)

**Ficheros en el ítem**



<b>Nombre:</b> Dataset_mayo_2023.pdf	<a href="#">Visualizar/Abrir</a>
<b>Tamaño:</b> 528.4Kb	
<b>Formato:</b> PDF	



<b>Descripción:</b> Dataset	
<b>Nombre:</b> Controls exposition October ... <a href="#">Embargado hasta: 2023-12-31</a>	<a href="#">Solicitar copia</a>
<b>Tamaño:</b> 6.140Mb	
<b>Formato:</b> Microsoft Excel 2007	



<b>Descripción:</b> Octubre 2021	
<b>Nombre:</b> Controls exposition Febrery 2022 ... <a href="#">Embargado hasta: 2023-12-31</a>	<a href="#">Solicitar copia</a>
<b>Tamaño:</b> 12.44Mb	
<b>Formato:</b> Microsoft Excel 2007	
<b>Descripción:</b> Febrero 2022	

# ¿Mis datos son FAIR?...



<https://ardc.edu.au/resources/working-with-data/fair-data/fair-self-assessment-tool/>

<https://fairaware.dans.knaw.nl/>



<https://satisfyd.dans.knaw.nl/>

SATIFYD



<https://wrco.ufpb.br/fair/>



## SATIFYD

### Self-Assessment Tool to Improve the FAIRness of Your Dataset

Welcome to SATIFYD: the DANS Self-Assessment Tool to Improve the FAIRness of Your Dataset. This tool will show you how FAIR (Findable, Accessible, Interoperable, Reusable) your dataset is and will provide you with tips to score (even) higher on FAIRness. Ideally, you use this tool prior to the deposit in EASY.

The 12 questions touch upon the FAIR data principles but do not strictly follow them. While answering the questions, the score per letter will be displayed underneath each letter. The more 'blue' the letters get, the more FAIR your dataset is. An overall score is provided at the end of the page.

Some questions are posed more than once (e.g. on metadata and data standards or usage licences), because the topics are relevant in more than one letter.

Want to know more? Please click [here](#) ←

If you have any questions, please let us know by sending an e-mail ✉



#### FINDABLE ?

1. Did you provide sufficient metadata (information) about your data for others to find, understand and reuse your data? ?

2. Did you use standards such as controlled vocabularies, taxonomies (thesauri) or ontologies to describe your dataset? ?

3. Did you provide rich and detailed additional documentation? ?

Controlled vocabularies

Taxonomies (thesauri)

Ontologies

There are no standards for my discipline

Readme file

Versioning

Provenance

#### ACCESSIBLE ?

4. Is the metadata publicly accessible even if the data is no longer available? ?

Yes  No  I can't find this information in EASY

F

A

I

R

## FAIR questions

### FINDABLE

1. Are you aware that a dataset should be assigned a globally unique persistent and resolvable identifier when deposited with a data repository?   Yes  No
2. Are you aware that when you deposit a dataset with a repository, you will need to provide some details (known as discovery metadata) in order to make the data findable, understandable and reusable to others?   Yes  No
3. Are you aware that the repository providing access to your dataset should make the metadata describing your datasets available in a format readable by machines as well as humans?   Yes  No

### ACCESSIBLE

4. Are you aware that access to your dataset may need to be controlled and that metadata should include licence information under which the data can be reused?   Yes  No
5. Are you aware that metadata should remain available over time, even if the data is no longer accessible?   Yes  No

### INTEROPERABLE

6. Are you aware that the metadata describing your datasets should use semantic vocabularies?   Yes  No

### REUSABLE

7. Are you aware that provenance information about the collection and/or generation of data should be included in the metadata?   Yes  No
8. Are you aware that metadata describing your data should follow the specifications of a community-endorsed standard?   Yes  No
9. Are you aware that data should be deposited preferably in a file format that is open – to support reuse – and supported by the repository for long-term preservation?   Yes  No
10. Are you aware that maintaining your dataset FAIR over time requires professional data curation and preservation?   Yes  No

## Formulario de FAIR AWARE

### Feedback

Please answer the four questions below to help us improve the tool and make it even more relevant for the community.

- Which of the following issues do you find hard to understand and/or to answer?

1. Globally unique persistent identifier (PID) 
2. Metadata for citation and discovery including PID 
3. Metadata available online is readable by humans and machines 
4. Metadata includes licence, level of access and conditions to access the data 
5. Persistence of metadata 
6. Use of controlled vocabularies in metadata 
7. Metadata includes provenance 
8. Community-endorsed metadata 
9. Data in a preferred format for reuse & preservation 
10. Digital curation and preservation 

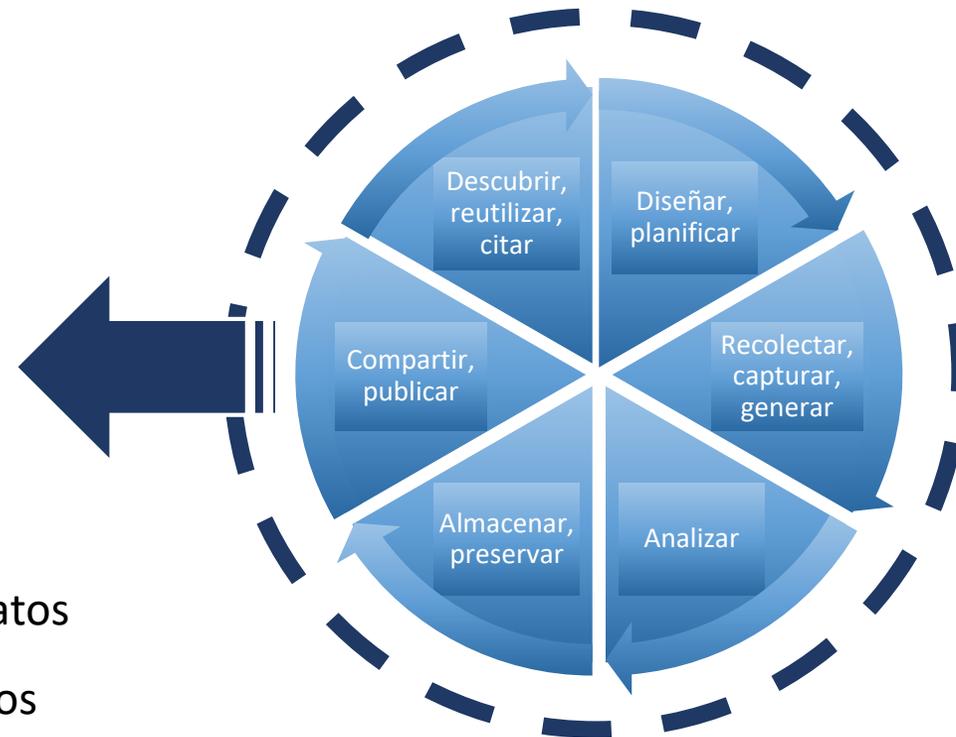
- Are there any issues relevant to your discipline and/or needed for enabling reusability of your dataset that are missing from this assessment? (Please do not include any personal data, such as your name and email address in your response.)

- Please submit any other feedback on how we might improve the FAIR-Aware assessment tool, including possible additional guidance. (Please do not include any personal data, such as your name and email address in your response.)

# ¿Qué debe contemplar un plan de gestión de datos?

(mínimos)

- Contexto
- Descripción de los datos que se van a tomar o crear
- La metodología y estándares para la recolección de datos
- Aspectos éticos y relacionados con la propiedad intelectual, si corresponde
- Vías para compartir y acceder a los datos
- Estrategia para la preservación de datos



**Las instituciones o agencias financiadoras pueden tener especificaciones propias**

# Aspectos básicos a responder en un plan de gestión de datos (1)

## 1. Descripción de los datos y recogida o reutilización de los datos existentes.

- a. ¿Cómo se recogerán o generarán los nuevos datos y/o cómo se reutilizarán los datos existentes?
- b. ¿Qué datos (por ejemplo, el tipo, los formatos y los volúmenes) se recogerán o producirán?

## 2. Documentación y calidad de los datos

- a. ¿Qué metadatos y documentación (por ejemplo, la metodología de recogida de datos y la forma de organizarlos) acompañarán a los datos?
- b. ¿Qué medidas de control de calidad de los datos se utilizarán?

## 3. Almacenamiento y copia de seguridad durante el proceso de investigación

- a. ¿Cómo se almacenarán los datos y los metadatos y se harán copias de seguridad durante el proceso de investigación?
- b. ¿Cómo se cuidará la seguridad de los datos y la protección de los datos sensibles durante la investigación?

## 4. Requisitos legales y éticos, códigos de conducta

- a. Si se tratan datos personales, ¿cómo se garantizará el cumplimiento de la legislación sobre datos personales y sobre seguridad de los datos?
- b. ¿Cómo se gestionarán otras cuestiones legales, como los derechos de propiedad intelectual y la titularidad? ¿Qué legislación es aplicable?
- c. ¿Cómo se tendrán en cuenta las posibles cuestiones éticas y se respetarán los códigos de conducta?

## Aspectos básicos a responder en un plan de gestión de datos (2)

### 5. Intercambio de datos y conservación a largo plazo

- a. ¿Cómo y cuándo se compartirán los datos? ¿Existen posibles restricciones para compartir los datos o razones de embargo?
- b. ¿Cómo se seleccionarán los datos para su conservación y dónde se conservarán a largo plazo (por ejemplo, en un depósito o archivo de datos)?
- c. ¿Qué métodos o herramientas informáticas se necesitarán para acceder a los datos y utilizarlos?
- d. ¿Cómo se garantizará la aplicación de un identificador único y persistente (p.e. un DOI) a cada conjunto de datos?

### 6. Responsabilidades y recursos para la gestión de los datos

- a. ¿Quién (por ejemplo, función, cargo e institución) será responsable de la gestión de los datos (es decir, el administrador de los datos)?
- b. ¿Qué recursos (por ejemplo, financieros y de tiempo) se dedicarán a la gestión de datos y a garantizar que los datos sean FAIR (Localizables, Accesibles, Interoperables, Reutilizables)?

# Herramientas para crear un Plan de Gestión de Datos

[https://rdmkit.elixir-europe.org/data\\_management\\_plan.html#relevant-tools-and-resources](https://rdmkit.elixir-europe.org/data_management_plan.html#relevant-tools-and-resources)



[https://dmponline.dcc.ac.uk/public\\_plans](https://dmponline.dcc.ac.uk/public_plans)



<https://pgd.consorciomadrono.es/home> → DMP → ARGOS



<https://argos.openaire.eu/splash/>



<https://dmp.csuc.cat/>

# Crear un plan de gestión de datos

Ejemplos de planes de libre acceso

Requerimientos de las agencias financiadoras según la convocatoria

The screenshot shows the DMPonline website interface. At the top, there is a navigation bar with the logo 'DMPonline' and links for 'Home', 'Public DMPs', 'Funder requirements', and 'Help'. Below the navigation bar, a notice states 'Signed out successfully.'. The main content area features the heading 'Plan to make data work for you' and a sub-heading 'Data Management Plans that meet institutional funder requirements.'. A central diagram illustrates the workflow of creating and sharing data management plans, involving users, data, and institutional requirements. To the right, a sign-in form is visible with fields for 'Email' and 'Password', a 'Forgot password?' link, a 'Remember email' checkbox, and buttons for 'Sign in' and 'Sign in with your institutional credentials'. At the bottom, a statistics bar displays: 114 456 Users, 323 Organisations, 126 033 Plans, and 89 Countries.

Registrarse como usuario para poder guardar y compartir

# Crear un plan de gestión de datos

## DMP Públicos

Los DMP públicos son planes creados con el servicio DMPonline y compartidos públicamente por sus propietarios. No se examinan por calidad, integridad o cumplimiento de las pautas de los financiadores.

Título del Proyecto	Modelo	Organización	Propietario	descargar
Head and neck paraganglioma registry	UMC Utrecht DMP	UMC Utrecht	Carolijn de Bresser	
3D Printing Our Imagination? Leveraging Research Insights And Organising Knowledge Translation For Impact-Cocreation With Digital Innovations To Address Grand Challenges	EPSRC Data Management Plan	University of Exeter	Stavros Polykarpou	
European Databases of Seismogenic Faults: EDSF Installation	Horizon Europe Template	Other	Roberto Basili	
Key Population HIV Estimates in Sub-Saharan Africa	Imperial College London Generic DMP	Imperial College London	Oliver Stevens	
GRAIL (Getting Responsible about AI and machine Learning in research funding and evaluation) - Workshop series	The University of Sheffield Research DMP	The University of Sheffield	Denis Newman-Griffis	
A highly automated InSAR detection system for nuclear infrastructure health monitoring	Imperial College London Generic DMP	Imperial College London	Stewart Agar	
Global Surgery Stockholm	Swedish Research Council Template	Karolinska Institutet	Jenny Löfgren	
Quantification of meltwater storage in Greenland using a statistically-optimal estimation of mass anomalies with satellite gravimetry data	Data Management Plan NWO (September 2020)	Delft University of Technology	Pavel Ditmar	
Burden of care: Incidence of surgical procedures on cleft patients	UMC Utrecht DMP	UMC Utrecht	Jette Boxem	
Model - DMP TPM Interviews	TU Delft Data Management Plan template (2021)	Delft University of Technology	Nicolas Dintzner	

Ejemplos de planes que nos pueden servir de ejemplos

# Crear un plan de gestión de datos

## Funder requirements

Templates for data management plans are based on the specific requirements listed in funder policy documents. The DCC maintains these templates, however, researchers should always consult the funder guidelines directly for authoritative information.

Q  búsqueda

Nombre de la plantilla <sup>▲</sup>	descargar	nombre de la organización <sup>◆</sup>	Última actualización <sup>◆</sup>	Enlaces de financiación	Crear un nuevo plan <sup>◆</sup>	Planes de muestra (si está disponible)
ERC DMP	 	European Research Council (ERC)	24/1/2023	<a href="https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf">https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf</a>	 	
Horizon 2020 DMP	 	European Commission	16/5/2019	<a href="#">Guidelines on FAIR data management</a>	 	ArchAIDE example from York Analysis of the distribution of the population of Austria by altitude Correlating LA Museum Visitors to Crimes in LA
Horizon Europe Template	 	European Commission	4/8/2021	<a href="https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/how-to-participate/reference-documents;programCode=HORIZON">https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/how-to-participate/reference-documents;programCode=HORIZON</a>	 	
Science Europe	 	Science Europe	17/2/2021	<a href="https://www.scienceurope.org/">https://www.scienceurope.org/</a>	 	

Resultados de búsqueda claros

Requerimientos de las agencias financiadoras según la convocatoria

# Crear un plan de gestión de datos

Si pinchamos en crear un plan, podemos empezar desde una plantilla estándar o escoger una preestablecida según el financiador

detalles del proyecto Contribuyentes Resumen del plan Escribir plan Compartir descargar

**\* Título del Proyecto**

Proyecto simulado para pruebas, práctica o fines educativos.

**Resumen del proyecto**

**B** *I*

**Inicio del proyecto**  **Fin del proyecto**

**ID**

**Financiador**

**Estado de financiación**

**Número de concesión / url**

**Guardar**

## Seleccione Orientación

Para ayudarlo a escribir su plan, DMPonline puede mostrarle la orientación de una variedad de organizaciones.

**Seleccione Hasta 6 Organizaciones Para Ver Su Orientación.**

Digital Curation Centre

Encuentre orientación de organizaciones adicionales a continuación

[Ver la lista completa](#)

**Guardar**

# Crear un plan de gestión de datos

detalles del proyecto   Contribuyentes   Resumen del plan   Escribir plan   Compartir   descargar

## Formato

pdf ▼

## Descargar ajustes

**Componentes Opcionales Del Plan**

- detalles del proyecto portada
- texto de la pregunta y encabezados de sección
- preguntas sin respuesta

## Formato PDF

<b>Fuente</b>			<b>Margen (mm)</b>			
<b>Estilo</b>	<b>Tamaño (pt)</b>	<b>Superior</b>	<b>Inferior</b>	<b>Izquierdo</b>	<b>Derecho</b>	
"Times New Roman", Times, Serif ▼	10 ▼	25 ▼	20 ▼	12 ▼	12 ▼	

Descargar Plan

Una vez completado podremos descargarlo en distintos formatos y añadir o modificar

REC1. Adoptar los principios FAIR y facilitar la producción científica en abierto, de forma legal y reutilizable.

REC2. Seleccionar aquellos datos que tengan interés para usos futuros.

REC3. Aplicar los principios FAIR también a los metadatos y al plan de gestión de datos.

REC4. Facilitar, en los informes de evaluación de los proyectos, los beneficios y ventajas de haber hecho que los datos de investigación sean FAIR.

REC5. Elaborar un plan de gestión de datos en el que se refleje cómo van a ser tus datos FAIR.

REC6. Indicar en el plan de gestión de datos si se han consultado o reutilizado datos ya existentes en lugar de obtenerlos de nuevo.

REC7. Utilizar estándares interoperables con el ecosistema de datos FAIR para facilitar la interoperabilidad entre sistemas.

REC8. Facilitar los datos de investigación de acuerdo con estándares reconocidos por una comunidad científica.

REC9. En la medida de lo posible, asignar licencias abiertas para propiciar la reutilización de los datos de investigación.

REC10. Facilitar una descripción detallada de las condiciones para que se puedan reproducir los ensayos experimentales.

REC11. Identificar el repositorio de confianza más adecuado para depositar los datos.

REC12. Verificar los requisitos de las políticas respecto a los datos de investigación.

REC13. Promover la citación de los datasets como cualquier otro recurso bibliográfico.

REC14. Tener en cuenta, sobre todo en aquellas disciplinas que impliquen trabajar con seres vivos, los códigos de conducta adecuados y reconocidos internacionalmente.

REC15. Contribuir al intercambio de las buenas prácticas respecto a la gestión de los datos y lecciones aprendidas entre tu comunidad científica.

REC16. Favorecer la colaboración entre profesionales y gestores de datos e información para participar y elaborar programas de formación.

## MareData (Red Española sobre Datos de Investigación en Abierto). Recomendaciones para la gestión de datos de investigación

**REC17. Recomendación final: propiciar que los investigadores y los centros para los que trabajan avalen el paradigma de la open science.**

## Rúbrica para evaluar PGD

<https://www.scienceeurope.org/media/22hpslfl/se-rdm-template-5-guidance-on-the-evaluation-of-data-management-plans.docx>

### Evaluation Rubric for Data Management Plans ¶

DMP-Question¶	DMP-Guidance¶	Performance-Level¶	
<b>Guidance for Researchers¶</b>		<b>Sufficiently-Addressed¶</b> The-DMP...¶	<b>Insufficiently-Addressed¶</b> The-DMP...¶
Administrative information¶	<ul style="list-style-type: none"> <li>• Provide information such as name of applicant, project number, funding programme, version of DMP. ¶</li> </ul>	<ul style="list-style-type: none"> <li>• contains the minimal information required to identify the applicant and the references of the project. ¶</li> </ul>	<ul style="list-style-type: none"> <li>• provides no or limited information, which makes it hard to identify who is responsible for the project. ¶</li> </ul>
<b>1--DATA-DESCRIPTION-AND-COLLECTION-OR-RE-USE-OF-EXISTING-DATA¶</b>			
<b>Guidance for Researchers¶</b>		<b>Sufficiently-Addressed¶</b> The-DMP...¶	<b>Insufficiently-Addressed¶</b> The-DMP...¶
<b>1a</b> -How will new data be collected or produced and/or how will existing data be re-used?¶	<ul style="list-style-type: none"> <li>• Explain which methodologies or software will be used if new data are collected or produced. ¶</li> <li>• State any constraints on re-use of existing data if there are any. ¶</li> <li>• Explain how data provenance will be documented. ¶</li> <li>• Briefly state the reasons if the re-use of any existing data sources has been considered but discarded. ¶</li> </ul>	<ul style="list-style-type: none"> <li>• gives clear details of where the existing data come from and how new data will be collected or produced. It clearly explains methods and software used. ¶</li> <li>• explains, if existing data are re-used, how these data will be accessed and any constraints on their re-use. ¶</li> <li>• explains clearly, if applicable, why new data must be collected, rather than re-using existing data. ¶</li> </ul>	<ul style="list-style-type: none"> <li>• provides little or no details on where the data come from and what data will be collected or re-used. ¶</li> <li>• does not, if applicable, provide sufficient rationale for generating new data. ¶</li> </ul>
<b>1b</b> -What data (for example the kind, formats, and volumes) will be collected or produced?¶	<ul style="list-style-type: none"> <li>• Give details on the kind of data: for example, numeric (databases, spreadsheets), textual (documents), image, audio, video, and/or mixed-media. ¶</li> <li>• Give details on the data format: the way in which the data is encoded for storage, often reflected by the filename extension (for example-</li> </ul>	<ul style="list-style-type: none"> <li>• clearly describes or lists what data types will be generated (for example numeric, textual, audio, or video) and their associated data formats, including, if needed, data conversion strategies. ¶</li> <li>• explains why certain formats have been chosen and indicates if they are in open and standard format. If a proprietary-</li> </ul>	<ul style="list-style-type: none"> <li>• provides no or little details on what data types will be generated and does not provide a valid reason for this omission (for example a statement that no data will be produced or generated). ¶</li> <li>• only lists/describes the kinds of data-</li> </ul>

## Rúbrica para evaluar PGD

Belmont Forum DMP Scorecard (v.20190819) <https://zenodo.org/record/3530933>

Criteria	Complete Response (score = 2)	Incomplete Response (score = 1)	No Response (score = 0)	SCORE
<b>1. What types of datasets and other digital outputs of long-term value do you expect the project will produce or reuse?</b>				
<p>1.1 Plan lists the <u>types</u> of data and other digital outputs of long-term value.</p> <p><i>(e.g. text, databases, images, 3D models, software, audio files, code, video files, reports, surveys, patient records, samples, and so forth)</i></p>	<p>Datasets and other digital outputs of long-term value are identified, including data type and encoding.</p> <p><i>"Environmental data will be delivered as NetCDF (Network Common Data Format) files. Raster files will use the raster2pgsql PostGIS module. Maps and other geographic data will use shapefiles."</i></p> <p><i>"Transcripts and coding will be provided in text files. Audio recording will be MP3 format."</i></p>	<p>Datasets and other digital outputs of long-term value are identified, but lack detail for users beyond the project to understand.</p> <p><i>"A combination of geo-referenced data at various spatial, temporal, and taxonomic scales (e.g., populations, regions, nations, circumpolar, biomes, habitats) will comprise our data of long-term value."</i></p> <p><i>"Long-term value data include data from anthropological field studies: transcripts of interviews and discussion workshops, associated metadata."</i></p>	No information about data types is included.	X
<p>1.2 Plan describes how the data and other digital outputs will be <u>collected, captured, or created</u>.</p> <p><i>(e.g., new observations, results from models, reuse of other data, or other)</i></p>	<p>Clearly defines how data will be collected, captured or created, including methods, instruments, software, or infrastructure where relevant.</p> <p><i>"The MIP compatible model runs to the ISI-MIP servers, where existing Fish-MIP runs are already stored. Scenario outputs will be archived."</i></p> <p><i>"Socio-economic data will include household food security, nutrition, and demographic data. Spatial data produced will include ground-truthed land use/land cover data ~3 km from 50 farms, land use scenario maps for 12 villages and 4 regions. All ecological/social data will be recorded on physical datasheets and entered directly into Excel or STATA."</i></p>	<p>Missing some details regarding how some of the data will be produced; makes assumptions about reviewer knowledge of methods or practices.</p> <p><i>"Models will produce a broad range of output simulation data."</i></p> <p><i>"Data collection includes gathering in-the-field various phytoplankton, zooplankton, fish and flooded forest biodiversity."</i></p>	No information about data collection, capture or creation.	X
1.3	Datasets and other digital outputs volume estimated.	Datasets and other digital outputs amount is vaguely estimated or	Amount of expected data is	X

## **PRÓXIMOS CURSOS:**

ISCIII. Plan de Formación interna 2024 (para investigadores).

Gestión de datos de investigación y Planes de Gestión de Datos (para investigadores)-2ª ed.

<https://sigade.isciii.es/publico/actual/VerCurso.asp?ID=4&CodProp=4450&CodEd=4549>

BNCS. Plan de Formación y difusión

Virginia Jiménez Planet (Coordinadora Repisalud)

BNCS-ISCIH

Tf: 918 222 339

Correo: [virginiaj@isciii.es](mailto:virginiaj@isciii.es)

Silvia Hernández Villegas (Técnico Repisalud)

BNCS-ISCIH

Tf: 918 222 117

Correo: [silviahernandez@isciii.es](mailto:silviahernandez@isciii.es)

Correo genérico: [repisalud@isciii.es](mailto:repisalud@isciii.es)

Gracias por su atención



Biblioteca Nacional  
de Ciencias de la Salud

*repisalud*  
Repositorio Institucional en Salud