

Gestión de datos en investigación

Olatz Alonso del Cura

Lucía Camacho Montaña

Olatz Alonso del Cura

- Grado en ingeniera biomédica
- Máster en bioinformática
- Actualmente trabajando como data manager en Investén



Lucia Camacho Montaña

- Grado en terapia ocupacional
- Máster en Neurocontrol motor
- Estudiante del grado de ciencia e ingeniería de datos
- Actualmente trabajando como Investigadora en Investén



Índice

Gestión de datos en investigación

1. Introducción a la gestión de datos de investigación
 1. Ciencia abierta
 2. Gestión de datos de investigación
 3. Datos de investigación
2. Ciclo de vida de los datos de investigación
3. Plan de Gestión de Datos (PGD)

1. Introducción a la gestión de datos de investigación (GDI)

Ciencia Abierta

Gestión de datos de investigación

Datos de investigación

1. Introducción a la gestión de datos de investigación (GDI)

Ciencia abierta (Open Science)

{ Gestión de datos de investigación



Fuente: Recomendación de la UNESCO sobre la Ciencia Abierta, https://unesdoc.unesco.org/ark:/48223/pf0000379949_spa

1. Introducción a la gestión de datos de investigación (GDI)

Los **datos de investigación** están tomando cada vez más valor:

- La mayoría de convocatorias, como parte de la propuesta de un proyecto de investigación, requieren la presentación de un **plan de gestión de los datos (PGD)**.
- Las revistas científicas ahora requieren que los conjuntos de datos se presenten con el manuscrito o se depositen en repositorios.

1. Introducción a la gestión de datos de investigación (GDI)

Gestión de datos garantizar que los datos generados en una investigación sean organizados, seguros, comprensibles y reutilizables de acuerdo con los principios FAIR.

Los datos se puedan encontrar y reutilizar rápidamente.

Conceptos clave:

- **Datos de investigación.**
- **Plan de gestión de los datos (PGD).**

1. Introducción a la gestión de datos de investigación (GDI)

El objetivo de la **gestión de datos** es planificar y cuidar todo el proceso relacionado con los datos de investigación:

- **Antes** de su creación
- **Durante** su creación y uso
- **A lo largo de su ciclo de vida**

Cómo y dónde se **recogen**
Cómo se **organizan** y nombran
Cómo se **asegura** su calidad
Cómo se **protegen** y **almacenan**
Cómo y cuándo se **comparten**
Cómo se **conservan** para el futuro



Fuente: <https://www.upo.es/biblioteca/servicios/investigabib/Gestion-de-datos-de-investigacion/>

1. Introducción a la gestión de datos de investigación (GDI)

Datos de investigación:

Todos aquellos materiales que son recolectados, observados o generados durante el transcurso de una investigación.

Datos abiertos (Research Open Data): aquellos datos legalmente accesibles, reproducibles y reutilizables.

Los datos pueden ser...

Según su origen:

- Datos primarios: Generados específicamente para el estudio.
- Datos secundarios: Recopilados de fuentes secundarias, repositorios, bases de datos existentes, HCE, estudios previos...

Datos de investigación

Los datos se tratan de forma distinta según su tipo.

Según su naturaleza:

- Cualitativo. Entrevistas
- Cuantitativo. Presión arterial

Según su nivel de procesamiento:

- Primarios (sin procesar). Niveles de glucosa
- Procesados. Nivel medio de glucosa
- Analizados. ANOVA comparando niveles de glucosa entre grupos

Según su la fuente de la que provienen:

- Experimentales. Datos procedentes de un ensayo clínico.
- Observacionales. Datos generados al registrar observaciones.
- Computacionales. Datos que acompañan a las simulaciones.

Datos de investigación

Según su formato:



Fuente: <https://datateam.mx/tipos-de-big-data/>

Datos estructurados

- Mediciones clínicas (glucemia)
- Resultados de escalas (dolor)
- Bases de datos en REDCap

Datos no estructurados

- Grabaciones de entrevistas o grupos focales
- Fotografías o imágenes clínicas
- Transcripciones sin etiquetas

Datos semiestructurados

- Formularios con preguntas cerradas y abiertas
- Encuestas con comentarios
- Transcripciones con etiquetas

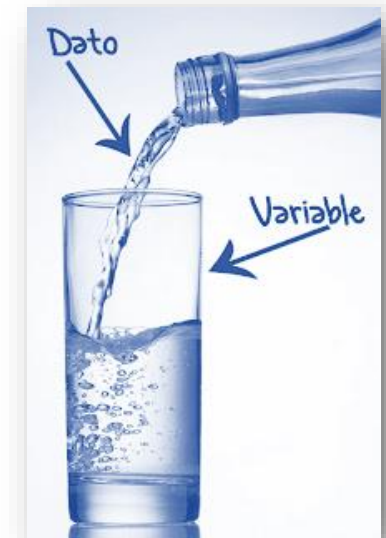
Datos de investigación

Variable:

- Es un atributo que puede tomar diferentes valores.
- Debe definirse en función del concepto que se pretende estudiar, incluyendo en su definición cómo se medirán y almacenarán dichos valores.

Los **datos** son los valores que toma la **variable** en un conjunto de observaciones o individuos.

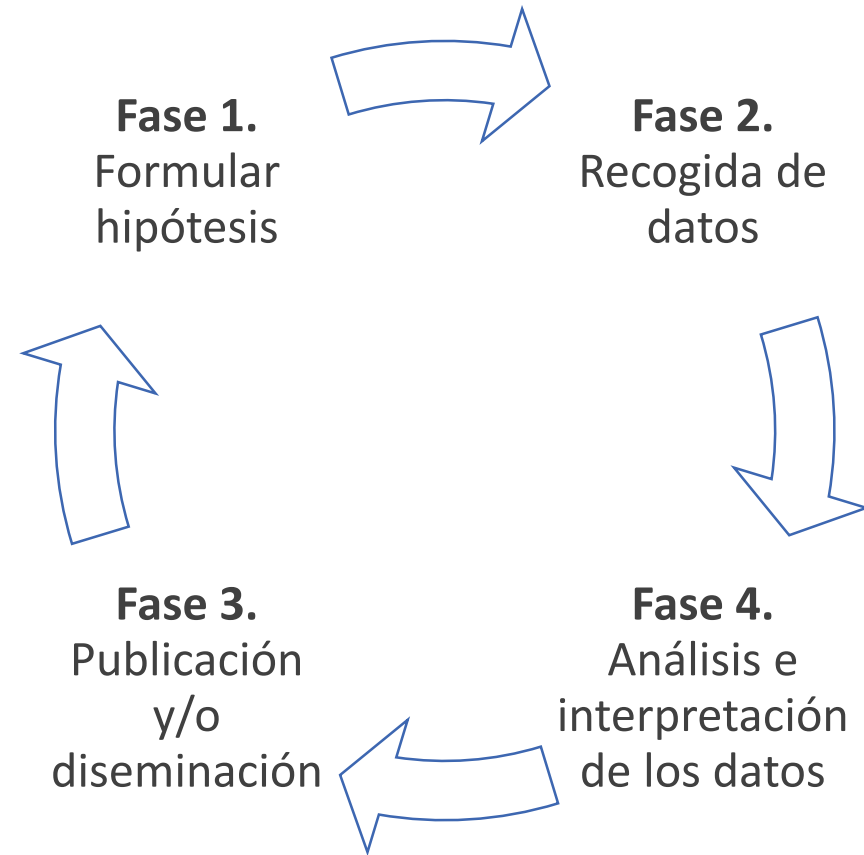
Campo (etiqueta)	Nombre variable	Tipo de dato	Formato	Unidad de medida	Valores permitidos	Herramienta estandarizada
------------------	-----------------	--------------	---------	------------------	--------------------	---------------------------



Fuente:
<https://estadisticaparaadministracion.blogspot.com/2012/05/de-los-datos-variables-cualitativa.html>

Datos de investigación

Ciclo de vida de la investigación



Michener WK. Ten Simple Rules for Creating a Good Data Management Plan. PLoS Comput Biol. 2015;11(10):e1004525.

Ciclo de vida de los datos



Los datos que descargué



Estructurados y limpiados



Ordenados y relacionados



Presentados visualmente

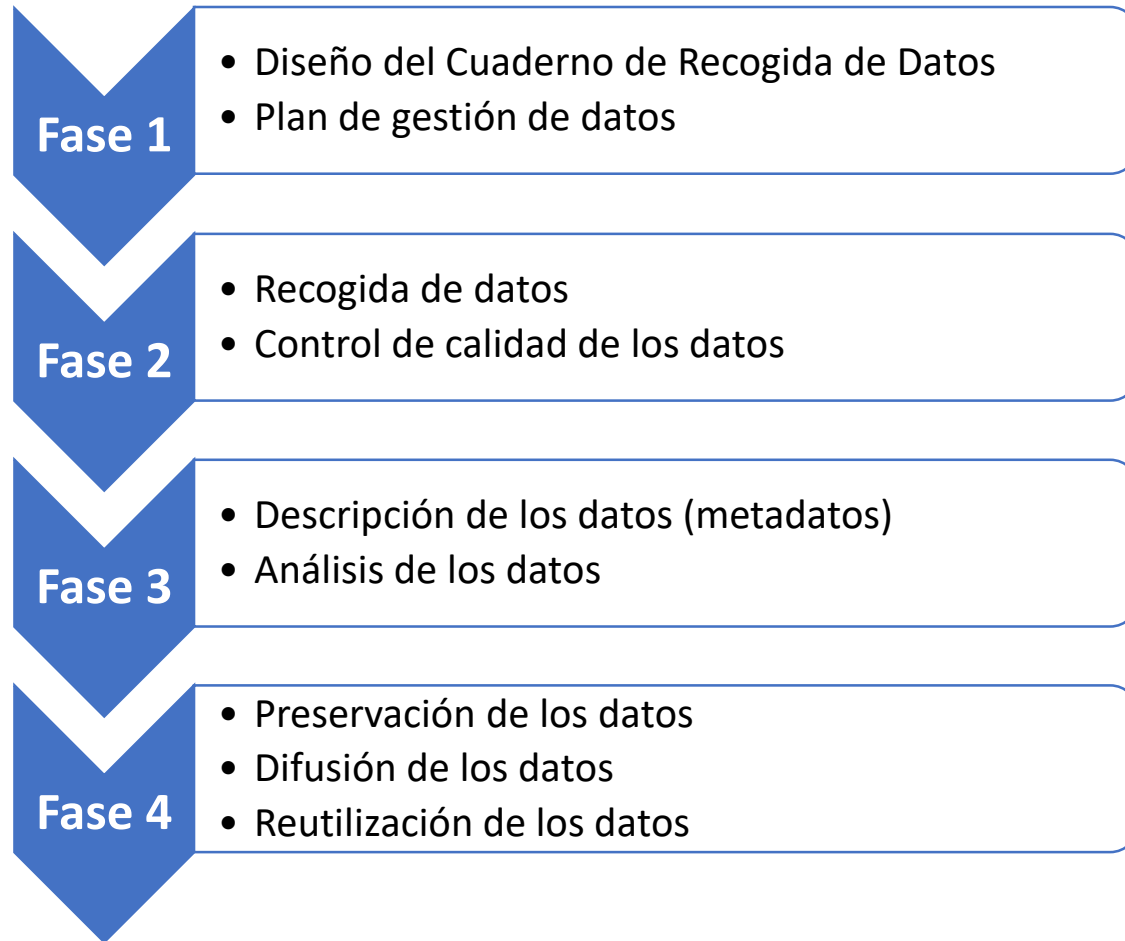


Fuente: <https://www.rebiun.org/acceso-abierto/el-ciclo-de-los-datos-cientificos>

Fuente: <https://e-archivo.uc3m.es/entities/publication/adf04df6-266f-4c0f-bf4b-11d91fd3eae6>

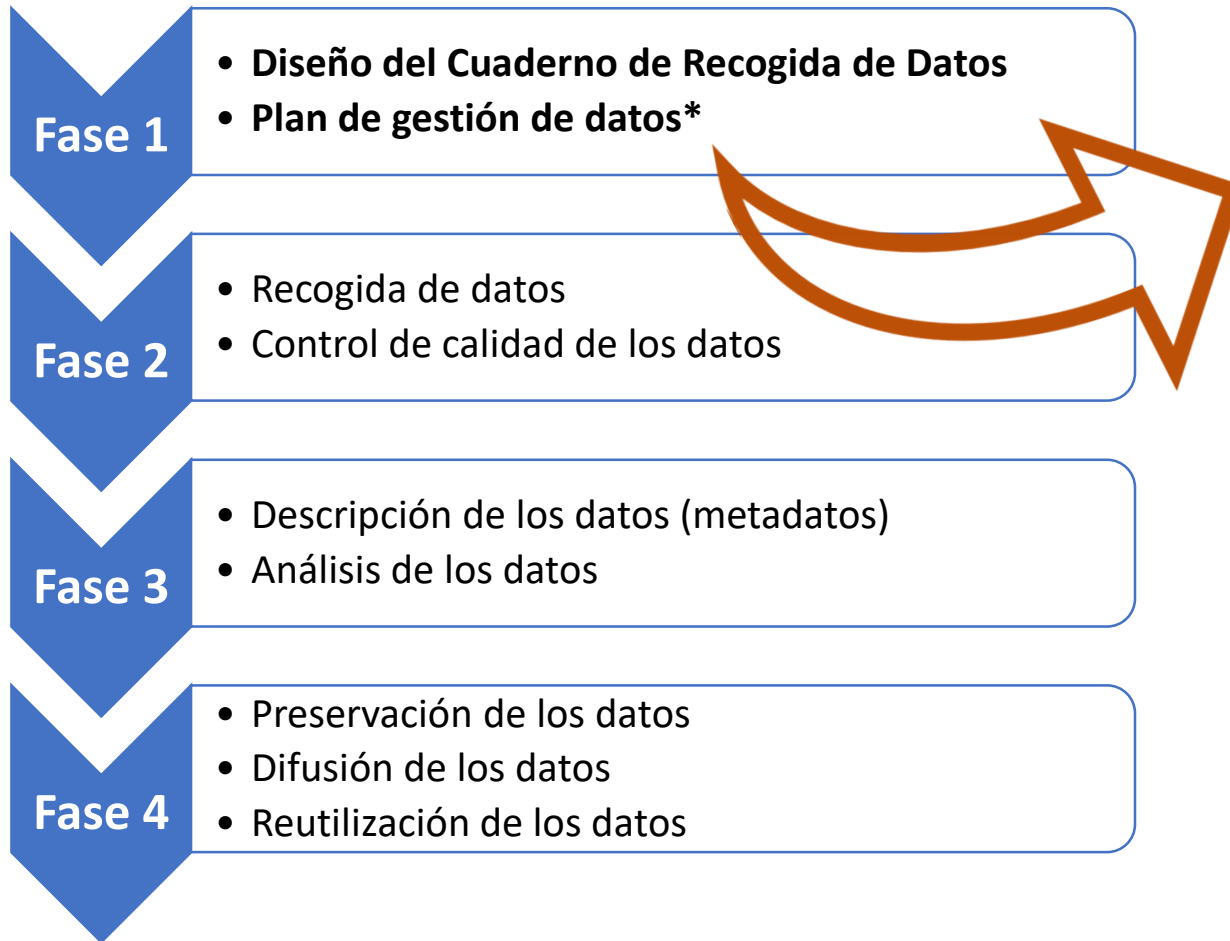
2. Ciclo de vida de los datos

Ciclo de vida de los datos



El ciclo de vida de los datos describe **todas las etapas que atraviesan los datos a lo largo de un proyecto.**

Ciclo de vida de los datos



Debe incluir aspectos relacionados con todo el ciclo de vida de los datos, por lo que lo veremos al final.

Diseño del Cuaderno de Recogida de Datos (CRD)

Fase 1

Objetivo principal

Recopilar todos los datos requeridos por el protocolo para asegurar que la información responda a los objetivos planteados

Pasos a seguir

1. Identificar las **variables** relevantes
2. Seleccionar el **formato** de recogida
3. Estandarizar su **codificación** y formato
4. Establecer mecanismos de **control de calidad**

Práctica: diseño de un CRD para un caso de uso en investigación

Bellary S, Krishnankutty B, Latha MS. Basics of case report form designing in clinical research. *Perspect Clin Res*. 2014;5(4):159–166. doi:10.4103/2229-3485.140555

Diseño del Cuaderno de Recogida de Datos (CRD)

Fase 1

1. Identificar las variables relevantes: QUÉ

- ¿Qué variables necesito recoger para responder a mi pregunta de investigación?
- ¿Puedo obtenerlos de registros clínicos, encuestas previas o bases de datos existentes?
- ¿Debo generarlos desde cero?
- ¿Cómo se definen? (tipo, unidades, formato...)

2. Seleccionar el formato de recogida: DÓNDE

- Formato **físico**: entrevistas, formularios en papel, notas de campo...
- Formato **electrónico**: formularios digitales, dispositivos clínicos, apps, HCE...
 - ¿Qué plataforma o sistema de recogida voy a utilizar? REDCap, Excel...



Prokscha S. *Practical Guide to Clinical Data Management*. 4th ed. CRC Press; 2024. doi:10.1201/9781003395621.

Diseño del Cuaderno de Recogida de Datos (CRD)

Fase 1

3. Estandarizar su codificación y formato: CÓMO

- ¿Hay instrumentos validados que puedo reutilizar?
- ¿Qué formatos facilitarán la recogida de datos homogénea y su posterior análisis?
- ¿Existen estándares que puedo aplicar? (SNOMED CT, CIE-10, CIAP2...)

4. Establecer mecanismos de control de calidad

- ¿Qué controles de validación necesito?
- ¿Cuáles son los valores válidos o rangos aceptables?

Table 1: Illustrating a well-designed and poorly designed data fields imparting the significance of visual cues to help the site personnel to understand the format

Poorly designed	Well designed
Date of visit: _____	Date of visit: □□ / □□ / □□□□ (DD/MM/YYYY)
Blood pressure: ____ / ____	Blood pressure: □□□ / □□□ (mmHg)
Pulse: _____	Pulse: □□□ (beats/min)
Temperature: _____	Temperature: □□ . □ (°C)
Respiration: _____	Respiration: □□ (/min)

Las preguntas deben ser claras e inequívocas para evitar confusiones

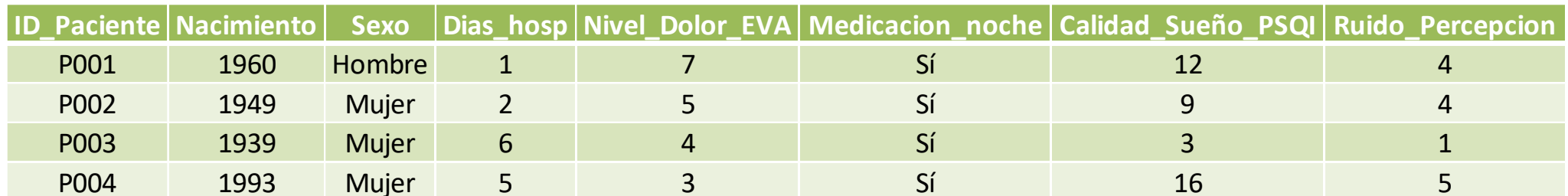
Bellary S, Krishnankutty B, Latha MS. Basics of case report form designing in clinical research. *Perspect Clin Res*. 2014;5(4):159–166. doi:10.4103/2229-3485.140555

Diseño del Cuaderno de Recogida de Datos (CRD)

- La estructura del formulario debe garantizar una **base de datos coherente y explotable**.
- Los datos deben estar organizados en un **formato que facilite y simplifique su análisis**.

Cada columna es una **variable o campo**

Cada fila es una **observación o registro**



ID_Paciente	Nacimiento	Sexo	Dias_hosp	Nivel_Dolor_EVA	Medicacion_noche	Calidad_Sueño_PSQI	Ruido_Percepcion
P001	1960	Hombre	1	7	Sí	12	4
P002	1949	Mujer	2	5	Sí	9	4
P003	1939	Mujer	6	4	Sí	3	1
P004	1993	Mujer	5	3	Sí	16	5

Diseño del Cuaderno de Recogida de Datos (CRD)

El **diccionario de datos** debe incluir toda la información necesaria para que la estructura del conjunto de datos sea entendible.

#	Variable/nombre del campo	Etiqueta de campo <i>Nota de campo</i>	Atributos de campo (tipo de campo, validación, opciones, lógica ramificada, cálculos, etc.)														
Formulario: Criterios Inclusion-Exclusion (criterios_inclusionexclusion)																	
1	[participante_id]	Record ID	text														
2	[caso_control]	Indique si el participante es caso o control	radio, Required <table border="1"><tr><td>1</td><td>Grupo de cohorte</td></tr><tr><td>2</td><td>Grupo de calibración</td></tr></table> alineación de encargo LH	1	Grupo de cohorte	2	Grupo de calibración										
1	Grupo de cohorte																
2	Grupo de calibración																
3	[localidad]	Localidad de estudio	radio, Required <table border="1"><tr><td>20</td><td>Albacete</td></tr><tr><td>30</td><td>Alicante</td></tr><tr><td>80</td><td>Barcelona</td></tr><tr><td>12</td><td>Huelva</td></tr><tr><td>52</td><td>Lleida</td></tr><tr><td>82</td><td>Madrid</td></tr><tr><td>84</td><td>Vizcaya</td></tr></table> alineación de encargo LV	20	Albacete	30	Alicante	80	Barcelona	12	Huelva	52	Lleida	82	Madrid	84	Vizcaya
20	Albacete																
30	Alicante																
80	Barcelona																
12	Huelva																
52	Lleida																
82	Madrid																
84	Vizcaya																

Diseño del Cuaderno de Recogida de Datos (CRD)

Fase 1

El diseño del CRD requiere un **enfoque colaborativo y multidisciplinar**:

Equipo investigador

Define variables a recoger
Selecciona método de recogida (entrevistas,
escalas...)
Aporta conocimiento científico y clínico

Delegado/a de Protección de Datos (DPD)

Asegura cumplimiento normativo
Evalúa riesgos
Propone medidas de seguridad y
anonimización

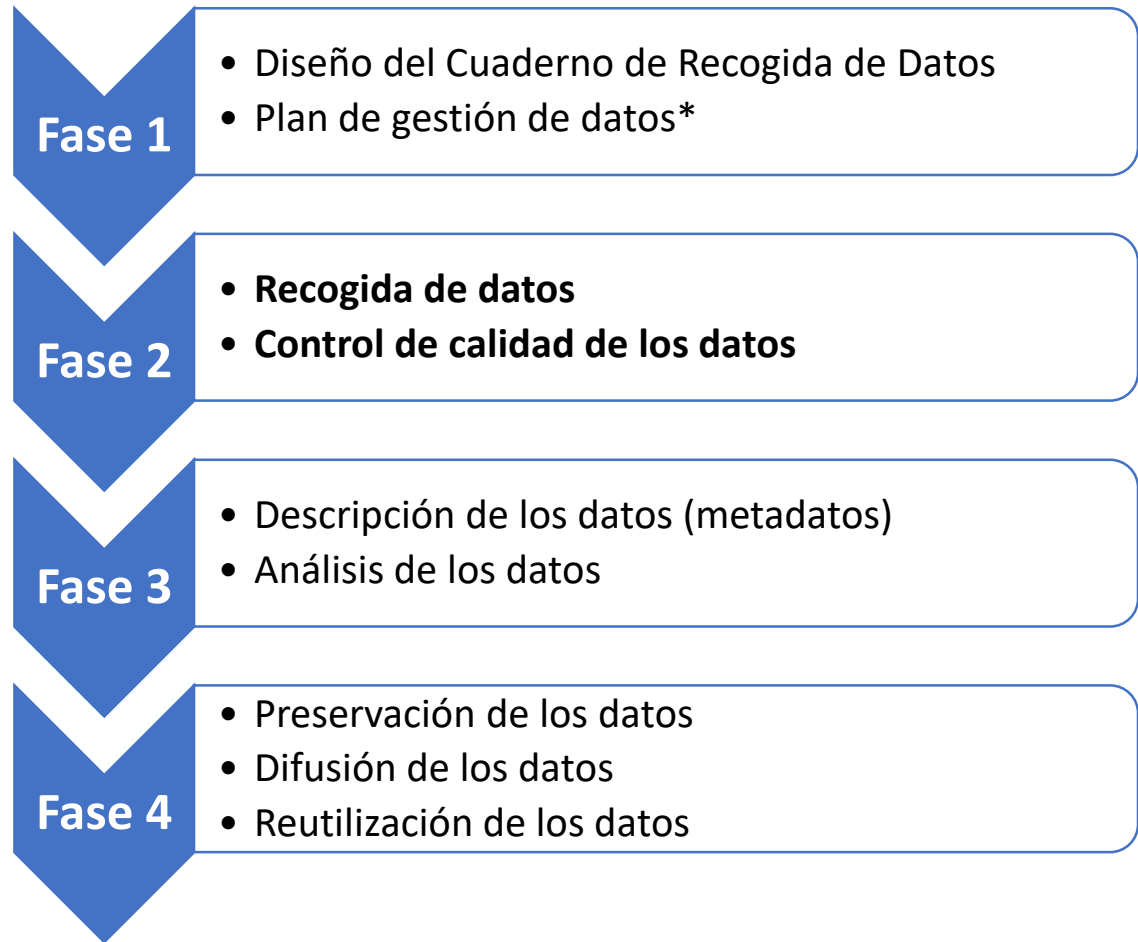
Soporte técnico en sistemas de información

Da soporte a herramientas digitales (REDCap...)
Integra entre sistemas
Realiza copias de seguridad

Gestor/a o responsable de datos

Diseña el flujo de datos
Supervisa la calidad
Controla los formatos y estándares utilizados

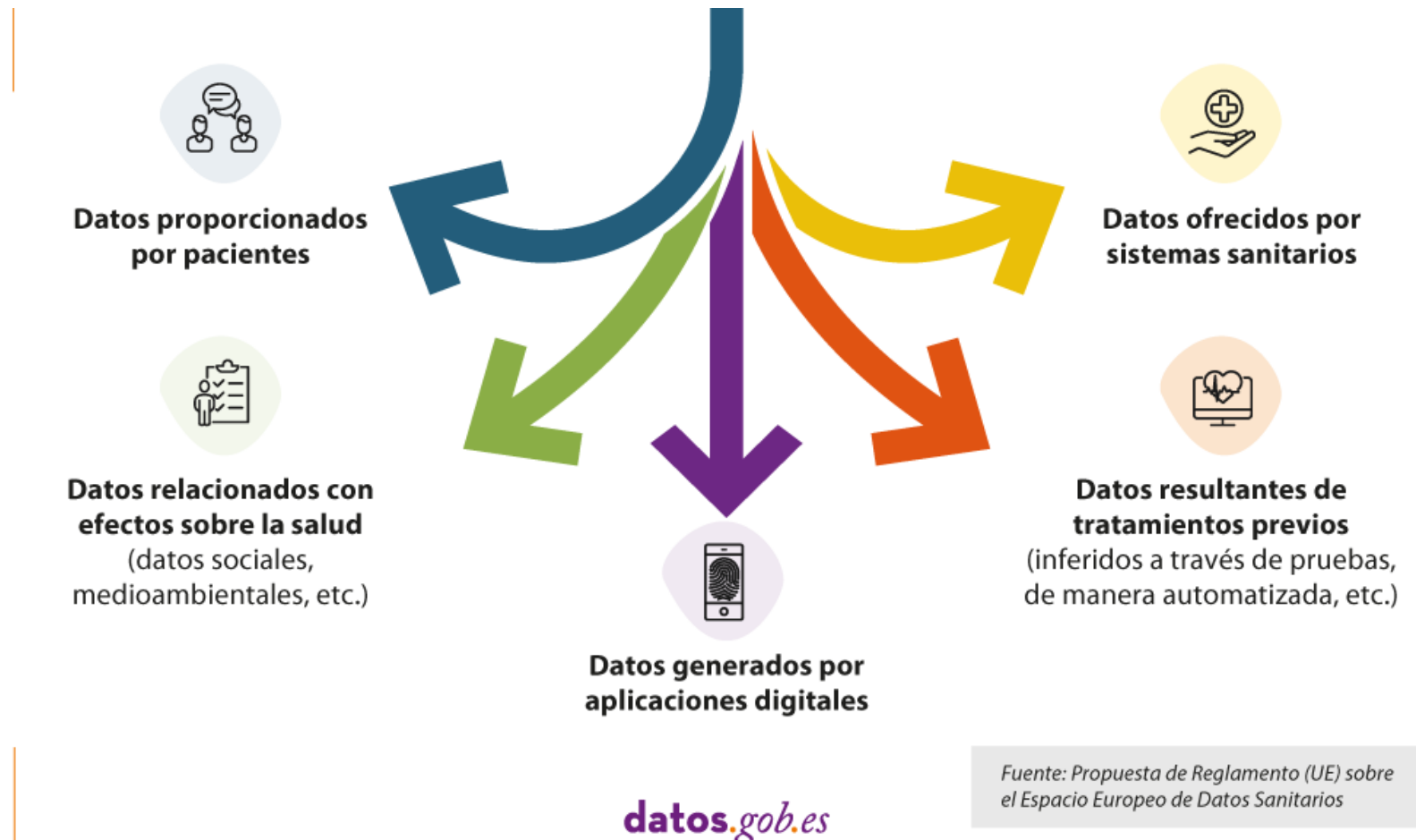
Ciclo de vida de los datos



Recogida de datos

Fase 2

Proceso de **generación o recopilación de datos** de investigación. Pueden proceder de distintas fuentes.



Fuente: Propuesta de Reglamento (UE) sobre el Espacio Europeo de Datos Sanitarios

Buenas prácticas

- Circuitos operativos claros
- Definición de roles y responsabilidades
- Formación adecuada del personal implicado
- Herramientas de trabajo colaborativo



Permite una **recogida de datos homogénea**



Recomendaciones

- Prueba piloto antes del inicio
- Cuaderno del investigador

Objetivo principal

Garantizar la **integridad y consistencia** de los datos, minimizando **errores, discrepancias y valores perdidos**.

Validación de datos

Proceso continuo que **verifica la validez de la información** conforme al protocolo del estudio.

Se realiza **a medida que se introducen los datos en el CRD**.

- **Valores fuera de rango:** paciente de 30 años cuando el criterio de inclusión es 55–70 años; PA 400 mmHg.
- **Datos inconsistentes:** *Fecha de alta anterior a Fecha de ingreso.*
- **Campos vacíos o incompletos:** *falta el código del paciente o la fecha de medición.*
- **Formatos incorrectos:** registro de fechas en formato no válido.

Un buen diseño del CRD reduce estos errores desde el inicio.

Limpieza de datos

Proceso de **revisión y corrección sistemática** de los datos antes del análisis estadístico.

- **Campos vacíos**
- **Valores atípicos**
- **Errores de transcripción**
- **Incoherencia temporal**
- **Valores duplicados**
- **Codificación incorrecta**

Cada corrección debe **documentarse** para mantener la **trazabilidad y transparencia** del proceso.

Roche C, McDonald S, et al. An overview of the fundamentals of data management, analysis, and interpretation in quantitative research. *Semin Oncol Nurs.* 2023;39:151398.

Control de calidad

El objetivo es obtener (y publicar) una base de datos de calidad



[Guía práctica para la mejora de la calidad de los datos abiertos](#)

PAUTAS GENERALES PARA GARANTIZAR LA CALIDAD DE LOS DATOS ABIERTOS

PROPORCIONAR UNA CANTIDAD ADECUADA DE DATOS PARA FACILITAR SU ANÁLISIS

Los publicadores deben asegurar que se publica una **cantidad razonable de datos** para que haya suficiente contexto y los usuarios puedan obtener valor de su explotación.

FORMATEO DE VARIABLES DE FECHA Y HORA

Las fechas deben codificarse siempre utilizando el estándar internacional de referencia **ISO 8601**.

EVITAR LA MEZCLA DE ESCALAS NUMÉRICAS

Intentar que la **escala no varíe** a lo largo del tiempo. En caso de ser necesario, proporcionar los datos en ambas escalas y documentar el cambio de escala.

EVITAR LA MEZCLA DE RANGOS EN UN MISMO CONJUNTO DE DATOS

Publicar los datos con el **mayor nivel de desagregación**. Si no es posible, mantener la consistencia en todos los valores de la variable.

EVITAR LA INCORPORACIÓN DE SUBTOTALES, TOTALES O AGRUPAMIENTOS

Presentar el **mayor nivel de desagregación posible** de los datos que contiene.

EVITAR LA FRAGMENTACIÓN DE DATOS Y DE DIFÍCIL LOCALIZACIÓN

Mejorar la **organización y etiquetado** de los contenidos, siendo necesario establecer conexiones entre los distintos conjuntos de datos.

ATRIBUTOS DE LA CALIDAD DE LOS DATOS



Exactitud



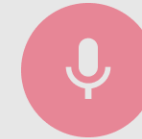
Completitud



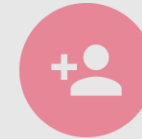
Consistencia



Credibilidad



Actualidad



Accesibilidad



Conformidad



Confidencialidad



Eficiencia



Precisión



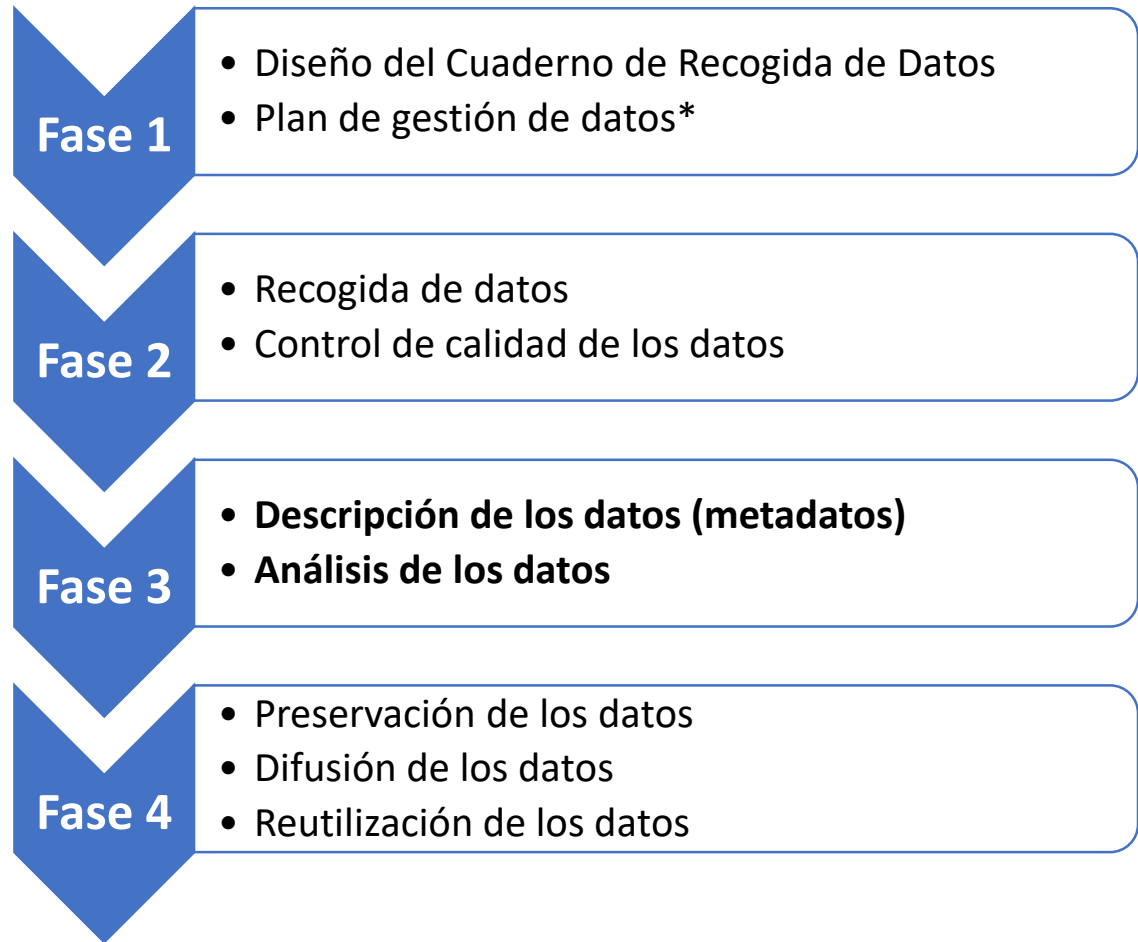
Trazabilidad



Comprensibilidad

Características de calidad de datos ISO/IEC 25012

Ciclo de vida de los datos



Descripción de los datos (metadatos)

“Los datos son inútiles a menos que puedan interpretarse”

Los metadatos son la forma de maximizar la interpretabilidad de los datos entre los potenciales usuarios.

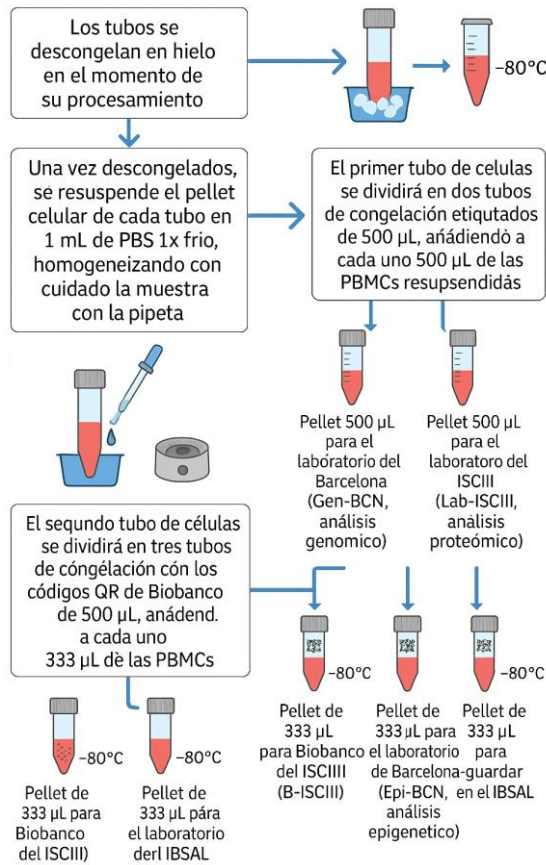
¿Qué son los metadatos?

- Información que **explica, describe y contextualiza** los datos.
- Permite que los datos sean **comprensibles y reutilizables**.

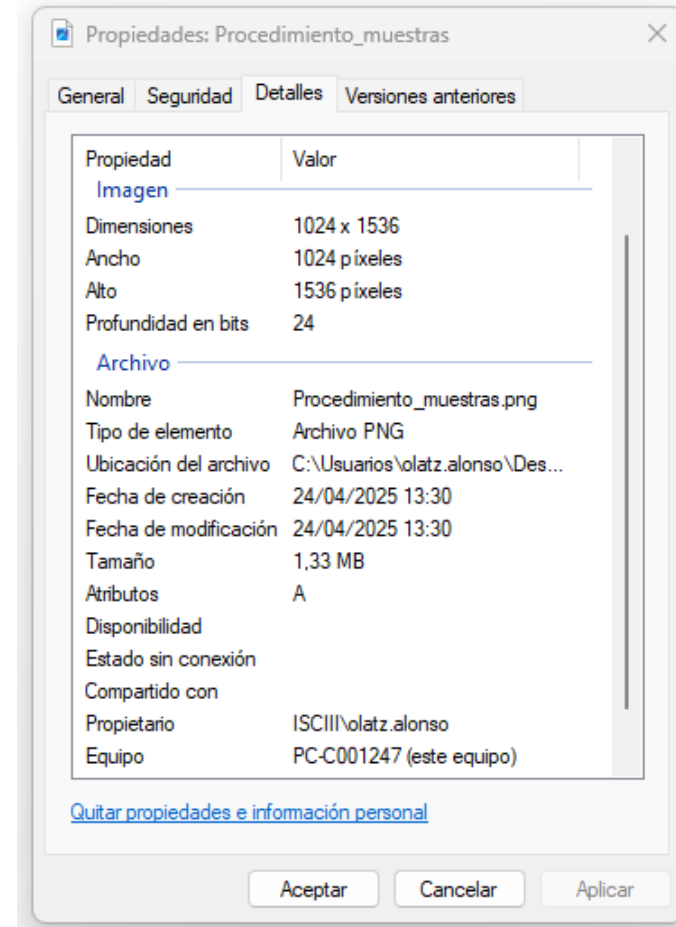
Alston JM, Rick JA. A beginner’s guide to conducting reproducible research. *Bull Ecol Soc Am*. 2020;102(2):e01801. doi:10.1002/bes2.1801

Descripción de los datos (metadatos)

Imagen "Procedimiento_muestras"



Metadatos para la imagen "Procedimiento_muestras"



En el contexto de una investigación...

2 categorías:

- Documentación del Proyecto
- Documentación del dataset

- Dónde y cómo se recolectan los datos.
- Cómo los ficheros se estructuran y organizan.
- Cómo se validan los datos y cómo se transforman.
- Quien accede y por qué.
- Cómo usarlos y en qué condiciones

- Nombres de las variables y descripciones.
- Explicación de códigos y esquemas de clasificación usados.
- Algoritmos usados para transformar los datos
- Formato de fichero y software

Fuente: [Taller] Jornada Aspectos prácticos de la preparación de propuestas en Horizon Europe. 18-19 octubre 2023, CITIC

Descripción de los datos (metadatos)

Fase 3

¿Dónde se almacenan esos metadatos?

Deben almacenarse junto con los datos, ya que facilitarán su interpretación posterior.

Varias formas:

1. Complimentar en el propio **repositorio** los metadatos que describen el dataset:

[Decision tree and glossary for classifying study designs: interactive version \(WCT EVI MAP Project\)](#)

2. Elaborar un **fichero readme.txt** que se incluirá junto al resto de ficheros del dataset



```
-----  
INFORMACIÓN GENERAL  
-----  
1. Título del documento:  
  
Decision tree and glossary for classifying study designs - interactive version (WCT EVI MAP Project)  
  
2. Contacto  
  
Investigador/a de contacto  
Nombre: Elena Plans-Beriso (autor principal)  
Filiación: National Center for Epidemiology. Instituto de Salud Carlos III, Madrid, Spain.  
Correo electrónico: elena.plans@isciii.es  
ORCID: https://orcid.org/0000-0002-9388-8744  
  
3. Descripción del proyecto  
  
The WCT EVI MAP project seeks to create a comprehensive map of the scientific evidence relevant to the World Health Organization's constrained timeframe. To address this, the review process will primarily focus on article abstracts, which, however, lack expert expertise. To mitigate these and other challenges, a decision tree accompanied by a glossary has been developed. This tool will be used to guide the review process.  
  
4. Descripción del documento  
A decision tree for classifying study designs and a related glossary are presented in an interactive platform for ease of use.  
  
5. Notas  
  
6. Fecha de depósito de los ficheros  
dic 2024  
  
7. Fecha de creación de los ficheros  
dic 2024  
  
8. Idioma  
English
```

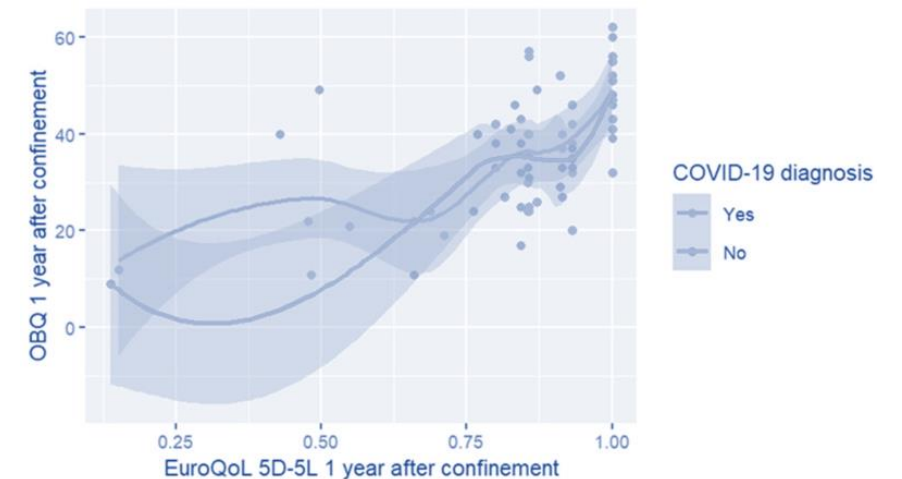
Fuente: <https://hdl.handle.net/20.500.12105/26132>

Plan de análisis estadístico (PAS)

- Define los **métodos estadísticos** que se aplicarán.
- Debe elaborarse **junto con el protocolo** del estudio.

Beneficios

- Evita recoger datos irrelevantes.
- Optimiza el diseño del CRD.
- Garantiza una base de datos coherente y lista para el análisis.



Fuente: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11545742/#sec3-healthcare-12-02164>

Miller G, Spiegel E. Guidelines for Research Data Integrity (GRDI). *Sci Data*. 2025;12(1):95. doi:10.1038/s41597-024-04312-x.

El análisis de datos debe documentarse de forma que garantice que los resultados sean reproducibles por otros investigadores.

Buenas prácticas

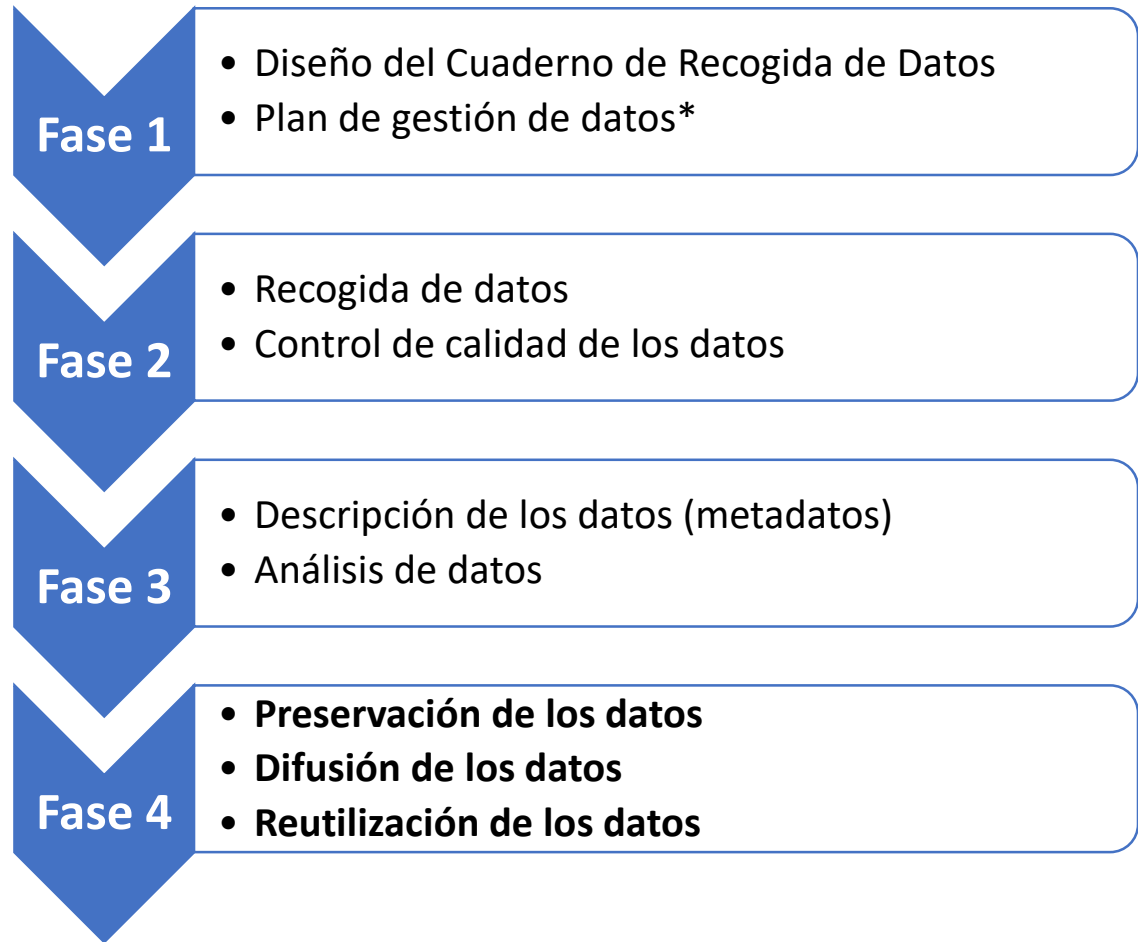
- Mantener la **versión original** de los datos y registrar cada cambio.
- Documentar los **métodos estadísticos** o cualitativos utilizados.
- Guardar **scripts, fórmulas o sintaxis** cuando se usan herramientas como R, SPSS o Excel.

El **objetivo** es que otros investigadores puedan:

- **Comprender** los resultados.
- **Verificarlos** o reproducirlos.
- **Reutilizarlos** en futuros estudios.

Miller G, Spiegel E. Guidelines for Research Data Integrity (GRDI). *Sci Data*. 2025;12(1):95. doi:10.1038/s41597-024-04312-x.

Ciclo de vida de los datos



Objetivo

Conservar los datos a largo plazo y asegurar que sigan siendo **comprensibles y reutilizables** en el futuro.

- **Gestión de archivos:** organizar los datos con nombres claros, versiones y estructura lógica:
 - Proyecto_Sueno_v01_2025-11-05.csv
- **Formatos abiertos:** guardar los datos en formatos como **CSV o TXT** .
- **Documentación con metadatos:** describiendo variables, código, limpieza, estructura y formato.
- **Almacenamiento seguro:** mantener copias de seguridad y acceso controlado.

UK Data Service. Preservation Recommended formats [Internet]. Essex: UK Data Service.

Objetivo

Publicar los datos en abierto para que otros investigadores puedan reutilizarlos.

Beneficios

- Aumenta la **colaboración** entre profesionales.
- Cumple con **requisitos** de revistas y financiadores.



“Quienes recojan y traten datos con financiación pública deberán ponerlos a disposición de los organismos competentes en materia de acceso a los datos sanitarios para facilitar su reutilización” (datos.Gob.es)

Repositorios

Plataforma digital que **almacena, preserva y pone a disposición pública** los datos de investigación.

Tipos de repositorios

- **Temáticos:** centrado en un área específica (clínicos, genómicos...)
- **Institucionales:** gestionados por universidades o centros de investigación.
- **Generales:** abiertos a cualquier disciplina (Zenodo).
- **Asociados a publicaciones:** permiten subir los datos como material suplementario de artículos científicos.

El acceso abierto a los datos no significa necesariamente acceso sin restricciones:
puede haber niveles de acceso, licencias específicas o requisitos de solicitud.

¿Qué beneficios proporciona un repositorio?



al investigador



- Permite publicar en abierto los documentos resultado de investigación y así cumplir los mandatos (de organismos financiadores).
- **Más visibilidad e impacto**, más citas.
- Garantiza una **correcta gestión de los derechos de autor**.
- **Acceso perpetuo a los trabajos** mediante enlaces permanentes.
- **Datos de uso** que facilitan la participación en evaluaciones.
- Permite el depósito de **todo tipo de documentos**, incluso inéditos (artículos, monografías, capítulos de monografías, comunicaciones en eventos, tesis, trabajos académicos, datasets, vídeos, etc.) **y en todo tipo de formatos**.



a la universidad



- **Reunir y difundir** al mundo la producción científica y académica de la institución.
- **Aumento de la visibilidad** de la institución a través de las obras de sus autores y mejora del posicionamiento en Google.
- **Preservación para el futuro** de la obra de los autores y de la actividad intelectual de la universidad.



a la sociedad



- **Acceso al conocimiento** para toda la sociedad y reutilización en beneficio de todos.
- Permite **visibilizar** y rendir cuentas de la inversión pública realizada en investigación.
- **Disminuye la brecha de acceso a la información** entre las instituciones y países.

<https://recolecta.fecyt.es/>

RECOLECTA

Agregador nacional de repositorios científicos de acceso abierto

En publicación ▾

BUSCAR



191

Repositorios



114

Repositorios institucionales



3.567 K

Resultados de investigación



1.447 K

Artículos



33.313

Set de datos

- **RECOLECTA:** [Database on holdover time of lightning-ignited wildfires](#)
- **HANDLE:** <https://ria.asturias.es/RIA/handle/123456789/14585>

Archivos en este documento:

Fichero	Descripción	Tamaño
0_readme.txt		1.97 kB
1_censored_data.csv		272.93 kB
2_non_censored_data.csv		201.96 kB
3_ancillary_data.csv		16.1 kB
4_description_variables.html		54.72 kB
5_references.html		42.16 kB
6_reproducible_examples.html		990.72 kB

Mostrar el registro Completo

```
-----  
Content of the database  
-----  
0_readme.txt --> This file  
  
1_censored_data.csv --> 42 frequency distributions of holdover time  
(2311 rows excluding headers and 12 columns)  
  
2_non_censored_data.csv --> 9 datasets of exact holdover times  
(6301 rows excluding headers and 9 columns)  
  
3_ancillary_data.csv --> Additional information on the original data  
sources (42 rows excluding headers and 33 columns)  
  
4_description_variables.html --> Description of the variables used  
in the database  
  
5_references.html --> Complete list of references  
  
6_reproducible_examples.html --> Examples for using the database  
with R
```

Uso secundario de datos

Aprovechar datos existentes para optimizar recursos y generar nuevo conocimiento.

- Ahorra **tiempo y recursos** de investigación.
- Permite **validar y reproducir** resultados.
- Facilita **nuevos análisis y descubrimientos** sin necesidad de recolectar más datos.

Licencias de uso

- Revisa siempre las condiciones de reutilización.
- Algunos datos son **abiertos**, otros **requieren permiso o no pueden reutilizarse**.

Algunos ejemplos de tipos de datos abiertos relacionados con la salud:



Datos de estudios e investigaciones médicas



Registros anonimizados de pacientes



Datos relacionados con los servicios de gestión sanitaria



Datos sobre hábitos de los ciudadanos

Fuente: https://datos.gob.es/sites/default/files/blog/file/reutilizacion_datos_abiertos_salud_y_bienestar.pdf

Buenas prácticas para reutilizar datos

- Usa **repositorios fiables** y reconocidos.
- Revisa la **documentación** para entender el contexto.
- Comprueba **licencias y restricciones** de acceso.
- Documenta tu **proceso** de reutilización.
- **Cita** correctamente los datos reutilizados.

3. Plan de Gestión de Datos (PGD)

¿Cómo elaborar un PGD?

Resumen de los datos

Datos FAIR

Asignación de recursos

Seguridad

Aspectos éticos

3. Plan de Gestión de Datos

El Plan de Gestión de Datos (Data Management):

Describe la gestión que se aplicará a los datos a lo largo del proyecto de investigación.

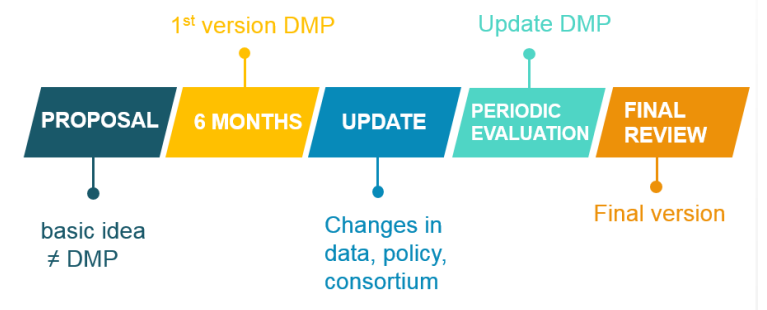


- Permite planificar, organizar y documentar los datos a lo largo de un proyecto de investigación.
- Incluye **todo el ciclo de vida de los datos**, desde su generación hasta su reutilización o eliminación.
- A modo de resumen, debería contemplar:



3. Plan de Gestión de Datos

- Es un **documento vivo**, elaborado al inicio del proyecto y actualizado periódicamente durante su desarrollo.
- En Horizonte Europa el PGD es obligatorio (Art. 35, p. 55)
- En el protocolo de investigación debe incluirse la gestión de datos.
- Requisitos según el programa Horizonte 2020:
 - Primera versión dentro de los primeros 6 meses del proyecto.
 - Actualizarse a mitad el proyecto (si su duración es superior a 12 meses).
 - Actualizarse al final del mismo (versión final).



Fuente: <https://www.openaire.eu/research-data-management-handbook>

3. Plan de Gestión de Datos

¿Por qué es esencial?

- **Recopilar y almacenar** datos de una manera **más estructurada**.
- Mejorar la protección de datos.
- **Minimizar** el riesgo de **pérdida de datos**.
- Compartir y reutilizar **datos FAIR** con garantías.
- La **verificabilidad y validación de los resultados**.
- Aumenta la **longevidad** del proyecto al estar los datos disponibles.
- **Identificar costes** asociados a un proyecto de investigación.
- Uso eficiente de los **recursos** disponibles.
- Cumplir las expectativas de agencias de financiamiento.

Garantía de calidad de los datos

Exactos, completos, auténticos y fiables

Proteger y preservar los datos



3. Plan de Gestión de Datos

Estructura propuesta



Plantilla de los programas de Horizonte Europa

1. Resumen de los datos
2. Datos FAIR
3. Asignación de recursos
4. Seguridad de los datos
5. Aspectos éticos

PROJECT	
Project number:	[project number]
Project acronym:	[acronym]
Project name:	[project title]

DATA MANAGEMENT PLAN	
Date:	[dd/mm/yyyy]
Version:	[DMP version]

1. Data Summary

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

What types and formats of data will the project generate or re-use?

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

What is the expected size of the data that you intend to generate or re-use?

What is the origin/provenance of the data, either generated or re-used?

To whom might your data be useful ('data utility'), outside your project?

https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/temp-form/report/data-management-plan_he_en.docx
<https://doi.org/10.5281/zenodo.8046772>

3. Plan de Gestión de Datos

<https://repisalud.isciii.es/entities/publication/3c01ac1a-8fa5-4c5a-9277-0cf0e0a8a9f0>

Publicación: Medicina Personalizada (MedPer) en la detección precoz del Deterioro Cognitivo (DC) preclínico (DENDRITE). Desarrollo de un modelo predictivo de riesgo. Plan de Gestión de Datos (PGD) v. 2.0



Archivos

[DENDRITE_PGD_v2_20250825.pdf \(619.9 KB\)](#)


Identificadores

URI: <https://hdl.handle.net/20.500.12105/26876>

Fecha de publicación

2025-08-25

Autores

Instituto de Salud Carlos III. Unidad de Investigación en Cuidados de Salud (Investén-ISCIII)
Instituto de Investigación Biomédica de Salamanca (IBSAL)
CIBER de Fragilidad y Envejecimiento Saludable, Instituto de Salud Carlos III (CIBERFES, ISCIII)
Alonso-del-Cura, Olatz 

Resumen

Plan de Gestión de Datos del proyecto DENDRITE. El objetivo principal del proyecto MedPer_DC es obtener un modelo predictivo para la detección preclínica del deterioro cognitivo en población de 55-70 años de edad, permitiendo así la implantación efectiva de la Medicina Personalizada de Precisión en Atención Primaria. El estudio involucra la participación de 6 Comunidades Autónomas y 7 centros de Atención Primaria en diferentes ubicaciones geográficas. Es una cohorte observacional, multicéntrica y no intervencional con métodos mixtos convergentes. Se desarrollará combinando una fase cuantitativa observacional y prospectiva con una fase cualitativa que involucrará grupos de discusión. Durante el estudio se recogerán datos multi-ómicos (moleculares, proteómicos, genómicos), asistenciales, sociales, ambientales, de voz y de conducta (provenientes de monitorización activa y pasiva) junto con información clínica complementaria. Todos estos datos se analizarán mediante diversas técnicas de análisis basadas en Inteligencia Artificial y en modelos estadísticos para sintetizar un perfil clínico, genético y molecular, de comportamiento y voz, y determinar su potencial para converger en un modelo predictivo global de DC.

Palabras clave

Deterioro cognitivo
Diagnóstico precoz
Medicina de precisión
Infraestructura de Medicina de Precisión asociada a la Ciencia y la Tecnología (IMPACT)
DENDRITE
Adultos de 55 a 70 años
Envejecimiento
Atención Primaria
Medicina Personalizada (MedPer)

3. Plan de Gestión de Datos

<https://repisalud.isciii.es/entities/publication/3c01ac1a-8fa5-4c5a-9277-0cf0e0a8a9f0>



Estructura propuesta →

Plantilla de los programas de Horizonte Europa

Índice de contenido

1. Resumen de los datos

1.	Resumen de los datos	5
1.1	Propósito de la generación de datos y su relación con los objetivos del proyecto	5
1.2	Tipo, formato y origen de los datos generados y recopilados	5
1.3	Reutilización de los datos	7
1.4	Tamaño esperado de los datos	7
1.5	Utilidad de los datos	8
2.	Fair data	8
2.1	Localización de los datos, incluyendo los metadatos (Findable Data)	8
2.1.1	Identificadores persistentes y únicos	8
2.1.2	Suministro de metadatos	8
2.1.3	Estándares para la creación de metadatos	8
2.1.4	Palabras clave	8
2.1.5	Recolección	8
2.2	Accesibilidad de los datos (Accesible Data)	9
2.2.1	Repositorios	9
2.2.2	Datos	9
2.2.3	Metadatos	11
2.3	Datos interoperables (Interoperable Data)	11
2.3.1	Vocabularios, estándares o metodologías para facilitar la interoperabilidad	11

2. Datos FAIR

2.4	Reutilización de los datos (Reusable Data)	11
2.4.1	Garantizar la reutilización	11
2.4.2	Tipo de licencia	11
2.4.3	Uso por terceras personas de los datos tras su finalización	11
2.4.4	Descripción de los procesos que aseguren la calidad de los datos	12
2.4.5	Período de tiempo durante el cual los datos permanecerán reutilizables	12
3.	Asignación de recursos	12
3.1	Coste destinado a adaptar los datos a los principios FAIR y a la preservación a largo plazo del dataset	12
3.2	Responsabilidades de la gestión de datos	12
4.	Seguridad	13
5.	Aspectos éticos	13
6.	Otros aspectos	14
6.1	Referencias	14

3. Asignación de recursos

4. Seguridad de los datos

5. Aspectos éticos







Plan de Gestión de Datos – Proyecto DENDRITE

4

3. Plan de Gestión de Datos

<https://zenodo.org/records/14024376>









TWEET-IE Deliverable 6.3: Data Management Plan

Prospathopoulos, John (Data manager)¹ ; Croce, Alessandro (Data manager)² ; Bouris, Demetri (Data manager)¹ ; Manolesos, Marinos (Data manager)¹ ; Riziotis, Vasilis (Data manager)¹ ; Chasapogiannis, Petros (Data manager)¹ 

Show affiliations

Contributors

Data collectors:

Vasiliki Pappa (orcid:0009-0007-5890-401X)¹ ; Pallas, Nikolaos-Petros¹ ; Kellaris, Konstantinos¹ ; Gromke, Christof² ; Mühle, Franz³ ; Tamaro, Simone³ ; Sciacchitano, Andrea⁴ ; DSouza, Brian⁴ 

Editor: Chondromatidis, Ioannis¹ 

Show affiliations

This deliverable describes the preliminary version of the Data Management Plan (DMP) of the TWEET-IE project. The aim of the DMP is to provide an analysis of the main elements of the data management policy that will be used throughout the project, with regard to all the datasets that will be generated and used. DMP is a living document that will be updated as the implementation of the project progresses and when significant changes occur. The format of DMP follows the guidelines suggested for Horizon Europe calls.


Up to this point (month 24), three datasets have been uploaded to Zenodo:

- The first is the dataset from the first twin test conducted in the KIT and NTUA wind tunnels (effect of vegetation on urban flows, <https://doi.org/10.5281/zenodo.10019001>), including LDV and PIV data
- The second dataset contains the wind tunnel measurement data from Phase I of the Blind test on wind turbine wake (<https://doi.org/10.5281/zenodo.10566400>), conducted in the closed-loop, low-speed boundary layer wind tunnel of TUM.
- The third is the dataset from the second twin test conducted in the TUM and NTUA wind tunnels (wake interactions of a cluster of turbines and wake steering techniques, <https://doi.org/10.5281/zenodo.14017240>). While both experimental set ups are detailed, only data from the TUM campaign are available at the time of writing, as the NTUA campaign results will form Phase II of an ongoing blind test campaign and cannot be published.

Files

TWEET-IE D6-3 v4-0.pdf

 VIEWS

 DOWNLOADS

► Show more details

Versions

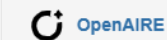
Version 4.0 10.5281/zenodo.14024376	Nov 1, 2024
Version 3.1 10.5281/zenodo.14023125	Jun 20, 2024
Version 3.0 10.5281/zenodo.14022681	Mar 31, 2024
Version 2.0 10.5281/zenodo.10061049	Nov 1, 2023

[View all 4 versions](#)

Cite all versions? You can cite all versions by using the DOI [10.5281/zenodo.10061048](https://doi.org/10.5281/zenodo.10061048). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

External resources

Indexed in



3.1 Resumen de los datos

1. ¿Cuál es el **propósito** de generar o reutilizar los datos **y su relación con los objetivos** del proyecto?
2. ¿Qué **tipos y formatos de datos** generará o reutilizará el proyecto?
3. ¿Cuál es el **tamaño esperado** de los datos que piensa generar o reutilizar?
4. ¿Cuál es el **origen o procedencia** de los datos, tanto generados como reutilizados?
5. ¿Reutilizará datos existentes y para qué? (si corresponde)
6. ¿Para quién pueden ser útiles sus datos ("**utilidad de los datos**"), fuera del proyecto?

Titulo del dataset
Propósito
Tipo
Formato
Tamaño
Origen
Reutilización
Utilidad de los datos

1. Nuevos datos se generarán para la validación clínica de la herramienta.
2. Datos estructurados (puntuaciones de las herramientas). Almacenados en formato CSV. Gráficos en formato pttx y/o Tablas en xlsx.
3. El tamaño aun no se conoce. Es probable que no sobrepase los 5 GB.
4. Se recogerán de mediciones en tiempo real de los participantes a partir de herramientas estandarizadas.
5. No se reutilizarán dataset generados por terceros.
6. Los datos pueden ser útiles para otros grupos de investigación que trabajen en (...)

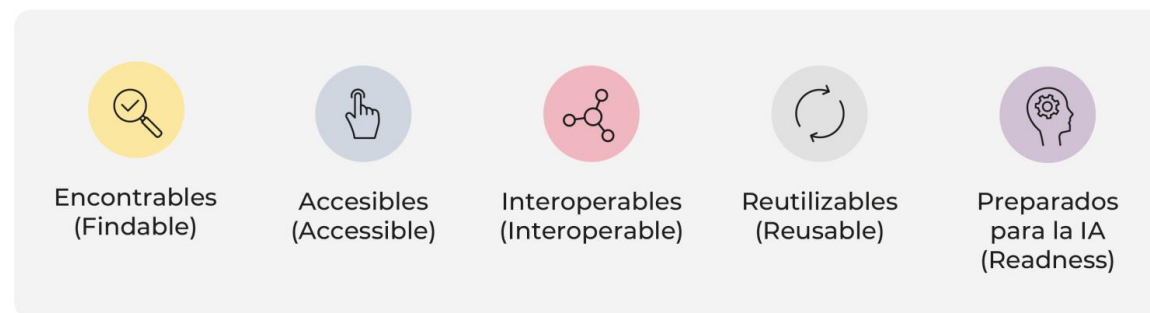


3.2 Datos FAIR

Principios FAIR

Son un conjunto de principios que describen los atributos que los datos deben tener para que se puedan localizar (F), acceder (A), entender (I) y reutilizar (R) fácilmente. Diseñados para **el intercambio y uso eficiente de los datos**.

- Guía para que los **datos**:
 - Sean útiles no solo para personas, sino también para **máquinas y sistemas automatizados**.
 - Deben estar bien descritos, accesibles y listos para ser reutilizados por otros profesionales.



Fuente: <https://datos.gob.es>

Proporcionan un marco común, medible y sostenible para la **gestión de datos de investigación**.

3.2 Datos FAIR

PRINCIPIOS FAIR



datos.gob.es
reutiliza la información pública

FINDABLE

Tiene que poder ser fácil encontrar los datos

ACCESIBLE

Tiene que ser fácil poder acceder a los datos para utilizarlos

INTEROPERABLE

Tiene que ser fácil poder interoperar con ellos

REUSABLE

Los datos tienen que poder ser reutilizados

[Principios FAIR: el secreto de los magos de los datos | datos.gob.es](https://datos.gob.es)

3.2 Datos FAIR

Como un libro...

F - Findable (Localizable)	<ul style="list-style-type: none">• Identificadores persistentes (DOI o handle).• Estándares/Esquemas indexables en repositorios.	Con título, autor y número de referencia en el catálogo
A - Accessible (Accesible)	<ul style="list-style-type: none">• Protocolo (métodos o herramientas de acceso) estandarizado, abierto, gratuito y universalmente aplicable.• Definir formas y niveles de acceso. Declaraciones de disponibilidad.• Restricciones: quién, cómo y por cuánto tiempo.	A consultar en una sala de lectura bajo ciertas condiciones
I – Interoperable (Interoperable)	<ul style="list-style-type: none">• Formatos (CSV) y vocabularios estandarizados (tesauros, CIE-10).• Metadatos que sigan estándares reconocidos (Dublin core).• Integración fluida con otros datos y sistemas.	Escrito en un idioma conocido y con estructura comprensible para distintos lectores
R - Reusable (Reutilizable)	<ul style="list-style-type: none">• Atributos precisos (bien descritos y conservados).• Licencias de uso y de reutilización clara.• Criterios de “provenance” (versiones).• Usar formatos abiertos para preservación a largo plazo.• Vocabularios reconocibles para su área de conocimiento.	Bien conservado y que otras personas pueden usar, citar y aprender de él

3.2 Datos FAIR

Para ser encontrado:

1. Identificador único y persistente
 - ✓ DOI o handle
2. Indexados en sistemas con motor de búsqueda.
 - ✓ Palabras clave de búsqueda
 - ✓ Nomenclatura
3. Datos descritos con metadatos enriquecidos.
 - ✓ Especificar estándares para la creación de metadatos
4. Especificar en los metadatos el identificador de los datos que se describen

indable = Fáciles de encontrar



Fuente: https://github.com/Open-Science-Training-Handbook/Open-Science-Training-Handbook_ES/blob/master/02OpenScienceBasics/02OpenResearchDataAndMaterials.md

3.2 Datos FAIR



e-cienciaDatos > Universidad Carlos III de Madrid > Federico-Tena World Population Historical Database >

Federico-Tena World Population Historical Database : Canada

Versión 1.0

1. DOI:

Federico, Giovanni; Tena Junguito, Antonio, 2023, "Federico-Tena World Population Historical Database : Canada", <https://doi.org/10.21950/FYZHE6>, e-cienciaDatos, V1

Citar dataset ▾

Revise los [Estándares de citas de datos](#).

Acceder al dataset ▾

Contactar con el propietario | Compartir

Estadísticas completas del dataset
Estadísticas "Make Data Count" (MDC) ?
desde 2020-03-01

Visualizaciones 969 | Descargas 12

0 Citas (desde Crossref) ?

Descripción ?

Project developed by Giovanni Federico (New York University Abu Dhabi) and Antonio Tena Junguito (Universidad Carlos III de Madrid). Dataset: Canada

Materia ?

Ciencias sociales

Palabra clave ?

World Population, Population by countries, New estimation Demographic transition

Publicación relacionada ?

Federico, G., Tena Junguito, A. (2023). How many people on earth? World population 1800-1938. Working Papers in Economic History. 23-02 [handle: 10016/36431](#)

Notas ?

Web site: <https://www.uc3m.es/investigacion/federico-tena-population>
SEE MAP

Licencia/Acuerdo de uso de los datos



Ficheros

Metadatos

Condiciones

Versiones

2. Palabras clave de búsqueda

(https://datos.gob.es/sites/default/files/doc/file/guia_calidad_de_datos_abiertos_1.pdf
<https://e-archivo.uc3m.es/entities/publication/86cf4b87-34ce-4339-94eb-a4d73a4bbf19>
<https://e-archivo.uc3m.es/entities/publication/adf04df6-266f-4c0f-bf4b-11d91fd3eae6>

Licencia CC BY-NC-SA 4.0

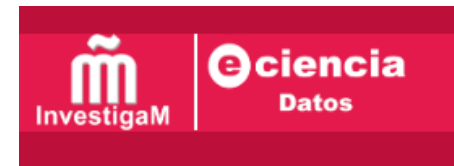
3.2 Datos FAIR



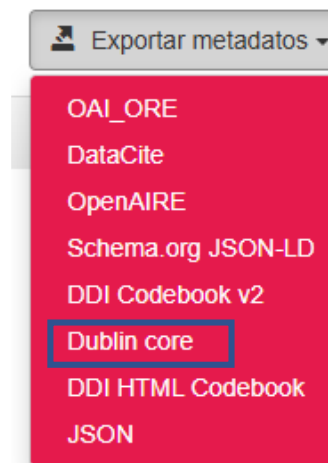
3. Metadatos enriquecidos.

4. Identificador de los datos descritos.

ID persistente	doi:10.21950/FYZHE6
Fecha de publicación	2023-04-14
Título	Federico-Tena World Population Historical Database : Canada
URL Alternativa	https://www.uc3m.es/investigacion/federico-tena-population
Autor	Federico, Giovanni (New York University Abu Dhabi) Tena-Junguito, Antonio (Universidad Carlos III de Madrid)
Contacto	Utilice el botón de e-mail de arriba para contactar. Tena Junguito, Antonio (Universidad Carlos III de Madrid)
Descripción	Project developed by Giovanni Federico (New York University Abu Dhabi) and Antonio Tena Junguito (Universidad Carlos III de Madrid). Dataset: Canada
Materia	Ciencias sociales
Palabra clave	World Population Population by countries New estimation Demographic transition
Publicación relacionada	Federico, G., Tena Junguito, A. (2023). How many people on earth? World population 1800-1938. Working Papers in Economic History. 23-02 handle 10016/36431 http://hdl.handle.net/10016/36431
Notas	Web site: https://www.uc3m.es/investigacion/federico-tena-population SEE MAP
Idioma	Inglés
Fecha de producción	2023-01
Información de la subvención	Agencia Estatal de Investigación (España): PID2021-124394NB-I00
Depositante	Tena Junguito, Antonio
Fecha de depósito	2023-03-31
Periodo de tiempo cubierto	Start Date: 1800 ; End Date: 1938
Tipo de datos	census/enumeration data
Software	Microsoft Excel
Dataset relacionado	Federico-Tena World Trade Historical Database Web site: https://www.uc3m.es/tradehist_db https://edatos.consorciodmadrono.es/dataverse/federico-tena



Estándares de metadatos:



Guía práctica para la mejora de la calidad de los datos (https://datos.gob.es/sites/default/files/doc/file/guia_calidad_de_datos_abiertos_1.pdf)

<https://e-archivo.uc3m.es/entities/publication/86cf4b87-34ce-4339-94eb-a4d73a4bbf19>

Licencia CC BY-NC-SA 4.0

<https://e-archivo.uc3m.es/entities/publication/adf04df6-266f-4c0f-bf4b-11d91fd3eae6>



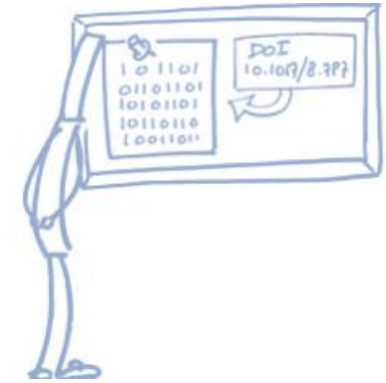
3.2 Datos FAIR

Para ser accesible:

1. Qué **datos** estarán **disponibles** de forma **abierta**.
 - ✓ Acceso restringido a datos sensibles.
 - ✓ **Metadatos** accesibles aunque los datos ya no estén disponibles.
2. Métodos o herramientas de **software** necesario para acceder a los datos. **CSV**
3. **Cómo** estarán disponibles los datos (**repositorio confiable “trusted repository”**).
4. **Dónde** se depositan los datos, los metadatos, y el código asociado.
5. Cómo se proporcionará el acceso en caso de restricciones
 - ✓ Autenticación y autorización (si necesario).

1. Todos excepto los datos sensibles serán de acceso abierto. Los datos sensibles (datos cualitativos brutos) serán anonimizados antes de ponerse a disposición.
2. Se archivarán en un formato CSV compatible con Hojas de cálculo de Google o Rstudio, software de código abierto.
3. Se pondrán a disposición mediante su depósito en repositorios Zenodo/e-cienciaDatos.

Accesible = accesible



Fuente: https://github.com/Open-Science-Training-Handbook/Open-Science-Training-Handbook_ES/blob/master/02OpenScienceBasics/02OpenResearchDataAndMaterials.md

Recuperables y descargados, por su identificador, sin coste para los usuarios y de aplicación universal a cualquier sistema



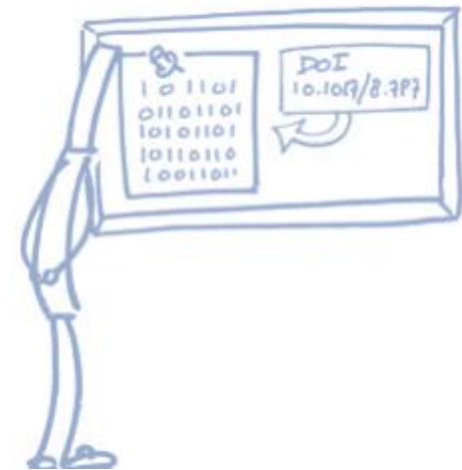
3.2 Datos FAIR

5. Cómo se proporcionará el acceso en caso de restricciones

03_DS_Quest_WEMAC_U4S_readme_v02.txt
 Texto plano - 9,0 KB
 Publicado 4 ago 2023
 17 Descargas
 MD5: 759...92c

03_DS_Quest_WEMAC_U4S_volunteers_ID.tab.cpt
 application/mac-compactpro - 1,2 KB
 Publicado 4 ago 2023
 16 Descargas
 MD5: a07...7f8
 IMPORTANT NOTE: This file is encrypted. To obtain the password, you must send the DUA (Data Usage Agreement) available on this website <https://www.uc3m.es/institute-gender-studies/DATASETS-to-uc3m4safety@uc3m.es>

A

ccesible = accesible

Condiciones de uso

Licencia/Acuerdo de uso de los datos Tanto por nuestras [Normas de la comunidad](#) como por las buenas prácticas científicas, se espera que se acredite su uso de forma correcta mediante una cita. Por favor, use la cita de datos mostrada en la página del dataset.

[Condiciones de uso personalizadas para el dataset](#) — Las siguientes condiciones de uso del dataset se han definido para este dataset.

Condiciones de uso This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

Creative Commons Attribution 4.0 International License

Requisitos de citas Publications: Publications include not only papers, but also presentations for conferences or educational purposes.

Renuncia de responsabilidad Warranty: The dataset comes without any warranty, the UC3M4Safety WEMAC dataset administrators cannot be held accountable for any damage (physical, financial, or otherwise) caused by the use of the database.

Nombre *

Fuente: https://github.com/Open-Science-Training-Handbook/Open-Science-Training-Handbook_ES/blob/master/02OpenScienceBasics/02OpenResearchDataAndMaterials.md



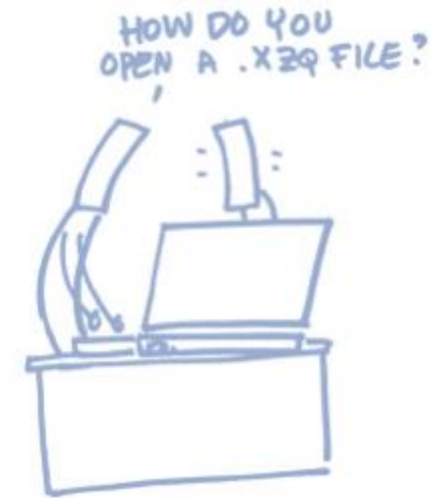
3.2 Datos FAIR

1. Qué datos y qué vocabularios de metadatos, estándares o metodologías seguirá.
2. ¿Vocabularios estandarizados para todos tipos de datos?
 - ✓ Si no, ¿proporcionará asignaciones entre los vocabularios específicos del proyecto y las ontologías más utilizadas?

Todos los conjuntos de datos generados seguirán estándares de datos y metadatos reconocidos:

- Vocabularios controlados MeSH/DeCS.
- Esquemas de metadatos Dublin Core.
- Representación de fechas ISO 8601.
- Formatos abiertos como CSV, TXT, XML.

Interoperable

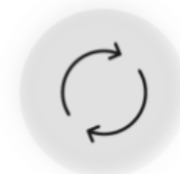


Fuente: https://github.com/Open-Science-Training-Handbook/Open-Science-Training-Handbook_ES/blob/master/02OpenScienceBasics/02OpenResearchDataAndMaterials.md



Fuente: https://www.researchgate.net/figure/Figura-2-Niveles-de-la-interoperabilidad-La-ventaja-de-la-herramienta-presentada-esta_fig2_262915194

3.2 Datos FAIR

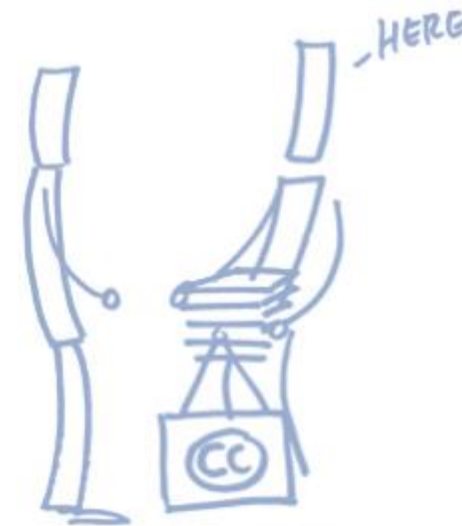


Fuente: <https://e-archivo.uc3m.es/entities/publication/adf04df6-266f-4c0f-bf4b-11d91fd3eae6>

1. **Licencia** de uso y reutilización, clara y accesible.
2. Cuándo estarán los datos **disponibles** para su reutilización.
 - ✓ **embargo de datos** (Si corresponde).
 - ✓ Horizonte Europa permite un máximo de tiempo de embargo de 12 meses.
3. Especifique si son utilizables **por terceros**.
 - ✓ Si restricciones, explique **por qué**.
4. Describa los procesos para asegurar la **calidad** de los datos.
5. Especifique el **periodo de tiempo** durante el cual los datos se mantendrán reutilizables.

1. Se publicarán bajo licencia CC-BY-NC-SA 4.0.
2. Los datos estarán bajo embargo, por un máx. de 12 meses desde la fecha de publicación
3. Estarán disponibles para terceros en acceso abierto (protección de datos personales).
4. El CGD garantizará la evaluación de la calidad de los datos.
5. Seguirán siendo reutilizables hasta que los repositorios dejen de dar soporte de calidad.

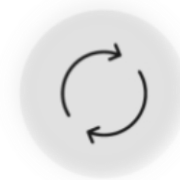
Reusable = reutilizable



Fuente: https://github.com/Open-Science-Training-Handbook/Open-Science-Training-Handbook_ES/blob/master/02OpenScienceBasics/02OpenResearchDataAndMaterials.md







Múltiples atributos precisos y relevantes que describen detalladamente los datos

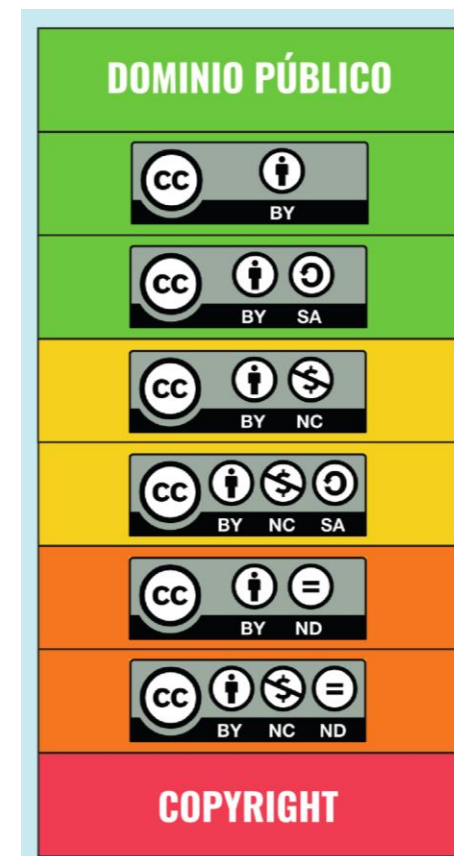
3.2 Datos FAIR



1. Licencia de uso y reutilización -> ¿CC-BY-NC-SA 4.0?

Reusable = reutilizable

LICENCIA	PERMITE	SIEMPRE QUE
 Reconocimiento (BY)	Compartir (copiar y redistribuir) y adaptar (remezclar, transformar y construir a partir del material), incluso para fines comerciales.	<ul style="list-style-type: none"> Se reconozca la autoría de la obra original de forma adecuada.
 Reconocimiento - Compartir Igual (BY-SA)	Compartir (copiar y redistribuir) y adaptar (remezclar, transformar y construir a partir del material), incluso para fines comerciales.	<ul style="list-style-type: none"> Se reconozca la autoría de la obra original de forma adecuada. Si se crea una obra derivada, esta se comparta bajo la misma licencia (BY-SA)
 Reconocimiento - Sin Obra Derivada (BY-ND)	Compartir (copiar y redistribuir) el material, incluso para fines comerciales.	<ul style="list-style-type: none"> Se reconozca la autoría de la obra original de forma adecuada. No se distribuyan modificaciones de la obra original.
 Reconocimiento - No Comercial (BY-NC)	Compartir (copiar y redistribuir) y adaptar (remezclar, transformar y construir a partir del material).	<ul style="list-style-type: none"> Se reconozca la autoría de la obra original de forma adecuada. No se utilice con propósitos comerciales.
 Reconocimiento - No Comercial - Compartir Igual (BY-NC-SA)	Compartir (copiar y redistribuir) y adaptar (remezclar, transformar y construir a partir del material).	<ul style="list-style-type: none"> Se reconozca la autoría de la obra original de forma adecuada. No se utilice con propósitos comerciales. Si se crea una obra derivada, esta se comparta bajo la misma licencia (BY-NC-SA)
 Reconocimiento - No Comercial - Sin Obra Derivada (BY-NC-ND)	Compartir (copiar y redistribuir) el material.	<ul style="list-style-type: none"> Se reconozca la autoría de la obra original de forma adecuada. No se utilice con propósitos comerciales. No se distribuyan modificaciones de la obra original.



3. Plan de Gestión de Datos

Conserva tus derechos de autor frente a las editoriales

Ten en cuenta antes de firmar el contrato

- Versión del documento** que se puede depositar en abierto
↓
Consulta
SHERPA RoMEO, DULCINEA
- Cuándo** se puede depositar en abierto
↓
Embargo
Tiempo que la editorial se reserva para distribuir en exclusiva el artículo.
- Qué derechos de explotación** conserva el autor sobre el documento cedido a la editorial
↓
Incluye una adenda
Al firmar el contrato añade que está sujeto a las condiciones de la adenda que adjuntes.

Publicar en abierto es totalmente **compatible** con la edición comercial

Si cedes todos tus derechos, es posible que no puedas...

- Icono de ordenador: Poner tu trabajo en tu propia web o en un repositorio.
- Icono de carrito: Usar una copia de tu trabajo para distribuir entre tus estudiantes o en un curso.
- Icono de libro: Usar tu publicación como base para futuros artículos.

Con ellas mostrarás qué condiciones pones para usar tu obra.

crue Universidades Españolas Red de Bibliotecas REBIUN

CC creative commons

Fuente: <https://www.rebiun.org/acceso-abierto/conserva-tus-derechos-de-autor-frente-las-editoriales>

3.3 Asignación de recursos

- **Coste económico** de la gestión de datos.

No hay costes asociados a los mecanismos para cumplir con los principios FAIR y preservar la base de datos a largo plazo.

- **Recursos humanos:** quién es el responsable de la gestión en cada fase.
- **Recursos para la preservación a largo plazo.**

Ej. Responsabilidades sobre la gestión de los datos:

- a. El coordinador del proyecto tiene la máxima responsabilidad y se encargará de preservarlos a largo plazo.
- b. Un investigador será responsable de la validación, el registro, así como del control de calidad.
- c. Otro investigador garantizará la integridad y compatibilidad para su uso por los miembros del equipo.
- d. El coordinador supervisará y asegurará la correcta coordinación entre todas estas responsabilidades.

- **Recursos tecnológicos:**

Ej. Se requerirá de un disco duro para la preservación del conjunto de datos.

3.4 Seguridad

- **Los datos deben estar protegidos frente a accesos, usos, modificaciones, divulgaciones o destrucciones no autorizadas.**
- Establecer protocolos claros para el almacenamiento, la realización de copias de seguridad, la preservación a largo plazo y la implementación de medidas de seguridad que aseguren la integridad y confidencialidad de la información.

Ej. Estrategia de seguridad:

- ❖ Restringir el acceso a materiales sensibles, como consentimiento informado.
- ❖ Mantener siempre los datos personales separados, seguros y cifrados.
- ❖ Utilizar cifrado en todos los dispositivos y soportes donde se almacene información sensible.

3.4 Seguridad

Aspectos legales:

- Tratamiento lícito de los datos.
- Minimización y anonimización de datos.
- Almacenamiento seguro y controles de acceso.
- Retención y eliminación de datos.
- Protocolos de intercambio de datos.
- Equidad, transparencia y responsabilidad.
- Derechos de propiedad intelectual.



Delegado
Protección
Datos

3.5 Aspectos éticos

¿**Qué** contenidos de tus datos están sujetos a **derechos de propiedad intelectual**?

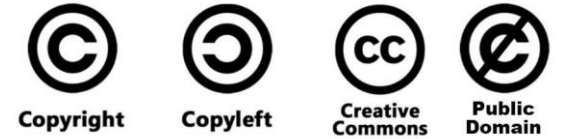
¿**Quién** será el propietario de los derechos sobre los datos que produzcas?

Si vas a reutilizar datos, ¿necesitas obtener **permiso o una licencia** para hacerlo?

¿Cuáles son las consideraciones éticas asociadas con los datos?

- comité de ética en investigación
- consentimiento informado
- Anonimización
- Legislación:
 - Normativa europea (Reglamento (UE) 2016/679). Reglamento general de protección de datos.
 - Ley Orgánica 3/2018 sobre protección de datos personales

¿Qué problemas éticos deben tenerse en cuenta al preservar y compartir los datos?



Fuente: [Derechos de autor, propiedad intelectual y copyright](#)

3.5 Aspectos éticos

Anonimización



Fuente: <https://nymiz.com/>

De la base de datos original:

- Eliminar **identificadores directos** (nombre y apellidos, NIF, dirección)
- Reducir el detalle de una variable mediante **agregación** (año de nacimiento en lugar de fecha de nacimiento, rango de edad)

Nombre	Apellidos	NIF	Dirección	Fecha nacimiento
--------	-----------	-----	-----------	------------------

Conjunto de datos original

ID aleatorizado	Año de nacimiento
-----------------	-------------------

Conjunto de datos anonimizado

3. Plan de Gestión de Datos

<https://repositoriorebiun.org/handle/20.500.11967/71>

Un **Plan de Gestión de Datos** (PGD) o Data Management Plan (DMP) es un **documento formal, que debe presentarse al inicio de la investigación, en el que se describe qué vas a hacer con tus datos durante y después de finalizar tu investigación** y que puede modificarse si se producen cambios en el proceso de la misma.

¿Por qué?

Es una **buena práctica**, es un **elemento clave de Open Science** y es **obligatorio** en los nuevos proyectos H2020.

Herramientas gratuitas para elaborar un PGD

-  **PGDonline** (Consortio Madroño) <http://dmp.consortiomadrono.es/>
-  **DMPonline** (Digital Curation Centre, UK) <https://dmponline.dcc.ac.uk/>

-  **01**
Revisa los **requerimientos** de la entidad financiadora (H2020).
-  **02**
Identifica los datos: tipología, procedencia, volumen, formatos y ficheros.
-  **03**
Define cómo se organizarán y gestionarán los datos: nombre de los ficheros, control de versiones, software necesario...
-  **04**
Explica cómo se documentarán los datos: identifica la información a procesar, consulta si hay estándares o esquemas de metadatos, identifica herramientas que permitan gestionarlos.
-  **05**
Describe los procesos que aseguran una **buena calidad de los datos**.
-  **06**
Prepara una estrategia de almacenamiento (durante el proceso) y de preservación de datos (repositorio).
-  **07**
Define las políticas de datos del proyecto: cuestiones sobre propiedad intelectual y cómo se tratarán los datos sensibles y personales.
-  **08**
Describe cómo se difundirán los datos: dónde, cuáles, cuándo se van a difundir. Si publicarás los datos en un repositorio, como información suplementaria del artículo o como un "data paper".
-  **09**
Asigna roles y responsabilidades para las personas y organizaciones participantes en el proyecto.
-  **10**
Prepara un presupuesto realista: la gestión de datos cuesta tiempo y dinero en términos de software, hardware, servicios y personal.

 **El personal de tu Biblioteca te puede asesorar**

Adaptado de Michener, WK. Ten simple rules for creating a good data management plan. PLOS Comp Biol. 2015; 11(10)

 **crue** Universidades Españolas  **Red de Bibliotecas REBIUN**



3. Plan de Gestión de Datos

Incluye todo el ciclo de vida de los datos

- **¿Qué datos se van a recoger?**

Tipo de datos, volumen, almacenamiento

- **¿Cómo se van a recoger o crear?**

Métodos de colección, estándares, organización de ficheros, control, protocolos de garantía de calidad durante y después de la colecta

- **¿Qué documentación y metadatos acompañarán a los datos?**

Describir la documentación para interpretar y usar: descripción de la metodología; metadatos; libros de códigos (codebooks); cuestionarios o instrumentos; análisis de los procedimientos

- **¿Cómo gestionar cualquier tema ético?**

Proteger confidencialidad: manejo y almacenamiento de datos sensibles, restricciones de acceso

- **¿Cómo gestionar el copyright y la propiedad intelectual?**

Identificar al propietario de los datos y las condiciones de uso, discutir permisos con los productores de datos

- **¿Cómo manejar el acceso y la seguridad?**

Definir el acceso restringido, describiendo medidas de seguridad y el cumplimiento de las normas de esas medidas

- **¿Qué datos debieran de retenerse, compartir y/o preservarse?**

Anotar el valor potencial de los datos, los esfuerzos que se requieren para preparar los datos para preservar y acceder

- **¿Cuáles son los plazos de preservación a largo plazo del dataset?**

Identificar un repositorio para archivar los datos, planes para preparar y documentar los datos y asegurarse de que serán preservados y usables en el futuro

- **¿Cómo compartir los datos?**

Identificar los mecanismos necesarios para compartir los datos, describir como otros pueden encontrar los datos y como los ficheros se entregarán, especificar el reconocimiento y requerimientos de citación

- **¿Se requieren restricciones a compartir los datos?**

Discutir preocupaciones de privacidad, habilitar la fecha en que se compartirán, anonimización de los datos, acuerdos de uso, explicar períodos de embargo

- **¿Quién será el responsable de la gestión de los datos?**

Identificar al encargado de supervisar e implementar el PGD

- **¿Qué recursos necesitarás para implementar tu plan?**

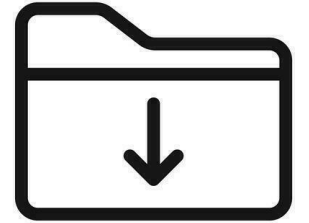
En función del tipo de dato (tamaño, geográficos...): identificar personal experto o equipo especializado, considerar los costes de la gestión durante y después del proyecto

Fuente: [Taller] Jornada Aspectos prácticos de la preparación de propuestas en Horizon Europe. 18-19 octubre 2023, CITIC

4. Material de apoyo – Repositorios

Repositorios “de confianza” para el depósito de los datos:

- <https://zenodo.org/>
- <https://data.mendeley.com/research-data/#/>
- <https://edatos.consorciomadrono.es/>
- <https://datadryad.org/>
- <https://www.re3data.org/>



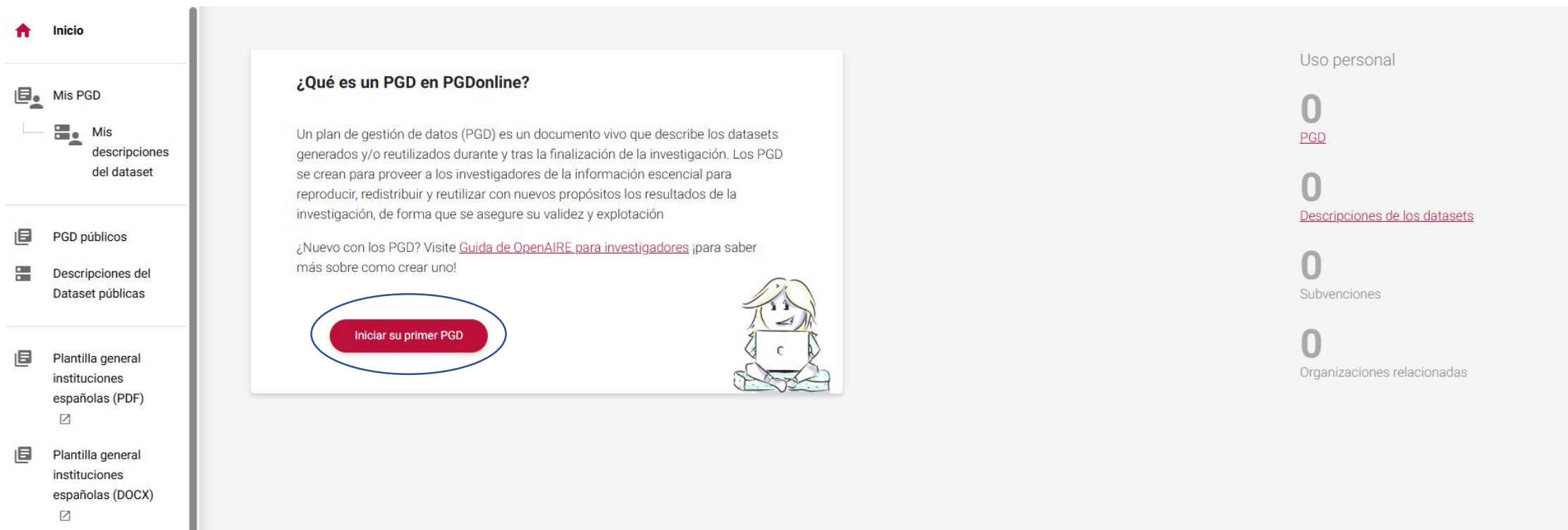
4. Material de apoyo – Plantillas para PGD

PGDonline: Crear un PGD

- Plantilla Horizonte Europa: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/temp-form/report/data-management-plan_he_en.docx
- Plantilla online del Consorcio Madroño, basado en Argos, desarrollado por OpenAIRE y adaptado por el Consorcio Madroño: <http://www.consorciomadrono.es/investigam/pgdonline/>
- DMPOnline: herramienta Comisión Europea. Contiene las plantillas de diversas agencias financiadoras de investigación para generar PGD.
- Ejemplos de PGD:
 - Depositados en Zenodo: https://zenodo.org/search?page=1&size=20&q=data%20management%20plan&access_restrictions=open
 - Depositados en ARGOS: <https://argos.openaire.eu/explore-plans>

4. Material de apoyo – Plantillas para PGD

<https://pgd.consorciomadrono.es>



The screenshot shows the PGDOnline website interface. On the left is a navigation menu with the following items:

- Inicio
- Mis PGD
- Mis descripciones del dataset
- PGD públicos
- Descripciones del Dataset públicas
- Plantilla general instituciones españolas (PDF)
- Plantilla general instituciones españolas (DOCX)

The main content area features a central white box with the heading "¿Qué es un PGD en PGDOnline?". Below the heading is a paragraph: "Un plan de gestión de datos (PGD) es un documento vivo que describe los datasets generados y/o reutilizados durante y tras la finalización de la investigación. Los PGD se crean para proveer a los investigadores de la información esencial para reproducir, redistribuir y reutilizar con nuevos propósitos los resultados de la investigación, de forma que se asegure su validez y explotación". Below this is another paragraph: "¿Nuevo con los PGD? Visite [Guía de OpenAIRE para investigadores](#) para saber más sobre como crear uno!". At the bottom of this box is a red button labeled "Iniciar su primer PGD" and a cartoon character sitting on a laptop. To the right of the main content area is a statistics section titled "Uso personal" with four rows, each showing a large "0" and a label: "PGD", "Descripciones de los datasets", "Subvenciones", and "Organizaciones relacionadas".

4. Material de apoyo – Plantillas para PGD



Un plan de gestión de datos (PGD) es un d...
generados y/o reutilizados durante y tras l...
se crean para proveer a los investigadores...
reproducir, redistribuir y reutilizar con nuev...
investigación, de forma que se asegure su...

¿Nuevo con los PGD? Visite [Guía de Oper...](#)
más sobre como crear uno!

Iniciar su primer PGD

PGDonline X

Iniciar un nuevo PGD

Un Asistente especial está disponible para
ayudarle a crear su PGD de forma sencilla. Tiene
una guía paso a paso con toda la información
necesaria para crear un PGD fácilmente.

Importar desde fichero

Lanzar el asistente

o crear uno nuevo en PGDonline.

¡Lo tengo!

Abandonar la visita virtual



4. Material de apoyo – Plantillas para PGD

Guía paso a paso

Un PGD en PGDOnline consiste en la información clave sobre una investigación. Para este propósito se incluyen además de objetivos e investigaciones la documentación sobre los datasets de la investigación. En concreto la descripción de los dataset que resalta los pasos seguidos y los medios usados en las actividades administrativas.

- 1. Información principal (1)
- 2. Información sobre financiación (2)
- 3. Información de la licencia
- 4. Información sobre el dataset (✓)

1.1 Título*

Título

1.2 Descripción

Describe brevemente el contexto y el propósito del PGD



Rellenar con la descripción

1.3 Investigadores

Añadir aquí los nombres de las personas que han producido, procesado, analizado los datos descritos en el PGD.

 Escriba más letras del nombre para que sea más probable encontrar el correcto

Seleccionar los investigadores

¿No se encontró el correcto? [Insértelo de forma manual](#)

1.4 Organizaciones

Añadir aquí los nombre de las organizaciones participantes en la creación y revisión de los PGD

 Escriba más letras del nombre para que sea más probable encontrar el correcto

Seleccionar la organización

¿No se encontró el correcto? [Insértelo de forma manual](#)

< Anterior

Siguiente >

4. Material de apoyo – Plantillas para PGD

Guía paso a paso

1. Información principal (1)
2. Información sobre financiación (2)
3. Información de la licencia
- 4. Información sobre el dataset (✓)**

Los datasets se documentan siguiendo plantillas predefinidas con el contenido de la descripción de los datasets. En PGDOnline un PGD puede incluir tantas descripciones de datasets como datasets se documenten.

4.1 Información sobre el dataset*

Seleccione una plantilla para describir sus datasets. Puede seleccionar una o más plantillas.

i Escriba más letras del nombre para que sea más probable encontrar el correcto

Seleccione una plantilla para describir su dataset *

Horizonte Europa. Versión española ✕

Si no encuentra una plantilla o si quiere crear una plantilla personalizada para su institución, comunidad de investigadores o pruebas, por favor, contacte con nosotros.

< Anterior

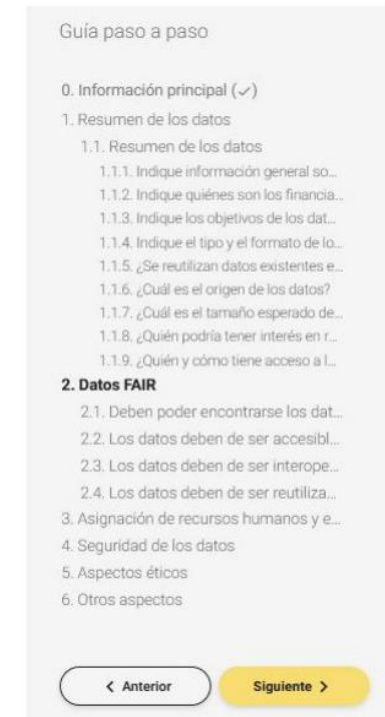
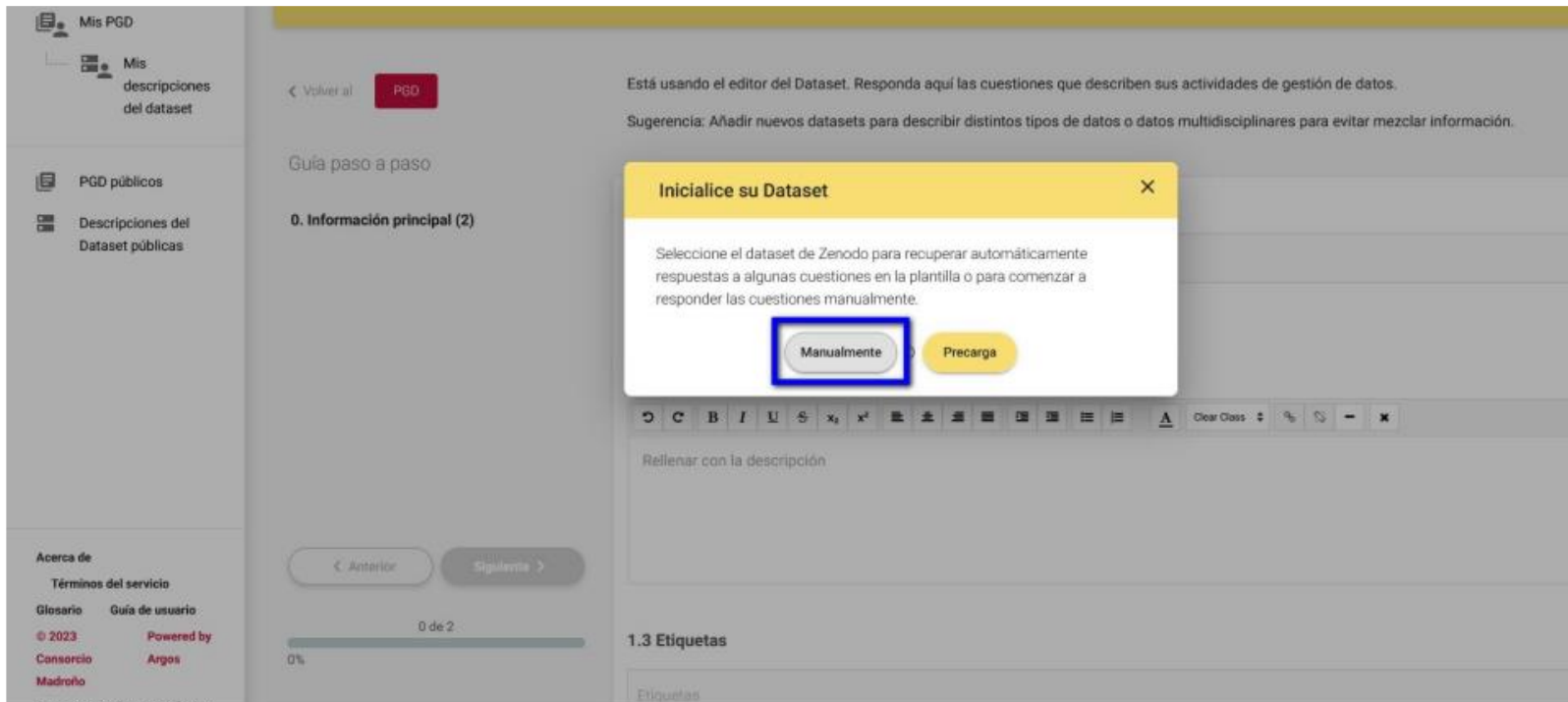
Grabar & Añadir dataset

3 de 5

60%

4. Material de apoyo – Plantillas para PGD

Rellenar los siguientes apartados:



Fuente: <https://e-archivo.uc3m.es/entities/publication/adf04df6-266f-4c0f-bf4b-11d91fd3eae6>

Muchas gracias por vuestra atención!

Alonso.Olatz@gmail.com
lcamacho@isciii.es

Referencias

- Krishnankutty B, Bellary S, Kumar NBR, Moodahadu LS. Data management in clinical research: an overview. *Indian J Pharmacol.* 2012;44(2):168–72. doi:10.4103/0253-7613.93842
- Lu Z, Su J. Clinical data management: Current status, challenges, and future directions from industry perspectives. *Open Access J Clin Trials.* 2010;93–105. doi:10.2147/OAJCT.S8172.
- Bellary S, Krishnankutty B, Latha MS. Basics of case report form designing in clinical research. *Perspect Clin Res.* 2014;5(4):159-166. doi:10.4103/2229-3485.140555.
- Prokscha S. *Practical Guide to Clinical Data Management.* 4th ed. Boca Raton: CRC Press; 2024. doi:10.1201/9781003395621.
- Miller G, Spiegel E. Guidelines for Research Data Integrity (GRDI). *Sci Data.* 2025;12(1):95. doi:10.1038/s41597-024-04312-x.
- Alston JM, Rick JA. A beginner’s guide to conducting reproducible research. *Bull Ecol Soc Am.* 2020;102(2):e01801. doi:10.1002/bes2.1801
- Roche C, McDonald S, et al. An overview of the fundamentals of data management, analysis, and interpretation in quantitative research. *Semin Oncol Nurs.* 2023;39:151398.
- Wilkinson M, Dumontier M, Aalbersberg I, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data.* 2016;3:160018. doi:10.1038/sdata.2016.18
- Michener WK. Ten Simple Rules for Creating a Good Data Management Plan. *PLoS Comput Biol.* 2015;11(10):e1004525.

Referencias

- Ministerio de Ciencia, Innovación y Universidades. Estrategia Nacional de Ciencia Abierta (ENCA) 2023–2027. Madrid: Gobierno de España; 2023. Disponible en: <https://ciencia.gob.es/enca>
- Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación. Boletín Oficial del Estado. 2011 (131): 50766–50797. Disponible en: <https://www.boe.es/eli/es/l/2011/06/01/14/con>
- Gobierno de España, Ministerio para la Transformación Digital y de la Función Pública. Guía práctica para la mejora de la calidad de los datos abiertos. Madrid: Gobierno de España; 2020. Disponible en: <https://datos.gob.es/es/conocimiento/guia-practica-para-la-mejora-de-la-calidad-de-datos-abiertos>
- Grupo de Trabajo de “Depósito y Gestión de datos en Acceso Abierto” del proyecto RECOLECTA. La conservación y reutilización de los datos científicos en España. Informe del grupo de trabajo de buenas prácticas. Madrid: Fundación Española para la Ciencia y la Tecnología (FECYT); 2012. Disponible en: <https://www.fecyt.es/>
- OECD. OECD Principles and Guidelines for Access to Research Data from Public Funding. Paris: OECD; 2007. Disponible en: <https://www.oecd.org/sti/sci-tech/38500813.pdf>
- European Commission. EU Grants: AGA — Annotated Grant Agreement: V2.0. Brussels: European Commission; 2025. Disponible en: <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/home>

Referencias

- Vlad A. Data Management Basics 1: Introduction to data management and sharing. Essex: UK Data Service, University of Essex; 2021. Disponible en: https://ukdataservice.ac.uk/app/uploads/2021_04_29_datamanagementbasics1webinarv2.pdf
- UK Data Service. Research data management. Learning Hub – UK Data Service. UK Data Service. Disponible en: <https://ukdataservice.ac.uk/learning-hub/research-data-management/> [UK Data Service](#)
- UK Data Service. Publishing and citing research data. UK Data Service. Disponible en: <https://ukdataservice.ac.uk/publishingcitingdata/>
- Tennant JP, Crane H, Crick T, Dávila Quintero JA, Enkhbayar A, Havemann J, et al. Ten myths around open scholarly publishing. PeerJ Preprints; 2019. Disponible en: <https://peerj.com/preprints/27580.pdf>
- O Bishop L. Anonymisation [Internet]. Colchester: UK Data Archive, University of Essex; 2013. Disponible en: <https://dam.ukdataservice.ac.uk/media/187709/anonymisationdm25april2013.pdf>
- Zhao A. Data Pre-processing: Clean, Reduce and Transform [Internet]. Colchester: UK Data Service; 2021. Disponible en: <https://ukdataservice.ac.uk/app/uploads/2021-01-21datapreprocessing2.pdf>
- Alonso-del-Cura O. Medicina Personalizada (MedPer) en la detección precoz del Deterioro Cognitivo (DC) preclínico (DENDRITE). Desarrollo de un modelo predictivo de riesgo V2.0 [PGD] [Internet]. Madrid: Instituto de Salud Carlos III (ISCIII), Unidad de Investigación en Cuidados de Salud (Investén); 2025. Disponible en: <https://hdl.handle.net/20.500.12105/26876>

Referencias

- Lenoir Improta R, Pérez Aliende ML. La Ciencia Abierta: curso impartido a los bibliotecarios de la Universidad Autónoma de Madrid [Internet]. Madrid: Universidad Autónoma de Madrid; 2023. Disponible en: <http://hdl.handle.net/10486/707884>
- Malo de Molina Martín-Montalvo T, Fernández-del-Pino Torres B. Taller virtual: ¿Cómo preparar un Plan de Gestión de Datos? Universidad Carlos III de Madrid, Biblioteca Universitaria; 2021. Disponible en: <https://hdl.handle.net/10016/43825>
- Marín del Campo R, Ortiz Uceta E. ¿Cómo preparar un Plan de Gestión de Datos? Universidad Carlos III de Madrid; 2024. Disponible en: <https://hdl.handle.net/10016/43893>
- European Commission. Horizon Europe Programme Guide 2021–2027 [Internet]. Brussels: European Commission; 2021 [updated 2025 Sep 15]. Disponible en: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf
- Heyard R, Currie G. improving Reproducibility In Science (iRISE). A Data Management Plan created using DMPonline. Disponible en: https://dmponline.dcc.ac.uk/plans/140490/export.pdf?export%5Bquestion_headings%5D=true
- Grupo de Trabajo de Apoyo a la Investigación del CSUC. Planes de Gestión de Datos (Versión 1, Enero 2020); 2020. Disponible en: <http://hdl.handle.net/2072/370510>
- Jones S, Grootveld M. ¿Cómo comprobar si sus datos son FAIR? [Internet]. Traducción al español por Gómez Sánchez AF; 2021. Disponible en: <https://doi.org/10.5281/zenodo.5508099>
- Autoridad Nacional de Protección de Datos de Singapur. Guía básica de anonimización. Traducción al español por Agencia Española de Protección de Datos (AEPD). Madrid: AEPD; 2022. Disponible en: <https://www.aepd.es/documento/guia-basica-anonimizacion.pdf>

Referencias

- <https://www.upo.es/biblioteca/servicios/investigabib/Gestion-de-datos-de-investigacion/>
- <https://biblioguias.ucm.es/datos-investigacion/pgd>
- <https://guias-tematicas.unavarra.es/sp/subjects/guide.php?subject=Datos#tab-1>
- https://bib.us.es/estudia_e_investiga/investigacion/estrategias/gdi
- <https://authorservices.taylorandfrancis.com/choose-open/open-data/>
- <https://edatos.consorciomadrono.es/readme.xhtml>