



# Open Science y gestión de datos de investigación

**Leticia Barrionuevo**

Biblioteca Universidad de León

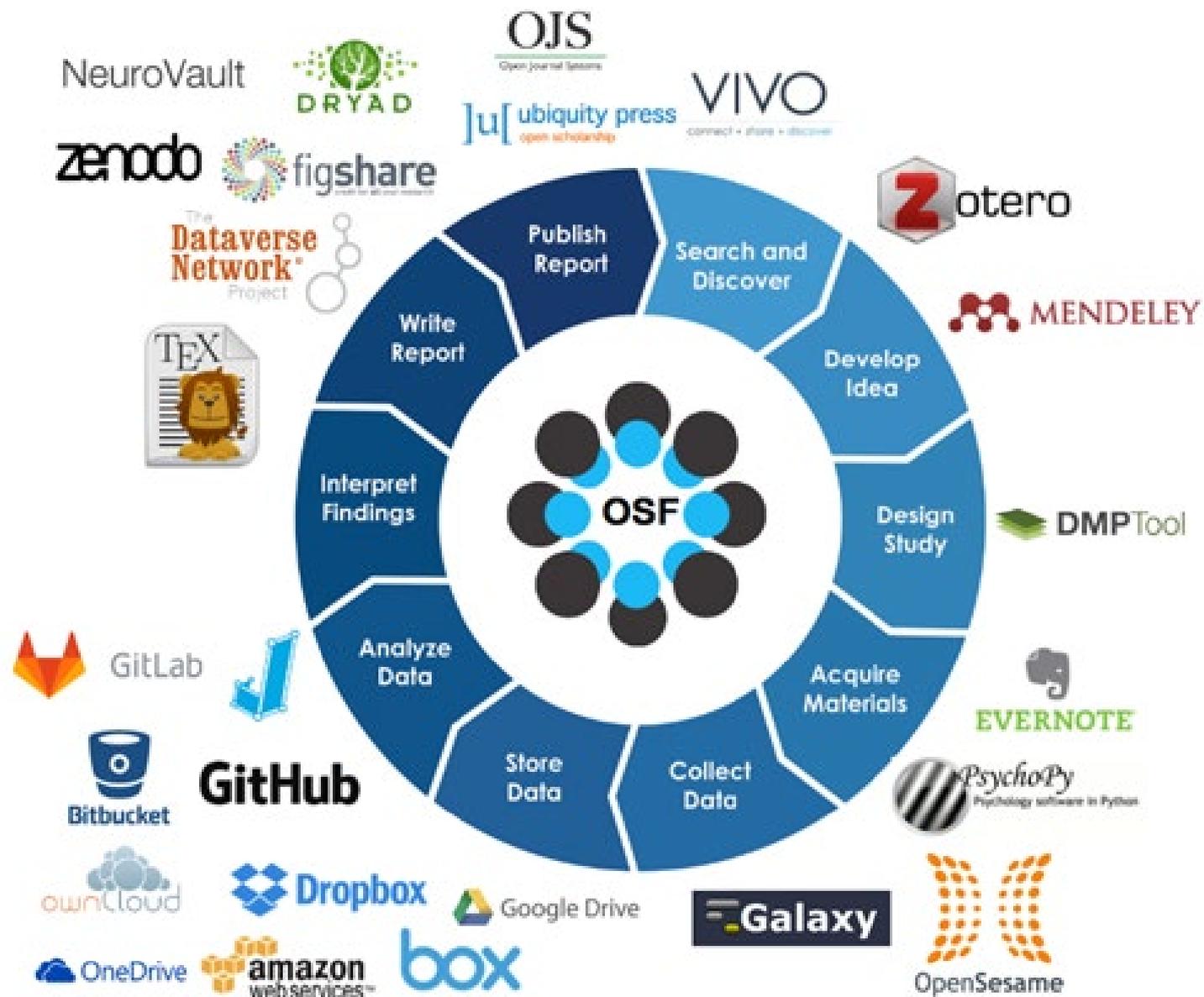
[buffl@unileon.es](mailto:buffl@unileon.es)

A pair of black-rimmed glasses with clear lenses is positioned diagonally across the top half of the image. Below the glasses, a piece of white paper is visible, featuring the word "RESEARCH" printed in a bold, blue, sans-serif font. The background is a light, neutral color, and the overall lighting is soft and even.

**RESEARCH**

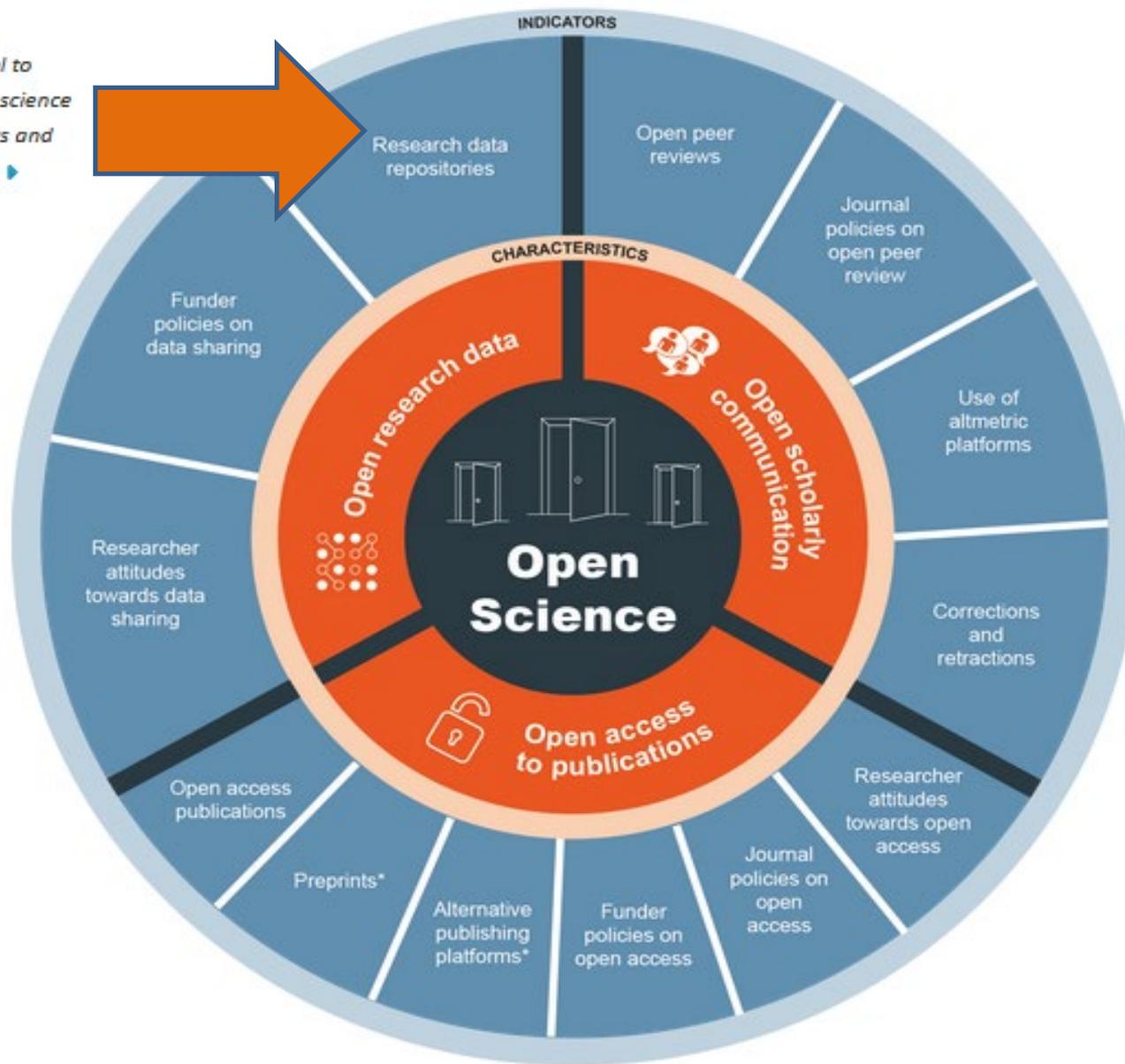
La ciencia abierta es un cambio de paradigma en la manera de hacer ciencia. Ésta no cambia sustancialmente con respecto a sus motivaciones y objetivos, pero sí lo hace (sustancialmente) en cuanto a sus métodos. El cambio **no está en lo que se hace, sino en cómo se hace.**

Lluís Anglada, Ernest Abadal, 2018



Fases de la investigación. “The Open Science Framework”. Center for Open Science

Use the wheel to explore open science characteristics and indicators. ▶▶



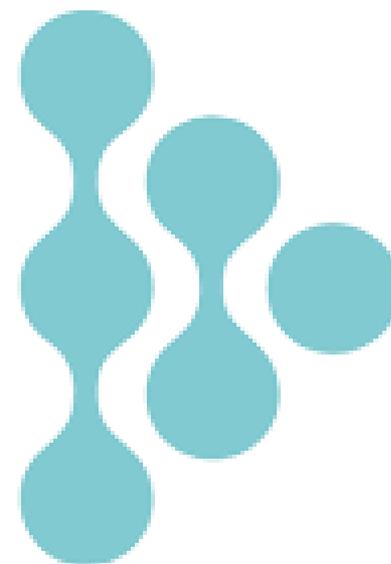
Los ejes de la ciencia abierta (RAND Europe, 2017)



Las “celdas del panel” de la Ciencia Abierta (FOSTER)

**FECYT**

**I N N O V A C I Ó N**







# ¿Qué son los datos de investigación?

"todo aquel material que sirve para certificar los resultados de la investigación que se realiza, que se han registrado durante ella y que se ha reconocido por la comunidad científica".

**cualquier información** reunida, utilizada o generada en la experimentación, observación, medición, simulación, cálculo, análisis, interpretación, estudio o cualquier otro proceso de indagación **que sustente y justifique las aportaciones científicas** que se difunden en las publicaciones de investigación. Se presentan **en cualquier formato y soporte.**



# ¿Qué son los datos de investigación?

“datos que son recolectados, observados o creados para ser analizados y producir resultados de investigación originales”

Puede haber grandes diferencias entre lo que constituye un dato para investigadores de la humanidades, de las artes o de las ciencias. En cualquiera de estos casos, es importante tener en cuenta que los datos pueden ser de tipo **cuantitativo o cualitativo**, y que pueden venir en muchos **formatos y soportes**, ya sean **físicos o digitales**.



# TIPOLOGY

## Según su **nivel de procesamiento** en el curso de la investigación

- **Datos primarios o sin procesar:** Datos originales que han sido recolectados pero aún no han sido procesados o analizados. Algunos ejemplos son los registros sonoros, observaciones, notas de campo o datos de experimentos.
- **Datos procesados:** Datos que han sido digitalizados, traducidos, transcritos, limpiados, validados, verificados y/o anonimizados.
- **Datos analizados:** Modelos, gráficos, tablas, textos u otros, que han sido creados a partir de los datos primarios y procesados, y que se pretende sean de ayuda en el descubrimiento de información útil, la presentación de conclusiones y la toma de decisiones.



# TIPOLOGY

## Según la **fuer**te de la que provienen

- **Datos canónicos o de referencia:** Sets de datos que pueden ser utilizados para validación, comparación o búsqueda de información (por ejemplo, secuencias del genoma humano, estructuras químicas o portales de datos espaciales)
- **Datos experimentales:** Datos generados en experimentos científicos. Generalmente son reproducibles y pueden ser generados por equipos de laboratorio.
- **Modelos o simulaciones:** Datos generados en computadores por algoritmos, modelos matemáticos, o simulaciones de experimentos.
- **Datos derivados:** Sets de datos creados al tomar datos ya existentes y realizar algún tipo de manipulación sobre los mismos.
- **Observaciones:** Datos generados al registrar observaciones de un evento específico y posiblemente irreplicable, en un lugar y tiempo dados.



# TIPOLOGY

## Según su **forma** o **tipo**

- Documentos de texto electrónicos
- Planillas de datos
- Cuadernos de campo o anotaciones de laboratorio
- Cuestionarios o transcripciones
- Fotografías o películas
- Registros sonoros
- Muestras, artefactos, especímenes
- Objetos digitales
- Modelos, algoritmos, scripts
- Bases de datos
- Metadatos
- Esquemas de metadatos
- Configuraciones de software
- Archivos de pre o post procesamiento de software



# TIPOLOGY

## Según su **formato**

- Textuales (Microsoft Word, PDF, RTF, ODT, etc.)
- Numéricos (Excel, CSV, etc.)
- Multimedia (JPEG, MPEG, WAV, etc.)
- Estructurados (XML, base de datos MySQL, etc.)
- Código de software (Java, C, etc.)
- Específicos de un software (Mesh, 3D CAD, modelo estadístico, etc.)
- Específicos de una disciplina (FITS en astronomía, CIF en química, etc.)
- Específicos de un instrumento (Olympus Confocal Microscope Data Format, Carl Zeiss Digital Microscopic Image Format)



# El ciclo de los datos científicos



Ayuda a planificar la investigación



El personal de tu Biblioteca te puede asesorar

# Gestión de Datos de Investigación (GDI)

«es el proceso activo de manejo de los datos generados en una investigación»

se realiza de forma continua y cubre todas las decisiones relacionadas con la gestión de los datos a lo largo de su **ciclo vital**, comenzando en la etapa de planificación de la investigación, abarcando su ejecución, la disseminación de sus resultados, la preservación de los conjuntos de datos de forma que estos sean precisos, completos, auténticos y fiables, y se mantengan accesibles y reutilizables a lo largo del tiempo.

- ✓ Los mecanismos de recolección de datos y la planificación de su uso.
- ✓ La definición de formas de identificar archivos de datos.
- ✓ Asegurar la calidad de los datos.
- ✓ La organización de los conjuntos de datos.
- ✓ El Almacenamiento de datos.
- ✓ La integridad, seguridad y privacidad de los datos.
- ✓ El depósito de datos y su preservación a largo plazo.
- ✓ La definición de mecanismos utilizados para compartir los datos.



# Gestión de Datos de Investigación (GDI)

**Ahorrar tiempo y hacer un uso eficiente de los recursos disponibles:**

Podrás encontrar y utilizar los datos en el momento que necesites.

**Facilitar la reutilización de los datos** que se han generado o recopilado.

**Cumplir con los mandatos de las agencias de financiación:** cada vez más, las instituciones financiadoras exigen la presentación de planes de gestión de datos y/o el depósito de datos en repositorios como requisitos para conceder subvenciones.

**Preservar los datos:** Al gestionar y depositar los datos en repositorios apropiados podrás resguardarlos de forma segura a lo largo del tiempo, protegiendo tu inversión de tiempo y recursos y permitiendo que puedan servir a nuevas investigaciones y descubrimientos en el futuro.



# Los investigadores

## En relación a **sus datos de investigación**:

- Tomar decisiones en cuanto a la apertura, depósito, selección y reutilización de los datos.
- Cumplir con los requerimientos éticos y legales asociados. **Plan de Gestión de Datos**

## En relación a los **datos creados por terceros**:

- Reutilizarlos ciñéndose a las condiciones propuestas por los creadores de los datos.
- Citar los datos cuando se utilicen. **Cómo se citan los Datos de Investigación**



# 10 pasos para elaborar un Plan de Gestión de Datos

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.consortiomadroño.es/>



DMPonline (Digital Curation  
Centre, UK)  
<https://dmponline.dcc.ac.uk/>



Revisa los **requerimientos** de la entidad financiadora (H2020).



**Identifica los datos:** tipología, procedencia, volumen, formatos y ficheros.



**Define cómo se organizarán y gestionarán los datos:** nombre de los ficheros, control de versiones, software necesario...



**Explica cómo se documentarán los datos:** identifica la información a procesar, consulta si hay estándares o esquemas de metadatos, identifica herramientas que permitan gestionarlos.



Describe los procesos que aseguran una **buena calidad de los datos**.



**Prepara una estrategia de almacenamiento** (durante el proceso) y de preservación de datos (repositorio).



**Define las políticas de datos del proyecto:** cuestiones sobre propiedad intelectual y cómo se tratarán los datos sensibles y personales.



**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".



**Asigna roles y responsabilidades** para las personas y organizaciones participantes en el proyecto.



**Prepara un presupuesto realista:** la gestión de datos cuesta tiempo y dinero en términos de software, hardware, servicios y personal.



## **IMPORTANTE**

**Los Planes de Gestión de Datos no son estáticos.**

**Deben mantenerse y actualizarse a lo largo  
del curso de una investigación**



**PGDonline**



**Research Data  
MANAGEMENT PLAN**

**PladeGestió de Dades de Recerca**

**EJEMPLO**



# Cita tus datos de investigación



## Por qué es importante citar los datos:

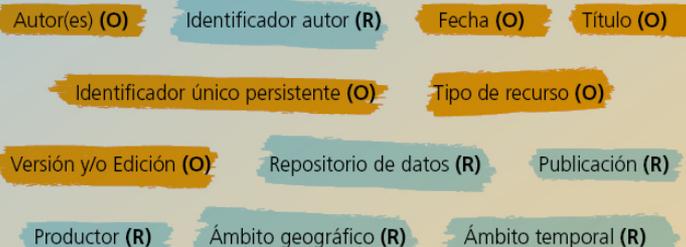
- Los conjuntos de datos también son resultados de investigación como los artículos, monografías, etc.
- Facilita la identificación y el acceso a los datos y de esta forma su localización, validación y reutilización.
- Permite reconocer la autoría de sus creadores.
- Facilita la métrica e impacto de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

## 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

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

Fecha    Repositorio    Tipo de recurso    Identificador único y persistente    Versión



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



El personal de tu Biblioteca te puede asesorar

# Ejemplos

- Irino, T.; Tada, R. (2009). Chemical and mineral compositions of sediments from ODP Site 127-797 [Dataset]. Geological Institute, University of Tokyo. <http://dx.doi.org/10.1594/PANGAEA.726855>.
- Geofon operator (2009). GEFON event gfz2009kciu (NW Balkan Region) [Dataset]. GeoForschungsZentrum Potsdam (GFZ). <http://dx.doi.org/10.1594/GFZ.GEOFON.gfz2009kciu>.
- Denhard, M. (2009). dphase\_mpeps: MicroPEPS LAF-Ensemble run by DWD for the MAP D-PHASE project [Dataset]. World Data Center for Climate. [http://dx.doi.org/10.1594/WDCC/dphase\\_mpeps](http://dx.doi.org/10.1594/WDCC/dphase_mpeps).



Dylanne Dearborn, Steve Marks, Leanne Trimble. [The Changing Influence of Journal Data Sharing Policies on Local RDM Practices](#). International Journal of Data Curation (IJDC) Vol. 12, No. 2 (2017)

Se analizaron las políticas de las diez revistas más importantes (clasificadas por factor de impacto según los Informes de citas de revistas) en 2014 y de nuevo en 2016 para determinar si se requerían o recomendaban depósitos de datos y qué métodos de depósito se enumeraban como opciones. Los resultados mostraron que el número de revistas de alto impacto en las ciencias físicas y aplicadas que requieren depósito de datos está creciendo.

# Administradores de datos

- Trabajar junto los investigadores en la recolección y análisis de datos.
- Vigilancia tecnológica.

# Servicios de Informática

- Ofrecer servicios y apoyo en almacenamiento, uso de software y soporte técnico.
- Soporte en el uso de almacenamiento y estructuración de datos

# Bibliotecas

- Apoyo en actividades como: depósito, publicación, preservación, acceso, localización y recuperación de los datos. **Repositorios de datos**
- Asesoramiento en materia de licencias, propiedad intelectual, citas de datos, etc.
- Apoyo a los investigadores en la creación de PGD.

# Portales

data.europa.eu

datos.gob.es  
reutiliza la información pública

The screenshot shows the homepage of the 'Datos Abiertos de Castilla y León' portal. At the top left is the logo of the Junta de Castilla y León. A search bar with the placeholder text 'Texto de búsqueda' is located at the top center. To the right of the search bar are icons for social media and a 'Otras herramientas' link. Below the search bar is a red navigation menu with the following items: 'DATOS ABIERTOS', 'INICIATIVA DE DATOS ABIERTOS', 'CATALOGO DE DATOS', 'ANÁLISIS DE DATOS ABIERTOS', 'APLICACIONES', 'CONCURSO DE DATOS ABIERTOS', and 'PARTICIPA'. The main content area features a large banner with the text 'DATOS ABIERTOS de Castilla y León' and an illustration of an open book with arrows pointing upwards. Below the banner is a section titled 'DATOS ABIERTOS DE CASTILLA Y LEÓN' containing four tiles: 'III Concurso Datos Abiertos', 'Novedades y suscripción', 'Preguntas frecuentes', and 'Análisis de datos'. At the bottom of the page is a section titled 'BÚSQUEDA DE DATOS'.

# Repositorios

re3data.org  
REGISTRY OF RESEARCH DATA REPOSITORIES



# Google Dataset Search Beta

Probar [boston education data](#) or [weather site:noaa.gov](#)

[Consulta más información sobre la inclusión de tus conjuntos de datos en Búsqueda de Datasets.](#)

Proporciona acceso a millones de conjuntos de datos en miles de repositorios de datos en la web de todo el mundo.

La búsqueda de set de datos muestra los resultados de los sitios de los editores, las bibliotecas digitales y las páginas web personales de los autores. Además, el nuevo motor de búsqueda se puede utilizar para buscar referencias a la mayoría de los conjuntos de datos en un principio esta compilando grupos de datos de ciencias naturales, ciencias sociales, datos gubernamentales y datos proporcionados por agencias de noticias, pero se espera que en breve estos conjuntos de datos se extiendan al resto de las disciplinas.



## PLOS Criteria for Recommended Data Repositories



# Recomendaciones de PLOS para repositorios de datos

# Equipo directivo

- Desarrollar políticas de GDI y promoverlas a nivel institucional.
- Crear una infraestructura de apoyo, teniendo en cuenta a todos los actores o agentes involucrado y formarles para su capacitación.
- Controlar el cumplimiento de los requisitos estipulados en la GDI.

**¿Está preparada mi institución para gestionar datos de investigación?**

[Proyecto LEARN. Encuesta de Autoevaluación](#)

INICIO

ACERCA DE ▾

PUBLICACIONES

EVENTOS ▾



**DATASEA EXTENDED**  
**GUÍA PARA UN MUNDO DE DATOS**

**Research Data**

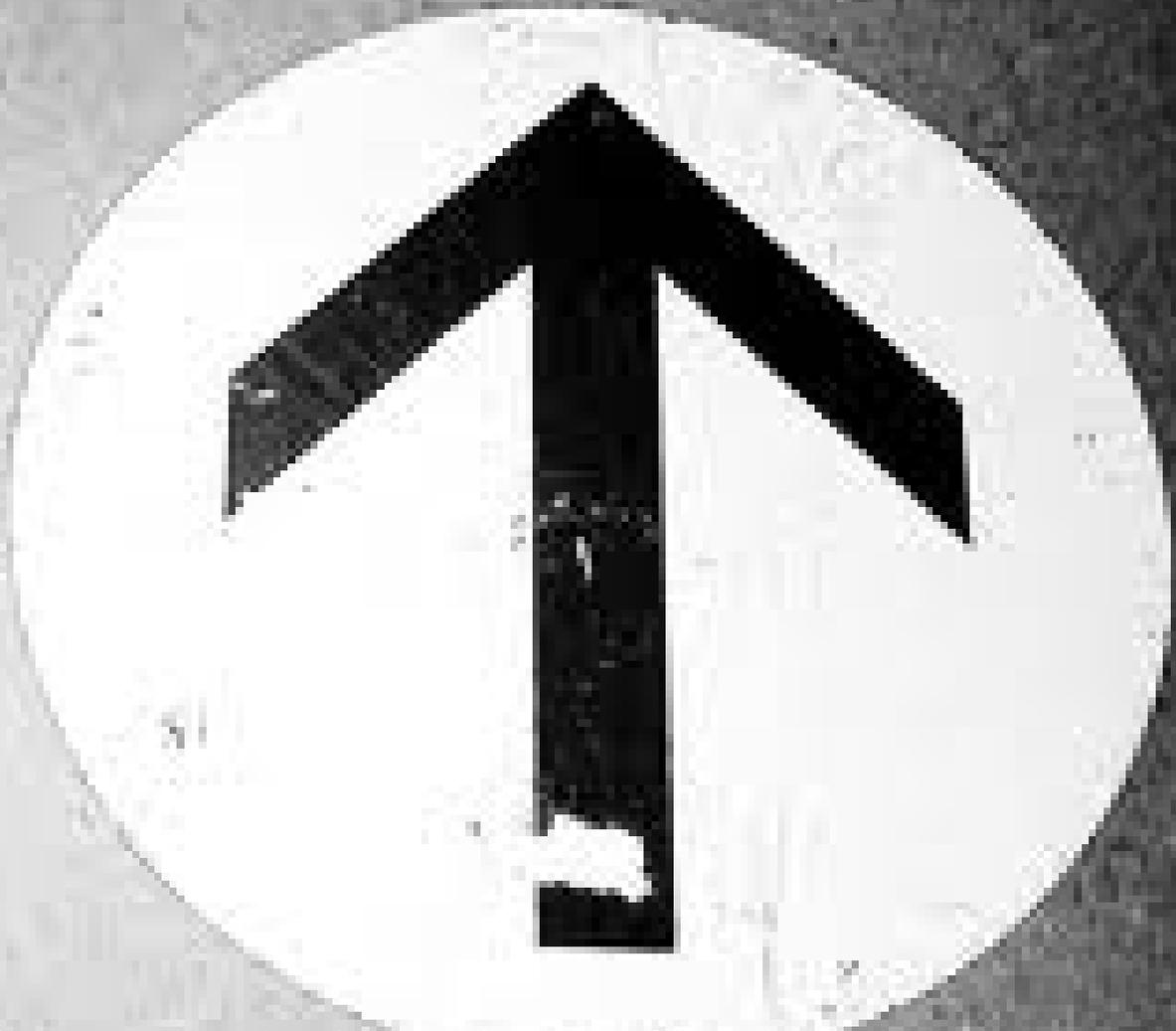
**Open Data**

Leticia Barrionuevo

[buffl@unileon.es](mailto:buffl@unileon.es)

Ext. 1004

follow  
me



Las fotos utilizadas han sido extraídas de Google Imágenes