

Gestión de Datos de Investigación, más que una buena práctica

Malgorzata Lisowska
Directora CRAI UR



SEMANA DEL ACCESO ABIERTO

- Universidad del Valle
- Universidad de Antioquia
- Universidad del Rosario

Contenido

- Panorama internacional
- Datos abiertos
- Plan de gestión de datos
- Experiencia UR
 - Estrategia
 - Servicios



“data is the
new oil for the
digital age”

Neelie Kroes, Vice-President of the EC
Responsible for the Digital Agenda for Europe



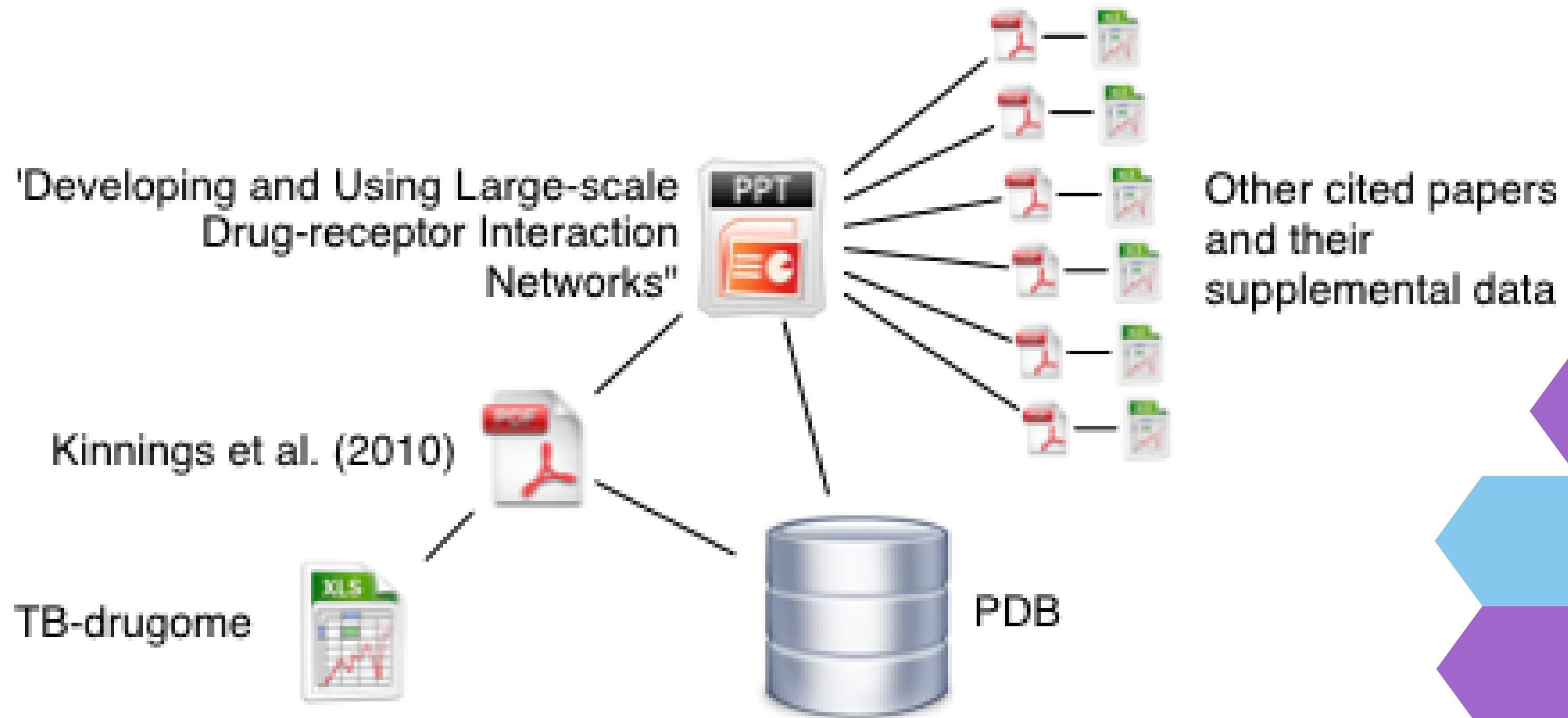


¿Qué son los datos de investigación?

Los datos de investigación son registros concretos, que pueden tomar forma de números, símbolos, texto, imágenes o sonidos, utilizados como fuentes primarias de investigación, que son comúnmente aceptados por la comunidad científica como necesarios para ***la validación de hallazgos de investigación***.

(Adaptada de OCDE,2017).

El nuevo objeto científico es complejo



- Datos perdidos
- Datos dispersos en diferentes medios
- Datos que no se preservan en el tiempo
- Datos incompletos y no fiables



Validación de resultados

“Cada vez que se publica una investigación en una revista de impacto, el material sirve de base para futuras investigaciones. Miles de científicos lo utilizan y lo citan. Las revisiones lo incluyen en sus bases de datos”

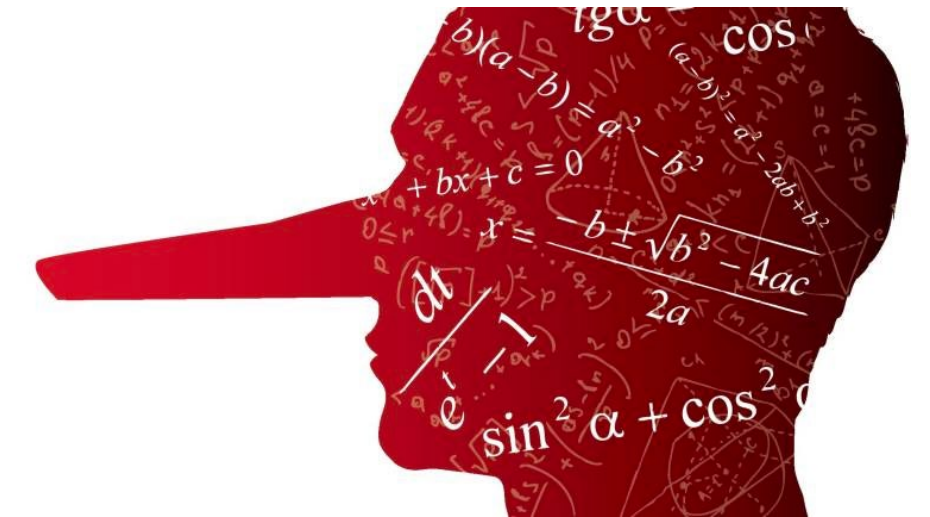
Fuente: <https://www.larazon.es/sociedad/plagio-y-fraude-tambien-entre-cientificos-GE19843858>

11% de casos de fabricación de datos; es decir, de falseamiento de las estadísticas o los resultados matemáticos para tratar de ajustarlos a la tesis.

17 % es autoplagio –repetir ideas o resultados de otras investigaciones propias sin advertirlo.

16 % es plagio de otros autores –por cierto, el primer caso se detectó en 1970–. Estas tres categorías son fraudes científicos y crecen como la espuma.

También aumentan los fallos no atribuibles a la mala fe, pero que desacreditan el trabajo. Aquí se incluyen errores de cálculo, ausencia de herramientas que permitan reproducir los datos o falta de explicación de los resultados.



El tiempo que pasa desde la publicación y la detección de un error o un plagio se está reduciendo (En 2002 se necesitaban 46 meses. Ahora menos de la mitad)

Caso Andrew Wakefield - El autismo y las vacunas

En 1998 este excirujano publicó un trabajo que relacionaba la aparición de autismo con la administración de la vacuna triple vírica.



Está más que demostrado que se trataba de una publicación con datos falseados, pero aún hoy este es uno de los principales argumentos esgrimidos por el movimiento antivacunas, que cuenta cada vez con más adeptos y supone un peligro para la sanidad pública

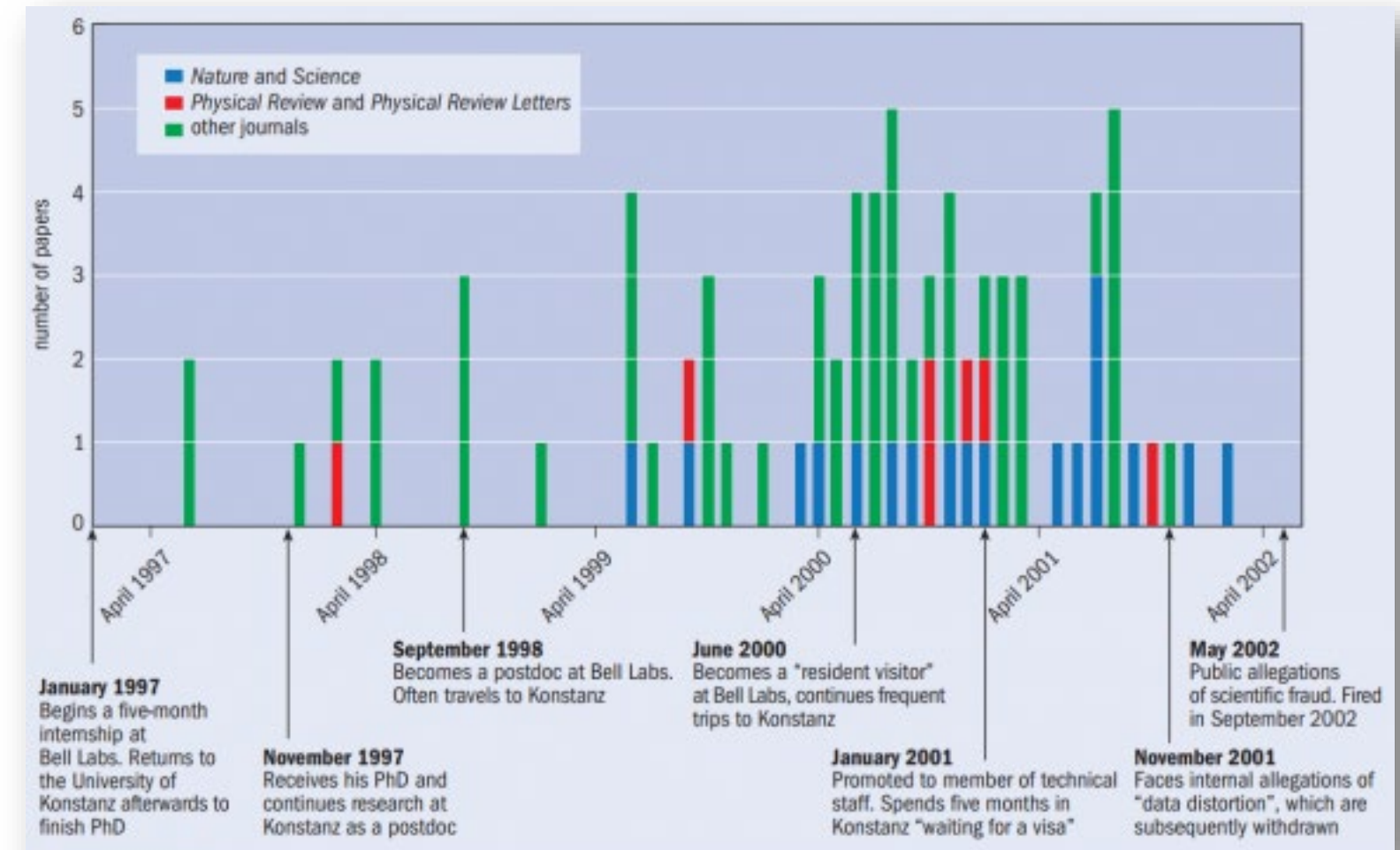
Caso Jan Hendrik Schön – fabricación de los resultados



Jan Hendrik Schön es experto en nanotecnología. En una auditoría de Bell Labs se descubrió que este prometedor científico se había inventado gran parte de los resultados de sus estudios.

Un día empezó a publicar resultados contradictorios entre sí, algo imposible, lo que hizo saltar todas las alarmas.

En el año 2002, era el científico joven más productivo del mundo, firme candidato al Nobel de Física. A sus 31 años de edad publicaba en las revistas más prestigiosas a un ritmo desaforado, hasta un artículo a la semana (por ejemplo, ocho artículos en *Nature* y *Science* en 2001). Sus logros en ciencia de los materiales y nanotecnología prometían revolucionar el mundo.



Caso Hwang Woo-suk - clonación de embriones humanos



Condenado

En el 2004 este científico surcoreano publicó un artículo en la revista Science, asegurando que había logrado clonar un embrión humano. En otro estudio posterior, afirmaba haber logrado extraer células madre del mismo, un hallazgo histórico (posible cura el párkinson o la diabetes). Poco tiempo después se demostró que este hallazgo fue un fraude y Hwang fue condenado a dos años de prisión por fraude y malversación de fondos dedicados a la investigación.

El mérito de Hwang que sí que parece ser verídico es la primera clonación de un perro, en el año 2005.



Disgraced Korean scientist **Hwang Woo-suk** loses legal battle over ...

The Korea Herald - 15 ago. 2017

Hwang Woo-suk, a South Korean stem cell scientist who caused a major stir in the scientific community for academic **fraud** in 2005, has ...



Another stem cell **fraud** in Japan

BioEdge - 27 ene. 2018

"[The **fraud**] is something that shakes the people's trust in research ... and falsification, from Korean researcher **Hwang Woo-suk** in 2004 to ...

Yoshitaka Fujii, el impostor más prolífico de la historia

Este médico investigador en el campo de la anestesiología ha falseado al menos 183 trabajos científicos y aun se sigue 'limpiando' su fraude.

Solo en el 2018 ha habido 21 retractaciones de sus artículos , 17 de ellos de la revista Clinical Therapeutics



Era muy difícil rastrear el fraude, pero los resultados “demasiado buenos” de sus investigaciones hicieron sospechar a otros científicos de la veracidad de sus datos.

Era muy ambiguo respecto a las fechas de los estudios y los nombres de las instituciones donde estos se realizaron; incluía como coautores a científicos de otras entidades (que a veces ni sabían de esto), dando la impresión de que los datos se recogían de distintos hospitales.

Caso Marc Hauser – coaccionó a sus alumnos para falsificar los datos para que se ajustaran a su propia tesis



Ha escrito más de doscientos trabajos de investigación y cinco libros, entre ellos: *The evolution of communication* (1996); *Mentes salvajes: ¿Qué piensan los animales?* (2000) y *Moral minds: The unconscious voice of right and wrong*.

Durante los últimos años estaba trabajando en un libro sobre el lenguaje junto a Noam Chomsky.

Licenciado en ciencias por la Universidad Bucknell y doctorado por la Universidad de California, fue profesor en la Universidad Harvard en los departamentos de Psicología, Organísmica y Biología



Caso Yoshimiro Sato – falsificación de los resultados clínicos en efecto del calcio en las fracturas de cadera

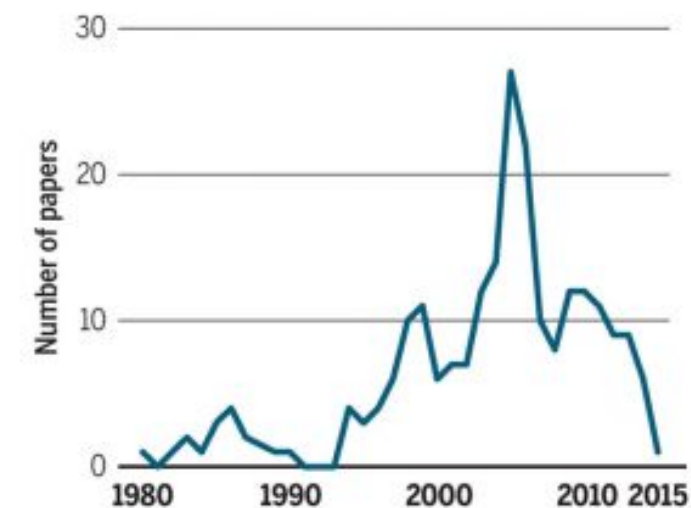
Este osteólogo japonés publicó más de 200 estudios sobre cómo reducir el riesgo en fracturas de hueso. Estudios que después se utilizaron como base para hacer meta-análisis y cuyas conclusiones tienen consecuencias en la práctica clínica.

Se suicidó en enero 2017 tras descubrirse que 33 de sus artículos (prestigiosas referencias para el estudio de la salud de los huesos) eran fraudulentos, de los cuales, hasta la fecha, solo han sido retractados 21.

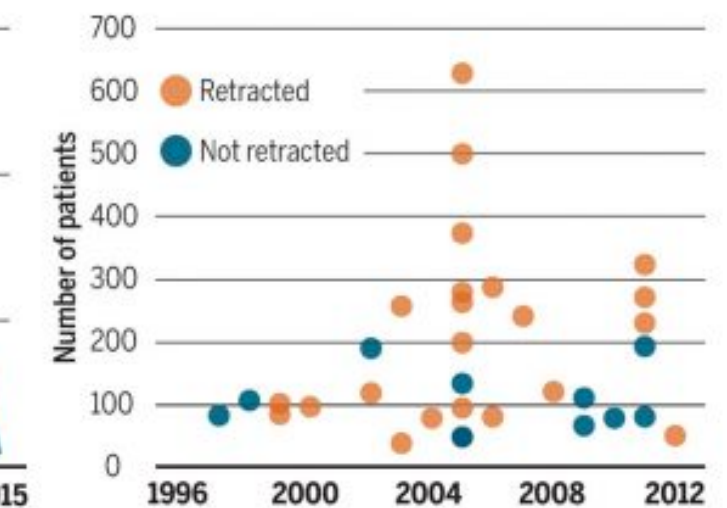
A far-reaching fraud

A team of four researchers has worked since 2012 to expose scientific misconduct by Japanese bone researcher Yoshihiro Sato, who published more than 200 papers before he died in 2016. The team has focused on Sato's 33 clinical trials, together involving 5894 patients.

Total scientific output



Clinical trials



Sato ha fallecido, pero surgen preguntas como:

¿Debe evitarse retractar sus artículos al no poderse recabar su opinión?

¿Se deben buscar responsabilidades entre sus coautores?

Datos abiertos



Apertura de los datos

- **Abierto**: contenido visible por todo usuario de internet. Su uso puede estar sujeto a distintas limitaciones según la licencia Creative Commons con la que se registre.
- **Cerrado**: contenido visible sólo para el personal encargado de la gestión.
- **Restringido**: contenido en principio cerrado, pero que puede ser puesto a disposición del público bajo ciertas condiciones y siempre bajo petición.
- **Embargado**: contenido cerrado que pasa automáticamente a ser abierto una vez transcurrido un período de tiempo determinado.

Tan abierto como sea posible y tan cerrado como sea necesario

Los datos abiertos hacen parte de un contexto más amplio, que involucra otros elementos y que en conjunto conforman lo que conocemos como Ciencia Abierta.

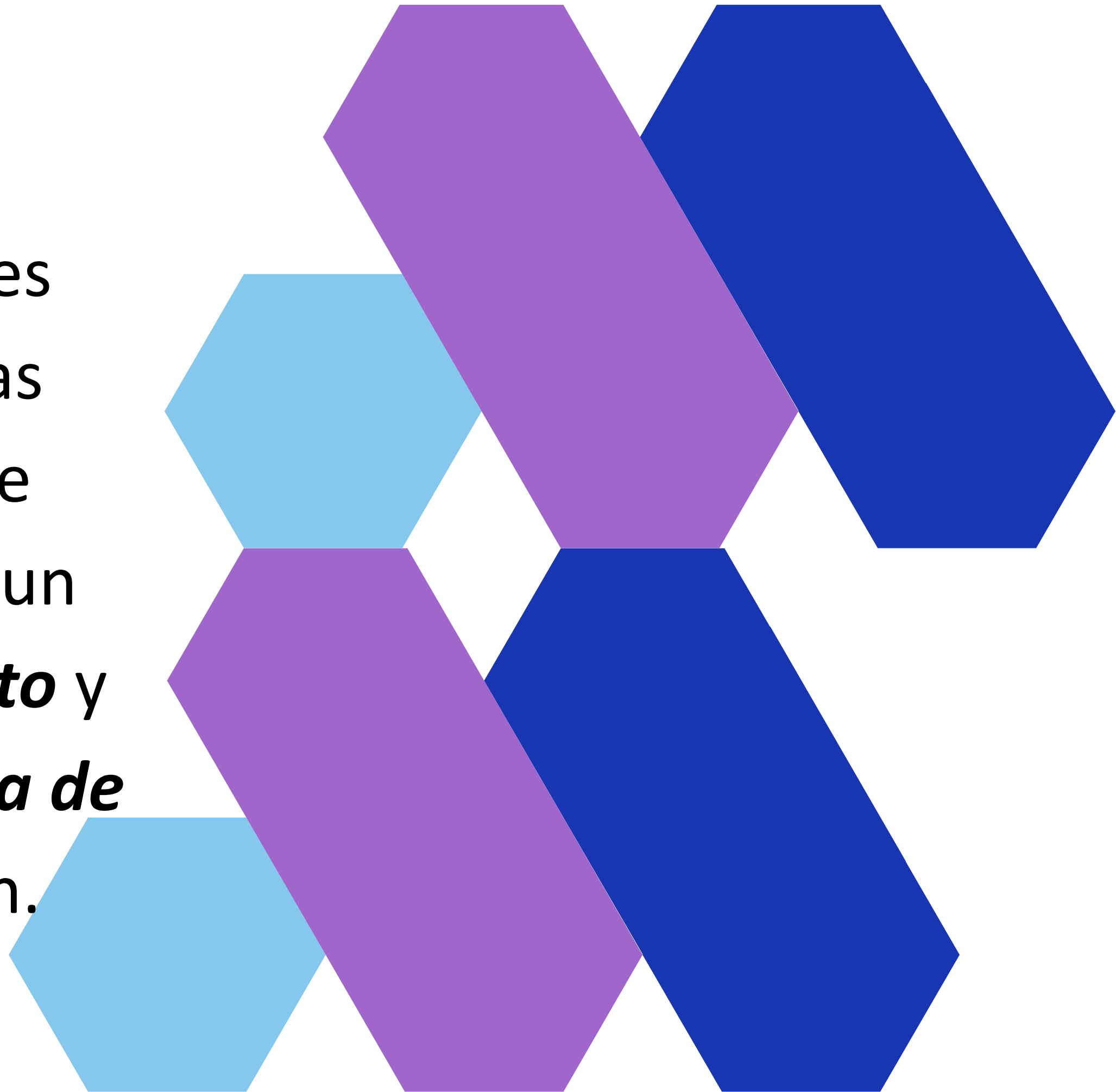


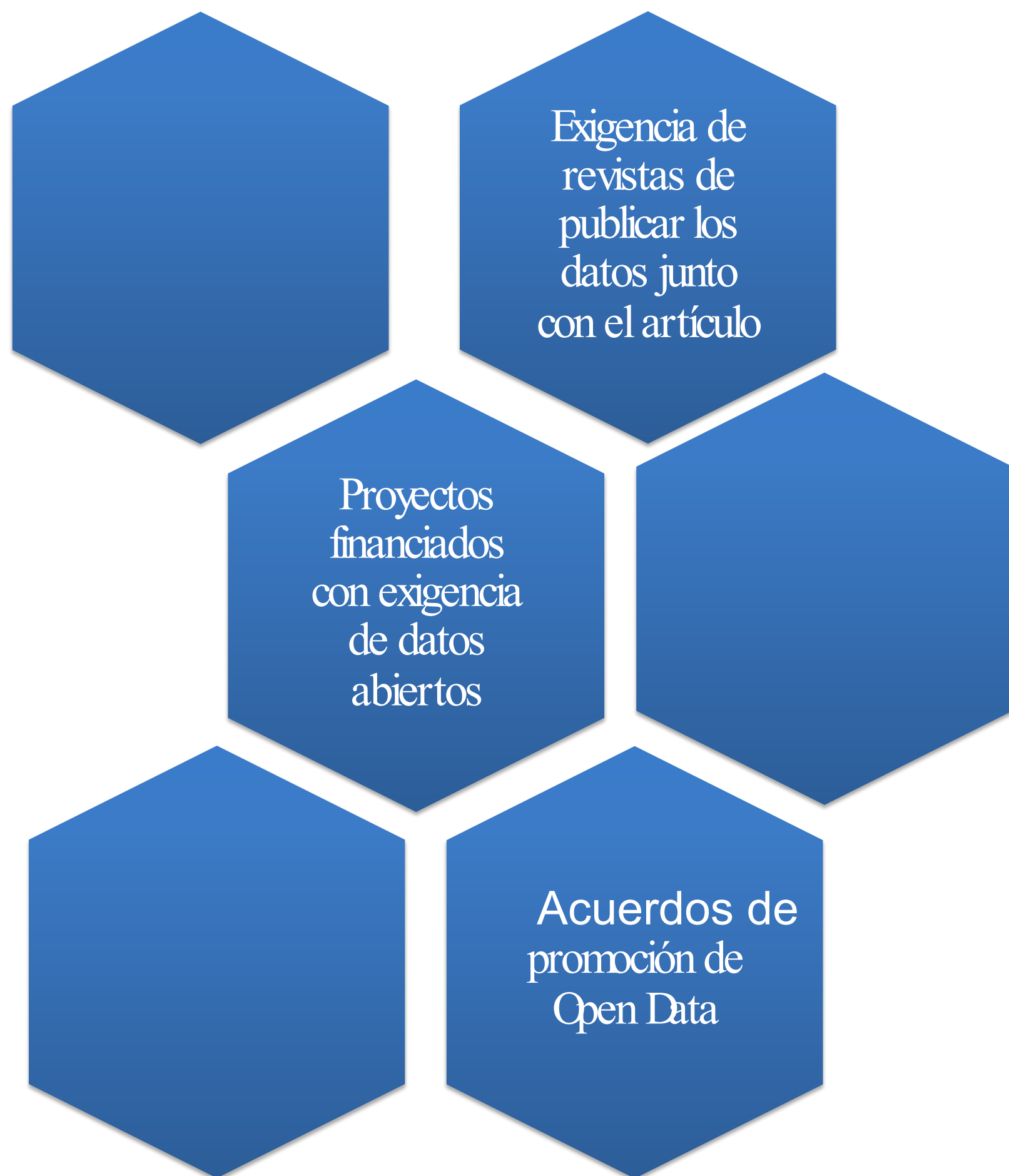
Datos abiertos: Son datos que cualquiera puede usar, modificar y compartir libremente para cualquier propósito.
(Opendefinition.org, 2018)

El principio es simple: alguien que no está familiarizado con tu proyecto debe ser capaz de ver los archivos de tu computadora y entender en detalle qué es lo que hiciste y por qué [...]. Pero por lo común resulta que **ese “alguien” eres tú**. En unos meses a partir de ahora, podrás no recordar lo que estabas haciendo cuando creaste un conjunto de archivos en particular, o podrás olvidar qué conclusiones esbozaste. Tendrás entonces que pasar tiempo reconstruyendo tus experimentos anteriores y percepciones adquiridas a partir de esos experimentos.

William Stafford Noble (2009) A Quick Guide to Organizing Computational Biology Projects. PLoSComputBiol 5(7): e1000424. doi:10.1371/journal.pcbi.1000424

Actualmente, las entidades
financieras y las revistas
académicas exigen que se
presente ***plan de gestión***, un
lugar seguro para ***el depósito*** y
en algunos casos ***la apertura de***
los datos de investigación.





Ejemplos de exigencias de organismos de financiación



Ciencia abierta a lo largo del programa

Ciencia abierta

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

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

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



Investigación abierta en Europa

[Navegar](#) [Puertas de enlace y colecciones](#) [Cómo publicar](#) [Acerca de](#) [Centro de recursos](#) [Blog](#)

[Hogar](#) > [Para autores](#) > [Directrices sobre datos, software y códigos abiertos](#)

Envíe su investigación

Mis envíos

Directrices del artículo

Directrices del artículo (nuevas versiones)

Directrices sobre datos, software y códigos abiertos

Directrices sobre datos abiertos y materiales fuente accesibles (HSS)

Cargos por procesamiento de

Directrices sobre datos, software y códigos abiertos

- [¿Qué se requiere al enviar un artículo?](#)
- [Orientación adicional](#)
- [Excepciones limitadas a estas pautas](#)
- [Los principios de datos FAIR](#)
- [Repositorios aprobados por Open Research Europe](#)

Estas directrices se relacionan con la política de Open Research Europe sobre disponibilidad de datos, que requiere que todos los autores compartan los datos subyacentes relacionados con su artículo. El texto de la política se puede leer [aquí](#). La Política de Datos Abiertos está alineada con los requisitos de Horizonte Europa para la gestión de datos de investigación, tal como se expresa en el artículo 17 del [Modelo de Acuerdo de Subvención](#).

Open science practices include early and open sharing of research (for example through preregistration, registered reports, pre-prints, or crowd-sourcing); research output¹⁷ management; measures to ensure reproducibility of research outputs; providing open access to research outputs (such as publications, data, software, models, algorithms, and workflows); participation in open peer-review; and involving all relevant knowledge actors including citizens, civil society and end users in the co-creation of R&I agendas and contents (such as citizen science).

These practices are explained and relevant resources provided in a separate section further below ([open science practices and resources](#)).

Mandatory open science practices

- Some open science practices are **mandatory for all beneficiaries per the grant agreement**. They concern:
 - open access to scientific publications under the conditions required by the grant agreement;
 - responsible management of research data in line with the FAIR principles of 'Findability', 'Accessibility', 'Interoperability' and 'Reusability', notably through the generalised use of data management plans, and open access to research data under the principle 'as open as possible, as closed as necessary', under the conditions required by the grant agreement;
 - information about the research outputs/tools/instruments needed to validate the conclusions of scientific publications or to validate/re-use research data;
 - digital or physical access to the results needed to validate the conclusions of scientific publications, unless exceptions apply;
 - in cases of public emergency, if requested by the granting authority, immediate open access to all research outputs under open licenses or, if exceptions apply, access under fair and reasonable conditions to legal entities that need the research outputs to address the public emergency¹⁸.

These obligations are described in the Model Grant Agreement (Article 17) and detailed guidelines on complying with them are provided in the Annotated Grant Agreement (Article 17).

Research data management (RDM): RDM is mandatory in Horizon Europe for projects generating or reusing data. If you expect to generate or reuse data and/or other research outputs (except for publications), you are required to outline in a maximum of one page how these will be managed. Further details on this are provided in the proposal template in the relevant section on open science. A full data management plan (DMP) is not required at submission stage. For those work programmes that require the use of the European Open Science Cloud (EOSC) federated repositories, proposers should explicitly discuss the use of such repositories in their proposals. By exception, in cases of a public emergency and if the work programme requires so, you should submit a full DMP already with submission of proposals or at the latest by the signature of the grant agreement. A [template for a DMP](#) is provided under the reporting templates in the Funding & Tenders Portal [Reference Documents](#) page.



Horizon Europe (HORIZON)

Programme Guide



Preparando su plan de gestión de datos

Financiamiento en NSF

[Encontrar financiación](#) +

[Preparando su propuesta](#) -

[Descripción general](#)

[Propuesta de presupuesto](#)

[Documentos de personal superior](#)

Plan de gestión de datos

[Investigación con animales vertebrados vivos](#)

[Investigación que involucra sujetos humanos](#)

[Presentando su propuesta](#)

[Cómo tomamos decisiones de financiación](#)

[Recursos adicionales](#) +

El plan de gestión de datos de dos páginas es una parte obligatoria de una propuesta a la Fundación Nacional de Ciencias de EE. UU. Describe cómo una propuesta seguirá la política de NSF sobre la gestión, difusión e intercambio de resultados de investigación.

Esta página proporciona una descripción general de los requisitos para el plan de gestión de datos. Consulte la [Guía de procedimientos y políticas de propuestas y adjudicaciones \(PAPPG\) XI.D.4](#) para obtener orientación completa y conocer la política de intercambio de datos de NSF.

En esta página

- [Política de intercambio de datos de NSF](#)
- [Orientación general](#)
- [Orientación de dirección y/o división](#)
- [¿Preguntas?](#)

International Development Research Centre IDRC



que hacemos ▾ Fondos ▾ Asociaciones Acerca del IDRC ▾ Noticias y Eventos ▾

Inglés | francés | Español

Prevención y preparación global para emergencias sanitarias

Aprende más

¿Cómo solicito una beca de investigación?

La información solicitada en el kit es la mínima requerida por el Centro, por lo que cada respuesta es obligatoria (a menos que se indique como opcional). Si no proporciona toda la información obligatoria y los documentos requeridos y/o no proporciona documentación sin firmar, se producirán retrasos en el procesamiento y/o su solicitud podrá ser rechazada. Consulte también los requisitos específicos de una convocatoria individual en la lista de [convocatorias abiertas actuales](#).

Los solicitantes deben consultar la oportunidad de financiamiento o seguir las pautas del IDRC sobre si la solicitud debe incluir un [Plan de gestión de datos \(DMP\)](#). Si es así, complete la Etapa 1 o la Etapa 2 del DMP.

[Kit de solicitud de subvención: un recorrido rápido](#)

Ejemplos de exigencias revistas

[Introducción](#)[Definición de conjunto de datos mínimo](#)[Métodos aceptables para compartir datos](#)[Restricciones de acceso a datos aceptables](#)[Restricciones inaceptables de acceso a datos](#)[Preguntas frecuentes](#)[Consejo asesor de datos de PLOS](#)

Disponibilidad de datos

La siguiente política se aplica a todas las revistas de PLOS, a menos que se indique lo contrario.

Introducción

Las revistas PLOS exigen que los autores pongan a disposición del público sin restricciones todos los datos necesarios para replicar los hallazgos de su estudio en el momento de la publicación. Cuando restricciones legales o éticas específicas prohíben compartir públicamente un conjunto de datos, los autores deben indicar cómo otros pueden obtener acceso a los datos.

Al enviar un manuscrito, los autores deben proporcionar una Declaración de disponibilidad de datos que describa el cumplimiento de la política de datos de PLOS. Si el artículo se acepta para publicación, la Declaración de disponibilidad de datos se publicará como parte del artículo.

Los métodos aceptables para compartir datos se enumeran [a continuación](#), acompañados de orientación para los autores sobre lo que debe incluirse en su Declaración de disponibilidad de datos y cómo seguir [las mejores prácticas en los informes de investigación](#).

PLOS cree que compartir datos fomenta el progreso científico. La disponibilidad de datos permite y facilita:

- › Validación, replicación, reanálisis, nuevo análisis, reinterpretación o inclusión en metanálisis;
- › Reproducibilidad de la investigación;
- › Esfuerzos para garantizar que los datos se archiven, aumentando el valor de la inversión realizada para financiar la investigación científica;
- › Reducción de la carga de los autores a la hora de preservar y encontrar datos antiguos y gestionar las solicitudes de acceso a los datos;
- › Citación y vinculación de datos de investigación y sus artículos asociados, mejorando la visibilidad y garantizando el reconocimiento de los autores, productores de datos y curadores.

< Datos de la investigación

Política de datos de investigación

Política de datos de investigación

Declaraciones de disponibilidad de datos

Orientación sobre el repositorio de datos

Información delicada

Preguntas frecuentes sobre política de datos

Servicio de asistencia técnica sobre datos de investigación

Política de datos de investigación

En Springer Nature avanzamos en los descubrimientos publicando investigaciones confiables, apoyando el desarrollo de nuevas ideas y defendiendo la ciencia abierta. También pretendemos facilitar el cumplimiento de los requisitos de las instituciones y los financiadores de la investigación para compartir datos.

Para ayudar a lograr esto, hemos establecido una política de datos de investigación estándar para nuestras revistas, basada en la transparencia en torno a los datos de respaldo. Esta política se aplica a todos los conjuntos de datos que son necesarios para interpretar y replicar las conclusiones reportadas en un artículo de investigación.

1. Todos los artículos originales deben incluir una declaración de disponibilidad de datos.

Las declaraciones de disponibilidad de datos deben incluir información sobre qué datos están disponibles, dónde se pueden encontrar y los términos de acceso aplicables. Esto se aplica tanto a los datos originales como a los reutilizados, y a si los datos se pueden compartir públicamente o no. Consulte nuestra guía sobre declaraciones de disponibilidad de datos para obtener más información.

SPRINGER NATURE

Preguntas frecuentes sobre política de datos

¿Tengo que compartir mis datos cuando los envío a una revista Springer Nature?

Recomendamos encarecidamente que todos los datos de la investigación se pongan a disposición de los lectores sin restricciones indebidas. Para algunos tipos de datos, es obligatorio enviarlos a un repositorio público respaldado por la comunidad. Consulte nuestra guía sobre la depósito obligatorio de datos para obtener más información .

Reconocemos que algunos datos, como los datos sobre participantes humanos identificables en investigaciones, no se pueden compartir abiertamente. Cuando los datos no se puedan compartir, se debe proporcionar una razón adecuada en la Declaración de disponibilidad de datos.

Los revisores y editores pueden solicitar acceso a conjuntos de datos que no están disponibles públicamente.

Escribir una declaración de disponibilidad de datos

Si envía su artículo a una revista de Taylor & Francis con una [política de intercambio de datos](#), se le pedirá que proporcione una declaración de disponibilidad de datos con su envío.

¿Qué es una declaración de disponibilidad de datos?

Una declaración de disponibilidad de datos (también llamada a veces "declaración de acceso a datos") le dice dónde están disponibles los datos de investigación asociados con un artículo y bajo qué condiciones se puede acceder a los datos. También incluyen enlaces (cuando corresponda) al conjunto de datos.



Nuestra [guía para compartir y citar datos](#) explora qué son los datos de investigación y los numerosos beneficios de compartirlos. Los siguientes pasos describen cómo puede compartir sus datos de manera efectiva y qué considerar a lo largo del camino.

1. Planifique compartir: escriba un plan de gestión de datos

Una buena práctica, y a menudo exigida por los financiadores, es elaborar un plan de gestión de datos (DMP) justo al inicio de su proyecto de investigación. El formato de estos planes varía, pero normalmente incluirá cómo recopilá y manejará de forma segura los datos recopilados durante todo el proyecto y más allá, así como quién es responsable de ellos. Vea estos ejemplos del [Centro de curación digital](#).

Piense también en cómo compartirá sus datos y detalle eso en el plan. En particular, debes considerar:

- 1. Consentimiento para compartir datos**
Los formularios de consentimiento de los participantes no solo deben incluir el permiso para utilizar sus datos en su investigación, sino también la publicación de los resultados (anonimizados).
- 2. Formato para compartir**
Si decide qué formato desea utilizar para compartir sus datos, le ahorrará tiempo más adelante si lo incorpora en las etapas de recopilación o procesamiento de datos.

2. Prepare sus datos para compartirlos, de acuerdo con las mejores prácticas éticas.

En primer lugar, deberá asegurarse de que no existan razones éticas, legales o de seguridad por las que no deba facilitar sus datos.



Incluso si puede compartir sus datos, es posible que deba anonimizarlos para proteger la privacidad de los participantes de la investigación. Consulte nuestra [introducción a la ética del intercambio de datos](#).

Es posible que existan estándares disciplinarios sobre cómo formatear sus datos y etiquetar archivos para compartirlos. Compruebe también si [el repositorio de datos](#) elegido tiene algún requisito específico.

Además de los datos en sí, también debe preparar los recursos que otros necesitarán para comprender sus datos. Esto podría incluir un diccionario de datos y los detalles de cualquier software necesario para ver los datos.

¿Ha oído hablar de los cambios en el proceso de solicitud de subvenciones de los Institutos Nacionales de Salud (NIH)?

Cualquier aplicación nueva o renovada deberá incluir un plan de gestión e intercambio de datos.

[Obtenga más información sobre las actualizaciones](#)

Consejos de publicación, directo a tu bandeja de entrada

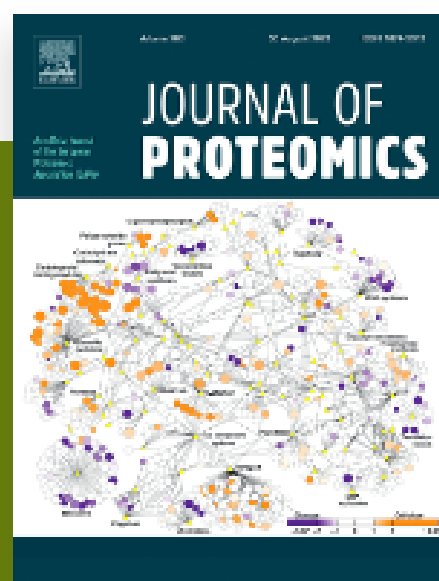
Consejos y orientación de expertos sobre cómo publicar y maximizar el impacto de su investigación. Regístrese ahora para recibir información semanal directamente en su bandeja de entrada.

[Regístrese hoy para Research Insights](#)

Vinculación del artículo, con set de datos de investigación publicado en Data in Brief



ELSEVIER



Journal of Proteomics

Supports open access

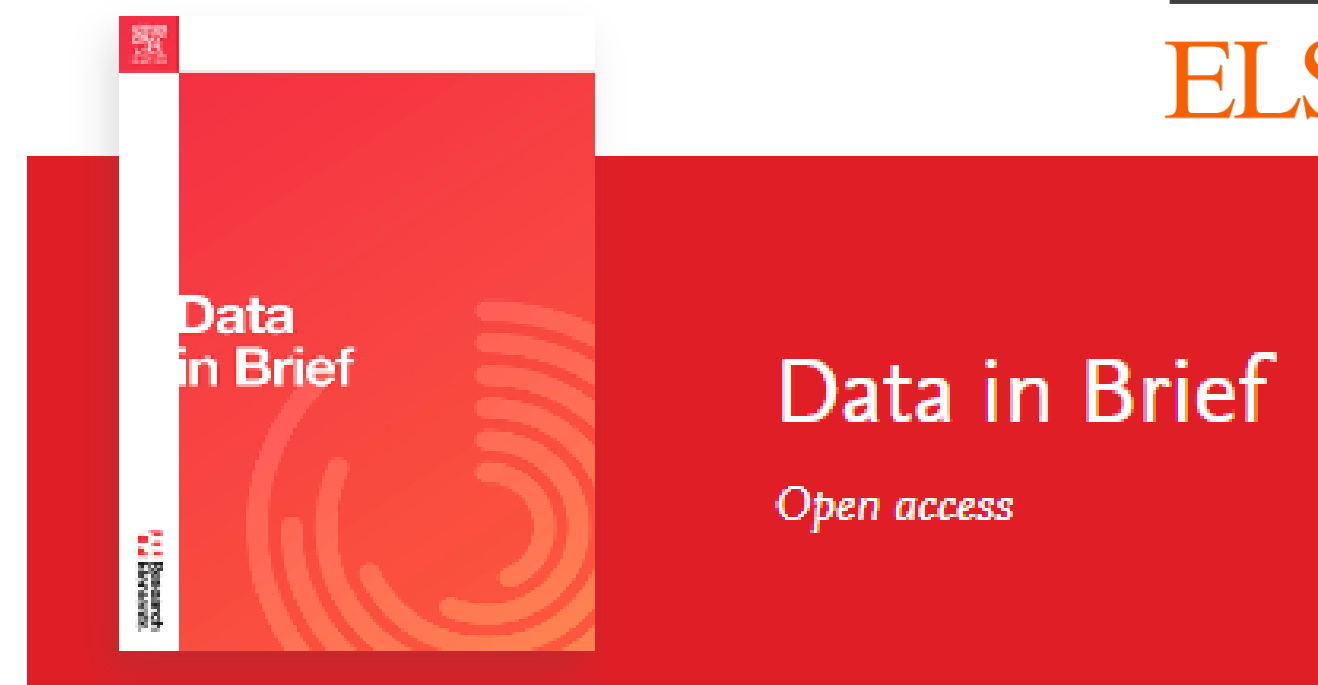


Journal of Proteomics

Volume 109, 23 September 2014, Pages 228-239



Proteomic analysis of the mitochondria from embryonic and postnatal rat brains reveals response to developmental changes in energy demands




Data in Brief

Volume 1, December 2014, Pages 42-45



Data Article

Data for mitochondrial proteomic alterations in the developing rat brain

[Lance M. Villeneuve](#), [Kelly L. Stauch](#), [Howard S. Fox](#)  

.....una revista multidisciplinaria, de acceso abierto y revisada por pares

Acuerdos de promoción de Operación Data

Open Government Data

Carta Internacional de los Datos Abiertos (Open Data Charter), 2015

Es un conjunto de principios y buenas prácticas para la liberación de datos abiertos gubernamentales.

Bogotá se compromete a trabajar en tres temas fundamentales que hacen parte de la agenda temática del Open Data Charter: Cambio Climático, lucha contra la corrupción y desarrollo económico.



<https://tic.bogota.gov.co/noticias/bogot%C3%A1-ingresa-al-open-data-charter>

Guía para la creación de Políticas de Datos Abiertos

La **Sunlight Foundation** (Washington, DC) ha creado este documento con el objetivo de presentar una amplia visión sobre el tipo de desafíos a los que la creación de Políticas de Datos Abiertos pueden enfrentarse.



Web Integrity Project

We monitor changes to government websites — revealing shifts in public information and access to Web resources, as well as changes in stated policies and priorities.



Open Cities

We make municipal governments across the country more transparent, accountable, and participatory. Specifically, we provide assistance with community-centered open data and open contracting.

<https://sunlightfoundation.com/>

Acuerdo Internacional

Promover los principios de datos abiertos



InterAcademy Partnership (IAP)

International Social Science Council (ISSC)

International Council for Science (ICSU)

World Academy of Sciences (TWAS)

1. Responsabilidades

- *Los investigadores*
- *Las instituciones de investigación y las universidades*
- *Los editores*
- *Los organismos de financiación*
- *Las asociaciones profesionales, sociedades científicas y academias*
- *Bibliotecas, archivos y repositorios*

2. Los límites de la apertura

3. Prácticas habilitantes

- *Citación y procedencia*
- *Interoperabilidad*
- *Reutilización no restrictiva*
- *Capacidad de vinculación (linked open data)*

Los datos



Gestionar

Procesar

Publicar

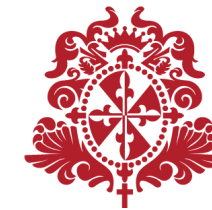
Preservar

¿Qué pasa cuando?

- Los investigadores se van de una institución
- No se hace un backup periódico
- No se actualizan adecuadamente los formatos de almacenamiento

Estrategia de gestión de datos de investigación





3 components principales de la estrategia



**Política
Institucional** de
gestión de datos de
investigación



**Repositorio de
datos** de investigación



**Servicios de
apoyo** a la gestión de
datos de investigación

Política para la gestión de datos de investigación en la UR

Decreto Rectoral No. 1576 del 27 de marzo de 2019



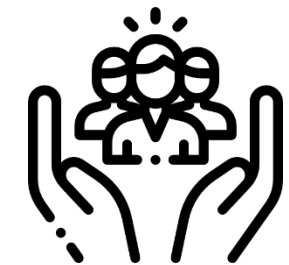
Visibiliza **la importancia** del proceso de gestión de datos de investigación.



Define que va a entender la Universidad por datos de investigación, cuál es la motivación y cuáles son sus **objetivos**.



Establece el compromiso de brindar un **apoyo integral** a los investigadores en este proceso a través de **recursos y servicios**.



Manifiesta **qué espera de los investigadores** como protagonistas del proceso.

Repositorio de Datos de Investigación

<https://research-data.urosario.edu.co/> 



Search - About User Guide Support English - Sign Up Log In

Universidad del Rosario | **dir** Repositorio de datos de investigación

El Repositorio de datos de investigación de la Universidad del Rosario es una plataforma online para administrar, almacenar y difundir datos de investigación de los investigadores de la UR y sus aliados, convirtiéndose también en un espacio de trabajo colaborativo que permite el acceso a datos publicados para su análisis y re-uso.

Buscar Explorar

Agricultural Sciences 1	Arts and Humanities 1	Astronomy and Astrophysics 1
Business and Management 27	Chemistry 1	Computer and Information Science 16
Earth and Environmental Sciences 1	Engineering 9	Law 13
Mathematical Sciences 4	Medicine, Health and Life Sciences 13	Physics 1
Social Sciences 38		

- 110 Investigadores
- 23 Proyectos de investigación
- 57 Datasets abiertos
- 391 Archivos



100

Repositorios en el mundo desarrollados con esta herramienta.



Repositorio de datos de investigación

Indexado en:



Google Dataset Search Beta



Type to search...

Pages Support

Works People Organizations Repositories

Repositorio de datos de investigación de la Universidad del Rosario

462 Works 2 Citations 29,020 Views 1,423 Downloads

The Universidad del Rosario Research data repository is an institutional initiative launched in 2019 to preserve, provide access and promote the use of data resulting from Universidad del Rosario research projects.

Data Access open embargoed restricted Persistent Identifier doi Certificates none Data Upload restricted Provider Type dataProvider

humanities and social sciences economics business administration jurisprudence social and behavioural sciences life sciences

Search for Datasets

Try boston.education data or weather.sitemasa.gov



Universidad del rosario

Última actualización Formato de descarga Derechos de uso Tema Gratis

Volumen 4 - Gráficos que reflejan las co-autorías... research-data.urosario.edu.co search.datacite.org Última actualización: Jun 17, 2020

Base de datos cuidadores a nivel global Scopus research-data.urosario.edu.co xlsx Última actualización: Aug 9, 2022

Data from: Extracting low-cost signals of quality control in... research-data.urosario.edu.co search.datacite.org application/x-stata+4 Última actualización: Sep 3, 2022



Base de datos cuidadores a nivel global Scopus

Ver en: Universidad del Rosario

xlsx(4137277)

Identificador único https://doi.org/10.34848/QTkZF1

Fecha de actualización del conjunto de datos Aug 9, 2022

Conjunto de datos proporcionado por Universidad del Rosario

Licencia Attribution-NonCommercial 4.0 (CC BY-NC 4.0) Se ha obtenido la información de la licencia automáticamente

Descripción



Search results for 'Universidad del Rosario' on Papyrus, showing a list of datasets from various institutions like Universidad Católica Luis Amigó and Fundación Universitaria del Área Andina.

¿Qué es la gestión de datos de investigación?



La Gestión de Datos de Investigación (GDI) es el proceso activo de manejo de los datos generados en una investigación. La GDI se realiza de forma continua y cubre todas las decisiones relacionadas con la gestión de los datos a lo largo de su **ciclo de vida**.



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.



Universidades
Españolas

Red de Bibliotecas
REBIUN

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



El personal de tu
Biblioteca
te puede asesorar





Tipos de datos

¿Qué tipo de datos generados (o que se van a generar) en el proyecto?

Textuales (DOC, DOCX, PDF, TXT, RTF, PDT, etc)

Multimedia (MP4, MP3, WAV, M4A)

Imágenes (JPEG, TIFF, PNG)

Tabulares (Shapefile de ArcGIS, STATA, Excel, CSV etc)

Estructurados (Bases de datos de genomas de organismos, XML, bases de datos MySQL, etc)

Código fuente con las rutinas o algoritmo (Java, C, R, Python, etc)

Específicos de programas (modelos matemáticos de simulación, modelos estadísticos, SPSS, etc)

Archivos netCDF, que almacenan datos científicos multidimensionales (variables) como la temperatura, la humedad, la presión, la velocidad del viento, la dirección, etc

Repositorio de datos de investigación

Es una plataforma online para administrar, almacenar y difundir datos de investigación, convirtiéndose también en un espacio de trabajo colaborativo que permite el acceso a datos publicados para su análisis y re-uso

Centraliza el acceso a los datos, permitiendo su correcta gestión.

Brinda un lugar seguro para preservar los datos a largo plazo, incluso años después de haberlos generado.

Permite generar un registro de información de todos los usuarios que descarguen la base de datos

Permite la asignación de licencias de uso para todos los sets de datos, lo cual facilita el re-uso adecuado de los mismos, y contribuye a la atribución de la propiedad intelectual a los autores.

Permite compartir los datos a la comunidad y **aumentar su visibilidad** a través de un identificador persistente.

¿Cómo organizará los conjuntos de datos (datasets) generados en el proyecto?

Un dataset, es un conjunto de datos que están relacionados.

Nombre del Dataset 1	
Encuesta Etiopia 2019	
Descripción	
Archivos que contiene	
Nombre	Descripción
Cuestionario Etiopia	
Datos recolectados Etiopia	
Reporte de resultados Etiopia	

Universidad del Rosario

↳ Proyecto Nutrición

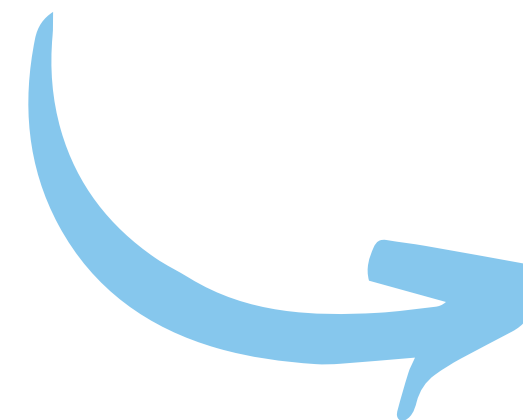
↳ Encuesta Etiopía

- ↳ ○ Cuestionario
- Datos recolectados
- Reporte de resultados

↳ Encuesta Vietnam

- ↳ ○ Cuestionario
- Datos recolectados
- Reporte de resultados

¿Cómo se van a asignar los permisos de acceso a los miembros del proyecto ?



Sistema de autenticación de usuarios y asignación de roles

Rol	Descripción	Nombre
Administrador	Persona que tiene todos los permisos para todos los datasets y archivos.	
Editor	Persona que puede agregar datasets	
Curador	Persona que puede editar los datasets almacenados por el administrador (no puede almacenarlos, ni publicarlos). Todos los cambios realizados deben ser enviados para revisión del admón.	
Miembro	Persona que puede ver dataverses y datasets no publicados pero NO editarlos	

3. Acceso y difusión

Identificadores persistentes

Handle

El sistema Handle proporciona un servicio de nombres global de propósito general que permite la resolución segura de nombres a través de Internet, diseñado para permitir que un amplio conjunto de comunidades utilice la tecnología para identificar contenido digital independientemente de la ubicación.

DOI

DOI (Digital Object Identifier) es un identificador único y permanente para las publicaciones electrónicas. El DOI proporciona información sobre la descripción del objeto digital a través de metadatos, (autor, título, datos de publicación,...) y su localización en internet para la identificación persistente y ***semánticamente interoperable***.

Acceso y difusión

Ventajas de los DOI para los datos de investigación

- ✓ Aumenta la visibilidad de los datos, ya que permiten su rápida identificación y acceso, lo cual contribuye al aumento de su impacto.
- ✓ Permiten la interoperabilidad con otras plataformas, repositorios o motores de búsqueda. Cuando se asigna un DOI, no solo se vuelve identificable, sino que, a través de los metadatos asociados y el índice de búsqueda central, se puede descubrir y conectar con otros resultados de la investigación.
- ✓ Los DOIs hacen que sus datos de investigación sean visibles y citables, lo cual contribuye a la preservación de la propiedad intelectual, ya que facilita dar crédito a los autores gracias a la información que lleva asociada en sus metadatos.
- ✓ Facilita el seguimiento del uso de datos, dado que permite obtener información sobre citas recibidas, vistas y descargas.

Ventajas de los DOI para los datos de investigación

The screenshot shows a Google search interface. At the top, the Google logo is on the left, and a search bar contains the text "Leishmania in the americas". To the right of the search bar are icons for help, notifications, and a user profile. Below the search bar, there are several filters: "Última actualización", "Formato de descarga", "Derechos de uso", "Tema", and "Gratis". On the far right, there is a button labeled "Conjuntos de datos guardados".

The search results are displayed in two columns. The left column shows a list of results, with the top one highlighted. The right column shows the details for the selected result.

Search Results:

- Leishmania in the Americas DB**
research-data.urosario.edu.co
Formato: tsv
Última actualización: Aug 25, 2021
- Metadata record for: An interactive database of...**
springernature.figshare.com
figshare.com
Formato: txt
Última actualización: Apr 2, 2020
- Data from: Global distribution mans of the Leishmaniases**

Details for "Leishmania in the Americas DB":

- Logo of Universidad del Rosario
- Button: **Ver en Universidad del Rosario**
- Text: 3 artículos académicos citan este conjunto de datos ([Ver en Google Académico](#))
- Formato: tsv(630337), tsv(694063), tsv(571488)
- Identificador único**
https://doi.org/10.34848/FK2_leshmania_ds
- Fecha de actualización del conjunto de datos**
Aug 25, 2021
- Conjunto de datos proporcionado por**
[Universidad del Rosario](#)

Ventajas de los DOI para los datos de investigación

Leishmania in the Americas



[Advanced search help](#)

Filter Results

7580 results

DATA TYPES

- Dataset (7430)
- Tabular Data (96)
- Image (91)
- Document (37)
- Collection (16)
- Text (6)
- Software/Code (5)
- Other (4)
- File Set (3)



Leishmania in the Americas DB

Contributors: Herrera, Giovanni, Bottin, Marius, Barragán, Natalia, Luna, Nicolás, Martínez, David et al

Date: 2020-01-01

Source: Repositorio de datos de investigación Universidad del Rosario

This database contains geo-referenced points of **Leishmania** species occurrence in **America**.

Data Types:

Dataset

Export: [APA](#) | [BibTeX](#) | [DataCite](#) | [RIS](#)

Tratamiento de datos sensibles

Protección de los datos sensibles generados en el proyecto



- Anonimización de las bases de datos que contengan datos sensibles
- Verificación de la protección de datos sensibles

Propiedad intelectual y re-uso de los datos



Los datos, al igual que otras creaciones de la mente, están sujetas a derechos de autor

Uso de datos de
terceros

Toma de
decisiones sobre
los datos del
proyecto

Propiedad intelectual y re-uso de los datos

Derechos morales

Son derechos personalísimos, a través de los cuales se busca salvaguardar el vínculo que se genera entre el autor y su obra, en tanto ésta constituye la expresión de su personalidad. Surgen en cabeza del autor por el sólo hecho de creación de la obra. Son inalienables, inembargables, intransferibles e irrenunciables.

Derechos patrimoniales

Son prerrogativas de naturaleza económico - patrimonial, con carácter exclusivo, que permiten a su titular controlar los distintos actos de explotación de que la obra puede ser objeto: Reproducción, transformación, distribución.

Fuentes:

Delgado, P.A. (2017). Derechos de autor en Colombia: Especial referencia a su transferencia y disposición jurídica en el ámbito universitario. Disponible en <http://www.scielo.org.co/pdf/cesd/v8n2/v8n2a04.pdf>

Dirección Nacional de Derechos de autor. Disponible en <http://derechodeautor.gov.co/>

Propiedad intelectual y re-uso de los datos

Licencias de uso



Permiten ceder algunos derechos de propiedad intelectual según las preferencias del autor.

Propiedad intelectual y re-uso de los datos

Se puede distribuir, adaptar , incluso para propósitos comerciales, dando crédito al autor



Se puede distribuir, adaptar , pero no para propósitos comerciales, y se debe compartir bajo los mismos términos



Se puede distribuir, adaptar , incluso para propósitos comerciales, dando crédito al autor y licenciando las nuevas creaciones en iguales términos



Se puede reusar la obra para cualquier propósito, incluso comercial, pero no se puede adaptar.

Se puede distribuir, adaptar , pero no para propósitos comerciales.



Es la más restrictiva. Solo permite que otros descarguen las obras y las compartan reconociendo la autoría.

Licencia CC0: Dominio público



Nuestra Experiencia



Universidad del
Rosario

CRAI

Centro de Recursos para el
Aprendizaje y la Investigación

Nuestra experiencia



Universidad del
Rosario



Lanzamiento de la estrategia de GDI UR

7 de noviembre de 2019

☰ YouTube ^{CO}



Lanzamiento de la estrategia de

Gestión de datos de investigación en la UR



<https://www.youtube.com/watch?v=hnro7tSYKuU>

Divulgación y sensibilización de la estrategia de GDI UR

2020 - 2023

Sesiones de sensibilización con las unidades académicas

Datos de investigación
¿Estás listo para gestionarlos?

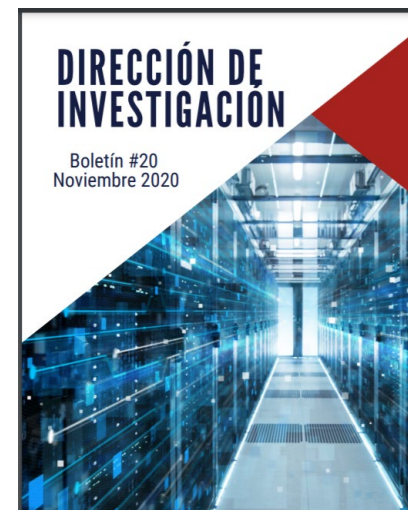


Redes sociales



Publicaciones en medios institucionales

NOVA *et*
VETERA



Voz a voz de investigadores
"embajadores"

Servicio de PGD

Construir el plan de gestión de datos de investigación

¿Cuándo se desarrolla este servicio?

Es punto de partida para el almacenamiento de los datos en el Repositorio de datos de investigación.

Desarrollo del PGD como requisito para una convocatoria o para publicación de un artículo científico.

Todos los proyectos que tienen sus datos almacenados en el Repositorio cuentan con un PGD

Todos los proyectos financiados por la UR deben tener un PGD

Depósito de datos, previa capacitación en el uso del repositorio para el equipo de investigación



- **Debe asistir el líder del proyecto**, y todos los miembros del equipo que vayan a utilizar el Repositorio.
- La capacitación se realiza únicamente cuando **el PGD ya ha sido construido y aprobado**.
- **Se realiza en dos fases:** (1) Fase teórica en la cual se presenta la estrategia y la plataforma y (2) Fase práctica en la cual se realiza un ejercicio paso a paso de almacenamiento y publicación de los datos.
- La Fase práctica se realiza en un **servidor de pruebas**, para evitar:
 - Publicación de datos no reales por error (No es posible des-publicar)
 - Asignación de DOIs a datos no reales

Entrega de un espacio exclusivo para el proyecto en el Repositorio de datos



Al finalizar la capacitación se realiza entrega **al líder del proyecto** el espacio exclusivo en el cual se ha incluido previamente:

- Nombre del proyecto
- Descripción del proyecto
- Plantilla de metadatos
- Definición de búsqueda facetadas



Responsabilidades del investigador

Recolectar, digitalizar y anonimizar los datos

Almacenar los datos de investigación

Publicar los datos en los casos en que se considere pertinente

El investigador se encarga de diligenciar los metadatos de cada dataset y de gestionar su Dataverse de manera independiente, de acuerdo al PGD.

Elementos del servicio:

- Acompañamiento en la elaboración y revisión de planes de gestión de datos de investigación.
- Apropiación de prácticas de Acceso Abierto al compartir los datos bajo las licencias adecuadas.
- Apropiación de prácticas de citación de datos.
- Formación sobre los recursos y orientación para la carga de datos del repositorio Institucional de datos de investigación.



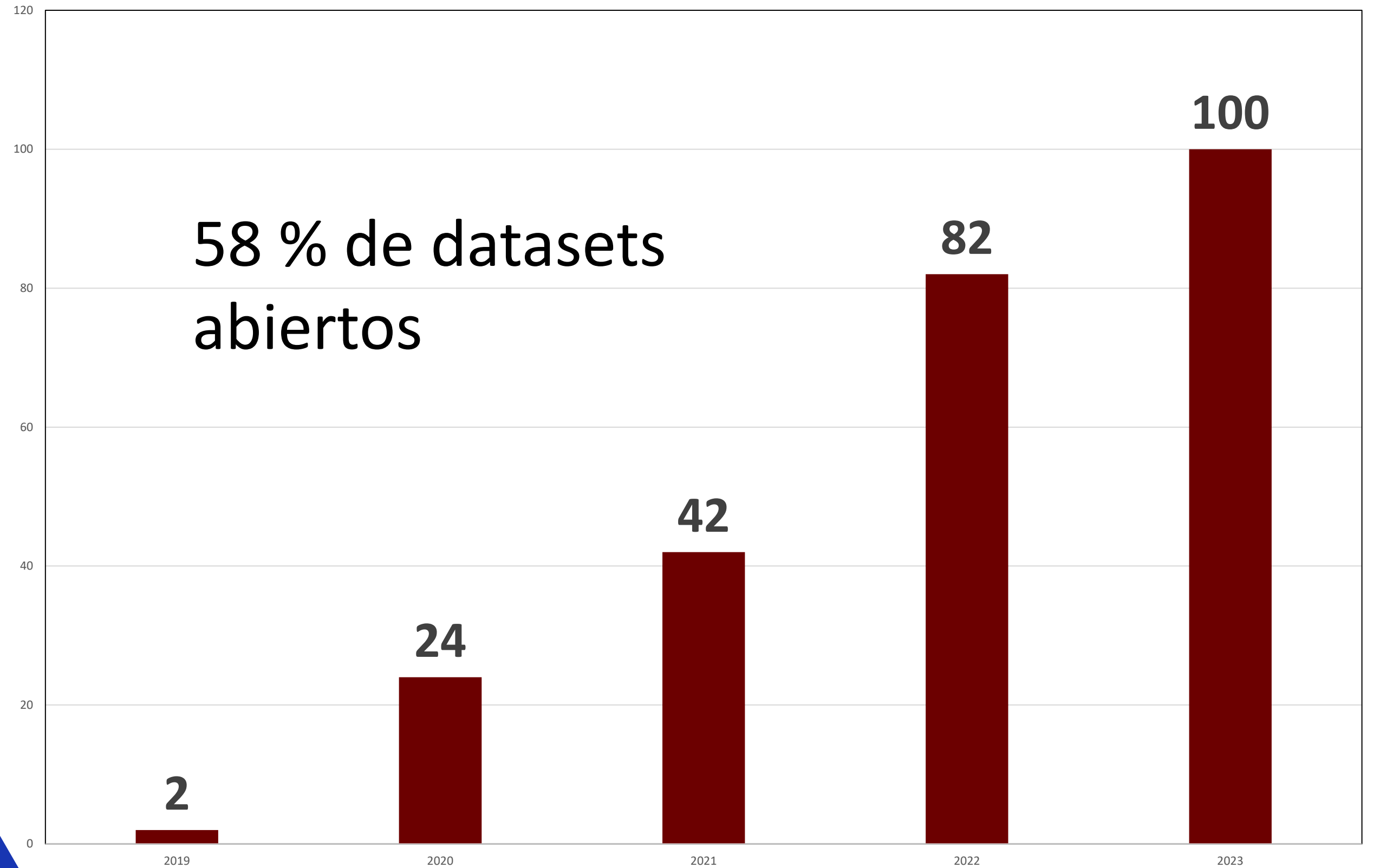
Desde el año 2020 hasta la fecha se han prestado más de 200 servicios en:

- Asesoría en la construcción del PGD
- Sensibilización sobre datos de investigación
- Uso del repositorio de datos

Servicios asociados a la gestión de datos de investigación

Apertura de datos

Datasets en el Repositorio de datos de investigación UR





Universidad del
Rosario

CRAI
Centro de Recursos para el
Aprendizaje y la Investigación

¡Gracias!

Malgorzata Lisowska Navarro,
margarita.lisowska@urosario.edu.co