

learn

Mini Taller LEARN

Martes 23 de mayo de 2017

9hs San José, CR | 10hs Bogotá, Lima, México DF |
11hs Santiago de Chile | 12hs Buenos Aires, Brasilia

Primeros para implementar un Plan De Gestión de Datos



Fernando Ariel López

(CAICYT – CONICET, Argentina)

flopez@conicet.gov.ar | [@fernando_lopez](https://twitter.com/@fernando_lopez)





Fernando Ariel López (CAICYT - CONICET, Argentina)

- Coordinador Comunicación y Formación / CAICYT-CONICET
- Director de Biblioteca / UMET
- Co-Fundador y Director / Aprender 3C
- Integra el proyecto de investigación sobre Acceso Abierto / IIGG-UBA
- Autor / InfoTecarios
- Evaluador en diversas revistas científicas y congresos profesionales.

CAICYT



CONICET

Expositor, Consultor y Formador en Acceso Abierto al conocimiento científico (Repositorios y Revistas Digitales), Ciencia Abierta (Datos científicos y Plan de Gestión de Datos), Ciencia 2.0, Web social, Tecnologías de Información y Comunicación, Alfabetización Informacional, Alfabetizador Digital, etc.

Más información en about.me/fernandoariellopez



INFORMATION



TIME

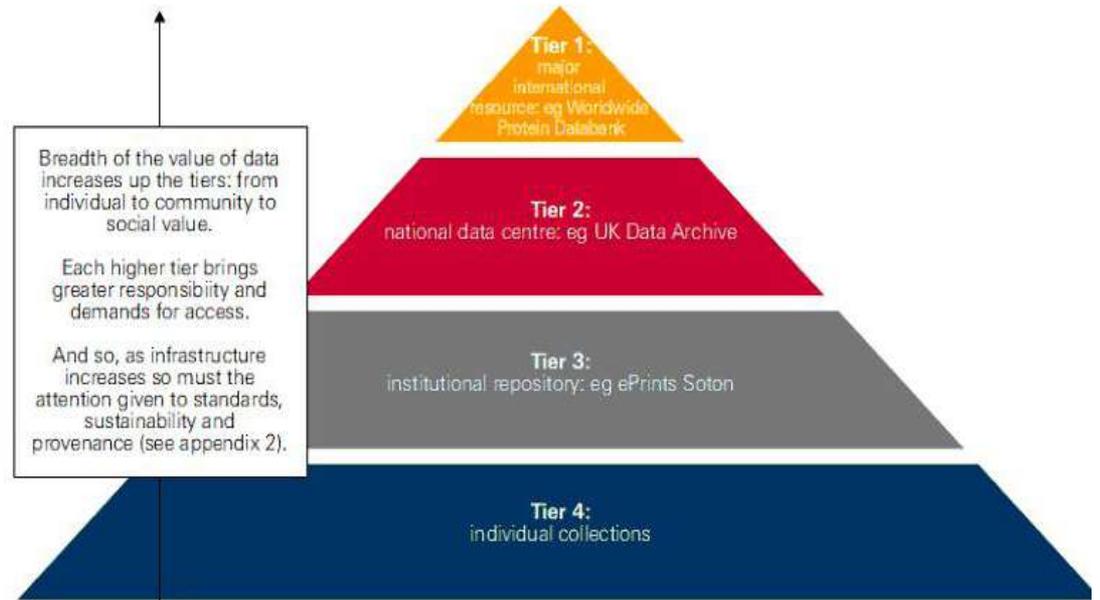


KNOWLEDGE

CIENCIA → Ciencia Abierta

GOBIERNO → Gobierno Abierto

**SOCIEDAD → Periodismo de Datos,
Hackativismo Cívico, Innovación abierta, etc.**



Líneas de Trabajo:

1. POLÍTICA DE DATOS CIENTÍFICOS
2. CICLO DE VIDA DE LOS DATOS CIENTÍFICOS
3. PLAN DE GESTIÓN DE DATOS CIENTÍFICOS
4. E-INFRAESTRUCTURAS
5. ROLES, COMPETENCIAS y FORMACIÓN

No vamos a explayarnos en...

- Open Science / Ciencia Abierta
- Research Data / Datos de Investigación
- Tipos y Formatos de Datos
- Curation / Curaduría
- E-Infraestructuras
- Roles, competencias y formación

Vayamos a centrarnos en...

Planes de Gestión de Datos (DMP)

- ¿Qué es un Data Management Plan (DMP)?
- ¿Para qué sirve un DMP?
- ¿Por dónde empezar?
- ¿Cómo comenzar a desarrollar un DMP?

¿Qué es un **Planes de Gestión de Datos**?

“... es un documento formal que describe todo el ciclo de vida de los datos, desde su recogida hasta la documentación completa del proceso de investigación y registra las decisiones tomadas en relación con estándares de metadatos, formatos, bases de datos, métodos, seguridad y periodos de almacenamiento, así como los costes asociados con la gestión de los datos...”

(Couto Corrêa, 2016).

UN PLAN de GESTIÓN de DATOS

Es un documento elaborado por el investigador o grupo de investigación donde se define:

- qué datos van a ser creados y cómo,
- cómo se van a describir, organizar, almacenar y gestionar los datos
- de qué forma van a ser compartidos, explicando cualquier restricción de uso que pueda ser aplicada.
- quién o quiénes serán los responsables de realizar cada una de estas actividades.
- de qué forma van a ser compartidos, explicando cualquier restricción de uso que pueda ser aplicada.

UN PLAN de GESTIÓN de DATOS

Es un documento vivo, va evolucionando durante la vida del proyecto. El PGD abarca desde la planificación del proyecto de investigación hasta el final de la investigación.

Momentos en los que se suele pedir un PGD:

- Acompañando al proyecto al momento de solicitar financiamiento.
- Iniciado el proyecto (a los 6 meses)
- A mitad del proyecto
- Finalizado el proyecto

¿PARA QUÉ SIRVE UN PGD / DMP?

Es una buena práctica que vienen realizando diversos organismos internacionales relacionados con el ámbito científico-tecnológico (Agencias Nacionales de Investigación, Financiadores, Consorcios de Universidades, etc.).

El PGD permite:

- organización de la gestión de los datos para el Investigador,
- capacidad de diagnóstico, caracterización y predicción acerca de la información que el PGD contiene.

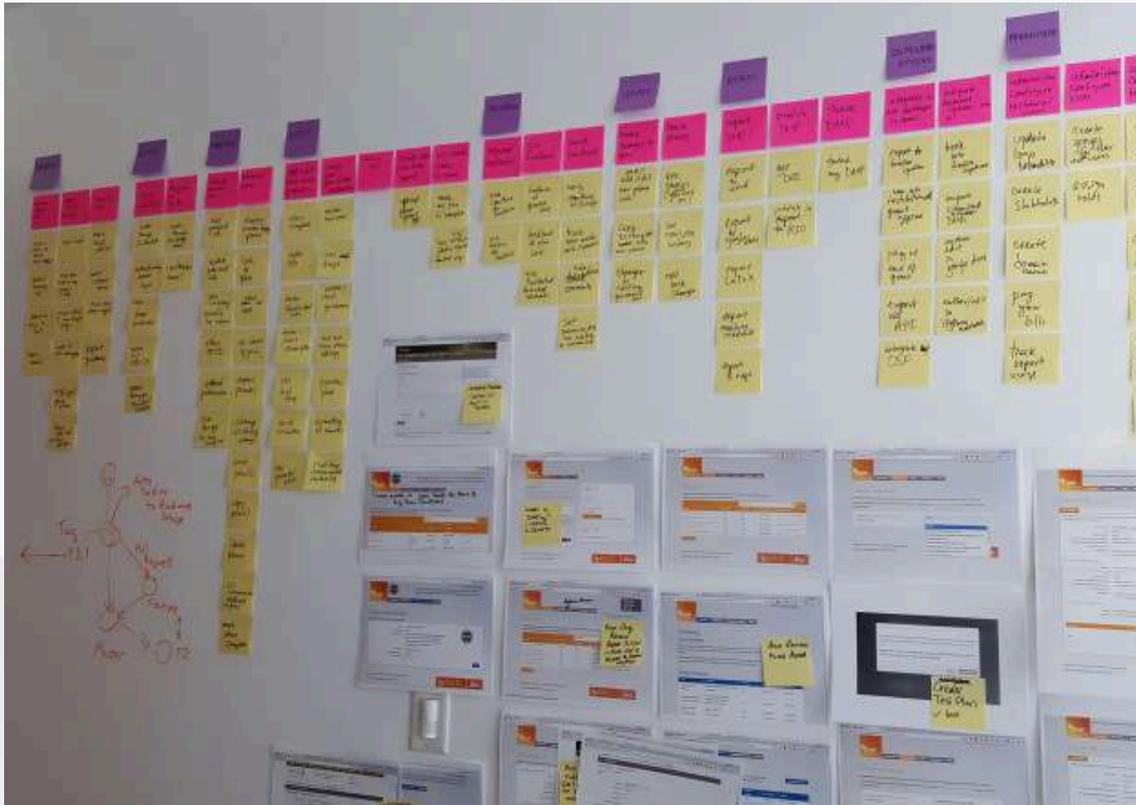
→ insumo de valor para el diagnóstico o predicción para una Institución que gestiona Ciencia y Tecnología.

¿POR DÓNDE EMPEZAMOS?

RECURSOS:

- **DMPTool (Data Management Planning Tool)** del Centro de Curado de la Universidad de California (EEUU) <https://dmptool.org>
 - Materiales de referencia, ejemplos
 - herramienta
- **DCC (Digital Center Curation)** del UK <http://www.dcc.ac.uk/>
 - [Materiales de referencia](#), ejemplos, [checklist](#)
 - [DMPonline](#)
 - Consorcio Madroño (España) - [PAGODA](#)
- **UK Data Archive** <http://www.data-archive.ac.uk/>
 - Materiales de referencia

ROADMAP DMPTool + DCC



Más información: <http://www.dcc.ac.uk/blog/dmponline-dmptool-roadmap-reciprocal-visits>

POLITICAS de RESEARCH DATA

Iniciativas internacionales en la Gestión de Datos de Investigación

[Cuadro comparativo]



Laboratorio de Información. CAICYT – CONICET

Antecedentes: Políticas en las Agencias de Financiamiento

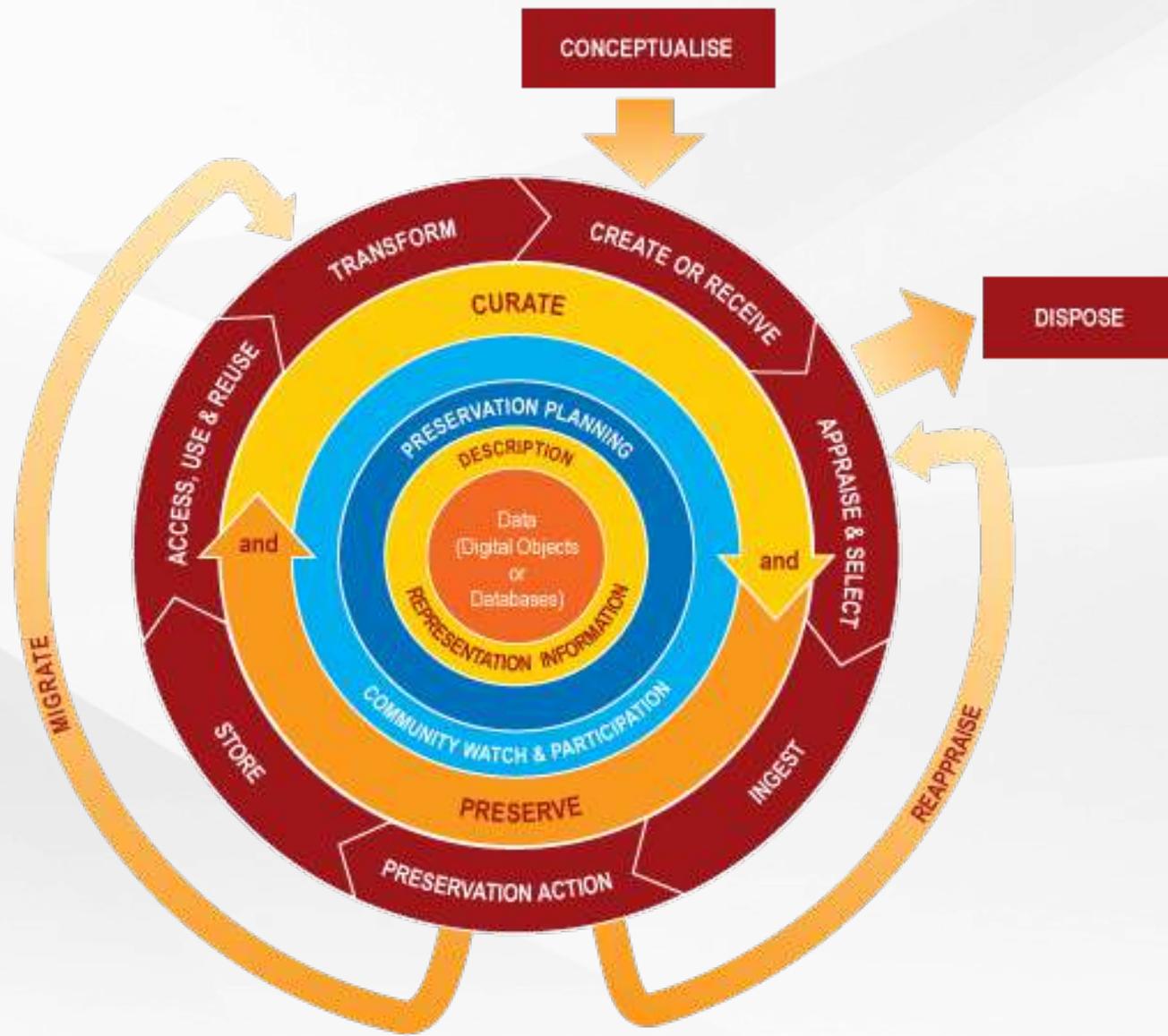
- **Estados Unidos:** NSF (National Science Foundation), NIH (National Institutes of Health), NASA y NODC (National Oceanographic Data Center).
- **Europa:** Horizon2020
- **Reino Unido:** AHRC, BBSRC, Cancer Research UK, EPSRC, ESRC, MRC, NERC, STFC, WellcomeTrust.
- **Australia:** ARC
- **Regional:** OCDE
- **Argentina:**
 - Iniciativas Nacionales de Datos
<http://sistemasnacionales.mincyt.gob.ar/>
 - **Ley Nacional 26.899.** Datos primarios en 5 años disponibles. Excepciones

CICLO DE VIDA DE LOS DATOS CIENTÍFICOS

1. Diseño y planificación de la creación/extracción de datos
2. Creación/Extracción de Datos
3. Limpieza, Normalización y Descripción de Datos
4. Almacenamiento y Preservación de Datos
5. Exploración, Explotación y Visualización de Datos
6. DataMining & Knowledge Discovery



Fuente: [infografía REBIUN - CRUE](#)



Análisis de las actuales Políticas y Requisitos
de importantes Agencias de Financiamiento:

- **DCC (UK)**
- **Horizon2020 (UE)**
- **NSF (EEUU)**
- **ARC (AU)**

LEGISLACIÓN (Argentina)

- **RESOLUCIÓN 538 de JFG (junio 2013)**
Sistema Nacional de Datos Públicos
- **LEY 26.899 (diciembre 2013)**
Creación de Repositorios Digitales Institucionales de Acceso Abierto, Propios o Compartidos
- **DECRETO 117 (enero 2016)**
Plan de Apertura de Datos
- **Ley Acceso a la Información Pública (2016)**
Acceso a la Información Pública



Observatorio Nacional de la
Degradación de Tierras
y Desertificación

**Diccionario de Datos para el Plan de Gestión
de Datos del Observatorio Nacional de la
Degradación de Tierras y Desertificación
(PGD - ONDTyD)**

Desarrollado por
**Centro Argentino de Información Científica y Tecnológica
del Consejo Nacional de Investigaciones Científicas y Técnicas
(CAICYT – CONICET)**

Financiado por
FUNDACIÓN WILLIAMS

1. ¿Qué es un Plan de Gestión de Datos Científicos (PGD)?
2. Diccionario de Datos para el Plan de Gestión de Datos del Observatorio Nacional de la Degradación de Tierras y Desertificación (PGD – ONDTyD)
3. Despliegue del Plan de Gestión de Datos del Observatorio Nacional de la Degradación de Tierras y Desertificación (PGD – ONDTyD)
4. Buenas prácticas en Formato de Datos
5. Buenas prácticas en la Estructura de Carpetas y Archivos
6. Buenas prácticas en el Control de Versiones
7. Buenas prácticas de Metadatos

DMP Argentina (CAICYT-CONICET)

Propuesta para Proyecto de Investigación:

- Bloque de Datos administrativos
- Bloque de Recolección de datos
- Bloque de Documentación y metadatos
- Bloque de Almacenamiento y copias de seguridad
- Bloque de Selección y preservación
- Bloque de Re-uso de Datos

EJEMPLO de los Planes de Gestión de Datos (PGD) desarrollados por el CAICYT – CONICET (Argentina):

- Observatorio Nacional de Degradación de Tierras y Desertificación (ONDTyD)
- Convocatoria a proyectos estratégicos del CONICET (julio 2016)

ENCUESTA DEL PILOTO - PGD CONICET

JULIO 2016

LANZAMIENTO DE UN PILOTO PGD CONICET

42 PROYECTOS UE INVITADOS

20 PROYECTOS UE VOLUNTARIOS

OBJETIVO

Evaluar el contenido del PGD

Conocer el tratamiento de los datos

Relevar el interés y las necesidades de los investigadores, agencias de investigación y financiadores

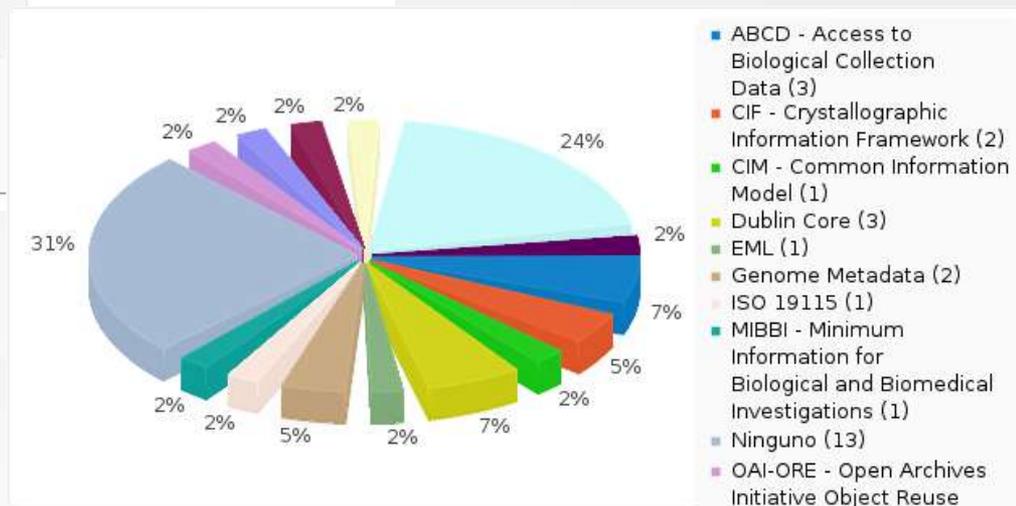
ALGUNOS RESULTADOS...

ENCUESTA DEL PILOTO - PGD CONICET

TIPOS DE DATOS

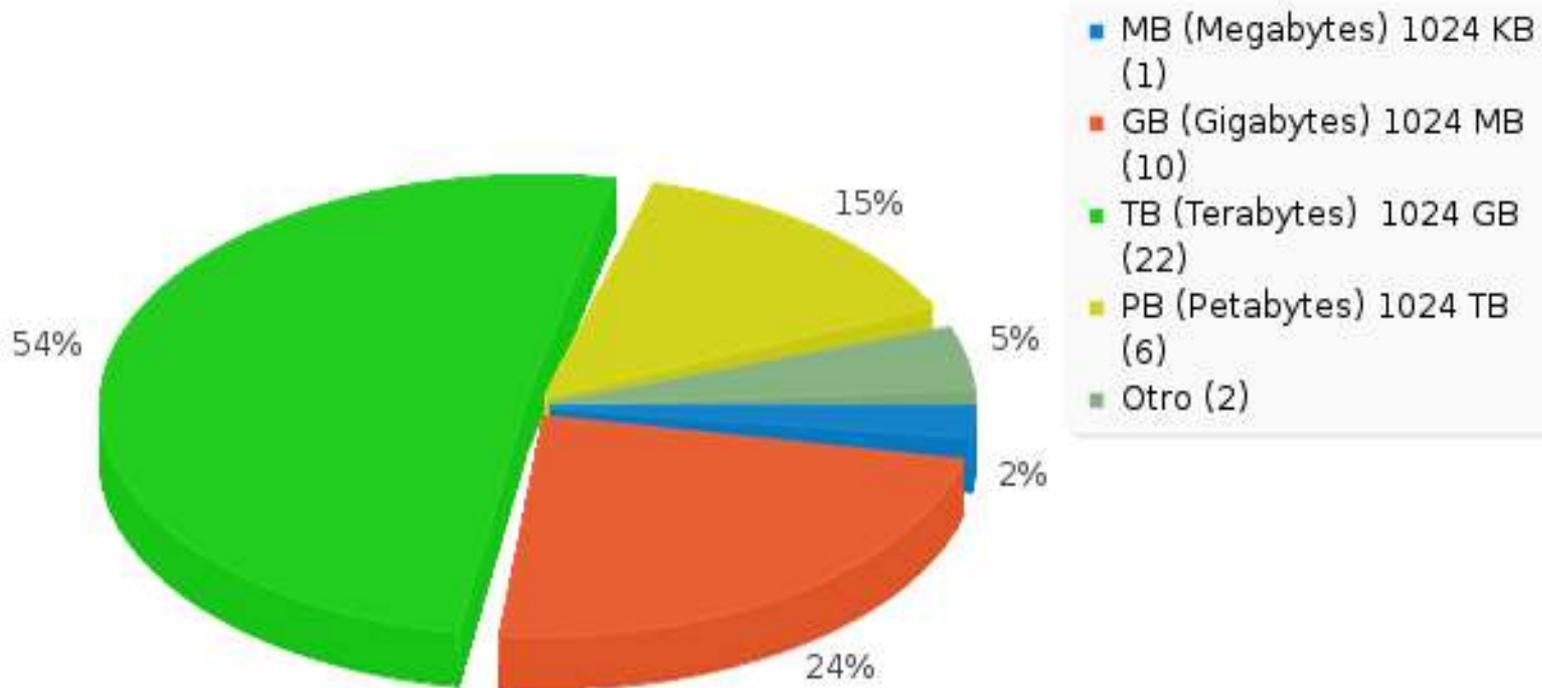


USO DE ESTANDARES



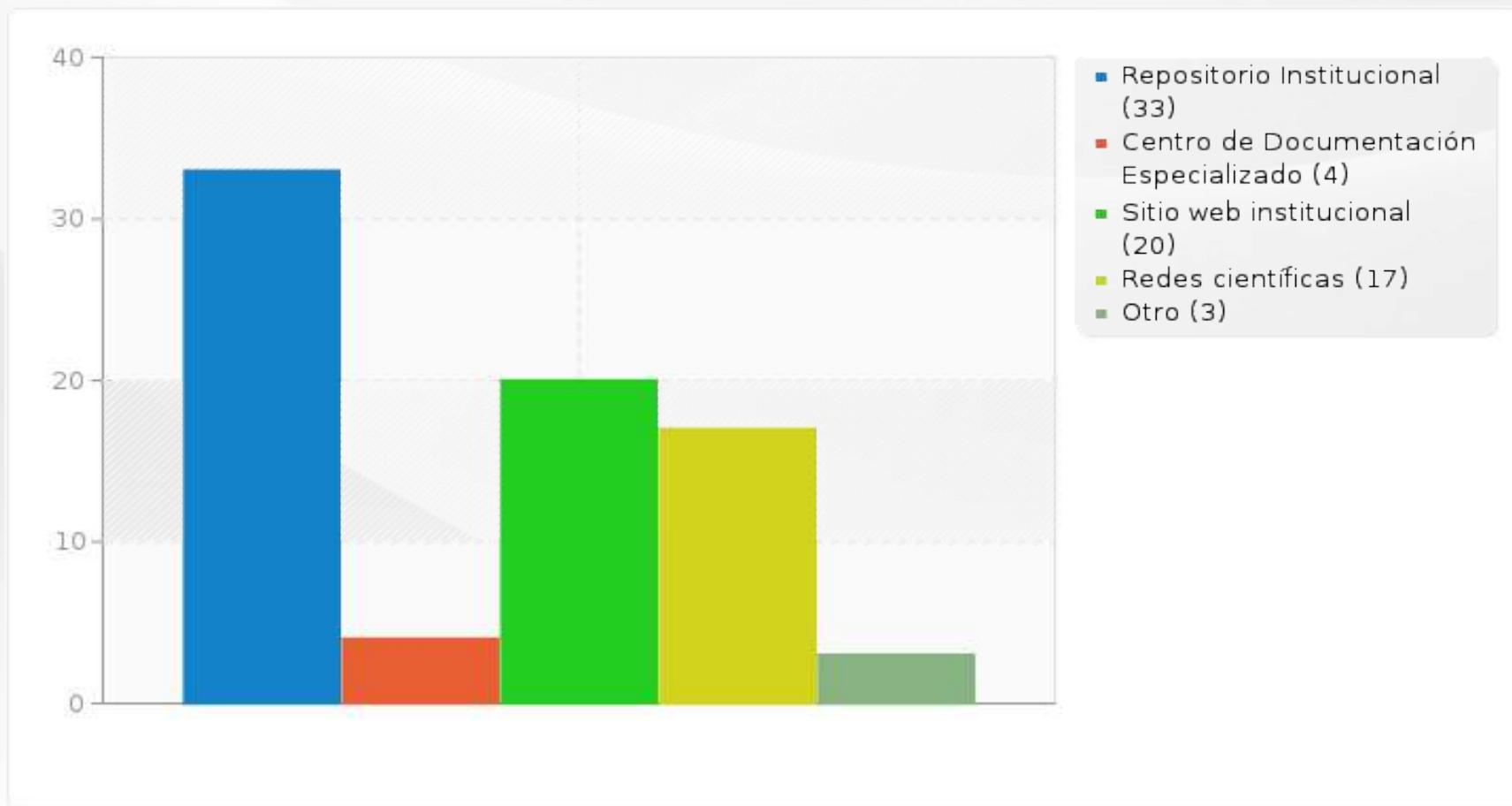
ENCUESTA DEL PILOTO - PGD CONICET

VOLUMEN



ENCUESTA DEL PILOTO - PGD CONICET

DONDE PONDRÁ ACCESIBLE SUS DATOS



ENCUESTA DEL PILOTO - PGD CONICET

75% CONSIDERA APROPIADA LA ORGANIZACIÓN del CONTENIDO

60% ASUMEN COMO NECESIDAD LA GESTION de un PGD

50% REQUIEREN SOPORTE O CAPACITACIÓN

50% REQUIERE MÁS INFORMACIÓN ACERCA del USO DE

ESTÁNDARES Y ESQUEMAS DE METADATOS

Mayor resistencia: **Cambio Cultural**



Los investigadores pueden ser reacios a compartir sus datos públicamente debido a los costos individuales reales y/o percibidos.



ACTORES IMPLICADOS

- Investigadores [productores de datos]
- Agencias de Financiamiento
- Universidades y Centros de Investigación
- Gestores de datos:
 - Repositorios Institucionales (corto plazo)
 - Centro de Datos (mediano-largo plazo)
- Usuarios (otros investigadores, ciudadanos, etc.)

Sociedad Civil (ONG) + Empresas

Comunidad Científica:

- **Acordar estándares de trabajo, normalización y gestión de datos**
- **Empoderar en Gestión de Datos Científicos**
- **Contribuir al avance de la ciencia, maximizando la visibilidad y el impacto de la inversión en Investigación**
- **Financiamiento y Desarrollo de e-infraestructuras para el aprovechamiento de la Big Data Científica**
- **Encuentro con otras comunidades de Gobierno Abierto, Periodismo de Datos, HackActivismo Cívico, IoT, etc.**
- **Repensar formas de producción, medición e impacto científico**

MUCHAS GRACIAS



Este obra está bajo una [licencia de Creative Commons Reconocimiento 4.0 Internacional](#).

Fernando Ariel López
(CAICYT – CONICET, Argentina)

flopez@conicet.gov.ar | [@fernando_lopez](https://www.instagram.com/fernando_lopez)