



VII Conferencia **Latinoamericana y Caribeña** de Ciencias Sociales

Transformaciones democráticas, justicia social y procesos de paz | 9 al 13 - Noviembre 2015 - Medellín, Colombia

 XXV Asamblea General de CLACSO

Plan de Gestión de Datos Científicos, una propuesta Argentina.



Fernando Ariel López
CAICYT – CONICET
@fernando_lopez

OPEN SCIENCE



Big Data: Volumen, Velocidad, Variedad y Veracidad
Explotación de Datos y Descubrimiento del Conocimiento
(Data Mining & Knowledge Discovery)

OPENNESS

Open Source

Las 4 libertades del Software Libre

- Libertad 0: Usar el programa para sus propios fines, sin restricciones.
- Libertad 1: Estudiar el código fuente y hacer las modificaciones necesarias.
- Libertad 2: Compartir el programa con quien se quiera.
- Libertad 3: Mejorar y modificar el programa, y publicar las mejoras.

OPEN DATA

- Gobierno Abierto
- Periodismo de Datos
- Ciencia Abierta



Open Hardware

Arduino
Raspberry
FRESADORAS
IMPRESORAS 3D
Balloon mapping

OPEN GOV

Open Government

Transparencia
Participación
Accesibilidad
Innovación



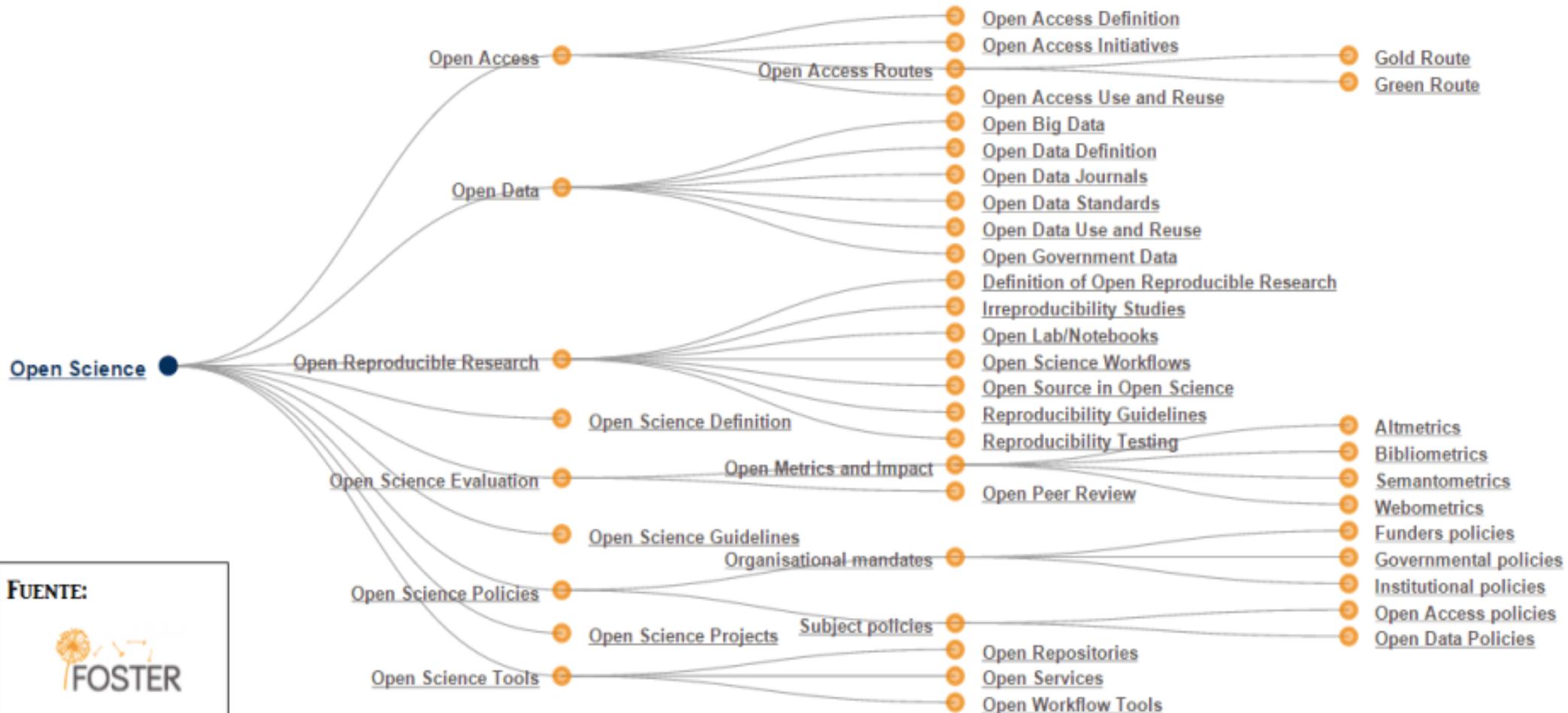
Revistas OA
Repositorios Digitales
Legislación
Comunicación Científica
Impact
Altmetric

Educación Abierta



OER / REA **MOOC** OWC

Ciencia Abierta = Open Science



FUENTE:



Principios de la CIENCIA ABIERTA



- 1) **Metodologías Abiertas** / Open Methodology (Métodos, procesos, documentos relevantes, etc.)
- 2) **Software y Hardware Libres** / Soft and Hard open
- 3) **Datos Abiertos** / Open Data (libres para reutilizar)
- 4) **Acceso Abierto** / Open Access (libre y gratuito)
- 5) **Revisión por pares Abierto** / Open Peer Review (transparencia en los criterios de evaluación y calidad)
- 6) **Recursos Educativos Abiertos** / Open Educational Resources (MOOC y REA)

¿Qué son los datos de investigación?

Los datos incluyen:

- **cuadernos de laboratorio o de campo,**
- **datos de investigación primaria (en papel o digital),**
- **cuestionarios,**
- **fotografías, audio y videos**
- **desarrollo de modelos**

Las colecciones de datos para la investigación pueden incluir:

- **diapositivas, diseños, muestras.**

Procedencia de los datos:

cómo, cuándo, donde se recogió y con qué (por ejemplo, instrumentos).



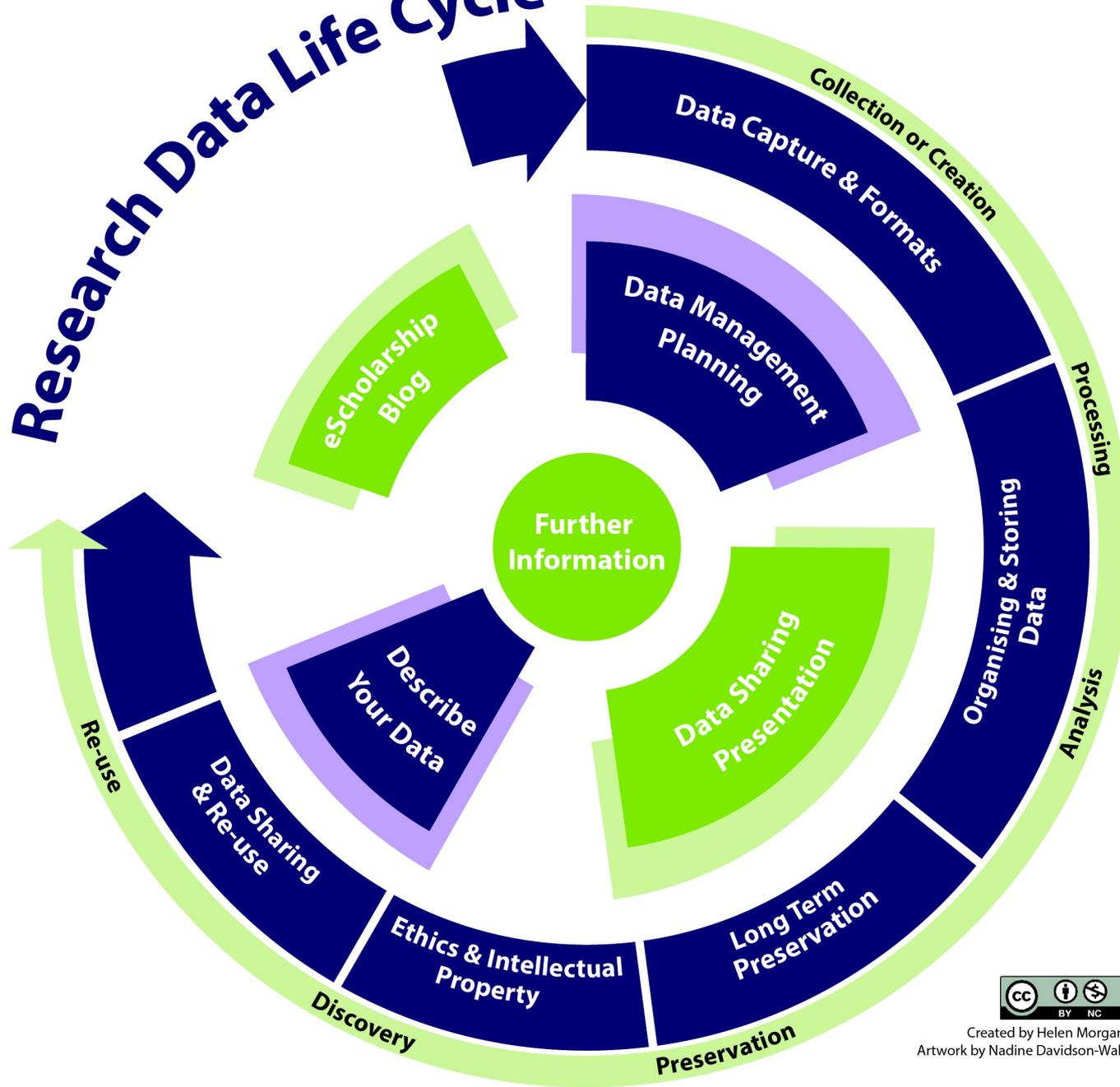


DATASET

Es el objeto específico de control, organización, descripción y preservación de datos científicos

- Es una colección de datos reunidos durante la ejecución de un proyecto de investigación.
- Son objetos digitales compuestos y heterogéneos.
- Constituye la base de la investigación y va asociado a una publicación científica (resultado de la investigación).
- Se almacena y gestiona en Repositorios Interoperables conforme a estándares internacionales.

Research Data Life Cycle



Created by Helen Morgan
Artwork by Nadine Davidson-Wall



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.)

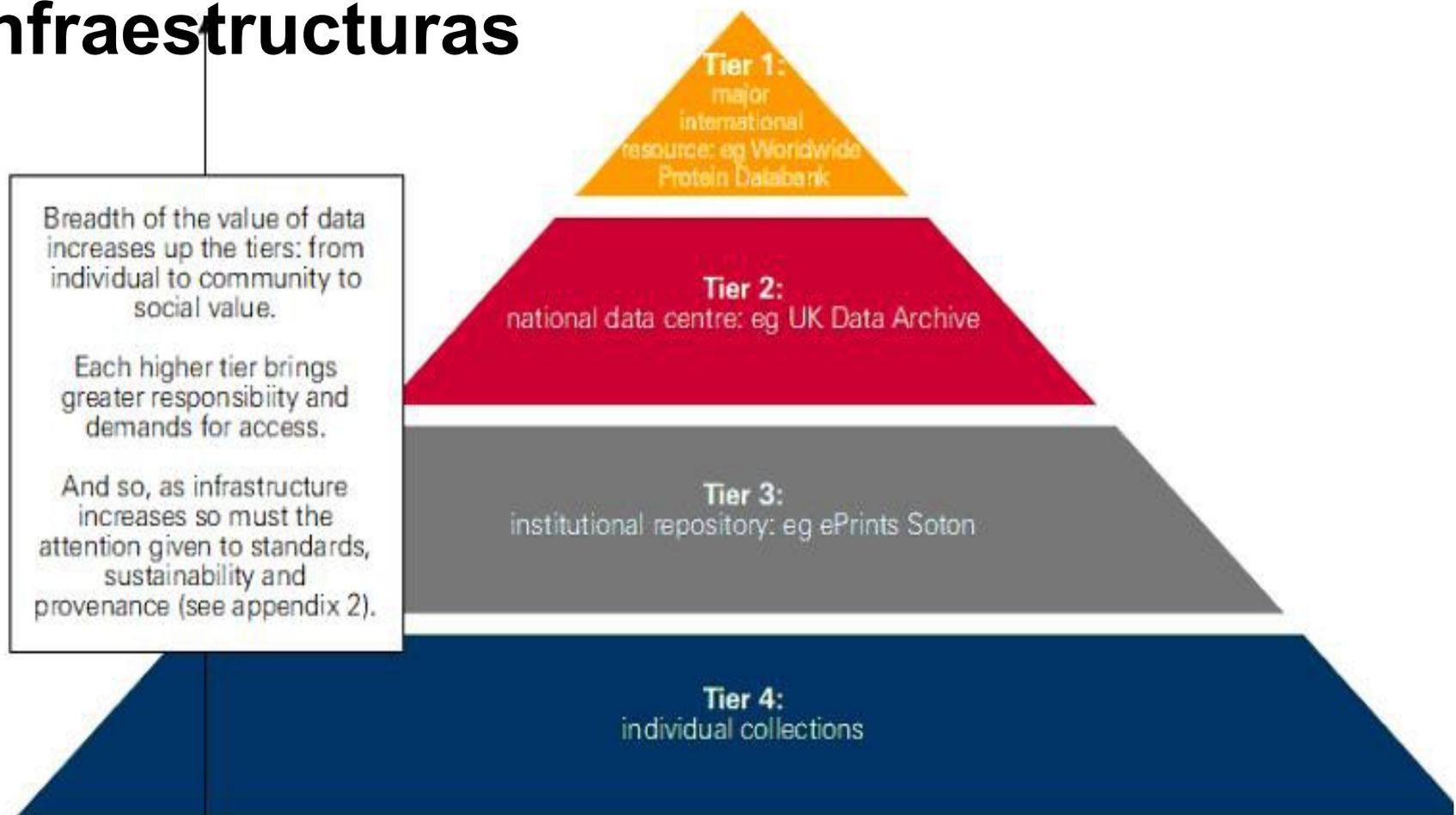
Sociedad Civil (ONG) + Empresas

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

Líneas de Trabajo

- **Plan de Gestión de Datos** / Data Management Plan (DMP)
- **e-Infraestructuras**





Centro Argentino de Información Científica y Tecnológica

- Es un *Centro de Servicios e Instituto de Investigación* perteneciente al Consejo Nacional de Investigaciones Científicas y Técnicas (**CONICET**) de **Argentina**.

Objetivos:

- Contribuir a la comprensión del desarrollo, evolución y transferencia del conocimiento mediante la investigación de la información publicada en ciencia y tecnología.
- Contribuir a la organización y acceso de la Información científico-tecnológica y a la calidad de su difusión a través de de las publicaciones Científico Tecnológicas nacionales.

Centro Argentino de Información Científica y Tecnológica

Linea de Trabajo Datos Científicos:

- Desarrollo de una plataforma y validador para apoyar a instituciones productoras de datos en el desarrollo de un Marco de Gestión de Datos (DMP).
- Marco de verificación de calidad de metadatos e infraestructura de datos científicos primarios.
- Formación en las necesidades y competencias de los Científicos de Datos, Editores Científicos y Bibliotecarios de Datos.



PLAN DE GESTIÓN DE DATOS [DMP]

Es un documento que describe el tratamiento que van a recibir los datos de investigación generados o recopilados en el transcurso de un proyecto de investigación.

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

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

DETALLES SOBRE:

- Proyecto / Institución / Datos administrativos
- Plan
- Datos
- Derechos y Licencias
- Preservación y Conservación Digital

Propuesta **CAICYT - CONICET:**

- 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

CAICYT- CONICET (Argentina)

Propuesta: Plan de Gestión de Datos Científicos

Bloque de Datos administrativos

- Lineamiento Generales de datos

Bloque de Recolección de datos

- Datos que se van a recolectar o crear
- Metodologías de Recolección y/o creación de datos

Bloque de Documentación y metadatos

- Documentación y metadatos que acompañan los set de datos
- Derechos de autor y derechos de propiedad intelectual

Bloque de Almacenamiento y copias de seguridad

- Almacenamiento y copias de seguridad durante la investigación
- Gestión del acceso y la seguridad

Bloque de Selección y preservación

- Retención compartido y preservación
- Preservación a largo plazo del set de datos

Bloque de Re-uso de Datos

- Compartido de datos
- Restricciones a los datos compartidos

Desarrollo:

- **Plataforma web**, basada en el desarrollo de la DCC (UK)
- **Plan de Gestión de Datos**, distintos momentos:
 - Al pedir Financiamiento
 - Al inicio del Proyecto, Revisión Intermedia y Revisión Final
- **Compartir** (incluir a los colaboradores y sus roles)
- **Exportar** (XML, HTML y PDF)

Por favor, rellene los detalles básicos del proyecto abajo y haga clic en 'Actualizar' para guardarlos

Cancelar

Nombre de proyecto	<input type="text" value="Mi proyecto (PIO. CONICET)"/>
ID	<input type="text" value="0029980"/>
Número de la subvención	<input type="text" value="001987/15"/>
Investigador principal	<input type="text" value="Fernando Ariel López"/>
ID del Investigador Principal	<input type="text" value="0000-0002-2467-3627"/>
Datos del contacto del proyecto	<input type="text" value="flopez@caicyt.gov.ar"/>
Descripción	<input type="text"/>

Preguntas a considerar:

- ¿Cual es la naturaleza de su proyecto de investigación?
- ¿Qué temas de investigación está tratando?
- ¿Para qué propósito se están recogiendo o creando los datos?

Orientación:

Resuma brevemente el tipo de estudio (o estudios) para ayudar a otros a comprender los propósitos para los cuales se están recogiendo o creando los datos.

Actualizar

Cancelar

Para cada dataset, especifique lo siguiente (5 questions, 0 answered)

El PGD debe centrarse en los puntos siguientes dataset por dataset y debe reflejar el estado actual de la reflexión sobre los datos que se producirán

Nombre y referencia del dataset

Orientación

Se creará un identificador para el dataset

Guardar

Aún no respondido/a

Descripción del dataset

B *I*    

Orientación

Descripción de los datos que serán generados o recolectados, su origen (si son recolectados), naturaleza, escala, por qué serán útiles y si respaldan a una publicación científica. Información de la existencia (o no) de datos similares y sus posibilidades de integración y reutilización.

Guardar

Aún no respondido/a

Estándares y metadatos

B *I*    

Orientación

Haga referencia a estándares existentes y apropiados para su disciplina. Si no existen, indicar un boceto de qué metadatos serán creados y cómo se crearán.

Mayor resistencia: Cambio Cultural



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

BENEFICIOS #DatosAbiertos

- Ayuda a **verificar los resultados.**
- **Evitar la fabricación y falsificación de datos.**
- **Diferentes interpretaciones o enfoques** aplicados a datos existentes contribuyen a los avances científicos.
- **Optimización en el uso de recursos.**
- **Preservación a largo plazo bien gestionada,** permite mantener la **integridad de los datos.**

TenopirC, Allard S, Douglass K, AydinogluAU, et al. (2011) Data Sharing by Scientists: Practices and Perceptions. PLoS ONE 6(6): e21101. doi:10.1371/journal.pone.0021101
<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0021101>

Debemos desarrollar Políticas y Acciones para Gestionar los Datos Científicos:

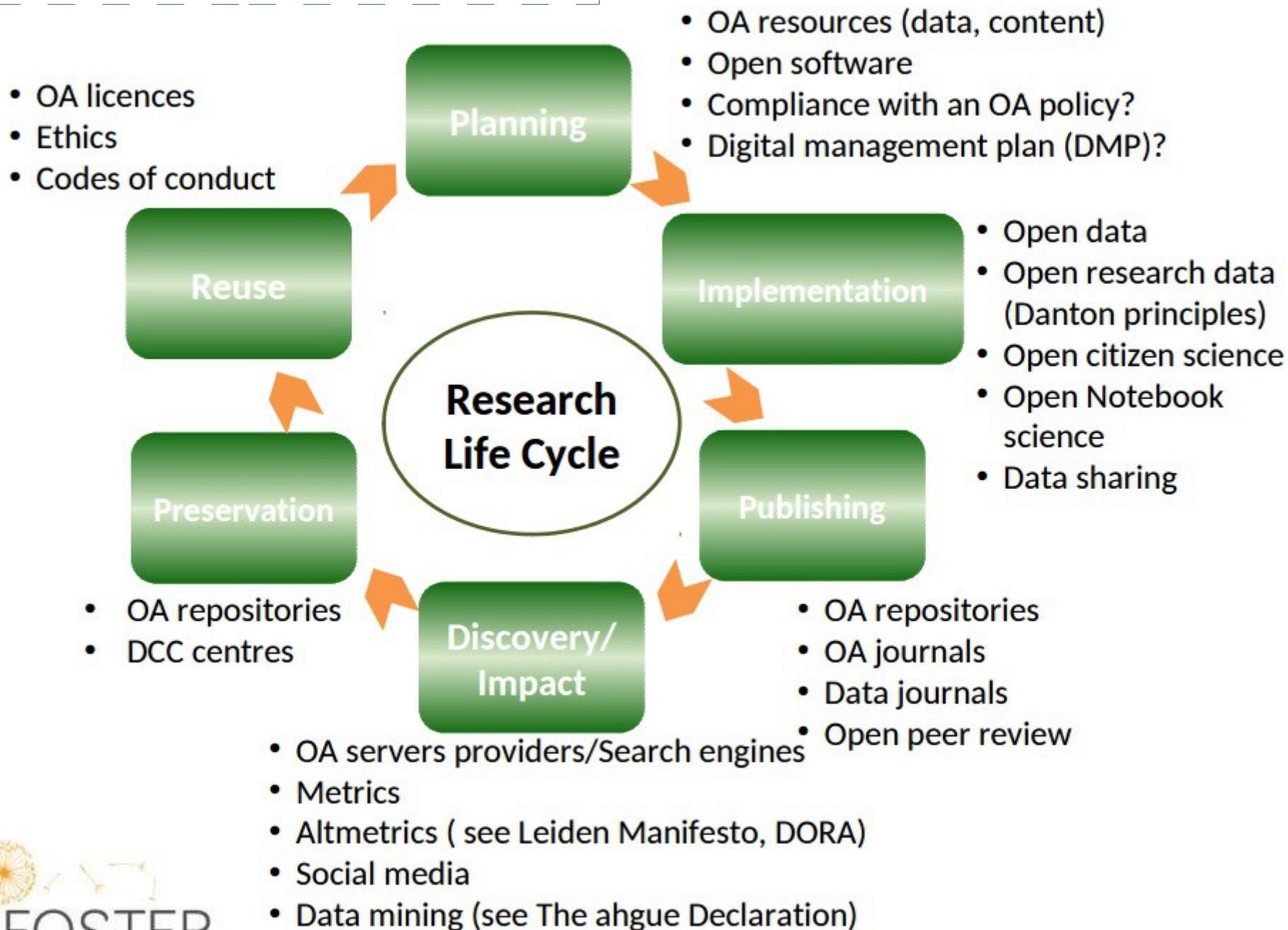
- **Políticas** a nivel de agencias de financiación e institucionales. Definición de roles/responsabilidades de los distintos actores.
- **Recursos financieros** a largo plazo ya que los datos son acumulativos y se preservan.
- **Recursos humanos** especializados (para generación de datos, normalización, explotación y preservación).
- **Infraestructuras** coordinadas para garantizar su interoperabilidad. Entre los requisitos de las infraestructuras destacar: *preservación, acceso, data curation, data processing, distribución.*

¿y los Investigadores?

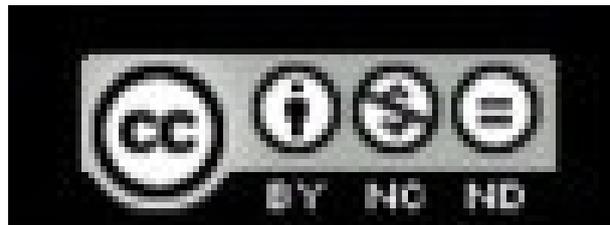
- Colaborar con su **formación**
- Proveer una **infraestructura de trabajo** (framework):
 - Almacenamiento de datos
 - Curación, Procesamiento, Explotación, Visualización de datos.
- Contar con **Repositorios / Centros de Datos**:
 - Identificadores únicos (Autor, Datos, etc.)
 - Estadísticas de impacto (uso, descarga, citación, etc.)
 - Preservación Digital

Ciclo de Vida de los Datos

ABIERTO



Muchas Gracias



CAICYT

CONICET

Fernando Ariel López
CAICYT – CONICET
@fernando_lopez