



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



Mirando al futuro, del acceso abierto a los datos abiertos de investigación



Fernando Ariel López
CAICYT – CONICET

OPENNESS

Open Source

Las 4 libertades del Software Libre

- Libertad 0: Usar el programa para cualquier propósito sin restricciones.
- Libertad 1: Estudiar y mejorar o adaptar el software para satisfacer sus necesidades.
- Libertad 2: Compartir el programa con otros para que también lo usen.
- Libertad 3: Mejorar y modificar el programa para mejorar su funcionamiento.

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



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



Openness → Open Science

Open Access + Open Data + Open Source
+ Innovación + Colaboración Abierta



OPEN DATA
DATOS
ABIERTOS

Gobierno Abierto
Tercer Sector + Ciudadano
Periodismo de Datos

Construcción de datos científicos: tipos

- **Experimentales:** datos provenientes de resultados experimentales
Ej: Aquellos que provienen de aparatos de medición en laboratorios, comúnmente reproducibles, pero caros.
- **Simulación:** datos generados de modelos de prueba donde el modelo y los metadatos pueden ser mas importantes que los datos de salida del modelo.
Ej: Modelos económicos o climáticos.
- **Desarrollados o compilados:** resultado de procesar y/o combinar datos “crudos”, comúnmente reproducibles pero caros.
Ej. Bases de datos compiladas, Resultados de text mining, Datos de censos consolidados.
- **Reference or canonical:** Una (estática u orgánica) conglomeración o colección de datasets mas pequeños (revisados por pares), la mayor parte de ellos publicados y “curados”
Ej. Bancos de datos genéticos, bases de datos cristalográficas.

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.

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>

Una gestión adecuada de los datos requiere al menos los siguientes aspectos:

- **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.*

Para dar respuesta a estos aspectos es necesaria una formación adecuada, equipamientos, sistemas de almacenamiento masivo de datos y redes de alta capacidad.

Mayor resistencia: **Cambio Cultural**



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

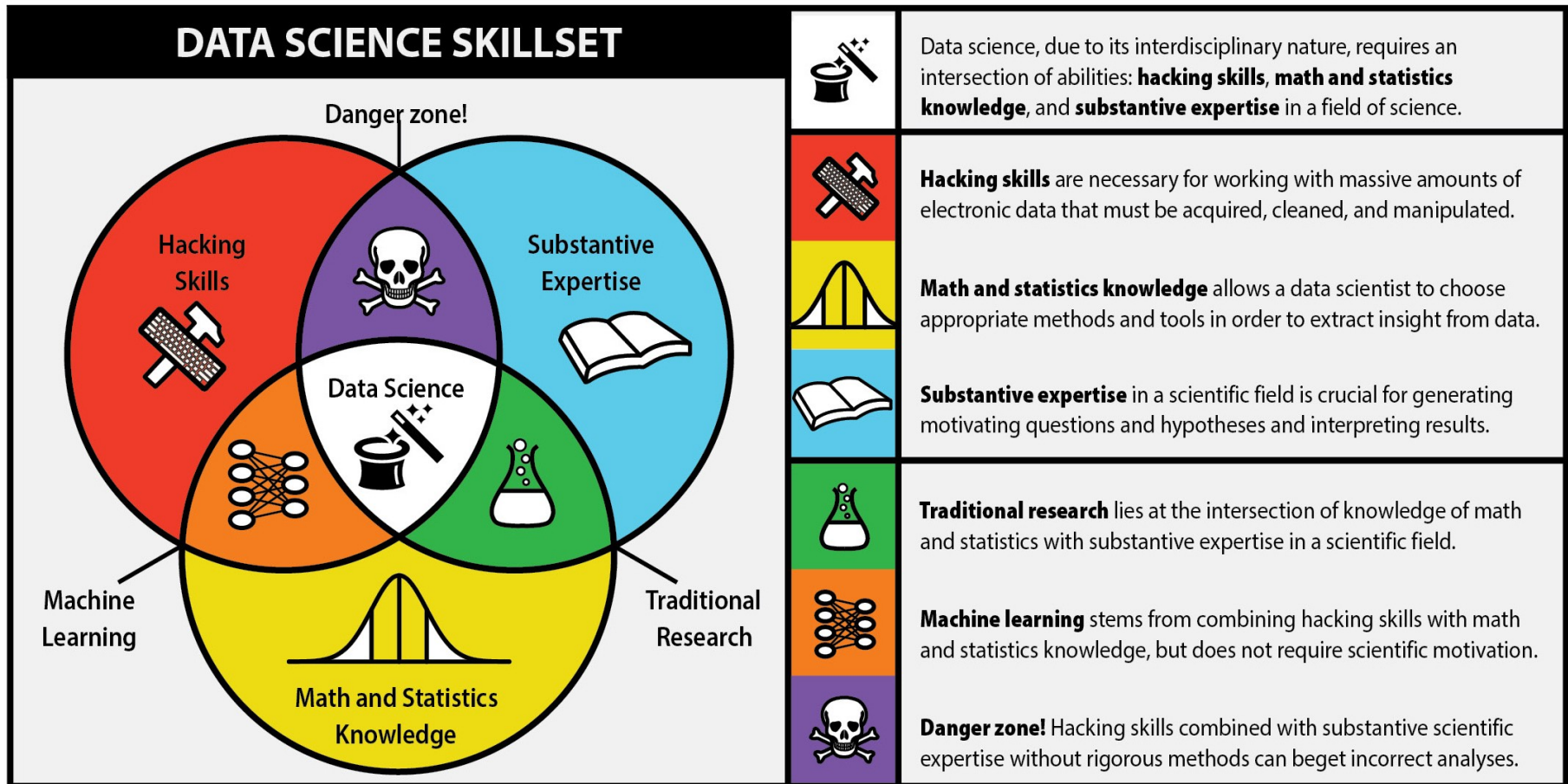


ACTORES IMPLICADOS

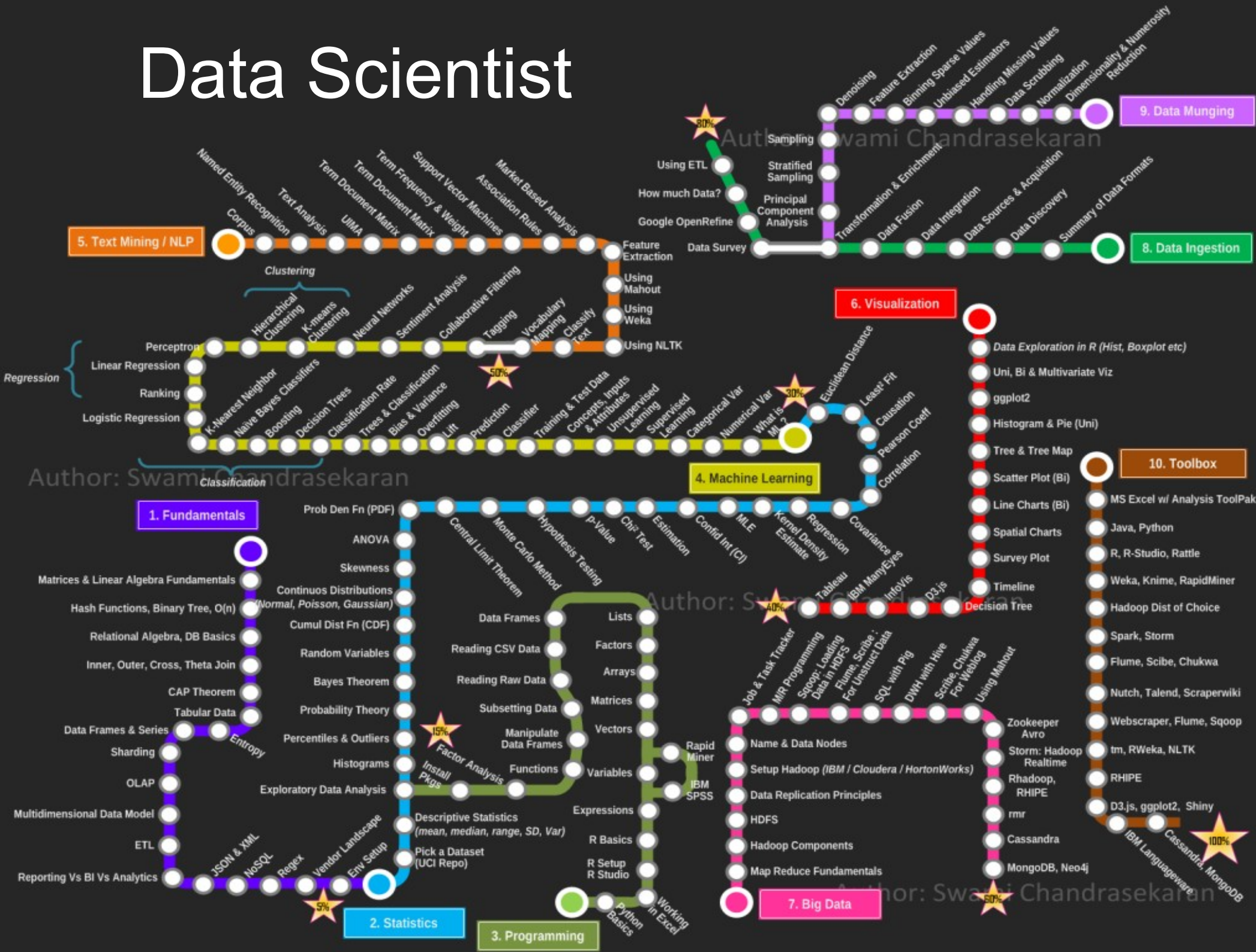
- Investigadores / productores de datos
- Universidades y Centros de Investigación
- Repositorios Institucionales (corto-mediano plazo)
- Centro de Datos (largo plazo)
- Gestores de datos
- Usuarios que reutilizan los datos
- Agencias de financiación

Tercer sector / Sociedad Civil

Científico de Datos: nuevos conocimientos y competencias



Data Scientist

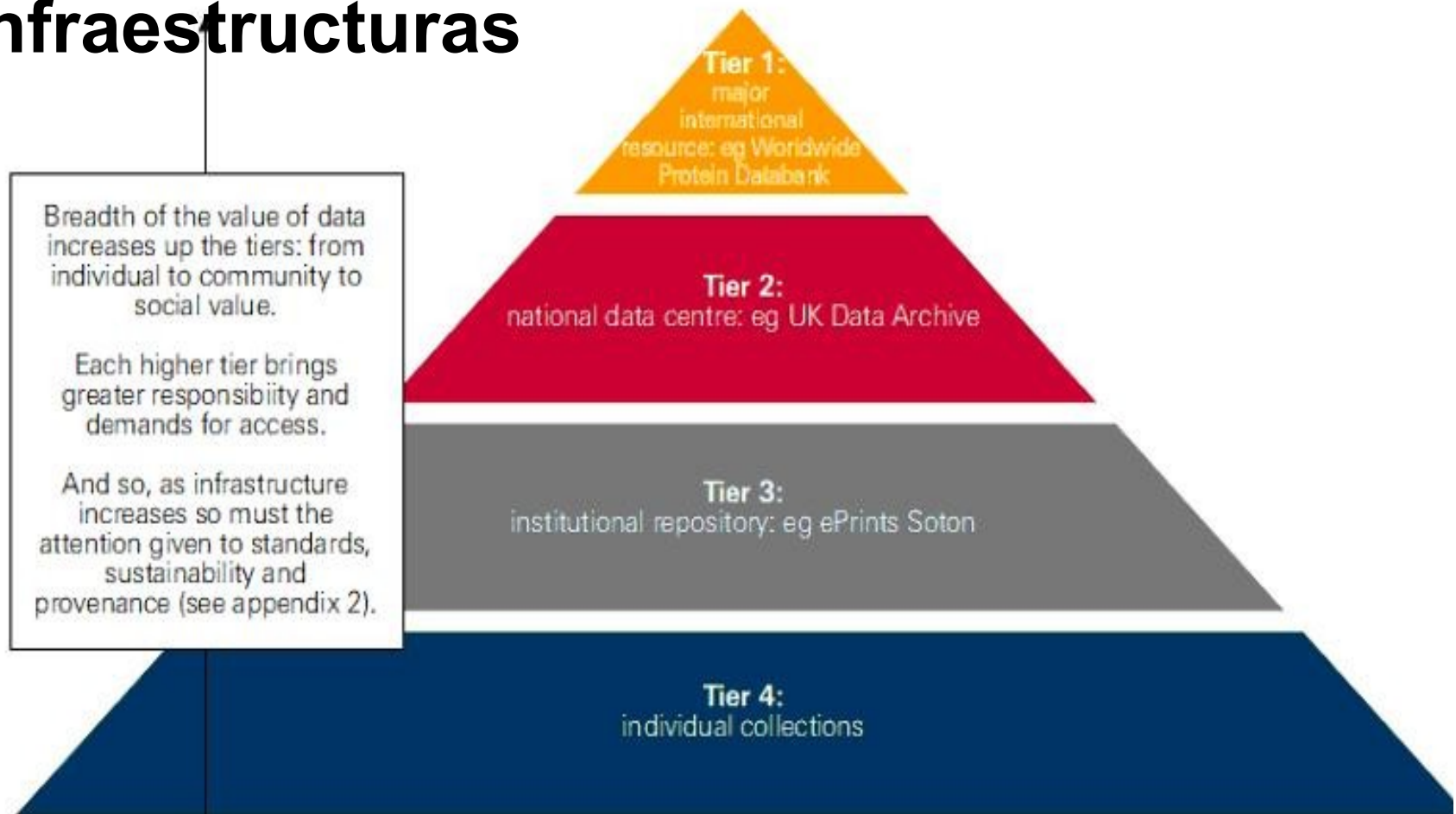


Antecedentes Políticas en las Agencias de Financiamiento

- **Estados Unidos:** National Science Foundation (NSF), National Aeronautics and Space Administration (NASA), National Oceanographic Data Center (NODC) y National Institutes of Health (NIH).
- **Europa:** Horizon2020
- **Reino Unido:** AHRC, BBSRC, Cancer Research UK, EPSRC, ESRC, MRC, NERC, STFC, WellcomeTrust.
- **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**



Plan de Gestión de Datos (DMP)

- **Referencia y nombre del set de datos**
- **Descripción del set de datos**
- **Estándares y metadatos**
- **Datos compartidos**
- **Archivo y preservación**
(incluyendo almacenamiento y copias de seguridad)
- **Formatos**
- **Metadatos**
- **Identificador digital de datos**
- **Marco legal relacionado con la gestión y divulgación de datos de investigación**
 - Acceso y datos
 - Privacidad y confidencialidad
 - Propiedad Intelectual y datos
 - Depósito de los datos
 - Licencias alternativas copyright
- **Preservación**

Horizon2020 (UE)

FECYT (España)

e-Infraestructura

- **Repositorio Interoperable de Datos**
 - Datos Públicos Argentina (CKAN, OKF),
 - Zenodo (Invenio, CERN)
 - DRYAD, PLICSS, LAGOS (DSpace, MIT & HP),
 - Harvard DATAVERSE (Eprints)
- **Plataforma de Trabajo para Investigadores**
 - OSF, HubZero, MyExperiment.org, etc.
- **Cluster de Almacenamiento y/o Procesamiento**

¿Preguntas, Dudas o Consultas?

Muchas Gracias



CAICYT

CONICET

Fernando Ariel López
CAICYT – CONICET

flopez@caicyt.gov.ar

[@fernando_lopez](https://twitter.com/fernando_lopez)