

LA Referencia

Red de repositorios de acceso abierto a la ciencia

Ciclo Virtual sobre Gestión de Datos Científicos

ALBERTO CABEZAS BULLEMORE

SECRETARIO EJECUTIVO LA REFERENCIA

Cabezas.alberto@redclara.net

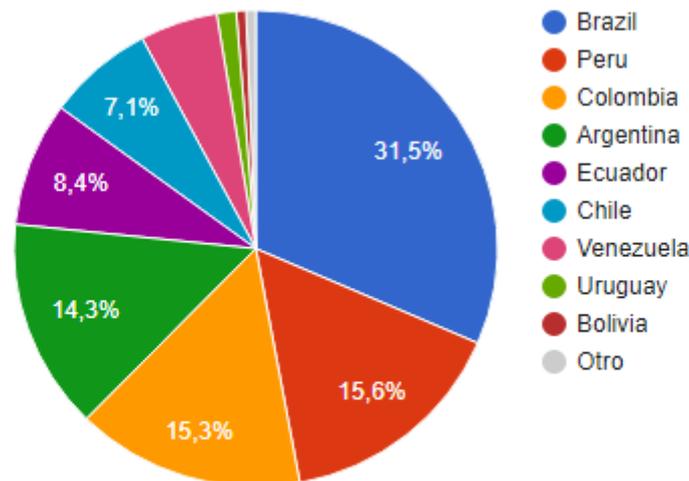
Temas

- **Introducción**
- **LA Referencia: Entorno y bienes públicos**
- **Directrices para repositorios de datos y publicaciones**
- **Ciencia Abierta**

Marco

- Las Universidades fueron pioneras en el uso de repositorios.
- Responde al ciclo de la innovación y a la misión.
- Visibilizar , preservar, compartir

Proportion of Repositories by Country - South America



Total = 308 repositories

OpenDOAR - 28-Nov-2017

LA Referencia: Principio

- El **objetivo final es simple**: Ofrecer gratuitamente y en acceso abierto a **toda la ciudadanía** la producción científica de excelencia regional con énfasis en aquella financiada total o parcialmente con fondos públicos.
- (La I+D es financiada mayoritariamente con fondos públicos en América Latina a diferencia de los países desarrollados y parte relevante de los resultados como los artículos tienen precio)
- **Misión**: Ofrecer la producción científica de América Latina, financiada con fondos públicos, mediante la cooperación y articulación de **una red federada de repositorios institucionales**, basada en acuerdos regionales y **estrategias nacionales** de acceso abierto.

Somos

- Cooperación Latinoamericana de Redes Avanzadas. RedCLARA
- Ministerio de Ciencia, Tecnología e Innovación Productiva, Mincyt; **Argentina**
- Instituto Brasileiro de Informação em Ciência e Tecnologia, Ibict; **Brasil**
- Colciencias, Ministerio de Educación, Renata; **Colombia**.
- Comisión Nacional de Investigación Científica y Tecnológica, Conicyt; **Chile**.
- Secretaria de Educación Superior, Ciencia y Tecnología, Senescyt; CEDIA. **Ecuador**.
- Viceministerio de Ciencia y Tecnología, Ministerio Educación; CBUES. **El Salvador**
- Consejo Nacional de Ciencia, Tecnología, Conacyt. Remeri; **México**
- Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica, Concytec; **Perú**
- Conare – Ministerio de Ciencia, Tecnología y Telecomunicaciones. **Costa-Rica**

Servicio 2014/2016

- **Sostenibilidad** / por cuotas de asociados desde fines del 2013.
- **Consolidación** de Consejo Directivo y Grupos Técnicos.
- **De piloto a plataforma de servicios.** De 3 países con cosechas a toda la red. De 320 mil registros a 1.4 millones.
- **Criterios** de calidad, generación proyectos y definición prioridades.
- **De definición** de directrices de metadatos a implantación de los nodos nacionales

9
NODOS NACIONALES

1 431 703
DOCUMENTOS

799 411
ARTÍCULOS
7 107
REPORTES

187 350
TESIS DE DOCTORADO
437 835
TESIS DE MAESTRÍA

Modelo



DESARROLLAMOS TECNOLOGÍA DE
COSECHADORES



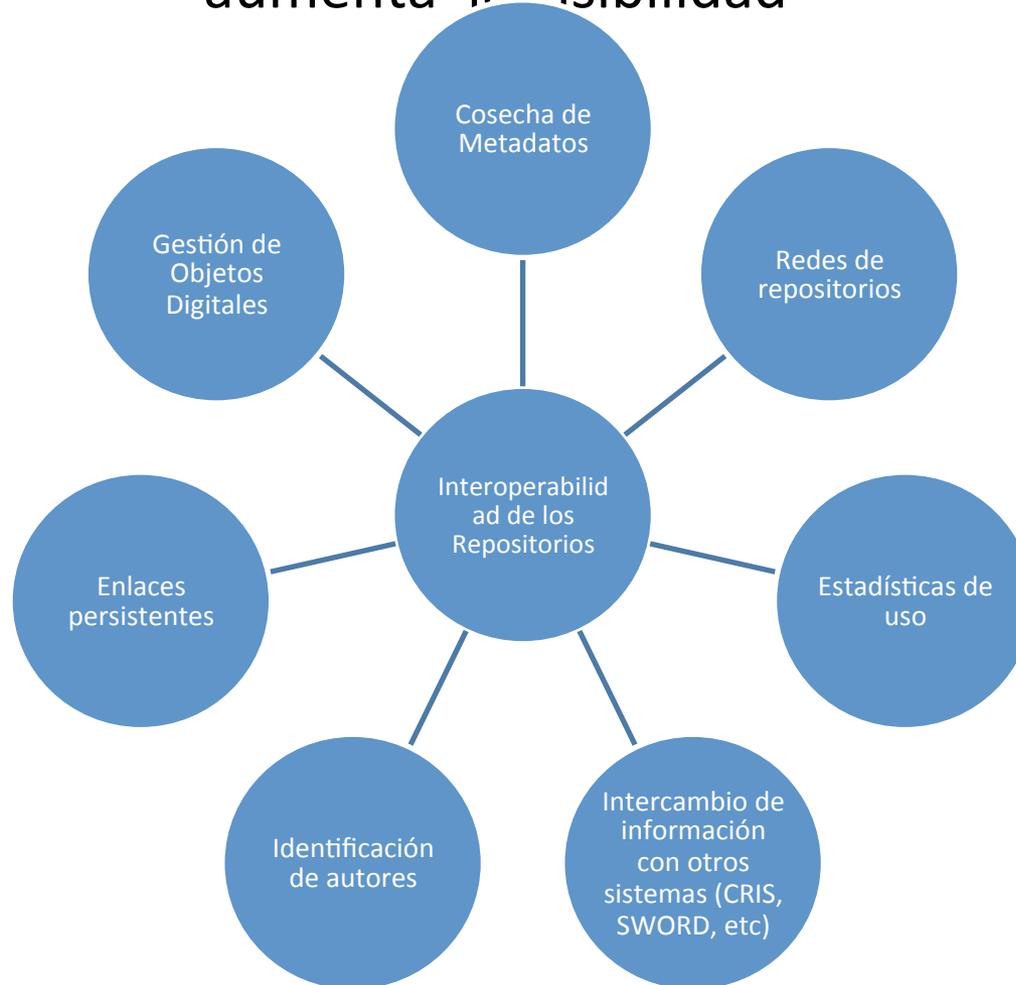
OFRECEMOS DIRECTRICES Y
RECOLECCIÓN DE METADATOS



IMPULSAMOS ACUERDOS DE
INTEROPERABILIDAD

Interoperabilidad

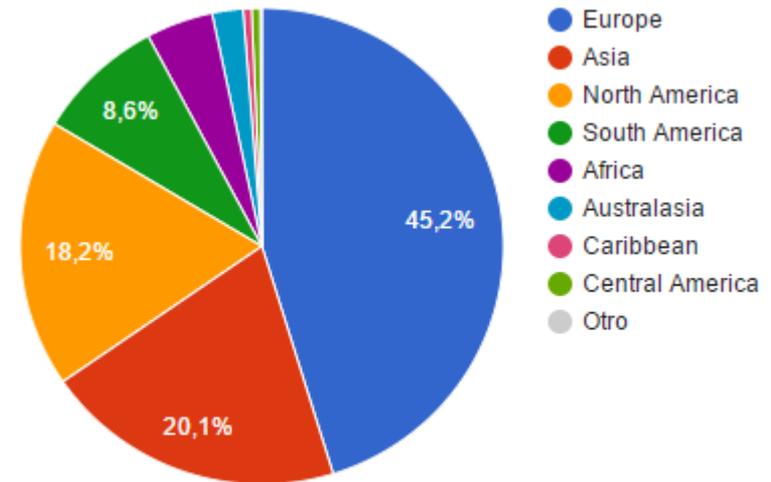
...un repositorio puede ser valioso, pero interconectado
aumenta la visibilidad



MODELO

- Existe un **modelo de acceso abierto** iniciado hace más de 15 -20 años
- Colaborativo
- No comercial
- Con apoyo público
- ...SciELO, Redalyc, Latindex...
- La producción en directorios de / más del doble que métricas tradic

Proportion of Repositories by Continent - Worldwide



Total = 3345 repositories

OpenDOAR - 29-May-2017

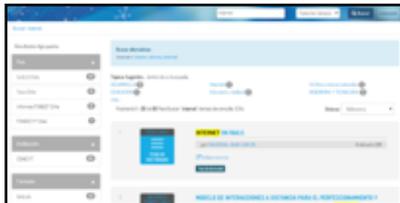
Bienes Públicos

- Estándares no propietarios que eviten la captura por un agente específico.
- Normas comunes (Ejemplo: Directrices).
- Modelo organizativo para la coordinación entre los asociados y acuerdos que benefician colectivamente. (Red federada basada en nodos nacionales).
- Desarrollo tecnológico de código abierto, transferible y que significa economías de escala para cada asociado.
- Servicios que representan economías de escala y pueden transferirse (por ejemplo, estadísticas, pruebas de concepto)
- Servicios que el sector privado no puede ofrecer dado su alto grado de incertidumbre y riesgo, escala, o falta de un mercado que permita la apropiación de los beneficios
- Acciones para que la información científica que fue financiada por fondos públicos, se ofrezca a los ciudadanos de forma gratuita.

Tecnología

LA Referencia

Red de repositorios de acceso abierto a la ciencia



En instalación



En test

Nueva Versión 3.0

- Beta exitoso en Brasil - Sept 2016.
- Reporte de errores para los RI.
- Administración Simplificada.
- Versión LR es la misma que los nodos.
- Soporta directrices actuales y futuras.
- Transferencia 2017.
- **Versión 3.2 Brasil. Sept 2017.**

TRANSFERENCIA



Proyectos



- Proyecto ambicioso con 50 instituciones
- OpenAIRE2020 represents a pivotal phase in the long-term effort to implement and strengthen the impact of the Open Access (OA) policies of the European Commission (EC), building on the achievements of the OpenAIRE projects. OpenAIRE2020 will expand and leverage its focus from (1) the agents and resources of scholarly communication to workflows and processes, (2) from publications to data, software, and other research outputs, and the links between them, and **(3) strengthen the relationship of European OA infrastructures with other regions of the world, in particular Latin America and the U.S.**

- 📍 **Inicio Proyecto 2015 / OBJETIVOS (3)**
- 📍 *“...fortalecer la relación de las infraestructuras europeas de OA con otras regiones del mundo, en particular América Latina y los EE.UU.”*
- 📍 *“...pilotos específicos se llevarán a cabo con CLARA/LA Referencia en América Latina para permitir un modelo y los servicios similares a OpenAIRE”*
- 📍 *“Dada la naturaleza internacional y de colaboración de la investigación, estas redes deben estar conectadas y alineadas en torno a temas tales como políticas, tecnologías y servicios...”*
- 📍 **Directrices comunes, tecnologías como validadores, portal OpenAIRE.**



- 📍 Estadísticas comunes (nuevas métricas)
- 📍 Transición de V. 3.0 a 4.0
- 📍 Servicios de notificación
- 📍 Identificación de resultados..
- 📍 Incremental
- 📍 Próxima generación.

Date: 2017/01/17

 **LAReferencia** – Red Federada de Repositorios
Institucionales de Publicaciones Científicas Latinoamericanas
PUBLICATION REPOSITORY AGGREGATOR



1.259.383 Open Access publications

- 267 EC (H2020/FP7) funded projects outputs identified
- 246 NSF (US) funded projects outputs identified
- 107 FCT (Portugal) funded projects outputs identified
- 56 Wellcome Trust funded projects outputs identified

<http://www.lareferencia.info>

 <https://www.openaire.eu/search/dataprovider?datasourceId=openaoar> ::dc6e224a8d74ce03bf301152d6e33e97

Datos son parte de las legislaciones

Artículos y datos

Estudios reportan que tasa de poblamiento voluntaria de repositorios es baja, si no existe algún tipo de mandato institucional o en el organismo financiador.

Legislaciones nacionales en Perú y Argentina (2013) y México (2014). Política de Datos en Chile. Manifiesto Brasil Ibiect
Muy pocos mandatos obligatorios de Universidades. Ayudaría mucho...

Ejemplos:

[Horizonte2020: Resultados en acceso abierto](#)

[Caso NIH](#)

[Caso EE.UU. Vía agencias federales.](#)

[Voluntaria: 15 %](#)

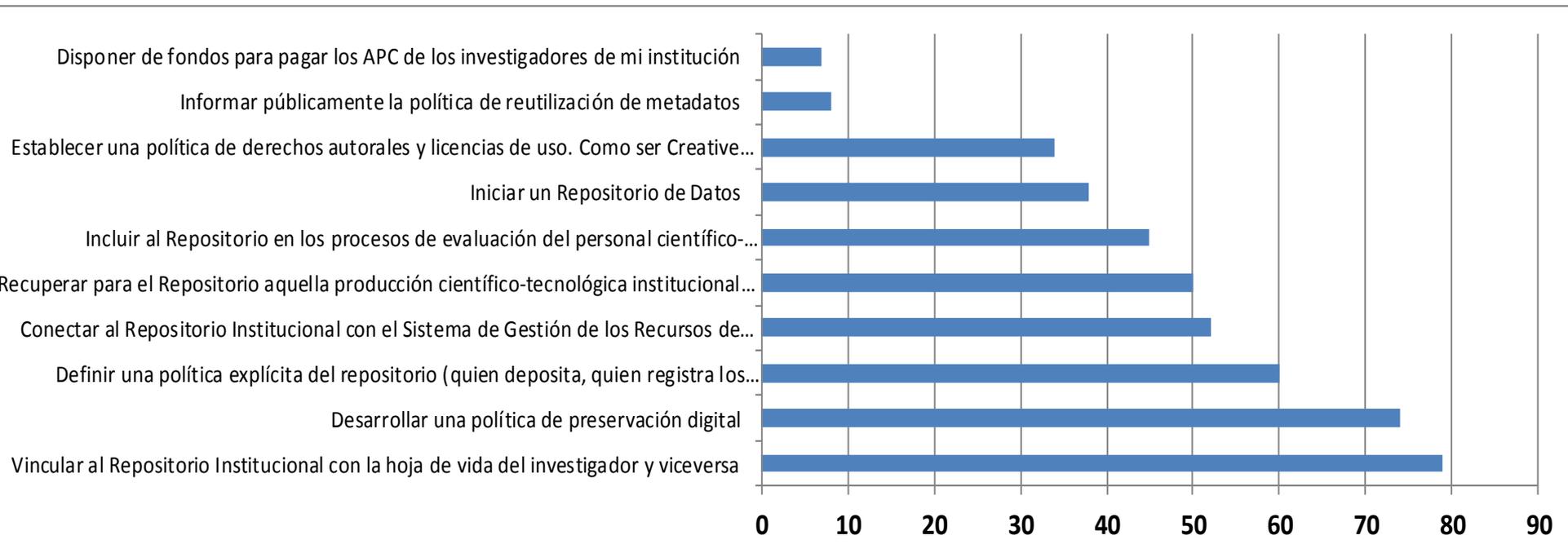
[Obligatoria: Sobre 70%](#)

[Embargos manuscritos no más 12 meses](#)

Qué desean del País



Qué desean de su institución





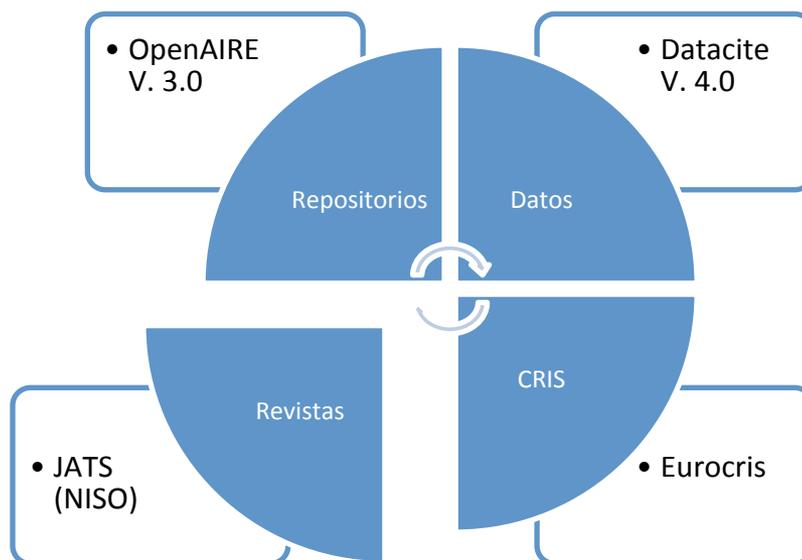
Directrices

Directrices para repositorios de datos y publicaciones; vocabularios controlados; identificadores. Datos primarios



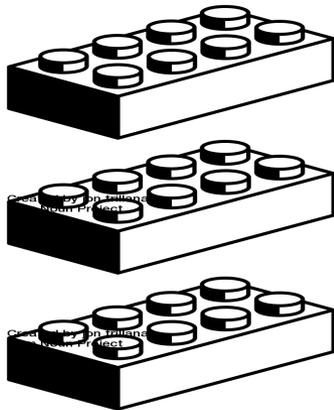
Directrices, Calidad y Posicionamiento

- Hoy el foco es literatura y datos
- Prioridad es AuthorID, ProjectID, Fin Embargo



Proposal metadata Layers

To be discuss with the communities: LA Referencia, Share, Jisc/RIOXX, COAR...

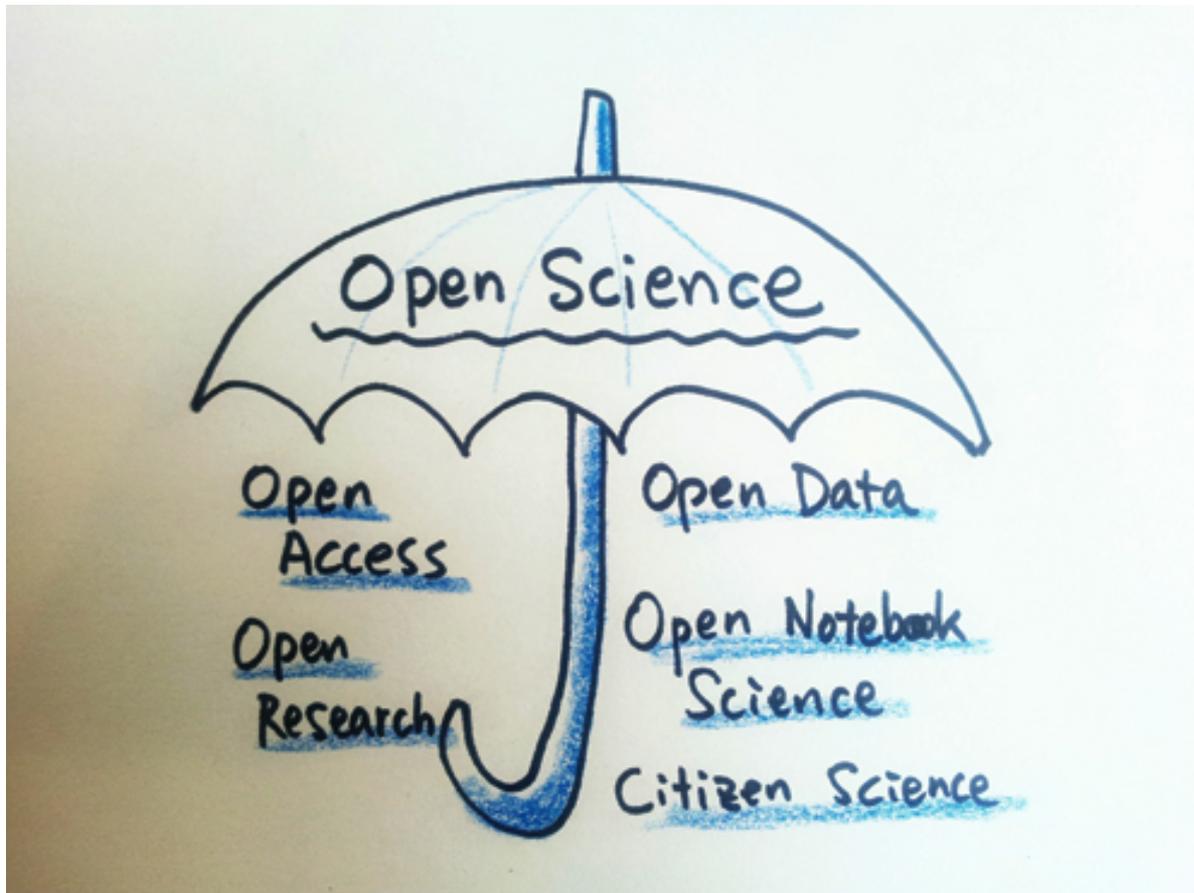


Created by Jon Trillana
from Noun Project

- Application Profile that defines**
- **basic layer (still Dublin Core)**
 - **enhanced layer (re-use DataCite)**
 - **extended layer (custom fields specific for regional networks / infrastructures)**

(Información provista por Pedro Príncipe)

...Perfil de aplicación.



CIENCIA ABIERTA



Ciencia Abierta

- OECD, tradición de la teoría de Bienes Públicos en este campo.
- Ciencia abierta comúnmente se refiere a los esfuerzos para que los resultados de la investigación financiada con fondos públicos sean más accesibles en formato digital a la comunidad científica, el sector empresarial, o la sociedad en general. Ciencia Open es el encuentro entre la tradición secular de la apertura en la ciencia y las herramientas de tecnologías de la información y la comunicación (TIC) que han transformado la empresa científica y requieren una mirada crítica de los responsables políticos que buscan promover la investigación a largo plazo, así como la innovación
- Es un medio, no un fin.
- Engloba más que publicaciones y datos, pero hoy ese aspecto es el foco central.

Por Qué

- La ciencia se basa en verificar y es acumulativa
- Datos permiten:
 - Verificar resultados
 - Reproducir la investigación
 - Combinar datos y reutilizarlos (nuevo conocimiento)
 - Preservar la integridad de la investigación
 - Evitar la duplicación de esfuerzos

En A. Latina, la I+D es financiada con fondos públicos.
Los datos deben estar en acceso abierto.



H2020 Programme

Guidelines on

FAIR Data Management in Horizon 2020

Principios

FAIR

Planes de manejo de datos

FAIR data management In general terms, your research data should be 'FAIR', that is **findable, accessible, interoperable and re-usable**. These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation solution.

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Comment | [OPEN](#)

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier [...] [Barend Mons](#)

Abstract

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the

Marco

- En América Latina, mayoritariamente la I&D se financia con fondos públicos
- Existe un modelo de acceso abierto iniciado hace más de 15 años en nuestras revistas. Se debe considerar.
- Asimetría: Algunos países con sus agencias de financiamiento pueden aportar en el futuro, pero hay naciones que necesitarán un fuerte apoyo.
- **Interoperabilidad es la demanda clave y es imperativo avanzar en propuestas que eviten la duplicación de esfuerzos bajo presupuestos limitados.**
- Un elemento de diseño debe ser el trabajo cotidiano de nuestros investigadores.

Tipos de datos

- Diferenciar open data o gobierno abierto de datos científicos. Foco posible de las acciones iniciales:
- *los datos que soportan y validan los resultados publicados,*
- *datos que se relevan DMP (Planes de Manejo de Datos) y que son parte de los proyectos financiados total o parcialmente con fondos públicos,*
- *software y herramientas necesarios para comprender y reusar los datos*

- Mediante... *metadata interoperable*
- *Respetando excepciones...(patentabilidad, datos sensibles, etc.), pero en principio todo público a menos que se justifique lo contrario.*

Revistas

- Sinergia con el modelo de LA de revistas de acceso abierto
- Fomentar **alternativas que permitan en el ámbito tecnológico y de estándares de metadatos una interoperabilidad entre los repositorios de datos científicos y las publicaciones**
- Inserta en los procesos de investigación para evitar la duplicación de esfuerzos (enviar paper y datos idealmente al mismo tiempo).

Tecnología y Mandatos

- Es conveniente un **modelo flexible o híbrido para la región que apunte a estructuras interoperables y federadas.**
- La tecnología subyacente **debe ser de acceso abierto y transferible** a instituciones de educación superior de tamaños diversos.
- En otros países puede elegirse un modelo más centralizado
- **Es clave para una política de RDM o DMP que las instituciones de investigación adopten recomendaciones explícitas o mandatos.** La experiencia muestra que en la medida que ello emerge, se crea o justifica también un impulso a las legislaciones
- El rol inicial para modificar los modelos de evaluación **debe provenir de la comunidad universitaria**

Costos

- En la medida que avance el uso de DMP existirán ítems financiados pertinentes al proyecto. Pero, ello sugiere también que la infraestructura debe ser provista por la institución como parte de su misión.
- La agencia debería colocar infraestructura para datos “huérfanos”, donde no existen repositorios en la institución del investigador.
- Hay agencias e instituciones de CTI en la región que simplemente no cuentan con los recursos adecuados. Es una situación de borde y requiere una fuerte colaboración intra-regional e internacional.

FAIR



To be Findable:

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

To be Accessible:

- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
 - A1.1 the protocol is open, free, and universally implementable.
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2 metadata are accessible, even when the data are no longer available.

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

To be Re-usable:

- R1. meta(data) have a plurality of accurate and relevant attributes.
 - R1.1. (meta)data are released with a clear and accessible data usage license.
 - R1.2. (meta)data are associated with their provenance.
 - R1.3. (meta)data meet domain-relevant community standards.
-

