



LA Referencia

Red de repositorios de acceso abierto a la ciencia

DATOS CIENTÍFICOS

ALBERTO CABEZAS B.

**Consejo Directivo, 2 y 3 Septiembre
2016**

TEMAS

- LA REFERENCIA Y LOS DATOS CIENTÍFICOS
- QUE ES CIENCIA ABIERTA
- LA RESPUESTA A CEPAL
- LOS PRINCIPIOS (FAIR)

Datos científicos

- 2014 surge la inquietud.
- Primer plan de trabajo menciona esta necesidad debido a legislaciones (Perú, Argentina)
- Un proyecto se presenta a IDRC.
- Emerge como una “deuda” de LA Referencia.
- Se suma legislación de México.
- Chile anuncia política.

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.

Ciencia Abierta

- OECD, tradición de la teoría de Bienes Públicos en este campo.
- *Open science commonly refers to efforts to make the output of publicly funded research more widely accessible in digital format to the scientific community, the business sector, or society more generally. Open science is the encounter between the age-old tradition of openness in science and the tools of information and communications technologies (ICTs) that have reshaped the scientific enterprise and require a critical look from policy makers seeking to promote long-term research as well as innovation.*
- Es un medio, no un fin.
- Engloba más que publicaciones y datos, pero hoy ese aspecto es el foco central.

Brussels, 27 May 2016
(OR. en)

9526/16

RECH 208
TELECOM 100

Unión Europea

Transición a un sistema de ciencia abierta

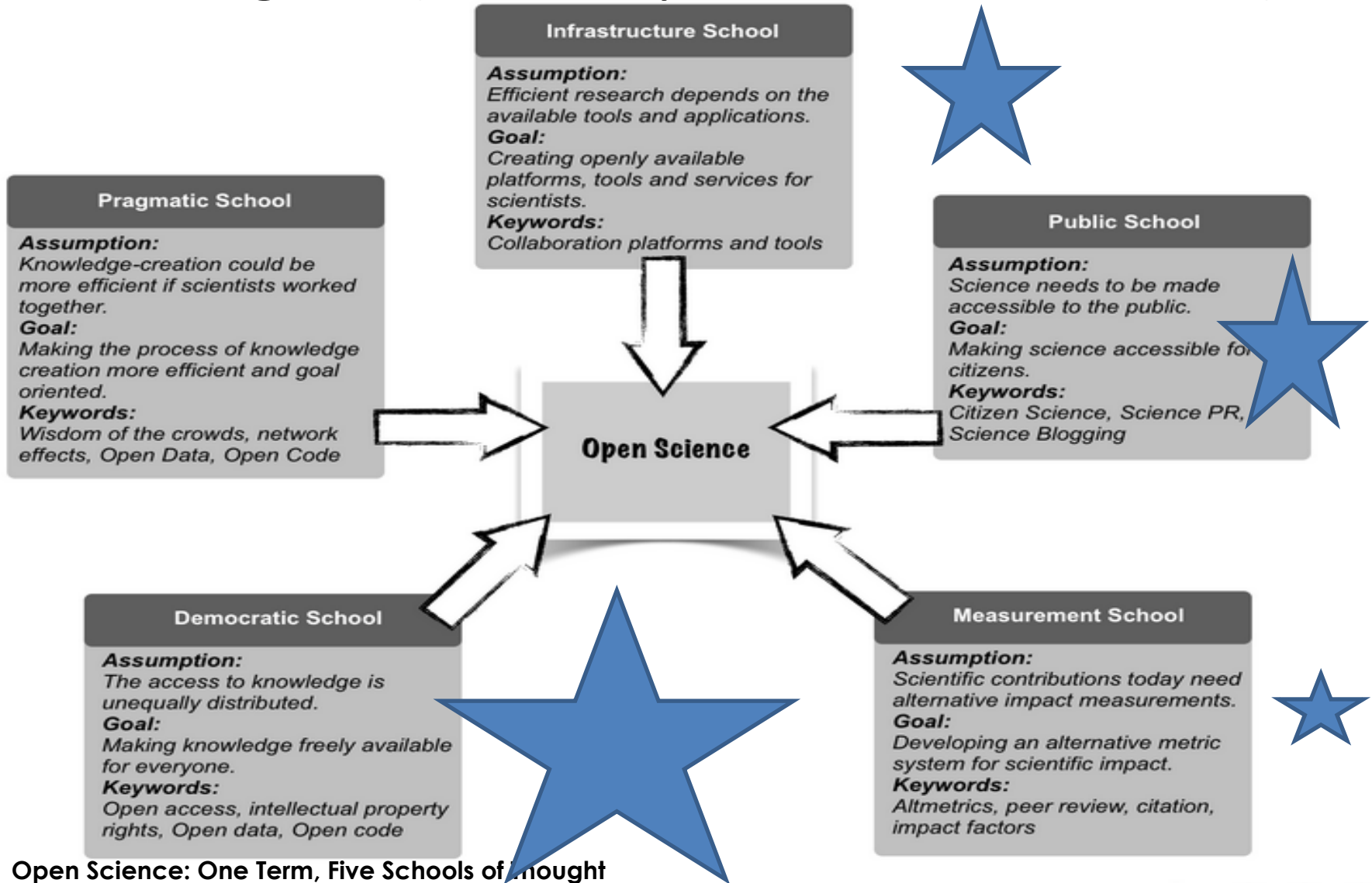
OUTCOME OF PROCEEDINGS

From:	General Secretariat of the Council
To:	Delegations
No. prev. doc.:	8791/16 RECH 133 TELECOM 74
Subject:	The transition towards an Open Science system - Council conclusions (adopted on 27/05/2016)

- 1. RECOGNISES that the exponential growth of data, the availability of increasingly powerful digital technologies, the globalisation of the scientific community, as well as the increasing demand from society to address the societal challenges of our times, are the bases of an ongoing transformation and opening up of science and research, referred to as "open science" affecting the modus operandi of doing research and organising science.
- 2. ACKNOWLEDGES that open science has the potential to increase the quality, impact and benefits of science and to accelerate advancement of knowledge by making it more reliable, more efficient and accurate, better understandable by society and responsive to societal challenges, and has the potential to enable growth and innovation through reuse of scientific results by all stakeholders at all levels of society, and ultimately contribute to growth and competitiveness of Europe.

- <http://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf>

Paradigmas (estrella: posición LA Referencia)



Open Science: One Term, Five Schools of Thought

Benedikt Fecher & Sascha Friesike

http://book.openingscience.org/basics_background/open_science_one_term_five_schools_of_thought.html

Promesa en la nube

Final draft

Monday 20 June 2016

A Cloud on the 2020 Horizon

Commission High Level Expert Group on the European Open Science Cloud

Realising the European Open Science Cloud: first report and recommendations

Preface by Barend Mons, Chair



This report aims to lay out a high level, living roadmap for the realisation of the European Open Science Cloud (EOSC). The High Level Expert Group, with ten members from European countries, Japan and Australia, discussed extensively in several meetings, conferences, policy events and met with key stakeholders (30 November 2015) and research funders (15 March 2016). Based on these consultations, on many 'white papers' and on a range of presentations and feed-back at international meetings, we are confident that our recommendations count on a high-level of consensus amongst all stakeholders. This was a solid basis to embark on this challenging journey with the Commission, the Member States and International partners in concert.

The title of this first report may have a slightly threatening ring to it and indeed, if we do not act, there might be a looming crisis on the Horizon. The vast majority of all data in the world (in fact up to 90%) has been generated in the last two years. Computers have long surpassed individuals in their ability to perform pattern recognition over large data sets. Scientific data is in dire need of openness, better handling, careful management, machine actionability and sheer re-use. One of the sobering conclusions of our consultations was that research infrastructure and communication appear to be stuck in the 20th century paradigm of data scarcity. We should see this step-change in science as an enormous opportunity and not as a threat. The EOSC is a positive 'Cloud on the Horizon' to be realised by 2020. Ultimately, actionable knowledge and translation of its benefits to society will be handled by humans in the 'machine era' for decades to come, machines are just made to serve us.

Mundo de iniciativas

- Organizaciones (RDA, Datacite...)
- Agencias que solicitan los planes de manejo de datos en los proyectos (Horizonte 2020, Agencias Federales de I+D Estados Unidos)
- Propuestas de organizaciones (ICSU, Unesco, OECD)
- Planes públicos nacionales
- Editoriales que exigen los datos científicos para el “paper”.
- Repositorios de datos (soft + hard + directrices)
- Etc. etc. etc.

RESPUESTA A CEPAL

La mirada de organismos de CyT

Proyecto LEARN, LEaders Activating
Research Networks: Implementando la Hoja
de Ruta y un conjunto de buenas prácticas
para la Gestión de Datos de Investigación
(GDI) Proyecto

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 de las acciones iniciales:
- ***los datos que soportan y validan los resultados publicados,***
- ***metadata interoperable,***
- ***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***

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).

Metadata

- Hay metadatos sujetos a la especificidad de disciplinas. Pero:
- Urgencia de un estándar genérico de descripción en áreas interdisciplinarias.
- Definir los elementos básicos de metadatos que facilitan la interoperabilidad entre repositorios de datos con las publicaciones.
- Una propuesta de elementos esenciales para la interoperabilidad entre repositorios de datos. Es decir, un set de metadatos “Obligatorios y Obligatorios si son Aplicables”.
- El modelo debe ser flexible y “agnóstico”: Debe evitar “Lock in”
- El camino está hoy en las sinergías entre OpenAIRE y Datacite

Licencias, RDM o DMP

- ¿Research Data Management (RDM) Policy , o partir de Planes de Manejo de Datos (DMP) en las postulaciones al financiamiento público?
- Práctica muy incipiente en América Latina, pero será una tendencia. El gran objetivo de una hoja de ruta podría ser **armonizar o definir un set común de elementos que se solicitan a los investigadores** (y que les simplificaría la vida)
- **Es conveniente incluir en la discusión una posición clara sobre el uso de licencias Creative Commons, que es el estándar de facto en América Latina**

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 financiables 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 cuenta con los recursos adecuados. Es una situación de borde y requiere una fuerte colaboración intra-regional e internacional.

H2020 Programme

Guidelines on

FAIR Data Management in Horizon 2020

Principios

FAIR

Planes de manejo de datos

 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

DPM y FAIR

- The ORD pilot applies primarily to the data needed to validate the results presented in scientific publications. Other data can also be provided by the beneficiaries on a voluntary basis, as stated in their Data Management Plans
- This Horizon 2020 FAIR DMP template has been designed to be applicable to any Horizon 2020 project that produces, collects or processes research data.
- 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.
- More information about FAIR: FAIR data principles (FORCE11 discussion forum) FAIR principles (article in Nature)

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.
-

