

Datos de investigación.

Universidad Politécnica de Valencia. ETSINF.

Salón de actos

16-17.30 h. 23 OCTUBRE 2014

Fernanda Peset (Universitat Politècnica de València)



FUNDACIÓN **MAPFRE**



Estructura

- **Introducción datos de investigación**
- **Conocimientos mínimos**
 - **Plan de gestión de datos**
 - **Organizar y documentar los datos**
 - **Almacenar los datos y seguridad**
 - **Aspectos éticos y de propiedad intelectual**
 - **Compartir datos, preservación y licencias**
- **Conclusiones**



Introducción datos de investigación

Conocimientos mínimos

Conclusiones



El contexto:



Un proyecto del Centre de Cultura Contemporània de Barcelona y Fundación Telefónica

Castellano

BIG BANG DATA

Barcelona 9 mayo - 26 octubre 2014 | Madrid 25 febrero - 24 mayo 2015

EXPOSICIÓN ESTACIÓN BETA ENTREVISTAS ARTÍCULOS AGENDA PROYECTO

INMERSOS EN EL TSUNAMI

Una historia de la infoexplosión

La idea de que vivimos bajo la onda expansiva de una gran explosión de información no es nueva. Sin embargo, el ritmo al que ha crecido en la última década la cantidad de datos que somos capaces de producir, transmitir y almacenar no tiene precedente. Hace cincuenta años, un disco duro era un enorme dispositivo, del tamaño de un coche utilitario, que almacenaba el equivalente a una canción en formato MP3, una ridícula porción de la capacidad del teléfono que llevamos hoy en el bolsillo.

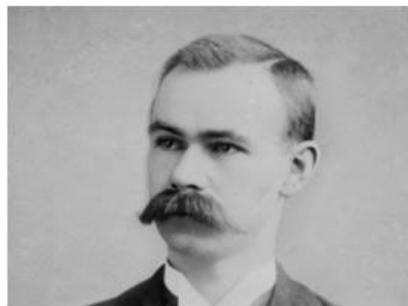
A medida que nuestra experiencia personal y nuestra historia colectiva se codifican en soportes digitales, el problema de su preservación a largo plazo se hace más urgente.



_BIG BANG DATA



Rosetta Project | Long Now Foundation



Después del diluvio



The Internet Archive | Jonathan Minard



<http://bigbangdata.cccb.org/es/>

Evolución

información	} DATOS
lectura	} ANALISIS
cerrado	} ABIERTO
adquisición	} ACCESO

Desde las instancias políticas, pasando por cada una de las administraciones, hasta los sectores académicos están interesados en conocerlos y analizarlos.



Realidad actual



Comisión Europea. **Hacia una economía de los datos próspera** (COM(2014) 442 final, 2.7.2014)



NSF financiación On March 29, 2012, six federal grant-making departments and agencies announced more than \$200 million in grant opportunities for the so-called “**Big Data research and Development Initiative**” (Office of Science and Technology Policy [OSTP] 2012).



DCC 10th international conference. Próxima CfP. London, 9 - 12 February 2015



Cultura abierta: conocimiento compartido



- Gijón, 28, 29 y 30 de mayo. <http://www.fesabid.com>

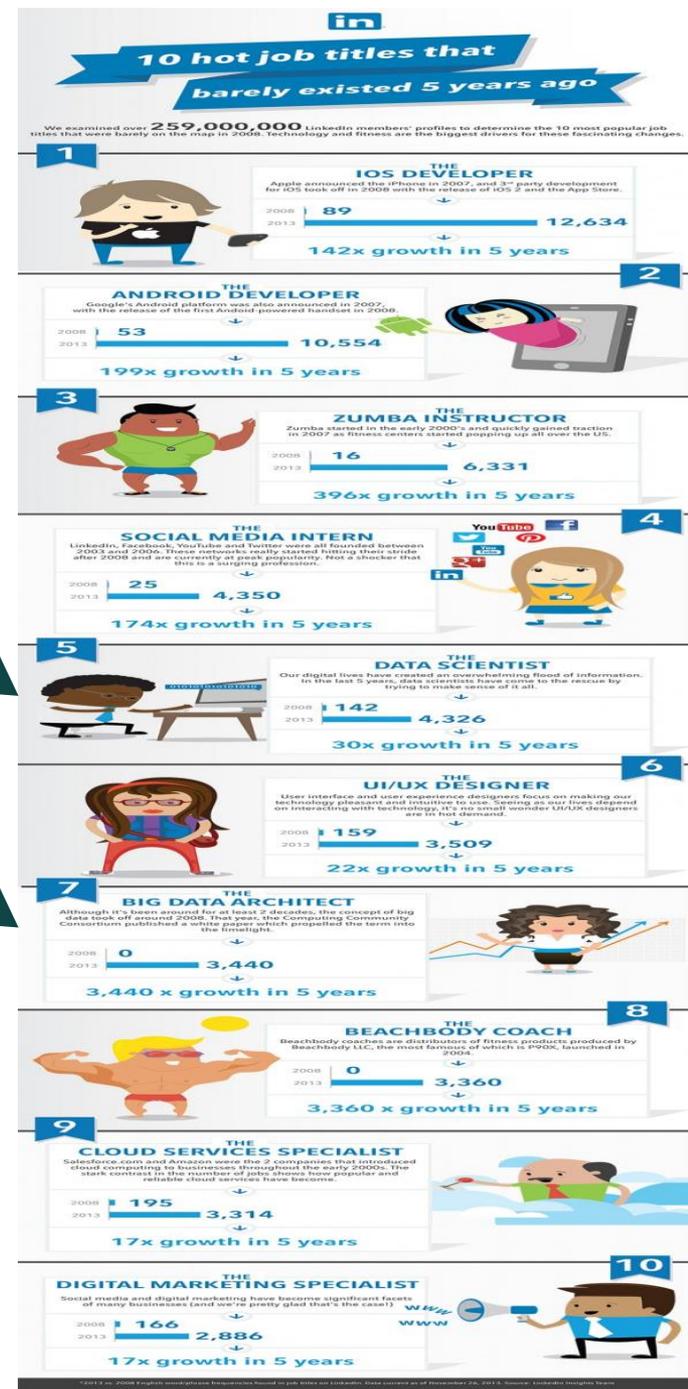
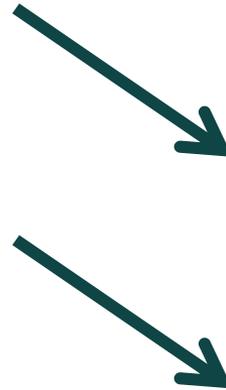
19 enero acaba la CfP

- FESABID 2015, se centra en la cultura “de lo abierto” y en el papel clave de los profesionales de la información para gestionar la información abierta y garantizar que el conocimiento generado se comparta de forma libre y democrática.



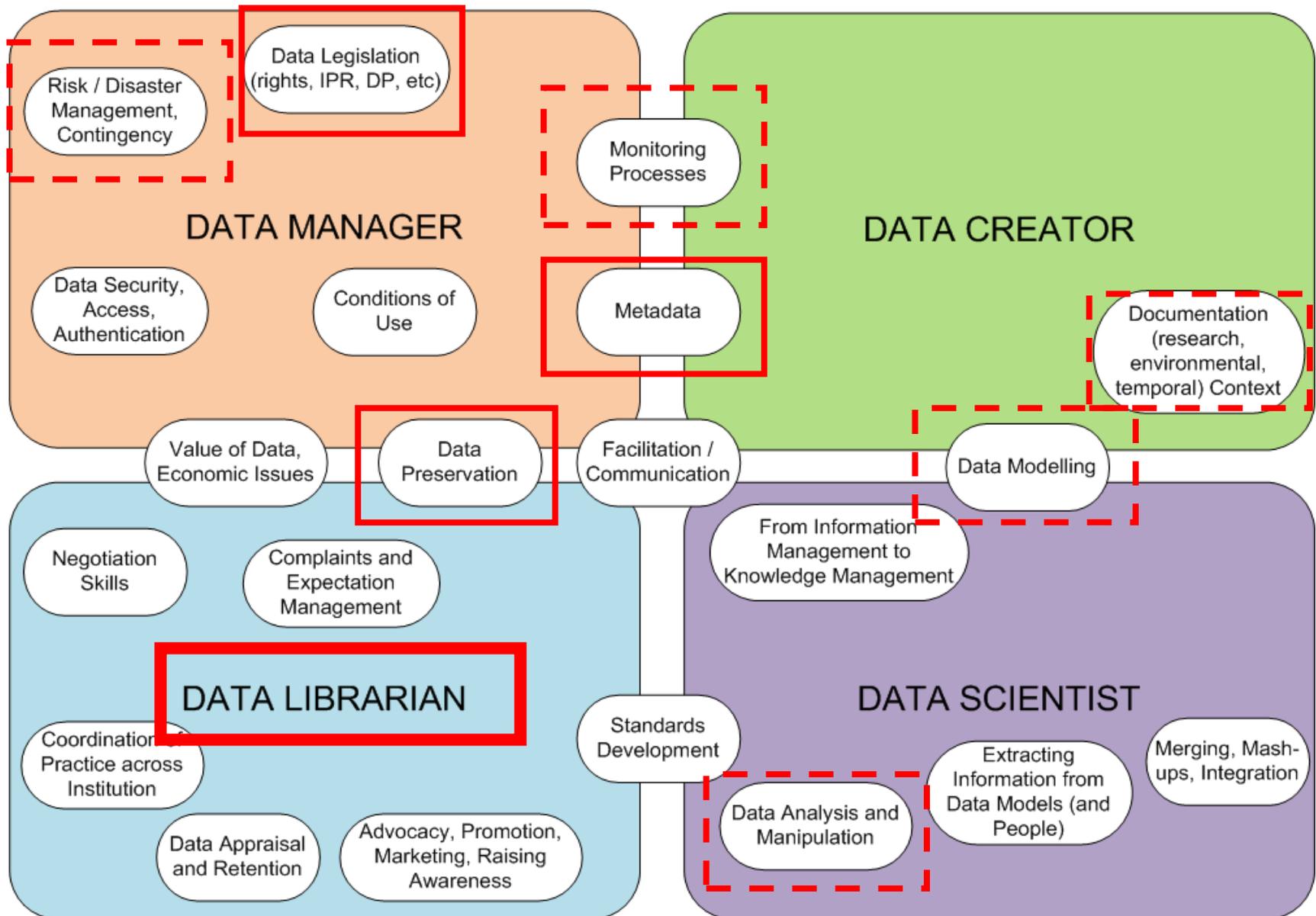


De hecho, el gráfico proporcionado por LinkedIn sobre los trabajos que antes no existían reflejan el Data Scientist y el Analista de BigData (LinkedIn, 2014).



CORE SKILLS FOR DATA MANAGEMENT

A follow-up from the second DCC Research Data Management Forum (November 2008)



En mi opinión...



Datos de investigación y lo abierto



- “datos de investigación [es] todo aquel material que ha sido **registrado durante la investigación**, reconocido por la comunidad científica y que sirve para **certificar** los resultados de la investigación que se realiza. [...] debe provenir de una **f fuente única** y deben ser **difíciles o imposibles de obtener de nuevo**”

Torres-Salinas; Robinson-García; Cabezas-Clavijo, 2012





Numerosos



- *Los datos son “una representación reinterpretable de información de una manera formalizada, adecuada para la comunicación, la interpretación o el procesamiento”*

ISO/IEC 2382-1, de Economía próspera (2014)

- *A research data set constitutes a systematic, **partial representation** of the subject being investigated.*

OECD (2006) recommendation for promoting the access to research data





- En algunas definiciones y herramientas de gestión/preservación no se distingue entre publicaciones y datos de investigación. Pero en el campo de las publicaciones se ha avanzado y en el de los datos menos
- Cuando hablemos de datos, nos referiremos a datasets u otro material suplementario, por contraposición a las publicaciones

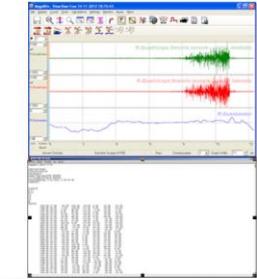


Figura 9. Datos recogidos de una electrografía de superficie (parte superior) y el correspondiente fichero en un formato plain text (parte inferior).

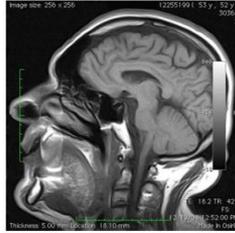


Figura 13. Imagen de tomografía axial con

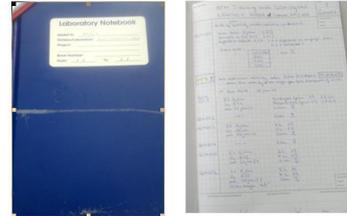
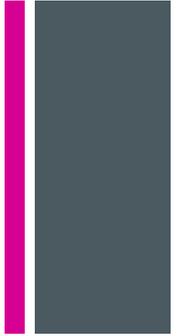


Figura 6. Ejemplo de cuaderno de laboratorio con anotaciones diarias. Laboratorio del prof. Dr. J.L. Tocá (BOKU)

Figura 11. Datos tratados y publicados (García-Massó et al. en prensa).

Los datos incluyen: “cuadernos de laboratorio, cuadernos de campo, datos de investigación primaria (incluidos los datos en papel o en soporte informático), cuestionarios, cintas de audio, videos, desarrollo de modelos, fotografías, películas, y las comprobaciones y las respuestas de la prueba. Las colecciones de datos para la investigación pueden incluir diapositivas; diseños y muestras. [...] El código de software “



¿Qué significa comunicar los datos de forma abierta?



Open data de Royal Society

- Criterios a cumplir cuando son liberados:
 - Accesibles (accessible): fáciles de encontrar y en una forma en que puedan ser usados [y preservados].
 - Evaluados/certificados (asessable): credibilidad para diferentes grupos de interés.
 - Inteligibles (intelligible): ser entendidos
 - Reutilizable (useable): en un formato para usar y con licencias adecuadas.

include the JPEG 2000, PNG and SVG standard image formats; ASCII, PDF, Open Document Format and Office Open XML format (the native format for recent versions of Microsoft Word) for text; HTML, XHTML, RSS and CSS for the web and NetCDF for some scientific data.



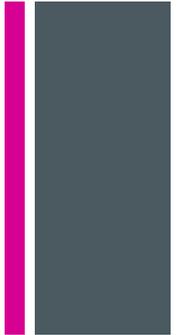
Lo libre, lo open... recordemos



- **Gratisito.** Sólo implica el acceso sin coste, pero no otras dimensiones de reutilización y difusión. “Free” tiene esa acepción.
- **Libre.** A menudo se mantiene en su forma española para diferenciarlo de lo simplemente gratis. En realidad este es el término que acoge otras connotaciones con la reutilización del contenido sin restricciones.
- **Abierto.** Desde su semántica inicial, de código accesible y manipulable, incluye actualmente la posibilidad técnica y legal de transformarse.



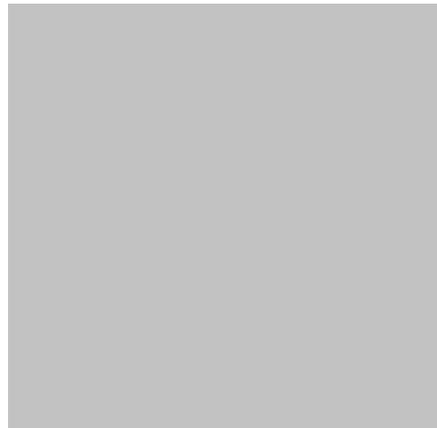
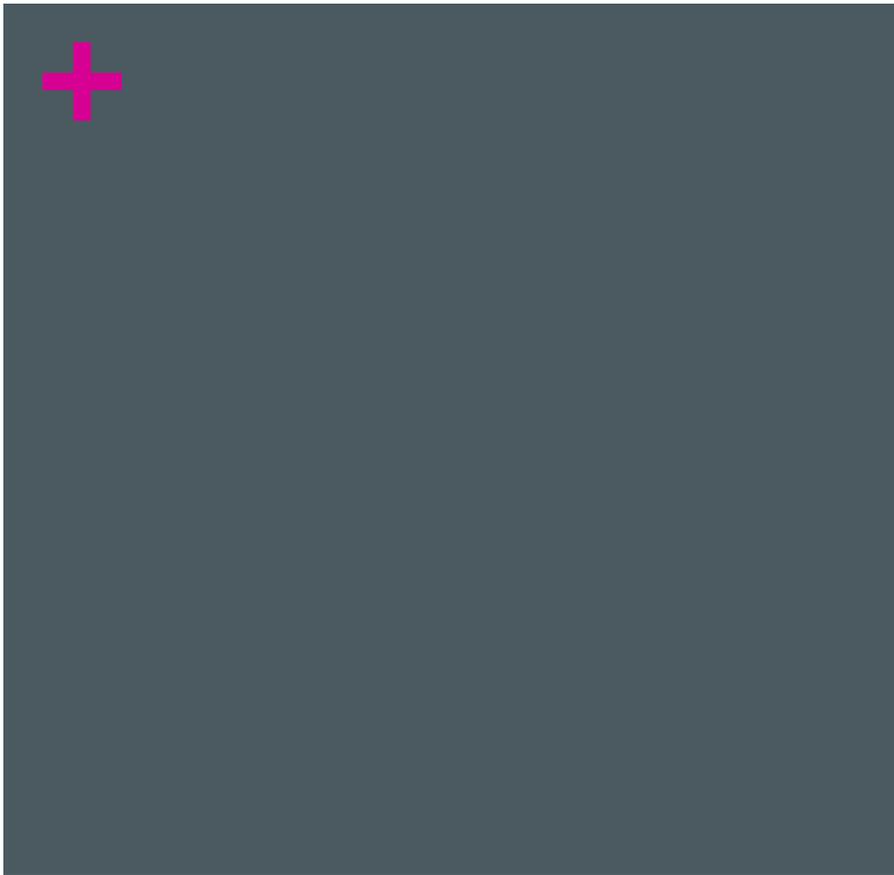
Veamos los Principios Panton (OKF)



- “En el ámbito de la ciencia, por **datos abiertos** nos referimos a su disponibilidad gratuita en Internet permitiendo a cualquier usuario **descargarlos, copiarlos, analizarlos, volverlos a procesar, transferirlos a un software, o utilizarlos** para cualquier otro propósito, sin barreras económicas, legales o técnicas fuera de aquellas que son inseparables del acceso a Internet.
- **Creative Commons’ (CCZero) y Licencia al Dominio Público (Public Domain Dedication & Licence PDDL-OKF).**



Open Knowledge
Foundation



Beneficios



Compartir datos tiene muchos beneficios, algunos nuevos/proprios

- responder de forma rápida y mucho más eficiente ante las emergencias (e-coli 2011)
- capacidad potencial de señalar los fraudes (datos inventados: reciente polémica del artículo de Science) o las malas prácticas (tensionar la integridad del sistema o de las personas).
- estimula nuevas y altamente creativas formas de colaboración científica
- estimula un movimiento social hacia la ciencia, quizá fundamental para el cambio de las dinámicas científicas: Galaxy Zoo, Fold-it, Ash-Tag, etc.



Ejemplo, Galaxy zoo



Classify



Hubble



Invert

Help

Restart

SHAPE

Is the galaxy simply smooth and rounded, with no sign of a disk?



Smooth



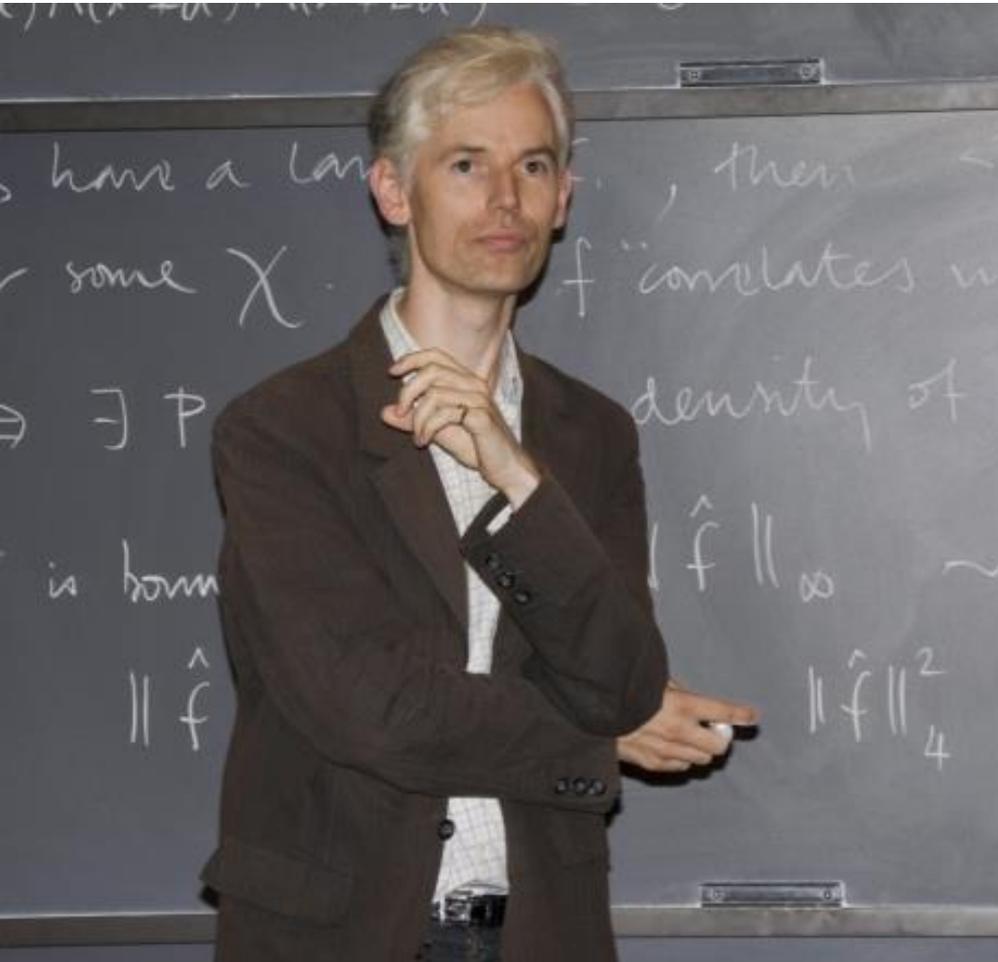
Features or disk



Star or artifact

Otro ejemplo

Tim Gowers, matemático, planteó un problema en su blog y fue resuelto por la comunidad — crowdsourcing— de manera inauditamente veloz.



Pero además...

- los **ciudadanos**, quienes pagan con sus impuestos la ciencia, demandan cada vez más las evidencias que dan lugar a las políticas que se les acaban imponiendo en todos los órdenes sociales
- única forma de afrontar **retos globales** que implican a toda la sociedad: cambio climático, crisis energética, las pandemias o residuos mundiales.

+ Objetivos

- Producir nuevo conocimiento a partir de datos previos
- Remezclar
- Mostrar patrones ocultos con las antiguas técnicas
- Inventar/intentar nuevas aproximaciones
- Visualizaciones...

+ Herramienta de análisis



- KRDS Benefits Analysis Toolkit <http://beagrie.com/krds-i2s2.php>





- *Increasing research productivity*
- *No loss of future research opportunities*
- *Input for future research*
- *Motivating new research*
- *Stimulating new networks and collaborations*
- *New research opportunities*
- *Short-term re-use of well curated data*
- *Lower future preservation costs*
- *No re-creation of data*
- *New research funding*
- *Value to current researchers and students*
- *Re-purposing data for new audiences*
- *No data lost from Post Doc turnover*
- *Enhancement of research tools and software by testing on a range of well-curated datasets*
- *Secure storage for data intensive research*
- *Knowledge transfer to other sectors*
- *Increasing economic growth*
- *Verification of research/research integrity*
- *Availability of data underpinning published findings*
- *Increased visibility/citation x*
- *Protecting returns on earlier investments*
- *Secures value to future researchers and students x*
- *Adds value over time as collection grows and develops critical mass*
- *Planned management from an early stage in the research life-cycle is ultimately more cost-effective than late intervention (providing proper selection of what to keep is done)*
- *Scholarly communication/access to data*
- *Use by new audiences*
- *Re-purposing methodologies*
- *Long-term re-use of well curated data x*
- *knowledge transfer to industry*
- *Increasing skills base of researchers/students/staff*
- *Fulfilling organisational mandate(s) x*
- *Fulfil research grant obligations x*
- *Commercialising research*
- *Catalysing new companies and high skills employment*



Resumen

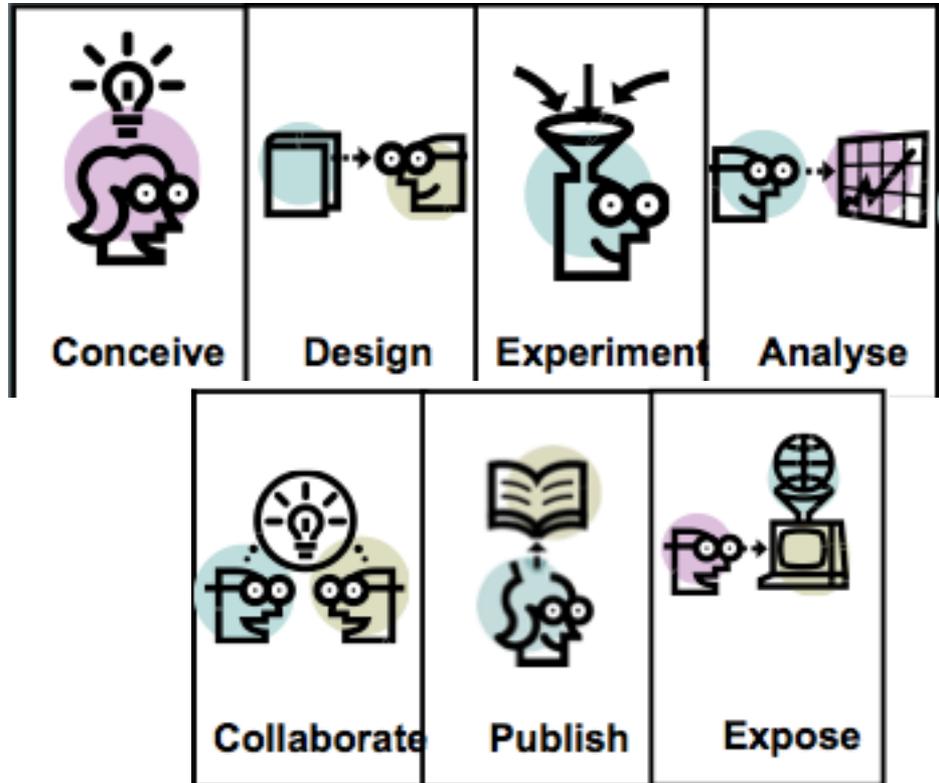
- Heterogéneos y variables
- Cerrados y abiertos
- Criterios de liberación bastante costosos

Es el momento de plantear preguntas esenciales:

- Qué y cómo son
- Cómo se distribuyen
- Cuál es su función



Vuestros usuarios



¿qué hace y piensa un **investigador**?

Producción
Explotación
Networkng/Datasharing

Su trabajo

Reconocimiento
Financiación
Rapidez

Sus intereses

Preservación
Compartir

**Nuevas
obligaciones**

+ Agencias de financiación y empleadores (públicos)

- Aprovechemos el potencial desaprovechado de los datos aislados (sin compartir)
- Los datos abiertos pueden ser el **combustible** de la **innovación, crecimiento y creación de trabajo** (Neelie Kroes, comisaria Agenda Digital)

Plan de gestión de datos en las solicitudes



Vosotros



+ Servicios de apoyo (Bibliotecas, OTRI, Vicerrectorados de investigación, Informática)

- Entender el trabajo de los científicos: producción y ciclo de vida
- Colaborar
- Prepararnos y conocer
- Asesorar
- Apoyar



Visión global y poliédrica

+ Data curation

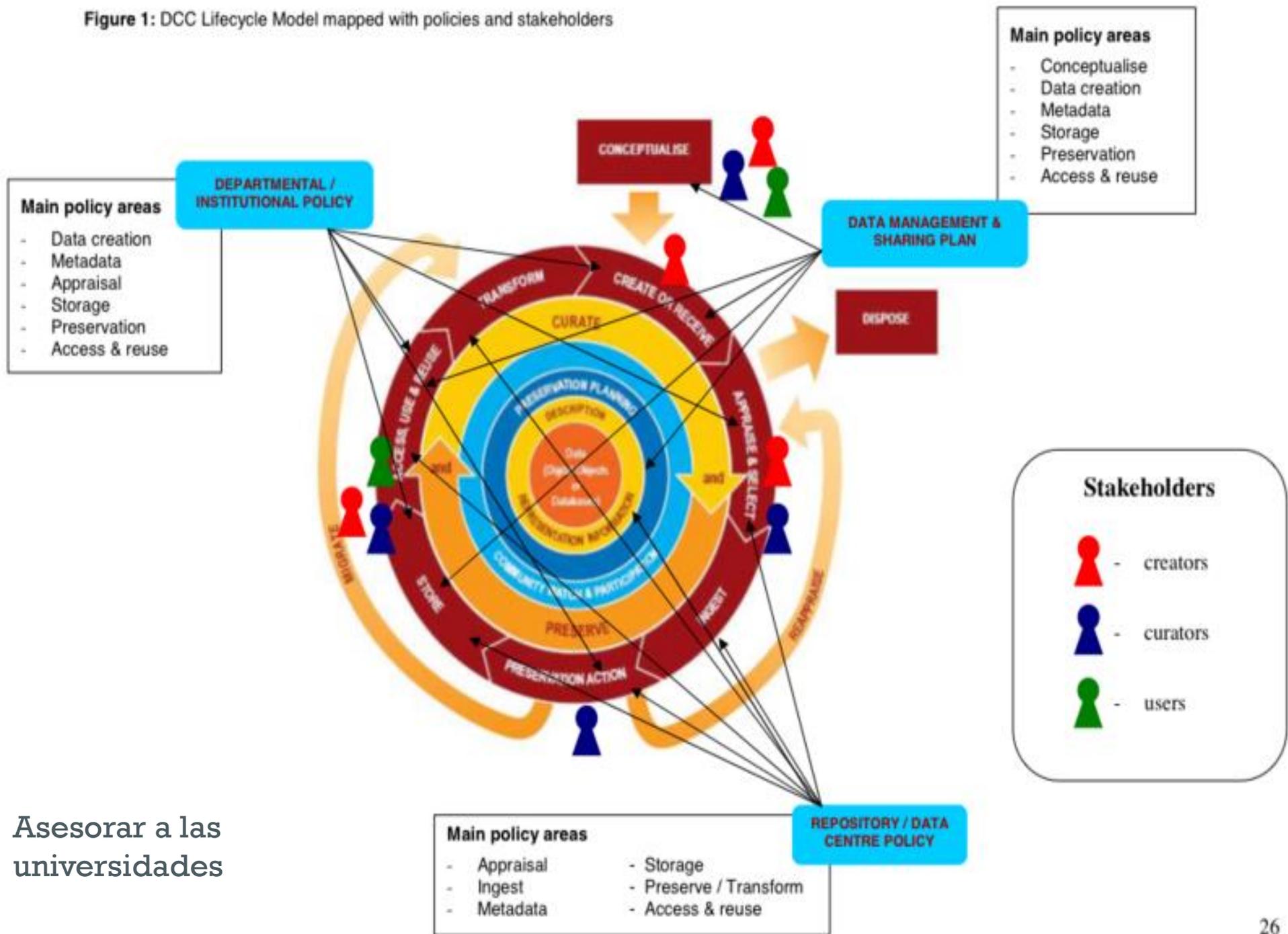


“any services related to the organization, management, or long-term preservation of data developed through scholarly research. These services encompass a range of activities, including **consultations on creating data management plans and strategies, physical or electronic archiving of datasets, and workshops**”

(Deards, 2013).



Figure 1: DCC Lifecycle Model mapped with policies and stakeholders



Asesorar a las universidades



Introducción datos de investigación

Conocimientos mínimos

Plan de gestión de datos

Organizar y documentar los datos

Almacenar los datos y seguridad

Aspectos éticos y de propiedad intelectual

Compartir datos, preservación y licencias

Conclusiones

+ Mínimos conocimientos propuestos

1. Plan de gestión de datos
2. Organizar y documentar los datos
3. Almacenar los datos y seguridad
4. Aspectos éticos y de propiedad intelectual
5. Compartir datos, preservación y licencias

<http://datalib.edina.ac.uk/mantra/libtraining.html>

EDINA and Data Library, University of Edinburgh.
DIY Research Data MANTRA Training Kit for Librarians





1. Plan de gestión de datos



- Obliga a explicitar qué datos serán creados y cómo, e indican de qué forma van a ser compartidos, explicando cualquier restricción de uso que pueda ser aplicada.
- Los investigadores están familiarizados con la solicitud de subvenciones, así que confían en los beneficios de planificar. No suelen conocer los servicios de apoyo, ni las buenas prácticas y estándares a utilizar.
 - Deben **justificar las elecciones** que se hacen
 - Deben dar cuenta de la **necesidad de crear esos datos** y por tanto de financiar esa investigación

Cubre generalmente

- i. Tipos de datos, formatos, estándares y métodos de captura
- ii. Aspectos éticos y propiedad intelectual
- iii. Acceso, compartido y reutilización
- iv. Almacenamiento a corto plazo y gestión de los datos (lo veremos en su apartado correspondiente: Almacenar los datos y seguridad)
- v. Depósito y preservación a largo plazo (lo veremos en su apartado correspondiente: Compartir datos, preservación y licencias)
- vi. Recursos necesarios

(Kerry EDINA MANTRA)

i. Tipos de datos, formatos, estándares y métodos de captura

- Volumen, tipo, contenido, calidad y formato. Se prefieren los formatos abiertos
- Metadatos, documentación y otro material de soporte para interpretarlos: facilitará que otros los descubran. Ha de explicar el porqué y el cómo se han creado
- Estándares y métodos para recolectarlos y gestionarlos. Se relacionan desde los identificadores y formas de denominarlos hasta los formatos de los archivos y esquemas de metadatos
- Relaciones con otros datos disponibles

ii. Implicaciones éticas y de propiedad intelectual

- Demostrar conocimiento
- Explicitar los consentimientos informados, o la anonimización
- Detallar con **contundencia** las restricciones para el intercambio (periodos de embargo) o las etapas hasta ponerlos en abierto
- Declarar el propietario de los datos. Si se compran licencias para la reutilización de datos ajenos deben tenerse en cuenta para el posterior intercambio y construcción de los datos nuevos.

iii. Acceso, intercambio y reutilización

- ¿Dónde se almacenarán (usando infraestructuras existentes indicando sus requisitos de calidad, estándares, metadatos) y cómo accederán otros investigadores (quién-puede-hacer-qué)?
- Calendario para liberarlos, cuando se publique como mínimo.
- Cómo posibilitará colaboraciones o las dificultades que se prevén para su intercambio, con indicación de las posibilidades estudiadas antes de aplicar cualquier restricción de uso.



iv. Almacenamiento a corto plazo y
gestión de los datos (lo veremos en
su apartado correspondiente
Almacenar los datos y seguridad)



v. Depósito y preservación a largo plazo (lo veremos en su apartado correspondiente Compartir datos, preservación y licencias)

[...] vi. Recursos necesarios



- indicar las capacidades técnicas necesarias, el apoyo y formación y cómo va a ser conseguido todo ello incluso si lo proporciona la institución
- No subestimar que documentar información es costoso en tiempo.
- utilidad para calcular los costes de estas actividades de conservación
 - [D3.1—Evaluation of Cost Models and Needs & Gaps](http://4cproject.eu/d3-1) <http://4cproject.eu/d3-1>
 - <http://www.data-archive.ac.uk/media/247429/costingtool.pdf>

Horizon2020 (UE)

- En el **artículo 29.3** del *Agreement* de los socios aparece *Open access to research data*, con las condiciones
- Parece voluntario, en una nueva iniciativa transversal, *Open Research Data Pilot* orientada a mejorar el acceso y reutilización. Las áreas que participan recibirán (2014-2015) sobre 3.000 millones de euros (Vlaeminck, 2013).

Los beneficiarios deben:

- (a) depositar en un repositorio abierto de datos de investigación: los datos necesarios para validar las publicaciones de investigación, con sus metadatos, tan pronto como sea posible; y otros datos, con sus metadatos, dentro del plazo previsto en sus planes de gestión de datos así como el momento de liberación que indicaron;
- (b) ofrecer información en el repositorio sobre las herramientas para validar la investigación e incluso, si es posible, proporcionarlas

Herramientas



- DMPonline <https://dmponline.dcc.ac.uk/>

A screenshot of the DMPonline web interface. The top left features the 'DMP ONLINE' logo. The top right shows the user is signed in as 'fernandapeset@gmail.com'. Below the logo are navigation buttons: 'View plans', 'Create plan', 'About', 'News', and 'Help'. The main heading is 'My project (DCC Template)'. Below this are tabs for 'Plan details', 'Generic DMP', 'Share', and 'Export'. The 'Plan details' tab is active, showing a form with the instruction: 'Please fill in the basic project details below and click 'Update' to save'. The form fields are: 'Project name' (Curso SEDIC), 'ID' (empty), 'Grant number' (empty), 'Principal Investigator/Researcher' (F Paset), 'Principal Investigator/Researcher ID' (empty), and 'Project data contact' (empty). A tooltip for the ID field shows 'E.g. ORCID http://orcid.org/'. A 'Cancel' button is located at the top right of the form.



Hasta aquí el plan de gestión de datos

Hasta los financiadores saben que es provisional y proponen su revisión a los seis meses



2. Organizar y documentar datos

(S.Macdonald)





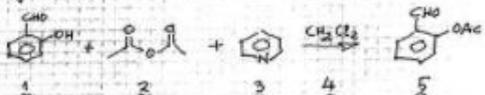
- **Control de Versiones (redactar manuales)**
 - Es imprescindible identificar las diferentes versiones, especialmente si el dataset es actualizado por múltiples usuarios
 - Usar un sistema de número secuencial: v1, v2, v3, etc.
 - No utilizar nombres confusos: revision, final, final2, etc.
 - Registrar todos los cambios, incluso los pequeños
 - Descartar versiones obsoletas (nunca la raw)
 - Usar backups automáticos
 - **Softwares: Bazaar, TortoiseSVN, SubVersion**

Experimento ESO-8

15-10-2003

ACETILACION DE SALICILALDEHIDO

Cornforth, J.; Du, M.H. J. Chem. Soc. Perkin Trans I, 2183 (199)



	1	2	3	4	5
Peso molecular	122.12	102.09	79.10	84.93	164.16
Peso (g)	2.29	2.97	2.94	—	2.9
Pureza (%)	99	98	99	98	—
milimols	18.56	20.79	36.76	—	17.67
Equivalentes	1.05	1.12	1.98	—	—
Densidad (g/ml)	1.146	1.082	0.978	1.325	—
Volumen (ml)	2.0	2.0	3.0	35	—

Materiales: equipo para calentamiento a reflujo

Procedimiento: Reactivos disueltos en CH_2Cl_2 y reflujo 4

Procesado: Adición de hielo (2g), agua (20 ml), la
siguiente se lava con HCl 2N (3 x 15 ml), con
 $NaHCO_3$ sol. sat. (3 x 15 ml) y NaCl sol. sat (1
Se seca con Na_2SO_4 , filtra y empuja
Peso obtenido: 3g

Destilación a presión reducida (trampa agua, 18 mmHg
140-142°C.

Rendimiento: $\frac{17.67 \text{ mmol producto}}{18.56 \text{ mmol reactivo limitante}} = 95\%$

Informe y 1H RMN (Clave ESO-8P1)

• Documentar Datos

Para recordar, ayudar a otros, verificar, replicar, archivar, reclamar autorías...

Ejemplos: cuadernos de laboratorio, notas de campo, metodologías cualitativas

- **Nivel de Proyecto:** documentar la base del estudio, métodos, instrumentos, hipótesis de trabajo
- **Nivel de archivo o dataset:** formatos, relaciones entre archivos
- **Nivel de variable o item:** como fue generada la variable y descripción de los campos



3. Almacenamiento y seguridad

(Brown)

- 
- Cuando es casi manual, en los dispositivos de los proyectos: pérdida de datos o de su integridad

- **¿Dónde almacenar?**

- **Local, Network (institución), Cloud**

- Jahnke y Asher (2013) : coste escondido de la externalización del almacenamiento: aumenta costes de gestión legal para el mantenimiento de contratos o litigios
 - Considerar la capacidad, y proporcionar acceso en cualquier momento y desde cualquier dispositivo; permitir copia y sincronizaciones; posibilidades de compartir e incluso preservar a largo plazo

¿Cómo asegurar los backup?

copias previstas (cómo, cuándo y dónde).

Seguridad e integridad de los datos

Pérdidas o archivos corruptos (hardware failure or data deletion)
Confidencialidad y propiedad intelectual (personal or intellectual property)

- si el grupo de trabajo es disperso, el acceso debe ser remoto con mecanismos que controlen la calidad de los datos, ya que son accedidos desde diferentes puntos.
- niveles de seguridad diferentes para los datos generados frente a los datos licenciados o comprados.

+ ¿Qué está sucediendo?

- **Opciones habituales:** dropbox, drive, pero también es posible en Figshare o Dataverse
- **Otras opciones:** muchos proyectos crean plataformas a medida (dispendio O investigación)



4. Aspectos éticos y de propiedad intelectual

(Rice)





- Las soc. Científicas tienen sus propias guías y consentimientos informados
- Cuando participan animales o personas o se hacen referencias individuales
- Acuerdos de confidencialidad obligan a cuidar de los datos (evitar envíos no encriptados...)

+ Propiedad intelectual

- Intellectual Property Rights (IPR) son inalienables al trabajo de los autores excepto que renuncien expresamente. Están limitados en el tiempo
- Permiten publicar, licenciar la distribución y defenderlos

(Rice)

En España la ley de la Ciencia atribuye la distribución a las universidades



5. Compartir datos, preservación y licencias

Datasharing

Data publication

(Anne Donnelly)



+ DataSharing

“... the practice of making data used for scholarly research available to others (sin ayuda del productor original). ” [Wikipedia]

Involucra a: productor de los datos, el repositorio y el personal de apoyo, y quien el usuario que los consume

Consideraciones: pensar en el futuro (formatos, software, anonimización/ética/confidencialidad, documentación); descripción y documentación; futura gestión de derechos y licencias; infraestructura

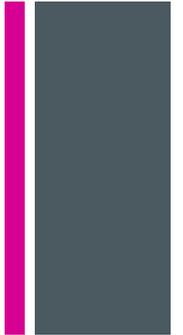
(Anne Donnelly)



- La selección toma tiempo pero ahorra costes de preservación y limita los riesgos a la institución (obligaciones legales)
- Identificar al personal adecuado de los repositorios o centros de datos para conocer los requisitos



1. Criterios DCC. *How to Select and Appraise Research Data for Curation*



1. Relevancia para la misión de la institución o financiador, incluyendo aquí si existen obligaciones legales
2. Valor científico o histórico, lo que indica que puede usarse en el futuro para la investigación o la formación
3. Unicidad y riesgo de pérdida
4. Potencial para la redistribución
5. Replicabilidad casi imposible
6. Costos de preservación razonables o justificables
7. documentación completa

<http://www.dcc.ac.uk/resources/how-guides/how-develop-rdm-services#Selection-criteria>



+ Las licencias (OKF)

- **OpenDataCommons Attribution License (ODC-BY)**, solo obliga a atribuir/reconocer el propietario original. Permite la copia, distribución y uso para producir obras derivadas sobre ella y modificar, transformar y construir sobre ella para cualquier propósito.
- **PDDL, Public Domain Dedication and License**, similar a la CC más permisiva (CC0) sin atribución ni limitación geográfica de uso.
- **Open Database License, ODbL**, es más restrictiva, pero permite compartir —copiar, distribuir y usar la base de datos—, producir trabajos sobre ella y modificar, transformar y construir otros productos sobre la base de datos. Mantiene la (Attribution y obliga a ceder el nuevo producto con esta misma licencia (Share-Alike). Añade condiciones a las bases de datos derivadas y puede aplicar restricciones tecnológicas (DRM) a los datos originales o los derivados. Por tanto, reduce la interoperabilidad y la reutilización



Preservación: Repositorios de datos



- No existe consenso en cómo preservar y compartir los datos, y parece que las soluciones mixtas entre los servicios universitarios y los externos
- Garantizar la procedencia, la persistencia de los identificadores, la utilización de estándares para los metadatos y formatos de citación y periodos de embargo, así como aspectos técnicos derivados de las posibilidades de recombinarlos, analizarlos automáticamente, migrarlos de un formato a otro, o los métodos de almacenamiento



Formas de compartir (Donnelly)



Repositorio de los financiadores

[ESRC Data Store-Reshare](#)

Repositorio Institucional

[Edinburgh DataShare](#)

Repositorio por Disciplina/Formato

Directorios: Databib (con Re3data)

[Figshare](#)

Y... en las revistas [ODiSEA](#)



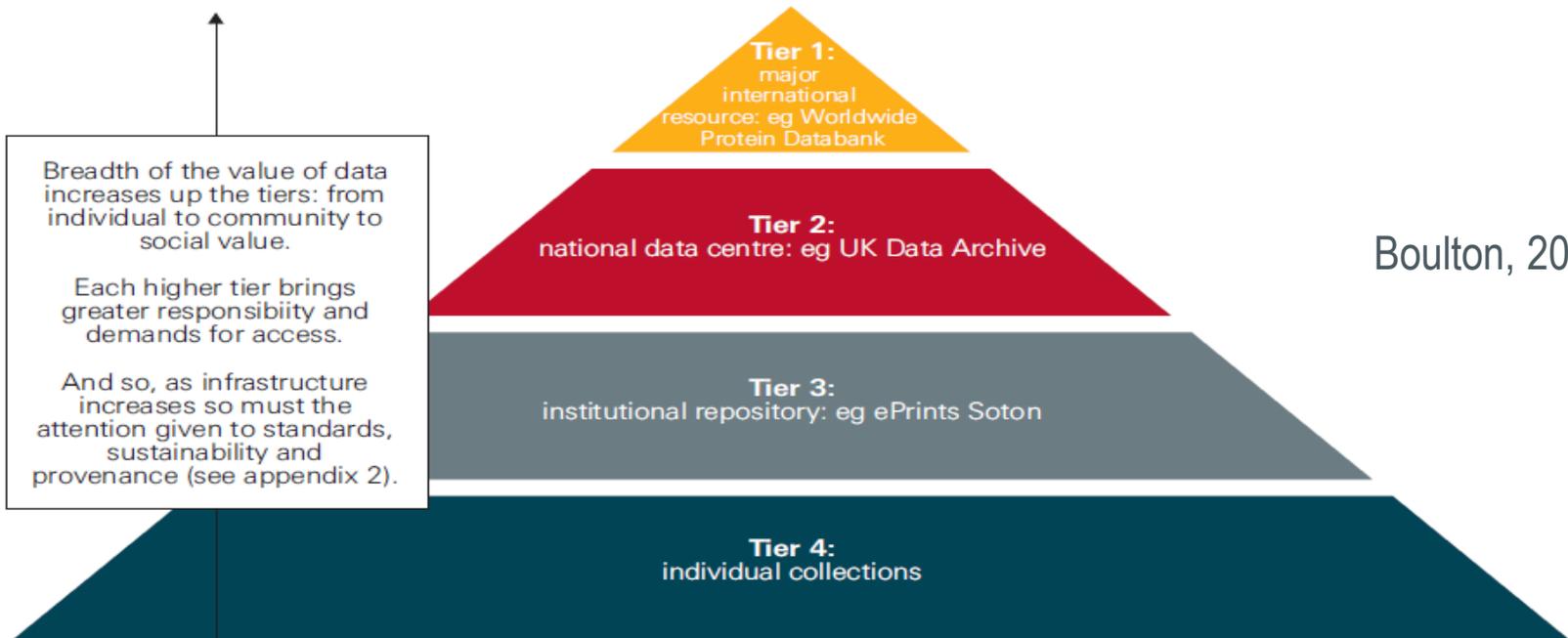
(Donnelly)



Herramientas: especializadas y huérfanos



- Figshare, Dryad, Dataverse o Zenodo (Pedro Príncipe)
- Pisos desiguales:



Boulton, 2013



Consejo....



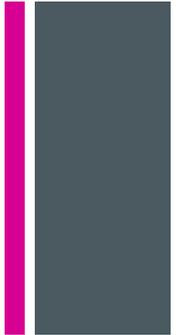
Monash University “adopta, adapta y en último término desarrolla”

“Developing a new product may be expensive, costly to support, and could split researchers from their community.”

La lealtad de los científicos con su comunidad es mayor que con su institución.



DRYAD <http://datadryad.org/>



- Admite datos relacionados con publicaciones, incluso durante la fase de revisión si la revista está asociada a Dryad. Se trata de un proyecto colaborativo en Dspace, que comenzó en el área de salud vinculado con el mundo de las bibliotecas, sociedades científicas y editores.
- Admite textos, hojas de cálculo, video, fotografías, código de software. Es de uso gratuito hasta 10Gb por fichero y asigna como DOI el del artículo. Garantiza la preservación de los datos al participar en CLOCKSS. Destaca la corrección de la cita bibliográfica, algo muy apreciado por los investigadores, y que cuenta con un proceso de validación adicional.

Describe publication

Submitting data to Dryad consists of three simple steps:

- 1. **Describe your publication**
- 2. Upload and describe your data files
- 3. Approve data for publication

Please describe your publication in as much detail as possible. Providing a detailed description will make it easier for others to find your data in Dryad. Please describe the **publication only**. Do not enter information specific to your data files on this page.

Fields marked with an asterisk (*) are required. For more information on expected contents for a field, hold your mouse over the field in question.

Publication metadata

Title*

Authors*

Last name, e.g. *Smith*

First name + initial, e.g. *Donald F.*

Please include the entire first name, and a middle initial if appropriate. Do not use only initials, as this makes the author more difficult to identify later.

Journal name*

Abstract

DOI

If your publication has not been assigned a DOI, please leave this blank.

Journal issue

Volume

Number

Year

If your publication has not been assigned a volume or issue number, please leave this blank.

Primary contact for data associated with this article

Subject keywords

Please enter general keywords associated with the data file. Keywords may be separated by commas, or added individually. For example:
adaptation
evolutionary contingency
founder effects

Taxonomic names

Please enter taxonomic names associated with the publication. These may be names at any level (species or higher-level clade). Names may be separated by commas, or added individually. For example:

Drosophila melanogaster
Geophila

+ Zenodo <https://zenodo.org/>

- Desarrollado como repositorio huérfano por el CERN sobre Invenio. Recientemente ha sido relanzado para todo tipo de datos dentro el proyecto europeo OpenAIREplus y probado con datasets del *Wellcome Trust* (Príncipe, 2013).
- Cabe destacar varios aspectos en diferentes momentos de la gestión de datos: verifica que los ficheros estén relacionados con una investigación; asegura un control descentralizado de los datos; y por último permite crear comunidades de usuarios finales.



Home / Upload / Edit

Delete

Save Submit

New upload

Instructions: (i) Press "save" to save your upload for editing later as many times you like. (ii) Upload and remove extra files in the bottom of the form. (iii) When ready, press "submit" to finalize and make your upload public (editing afterwards only possible via submitting changes to info@zenodo.org).

Type of file(s) required

- Publication
- Poster
- Presentation
- Dataset
- Image
- Video/Audio

Type of publication: Book

Type of image: Figure

Basic information required

Digital Object Identifier: e.g. 10.1234/foo.bar

Optional: Did your publisher already assign a DOI to your upload? If not, leave the field empty and we will register a new DOI for you. A DOI allow others to easily and unambiguously cite your upload.

Pre-reserve DOI

Publication date: 2013-06-26

Required. Format: YYYYMM-DD. The date your upload was made available in case it was already published elsewhere.

Title

Required.

Authors: Family name, First name: Affiliation (one author per line)

Required. Format: Family name, First name: Affiliation (one author per line)

Description

Required.

Keywords: One keyword per line...

Optional. Format: One keyword per line.

Additional notes

Optional.

License required

Unless you explicitly specify the license conditions below for Open Access and Embargoed Access uploads, you agree to release your data files under the terms of the Creative Commons Zero (CC0) waiver. All authors of the data and publications have agreed to the terms of this waiver and license.

- Access right
- Open Access
 - Embargoed Access
 - Restricted Access
 - Closed Access

Required. Open access uploads have considerably higher visibility on ZENODO.

Embargo date: 2013-06-26

My uploads

Unsubmitted

Untitled 21 Jun 2013, 19:40

Submitted

JCR Journals, sorted by impact factor 2011 with the JCR edition indication 07 Jun 2013, 19:54

+ Figshare <http://figshare.com/>

- Es un proyecto que permite publicar los datos de investigación en abierto iniciado por un estudiante de doctorado y ahora está apoyado por Digital Science–Macmillan Publishers Company. Observamos que se va posicionando en el mercado de las aplicaciones, estando presente en el Forum establecido por Thomson Reuters y es tomada en cuenta su opinión en el DCC.
- Orientada a quien deposita la información, el investigador, preocupado especialmente por no perder el reconocimiento de su trabajo.
- Lo destacable de este repositorio son los tipos de licencias que utiliza: publica todo bajo la licencia CC-BY, pero distribuye los datasets con la CC0. También está asociado a CLOCKSS. Si los ficheros se depositan en abierto, el espacio disponible es ilimitado; en caso contrario proporciona un 1 Gb. Otro aspecto que cabe mencionar es que interopera con ORCID para el registro de un nuevo usuario. Una vez subido el fichero, del que se demandan escasos metadatos, no hay ninguna validación posterior



JCR Journals, sorted by Impact Factor 2011 wi...

Select type of file:

Dataset



JCRjourn.xls

[Replace](#)

[Download](#)

Title: JCR Journals, sorted by Impact Factor 2011 with t

Author(s): Fernanda Peset | Yolanda Blasco
Luis-Millán González

Categories: 2 selected

Tags: jcr | Indexed journals 2011 | Impact Factor
bibliometrics | journals

Description: **Description of the spreadsheet:**
"Journals in JCR sorted by IF'11" lists the journals from Thomson Reuters JCR website; it's sorted by edition (science and social science) and Impact Factor 2011 descending (but not difunded). Fields: Abbreviated Journal title, ISSN, JCR ed.

Methodology:

B *I* U

Links: <http://www.datasea.es>

Helpful tips

Use this form to edit all the information related to your data. Please be as descriptive as possible. You can always edit this information later.

This message will be replaced with helpful tips and suggestions as you will be interacting with the form.

Use the left or right arrows on your keyboard to navigate through articles.

[Preview article](#)

This article is public. Making changes now will generate a new version of this article.

[Save changes](#)



Dataverse <http://thedata.harvard.edu/dvn/>



- Alojado en el Institute for Quantitative Social Science (IQSS) de la Universidad de Harvard. Desde 2006 se encuentra disponible para todas las disciplinas, aunque la idea se remonta a 1987.



The Data Citation is generated automatically by the Dataservice based on the information entered below.

Title *	<input type="text"/>
Study Global ID *	hdl:1902.1/21779
Author *	Name <input type="text"/> Affiliation <input type="text"/> (LastName, FirstName)
Producer *	Producer <input type="text"/> Abbreviation <input type="text"/> Affiliation <input type="text"/> URL <input type="text"/> (http://example-domain.edu) Logo URL <input type="text"/> (http://example-domain.edu/image)
Production Date *	<input type="text"/> (YYYY or YYYY-MM or YYYY-MM-DD, AD or BC optional)
Funding Agency *	<input type="text"/>
Distributor *	Distributor <input type="text"/> Abbreviation <input type="text"/> Harvard Dataverse Network Affiliation <input type="text"/> URL <input type="text"/> http://thedata.harvard.edu/dvn/ (http://example-domain.edu) Logo URL <input type="text"/> http://thedata.harvard.edu/dvn/resources/images (http://example-domain.edu/image)
Contact *	Contact <input type="text"/> Affiliation <input type="text"/> E-mail <input type="text"/>
Distribution Date *	<input type="text"/> (YYYY or YYYY-MM or YYYY-MM-DD, AD or BC optional)
Deposit Date *	2013-07-05 (YYYY or YYYY-MM or YYYY-MM-DD, AD or BC optional)
Description and Scope	
Description *	<p> Copying and pasting from a Word document can create errors when you save this page.</p> <p>Description</p> <div style="border: 1px solid #ccc; height: 100px; width: 100%;"></div> <p>Description Date</p> <input type="text"/> (YYYY or YYYY-MM or YYYY-MM-DD, AD or BC optional)
Keyword *	Keyword <input type="text"/> Vocabulary <input type="text"/> URL <input type="text"/> (http://example-domain.edu)
Topic Classification *	Topic Classification <input type="text"/> Vocabulary <input type="text"/> URL <input type="text"/> (http://example-domain.edu)
Time Period Covered - Start *	<input type="text"/> (YYYY or YYYY-MM or YYYY-MM-DD, AD or BC optional)
Time Period Covered - End *	<input type="text"/> (YYYY or YYYY-MM or YYYY-MM-DD, AD or BC optional)
Geographic Coverage *	<input type="text"/>
Geographic Unit *	<input type="text"/>





Introducción datos de investigación

Conocimientos mínimos

Conclusiones



Royal Society y ThomsonReuters



- Desechar una cultura donde se contemplan los datos como un bien privado
- Aumentar los criterios usados en la evaluación de la carrera de los investigadores para reconocer el esfuerzo que supone la liberación de datos útiles y las nuevas formas de colaboración, tal y como recomienda la *San Francisco Declaration on Research Assessment, Putting science into the assessment of research*
- Desarrollar estándares comunes para comunicar esos datos, también incluyendo criterios de calidad de los datos y cuestiones relacionadas con las licencias y derechos



Royal Society y ThomsonReuters



- Ordenar mandatos que promuevan la liberación de los datos relevantes de las publicaciones científicas
- Reforzar el número de científicos de datos que son necesarios para gestionar y apoyar el uso de los datos digitales (lo que es también crucial para el sector privado de análisis de los datos y las estrategias open data de los gobiernos), sobre todo en las tareas de evaluación, que garantizan la calidad de los datos
- Desarrollar y utilizar nuevas herramientas de software para automatizar y simplificar la creación y explotación de conjuntos de datos, proporcionando un acceso uniforme a la variedad de *outputs* de la investigación

Recomendaciones concretas

- servicios de consultoría para el desarrollo de los planes de gestión,
- acceso web a las políticas y buenas prácticas,
- información y apoyo práctico sobre infraestructura de almacenamiento y repositorios de datos,
- acceso a los recursos informáticos y herramientas para gestionar sus propios datos,
- acceso a los datos de investigación...



Nada más

¿Cómo lo ven ustedes?

<http://www.datasea.es>

mpesetm@upv.es en nombre del

Grupo DataSEA

■ Peset, F. (2014). Datasea.

<https://polimedia.upv.es/visor/?id=a48abcad-21bc-ab4a-ac6a-af987b3e7f1a>



The image shows a video player interface with a slide titled "DATASEA - Research Open Data - Open Research Data". The slide content includes:

- Best methods for sharing and publishing research data**
- OPEN DATA RESEARCH**: Encouraging research sharing with scientific freedom and reproducibility.
- OPEN DATA**: Guidelines and Advisory for Spanish agents: researchers, libraries, journals, publishers, managers and editors, users.
- Open Journals**: Finding which journals or accepting data are "Open Access" journals.
- Funding**: CERCA, OpenAccessBarCELONA.
- Funded by**: Spanish Ministerio de Economía and Innovation, CERCA, OpenAccessBarCELONA.
- Contact**: open@upv.es, <http://www.datasea.es>

The video player has a play button in the center and a "poliMedia" logo in the bottom left corner. A woman is visible in the bottom right corner of the video frame.