



COMPARTIR LOS DATOS DE INVESTIGACIÓN EN CIENCIA: INTRODUCCIÓN AL *DATA SHARING*



Daniel Torres-Salinas, Nicolás Robinson-García y Álvaro Cabezas-Clavijo



Daniel Torres-Salinas, doctor en documentación científica, trabaja como técnico de gestión de la investigación en la *Universidad de Navarra*, donde realiza auditorías sobre la calidad y el impacto de la investigación desde una perspectiva bibliométrica. Asimismo es miembro del grupo *EC3 (Evaluación de la Ciencia y de la Comunicación Científica)* de la *Universidad de Granada* donde participa en diferentes proyectos. Es miembro del think tank *ThinkEPI*.

Universidad de Navarra
Centro de Investigación Médica Aplicada
Avda. Pío XII, 55. 31008 Pamplona
torressalinas@gmail.com

Nicolás Robinson-García es licenciado en documentación y máster en información científica por la *Universidad de Granada*. Es miembro del grupo *EC3 (Evaluación de la Ciencia y de la Comunicación Científica)* de la misma universidad, donde se encuentra haciendo el doctorado. En la actualidad sus intereses se centran en la investigación sobre bibliometría y la evaluación de la calidad investigadora.

Universidad de Granada
Facultad de Comunicación y Documentación
Colegio Máximo de Cartuja - Campus de Cartuja, s/n. 18071 Granada
elrobin@ugr.es

Álvaro Cabezas-Clavijo es licenciado en documentación y máster en información científica por la *Universidad de Granada*, donde prepara su tesis doctoral. Forma parte del grupo de investigación *EC3 (Evaluación de la Ciencia y de la Comunicación Científica)* en el que realiza evaluaciones bibliométricas de agentes investigadores y estudia el impacto de las herramientas de la web 2.0 en la actividad científica.

Universidad de Granada
Facultad de Comunicación y Documentación
Colegio Máximo de Cartuja - Campus de Cartuja, s/n. 18071 Granada
acabezasclavijo@gmail.com

Resumen

Se analiza el movimiento conocido en el mundo científico como *data sharing*, consistente en compartir los datos finales de investigación entre los científicos, con el objetivo de maximizar esfuerzos y recursos. En primer lugar, se revisa el concepto de datos de investigación y las dificultades técnicas que plantean según la disciplina de la que provengan. Se examinan las motivaciones, el origen del movimiento *data sharing* y la magnitud que está alcanzando dentro de la comunidad científica mediante la creación de repositorios y bancos de datos, planteando retos tanto técnicos como sociales. A continuación, se estudian las iniciativas y las políticas editoriales que las agencias financiadoras y las revistas científicas están llevando a cabo para incentivar esta práctica. Finalmente se examina el impacto que este cambio en los hábitos de los investigadores tiene dentro de la profesión bibliotecaria, dando lugar a la aparición de nuevos perfiles profesionales.

Palabras clave

Data sharing, Datos de investigación, Acceso abierto, Investigadores, Revistas científicas, Repositorios.

Title: Sharing scientific research data: introduction to *data sharing*

Abstract

The emergence in the scientific community of an initiative known as data sharing, consisting of sharing research data among researchers and aiming to maximize efforts and resources, is analysed. First, the concept of research data and the related

Artículo recibido el 03-10-11
Aceptación definitiva: 16-12-11

technical difficulties depending on the discipline are reviewed. We also examine the motivations, origins and growth of this movement, which has had an important impact on the scientific community's behaviour through the creation of repositories and data banks, raising both technical and social challenges. Then we discuss leading funding agencies' initiatives and scientific journals' editorial policies promoting these practices. Finally, we examine the impact these major changes in researchers' habits have for librarians, including the emergence of new professional profiles.

Keywords

Data sharing, Research data, Open access, Researchers, Scientific journals, Repositories.

Torres-Salinas, Daniel; Robinson-García, Nicolás; Cabezas-Clavijo, Álvaro. "Compartir los datos de investigación: introducción al *data sharing*". *El profesional de la información*, 2012, marzo-abril, v. 21, n. 2, pp. 173-184.

<http://dx.doi.org/10.3145/epi.2012.mar.08>

1. Introducción

Durante la última década la comunidad científica está debatiendo intensamente la necesidad de compartir en acceso abierto los datos fruto de las investigaciones para que éstos puedan ser reutilizados con diferentes propósitos por el resto de investigadores, una tendencia que se conoce en inglés como *data sharing*. La disponibilidad de los datos es uno de los pilares esenciales de la e-Ciencia (Shneiderman, 2008), por ello, las más prestigiosas revistas del mundo han dedicado a esta cuestión números especiales (*Nature*, 2009; *Science*, 2011) y son variadas las editoriales sobre el tema (por ejemplo: Kirwan, 1997; Nagelkerke; Bernsen; Rizk, 2007; Groves, 2009). Asimismo, las más importantes y diversas instituciones científicas^{1, 2, 3}, así como diferentes organismos nacionales y supranacionales (Axelson; Schroede, 2009) vienen desarrollando el marco legal, las políticas y las infraestructuras necesarias para poder compartir los datos adecuadamente debido, sobre todo, a las enormes ventajas y oportunidades que puede generar la adopción generalizada de una actitud de este tipo por parte de los científicos.

Los beneficios del *data sharing* están bien identificados y concretados (Arzberger *et al.*, 2004; Vickers, 2006). En primer lugar en el contexto de crisis económica actual, es necesario buscar un mayor rendimiento del dinero invertido en ciencia (OECD, 2010) y prácticas de este tipo contribuirían a un mayor aprovechamiento de los recursos, especialmente en aquellos proyectos financiados con fondos públicos. La libre disposición de los datos permitiría realizar nuevos estudios, por ejemplo metaanálisis (Ramamy *et al.*, 2008), al tiempo que supondría un ahorro de costes ya que podría acabarse con la duplicación de proyectos destinados a obtener resultados similares. Junto a la optimización de la financiación, compartir los datos facilitaría una mayor transparencia por parte de los investigadores sobre todo en la lucha contra el fraude, ya que permitiría replicar fácilmente los experimentos y la verificación de hipótesis (Renolls, 1997).

“ El *data sharing* permitiría obtener un mayor rendimiento del dinero invertido en ciencia ”

Además de esos beneficios generales, representa un beneficio a nivel personal, puesto que esta práctica contribuye a

augmentar el número de citas de los trabajos (Piwowar; Day; Fridsma, 2007).

A este conjunto de ventajas podemos añadir dos circunstancias del contexto científico actual que han permitido y están fomentando su expansión. Por un lado, los grandes avances tecnológicos que han tenido lugar en los últimos años ofrecen una posibilidades que hasta hace unas décadas eran impensables (Pennisi, 2011); en segundo lugar, la tendencia de los científicos a mostrar una actitud más abierta respecto a sus hallazgos ha ido calando progresivamente gracias sobre todo a movimientos como el *open access*, con casos de éxito rotundos como los de los repositorios –y sobre todo *arXiv* (Ginsparg, 2011)–, o más recientemente el de la Ciencia 2.0 (Cabezas-Clavijo; Torres-Salinas; Delgado-López-Cózar, 2010).

“ Los científicos tienden a mostrar una actitud más abierta respecto a sus trabajos gracias sobre todo a movimientos como el OA ”

Sin embargo, aunque tanto las ventajas como los factores señalados anteriormente han provocado que el debate sobre *data sharing* se encuentre en su máximo apogeo, hay que matizar que esta cuestión siempre ha estado presente en la ciencia, pudiéndose rastrear sus inicios hasta 1901. Así Galton, a quien también se asocia al nacimiento de la cienciometría (Godin, 2007), se expresaba en la revista *Biometrika* en los siguientes términos (Pernerger, 2011): “Pienso que nadie debiera publicar resultados biométricos sin depositar una copia de sus datos bien redactada y presentada en algún lugar donde todo aquel que lo deseara pudiera verificar su trabajo”⁴. Más allá de las declaraciones, el acontecimiento más reseñable y el hito que verdaderamente pone en marcha una forma eficiente de compartir y distribuir datos es la creación de los bancos de datos para esta finalidad. Esto sucedió a partir 1971 con la creación del *Protein data bank*⁵ (PDB) bajo los auspicios de la *American Crystallographic Association* (Crawford; Hurd; Weller, 1996; Berman, 2007). Así, el propio desarrollo del PDB puede ilustrar la evolución del *data sharing*, puesto que desde su creación este banco de datos ha mantenido un crecimiento exponencial pasando de apenas 500 estructuras a finales de los ochenta a más de 75.000⁶ en 2010 (figura 1).

Basándose en el modelo del *PDB* han ido surgiendo repositorios similares en diversas disciplinas como por ejemplo *Neurodatabase* (neurociencias)⁷, *The cancer genome atlas* (Oncología)⁸ o el *Data observation network for Earth* (ciencias de la Tierra)⁹. La razón del éxito de estos repositorios disciplinares no sólo radica en el compromiso y la necesidad común de los investigadores de acelerar el progreso científico sino que detrás de estas infraestructuras se esconden años de esfuerzo de otros agentes como son las propias agencias de financiación o las sociedades y revistas científicas; todas contribuyen sustancialmente para que un banco de datos se convierta en la herramienta de trabajo estándar de una disciplina.

Un claro ejemplo de estos esfuerzos son por ejemplo las *Bermuda rules* de 1996 para liberar los datos sobre el genoma humano y que dio lugar al *GenBank*¹⁰ (Marshall, 2001). Estas circunstancias han provocado que en determinados frentes de investigación como por ejemplo aquellos que hacen uso de las expresiones génicas, un 45% de los trabajos hagan públicos sus datos (Piwowar, 2011). Sin embargo, aunque hay ejemplos de éxito que afectan a especialidades muy concretas, éstos son la excepción más que la regla (Nelson, 2009) y en el resto de la ciencia aún nos encontramos en una fase muy embrionaria, algo lógico por otra parte, ya que la cuestión de compartir datos es mucho más compleja y poliédrica de lo que pudiera parecer en principio.

El *data sharing* tendrá todavía mayor proyección y determinará en buena medida nuestro futuro profesional, especialmente en bibliotecas universitarias

Al igual que ocurre en la ciencia, en el ámbito de la documentación científica el *data sharing* también está consolidándose con fuerza sobre todo en relación con los roles y funciones que podemos desempeñar los profesionales de la información (Torres-Salinas, 2010a), por lo que términos como el *data curation* o los *data center* (Martínez-Uribe; MacDonald, 2008; RIN, 2011) empiezan a sernos más que habituales. Por tanto consideramos que nos encontramos ante un tema que tendrá todavía mayor proyección y que determinará en buena medida nuestro futuro profesional, especialmente en ámbitos concretos como las bibliotecas universitarias (Newton; Miller; Bracke, 2010). No obstante, aún no existe en nuestra bibliografía profesional ningún trabajo que ofrezca una visión general del *data sharing* por lo que el objetivo principal del texto que se presenta es introducir de forma básica este movimiento. Para ello el artículo se organiza como sigue: en el próximo apartado nos aproximamos a la definición de datos de investigación para, a continuación analizar los medios y los modos a través de los que se comparte; en el siguiente apartado nos centraremos en el rol de diferentes agentes en el desarrollo e impulso

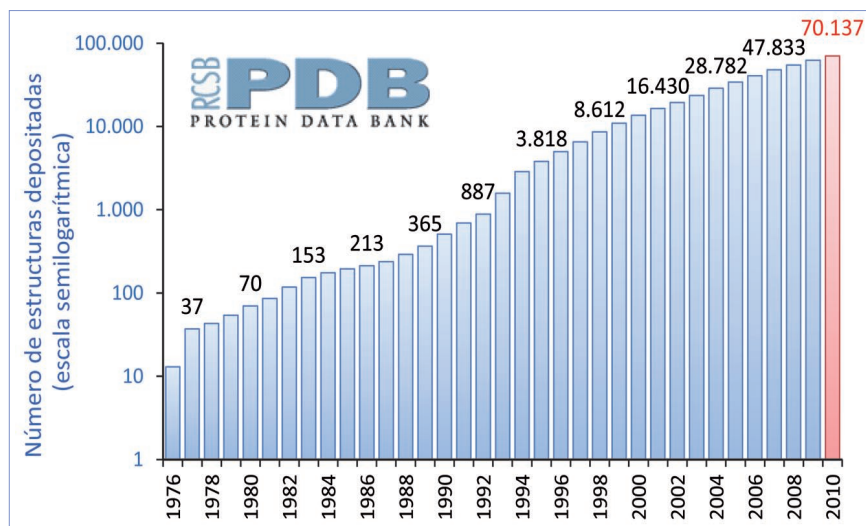


Figura 1. Evolución del número de estructuras depositadas en el *Protein data bank* en el período 1976-2010

al *data sharing*, para finalmente concluir con una reflexión sobre las dificultades técnicas para llevarlo a cabo y reiterar la necesidad de que los documentalistas afronten estos nuevos retos.

2. Hacia una definición de dato de investigación

Uno de los aspectos fundamentales para compartir de forma efectiva los datos es conocer con claridad qué se entiende por dato de investigación, ya que existen múltiples definiciones y clasificaciones de los mismos. En lo que a definiciones respecta, la que más consenso aún al haber sido adoptada por entidades como los *National Institutes of Health (NIH)* de Estados Unidos¹¹ o la *OECD* (2007), es la que considera datos de investigación 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. Además de estas características también se especifica que debe provenir de una fuente única y deben ser difíciles o imposibles de obtener de nuevo por ser propios de un momento o circunstancias irrepetibles de una forma exactamente igual. Todos los datos que cumplen las características anteriores se conocen como datos finales de investigación. Por tanto, fuera de esta definición quedarían materiales que, aunque también se producen dentro del proceso de investigación, no se tienen en cuenta, como es el caso de las notas de laboratorio, análisis preliminares, borradores, informes o conversaciones informales con otros colegas.

Junto a esta definición existen algunos intentos de hacer una taxonomía de los datos de investigación. Una de las clasificaciones más elementales es la que se refiere a los propios formatos (textual, imágenes, vídeo, etc.), sin embargo es poco operativa para determinar qué compartir. La clasificación que ofrece la *Research Information Network (RIN)* (Swan; Brown, 2008) sí es más operativa en este sentido y establece tres criterios no excluyentes para determinar qué son los datos de investigación y cómo clasificarlos:

1. Según su proceso de obtención

Dentro de este grupo se pueden establecer tres tipos: experimentales, simulaciones y observaciones. Los datos experi-

mentales serían los obtenidos como fruto de un experimento normalmente reproducible como por ejemplo secuencias genéticas o cromatografías. Por simulaciones entenderíamos todos los generados como resultado de la aplicación de un modelo (climatológico, económico, etc.), donde los metadatos y el modelo en sí suelen ser más interesantes que los propios datos obtenidos. Finalmente, datos de observación serían resultado de la observación directa de un fenómeno en tiempo real, y son únicos e irremplazables como por ejemplo los resultados de una encuesta.

2. Según el objetivo de su recogida

Quedarían divididos en tres tipos: específicos, entendiendo como tales aquellos que responden a las necesidades de un proyecto de investigación concreto y cuyo valor es muy limitado más allá de los objetivos iniciales por los que son recogidos; datos de alcance medio, es decir, los de los que puede beneficiarse una comunidad muy limitada de investigadores pertenecientes a una determinada especialidad; y por último, datos de interés general. Se trataría de aquellos cuya importancia es vital para el desarrollo de la ciencia y que despierta gran interés en toda la comunidad científica.

3. Según el tratamiento que hayan recibido

Atendiendo a la fase de la investigación se distingue entre datos preliminares y datos finales. Los primeros serían los recién extraídos que no han recibido ningún tipo de tratamiento por parte del investigador (*raw data*). Los datos finales corresponden con lo que los *NIH* definen también como *final research data* y que son el resultado de combinar o procesar los datos preliminares. Siguiendo el criterio del tratamiento a lo largo del ciclo de investigación, **Green** (2009) distingue cinco tipos de datos: preliminares, elaborados y listos para su uso, preparados de acuerdo con un estándar para ser compartidos, datos tabulados y datos derivados.

Establecer de forma genérica qué se debe compartir o no debe hacerse dentro de las propias disciplinas, y además teniendo en cuenta los métodos de investigación utilizados

Aunque estas definiciones y taxonomías nos ayudan a establecer el marco en el cual se pueden generar los datos y algunas de sus características, si tuviéramos en cuenta todas las especialidades del conocimiento existiría una enorme variedad de ellos que quedarían fuera de las mismas. Hemos de tener en cuenta que algunos de los datos son claramente identificables como por ejemplo los rayos-X en medicina, los estudios espectrales en astronomía, las estructuras de proteínas en bioquímica o los datos registrados por los sensores en ecología. No en vano uno de los problemas de la definición de los *NIH*, por la propia naturaleza de la institución, es que está muy apegada a la investigación biomédica donde es relativamente fácil contextualizar qué son los datos. En el otro extremo estarían las humanidades, donde según **Borgman** (2008) cualquier registro fruto de la experiencia humana es susceptible de ser un dato: desde un

fragmento de tablilla sumeria hasta el registro de una entrevista antropológica, aunque evidentemente en nuestro contexto sólo nos valdría aquello que estuviera digitalizado. Por ello, establecer de forma genérica y común qué se debe compartir o no debe hacerse dentro de las propias disciplinas, y además teniendo en cuenta los métodos de investigación que los generan ya que los datos van estrechamente asociados a los mismos.

3. Cómo se comparten los datos de investigación

El panorama actual ofrece un amplio abanico de formas para compartir datos entre los investigadores. A un nivel muy general se pueden distinguir dos modalidades en función del tipo de canal empleado: formal o informal. El primero lo constituyen los ya mencionados bancos de datos (*data-banks*) que son la herramienta fundamental así como el objetivo a alcanzar en toda disciplina. Su filosofía es similar a la de los repositorios de trabajos científicos, esto es: una base de datos en línea, que puede tener diferentes niveles de acceso, donde los datos son depositados y descritos conforme

A. Definición de datos de los National Institutes of Health	
Por datos finales de investigación entendemos material factual registrado, aceptado por la comunidad científica y necesario para validar los resultados de la investigación. No son datos finales de investigación: notas de laboratorio, sets de datos parciales, análisis preliminares, borradores de trabajos científicos, planes para investigaciones futuras, informes que han tenido un proceso de revisión por pares, comunicaciones con colegas, u objetos físicos como geles o ejemplares de laboratorio.	
B. Diferentes taxonomías de los datos de investigación	
B.1 Según el formato	B.2. Proceso de obtención
<ul style="list-style-type: none"> - Textos - Números - Imágenes - etc. 	Experimentales <ul style="list-style-type: none"> - Secuencias genéticas - Cromatografías Simulaciones <ul style="list-style-type: none"> - Modelos climáticos - Modelos económicos Observacionales <ul style="list-style-type: none"> - Encuestas - Experimentos irrepetibles
B.3. Según objetivo recogida	B.4. Según fase de investigación
Específicos <ul style="list-style-type: none"> - Solo de interés para un proyecto de investigación. Alcance medio <ul style="list-style-type: none"> - De interés para una disciplina concreta. De interés general <ul style="list-style-type: none"> - De interés para la ciencia en su conjunto e incluso de interés social. 	Datos preliminares <ul style="list-style-type: none"> - Datos recién extraídos sin ningún tipo de procesamiento. Denominados en inglés <i>raw data</i>. Datos finales <ul style="list-style-type: none"> - Datos que ya han sido procesados y combinados con otros. Denominados en inglés <i>final research data</i>.
C. Medios para compartir los datos de investigación	
C.1. Comunicación formal	C.2. Comunicación informal
Repositorios de datos / bancos de datos <ul style="list-style-type: none"> - Centralizados - Descentralizados - Federados - Ciberestructuras 	A petición <ul style="list-style-type: none"> - Por ejemplo vía correo electrónico Descentralizada <ul style="list-style-type: none"> - Por ejemplo a través de webs personales

Tabla 1. Definición, taxonomías y formas de compartir los datos de investigación

a un estándar quedando listos para su posterior identificación y recuperación por otros científicos. A diferencia de los repositorios de pre-prints/post-prints, que tienen un único formato bibliográfico, existe una gran variedad de bancos de datos y las soluciones adoptadas en cada ámbito son muy diversas, por lo que requieren conocimientos del mismo para poder usarlos. Asimismo también se diferencian en una mayor complejidad, tanto en el depósito como en la recuperación, y el usuario final debe especializarse en ellos. Esta situación no impide que los bancos de datos estén ahora en pleno desarrollo y que, como se muestra en la tabla 2, diversos campos del conocimiento cuenten ya con un repositorio de referencia. Pero no existe una solución única y así por ejemplo, al margen de la disciplina para la que se creen, es interesante la clasificación de **Kowalczyk (2011)**, que distingue entre cuatro modelos que responden a la propiedad de los datos depositados: centralizados, descentralizados, federados y ciberestructuras.

El *modelo centralizado* se corresponde con los bancos de datos bajo el control de una sola institución (universidad, centro de investigación, etc.) y presenta el problema de la

dependencia de la institución que alberga los datos y del monopolio que ésta pueda ejercer sobre los mismos. Para evitar esto, se presenta el *modelo descentralizado*, en el cual el mantenimiento y la financiación dependen de varias instituciones, aunque existe una única ubicación física de los datos, por lo que también se podrían derivar problemas de control. Sin embargo, en el *modelo federado* también participan distintas instituciones pero ahora los datos están físicamente distribuidos entre los participantes, aunque virtualmente sean accesibles a través de una plataforma común. El último caso, la *ciberestructura*, representa un paso más allá ya que la estructura tecnológica es más compleja sirviéndose como modelo de la propia configuración de internet, donde los datos se encuentran distribuidos en redes de ordenadores en múltiples ubicaciones y permanentemente accesibles.

A las formas reseñadas además habría que unir todos los tipos de repositorios que se pueden formar en función del formato de datos, que como comentamos en el apartado anterior, dependen tanto de las disciplinas como de los métodos de investigación que los generan. Por ejemplo (tabla

Nombre del banco de datos	Área	Objetivo y tipo de datos compartidos	Estadísticas
<i>Worldwide protein data bank</i>	Proteómica	Estructuras de macromoléculas, gratuito y accesible a todo el mundo.	72.749 depósitos entre los años 2000-2011 en tres bases de datos: 54.153 en RCSB PDB, 7.966 en PDBj y 10.630 en PDBe
<i>Biological magnetic resonance data bank</i>	Biología	Colecciona, anota, archiva y difunde los datos cuantitativos de investigaciones espectroscópicas de macromoléculas biológicas y metabolitos. Es de dominio público.	Combinaciones genéticas 4.852.279 (proteínas), 24.441 (ADN) y 40.233 (ARN)
<i>GenBank</i>	Genómica	Secuencias genéticas. Producido por los NIH, es de acceso público.	126.552 millones de bases en 135 millones de secuencias en <i>GenBank</i> y 191.402 millones de bases en 62 millones de registros de secuencias en la división WGS
<i>UniProtKB/Swiss-prot</i>	Proteómica	Secuencias proteicas y funcionales, gratuito.	531.473 secuencias que suponen 188 millones de aminoácidos resumidos en 200.346 referencias
<i>ArrayExpress</i>	Genómica	Experimentos genómicos que incluyen expresiones génicas. Sigue los estándares <i>Miame</i> y <i>Minseqe</i> .	863.732 experimentos y ensayos
<i>Dryad</i>	Biociencias	Datos referenciados en artículos científicos de biociencias aplicadas y básicas. Permite validar artículos publicados, explorar nuevas metodologías de análisis, reutilizar datos para nuevas investigaciones que pasaron por alto los autores originales, y para realizar análisis sintéticos.	902 paquetes de datos y 2.157 ficheros de datos
<i>ClinicalTrials.gov</i>	Medicina	Registros y resultados de ensayos clínicos realizados en los EUA y en otro países financiados con dinero federal o privado.	113.224 ensayos
<i>Proteomics identifications database</i>	Proteómica	Repositorio público de proteínas y péptidos, con las evidencias que justifican sus identificaciones.	18.137 experimentos; 6,3 millones de proteínas; 32,5 millones de péptidos; 3,8 millones de péptidos únicos; y 199 millones de espectros
<i>Pangaea</i>	Geociencias	Datos georeferenciados sobre investigaciones geológicas, en acceso abierto.	184 proyectos de investigación; entre ellos 55 proyectos europeos
<i>Simbad astronomical database</i>	Astronomía	Información básica, identificaciones cruzadas, bibliografía y medidas de objetos astronómicos de fuera del sistema solar.	5.433.523 objetos, 15.198.239 identificadores, 256.720 referencias bibliográficas y 8.272.599 citas de objetos en publicaciones
<i>Species 2000 - Catalogue of life</i>	Biología	Federación de organizaciones de bases de datos que trabajan estrechamente con sus usuarios, taxonomistas y agencias financiadores. El objetivo es crear una lista validada de especies del mundo (plantas, animales, hongos y microbios).	1.368.009 especies de 100 bases de datos

Tabla 2. Ejemplo de algunos de los principales repositorios de datos empleados en diferentes disciplinas científicas

2) *Pangaea* busca una “lista validada de especies del mundo” mientras que *ArrayExpress* tiene por objetivo el almacenamiento de expresiones génicas. Evidentemente, el tipo de dato generado en una y otra difiere totalmente por lo que da lugar a repositorios ad hoc y hace difícil trasladar un tipo de banco de datos de una disciplina científica a otra. Una situación muy diferente de la ocurre con los repositorios de documentos bibliográficos, que cuentan con soluciones parecidas para su creación como *DSpace* o *E-Prints* (Tramullas; Garrido-Picazo, 2006).

Aunque los repositorios son el eje y motor en torno al cual se articula el *data sharing*, no son el único medio del que disponen los investigadores. Mucho antes de la existencia de los mismos, la práctica de compartir datos estuvo presente en las pautas de comportamiento de la comunidad científica, al ser inherente a la propia ciencia (Hrynaskiewicz; Altman, 2009). Tradicionalmente se han venido compartiendo datos a través de canales no controlados y sin un vehículo preestablecido. Además no todas las disciplinas cuentan con un repositorio, especialmente en ciencias sociales y humanas, por lo que es habitual que se compartan los datos por vías informales. Piwowar (2008a) señala dos canales informales básicos:

- a petición, es decir, respondiendo a demandas concretas de otros colegas y sin llegar a ser compartidos en ningún caso de manera pública;
- manera descentralizada, colgando los datos en las webs personales de los investigadores o de los grupos de investigación.

La valorización de los bancos de datos se debe a dos agentes principales: los organismos financiadores de la investigación y las editoriales científicas

De esta última modalidad se deduce un grave impedimento que obstaculiza la reutilización permanente de los datos, y es el referente al formato en que se presentan ya que su obsolescencia puede complicar la conservación y disponibilidad de los datos en el futuro. Por otro lado, los datos compartidos de manera informal cumplen únicamente el criterio de normalización del autor, que a su vez se adecua a su forma de trabajo, impidiendo en muchos casos que puedan ser empleados con provecho.

Compartir datos a través de canales informales no deja de ser una solución circunstancial y por ello, los repositorios son la solución idónea. Sin embargo, en el caso de estos últimos, es difícil establecer y consensuar directrices para su utilización, por lo que hace necesario el concurso de otros agentes que promuevan, favorezcan y gestionen su uso.

4. Rol de las agencias de financiación y las revistas científicas

El valor de los bancos de datos como la vía óptima para compartir datos y sobre todo el reconocimiento que éstos han

tenido en diversas comunidades científicas altamente especializadas se puede atribuir, gracias al establecimiento de políticas y al fomento de su uso, a dos agentes principales: los organismos financiadores de la investigación y las editoriales científicas. Los primeros, casi todos ellos de carácter público, empiezan tímidamente a tomar conciencia desde mediados de los noventa cuando el *Economic and Social Research Council* del Reino Unido establece las políticas y estándares a seguir para compartir datos: *ESRC data policies and standards*¹² (1994, actualizado en 2000).

Este tipo de documentos empiezan a proliferar especialmente a partir del año 2000, gracias sobre todo al publicado por un organismo tan influyente como la *National Science Foundation*, titulado *NSF data sharing policy and data management plan requirements*¹³. Desde entonces los diferentes centros de financiación, casi siempre de Estados Unidos y Reino Unido, han ido lanzando y actualizando sus políticas tal y como se muestra en la recopilación realizada en la tabla 3.

Paralelamente se han publicado directrices y recomendaciones dirigidas a los propios investigadores para ilustrarles cómo deberán compartir sus datos de investigación (*UK Data Archive*, 2011). En este sentido, la razón que justifica compartir los datos desde el punto de vista institucional es la misma por la cual se exige compartir los trabajos de investigación. Los datos que provienen de proyectos financiados con dinero público también son resultado de la investigación y por tanto, al igual que los artículos publicados, deben hacerse públicos.

Destaca sobre todo la política emprendida por los *NIH*, organismo encargado de distribuir los recursos económicos para la investigación biomédica en Estados Unidos, que fueron los pioneros en establecer prácticas de obligado cumplimiento. En 2003 dictaron una firme política a favor del *data sharing* (*NIH*, 2003), animando a todo investigador que solicitaba un proyecto a declarar cuál sería su plan para poner en acceso abierto los datos que se derivaran del mismo, siendo éste un requisito imprescindible para aquellos que solicitaran subvenciones por encima del medio millón de dólares. Además, fijaban como fecha límite para la liberación de los datos el día en el que una revista aceptara el artículo con los resultados de la investigación. Esta política obligó además a los *NIH* a crear diversos repositorios y herramientas para dar cobertura y ayudar al cumplimiento de sus propias exigencias¹⁴. En concreto, los artículos deben depositarse en *PubMed Central*.

Los datos que provienen de proyectos financiados con dinero público también son resultado de la investigación y por tanto, al igual que los artículos, deben hacerse públicos

En España existen pocos ejemplos referentes a políticas de gestión de datos científicos, y cabe resaltar tan sólo el mandato existente en el ámbito de la investigación antártica.

Organización	País	Rol	Normativa	Año
<i>Biotechnology and Biological Sciences Research Council</i>	Reino Unido	Financiación	<i>Bbsrc data sharing policy</i>	2007
<i>Cancer Research UK</i>	Reino Unido	Financiación	<i>Policy on data sharing and preservation</i>	2009
<i>Center for Drug Evaluation and Research</i>	Reino Unido	Regulador	<i>CDER data standards plan, version 1.0</i>	2010 (desde 2003)
<i>Centro Nacional de Datos Polares</i>	España	Coordinador	<i>Protocolo de remisión, almacenamiento y difusión de datos antárticos</i>	2004
<i>Department of Energy</i>	Estados Unidos	Financiación	<i>Genomics: GTL program information and data sharing policy</i>	2008
<i>Economic and Social Research Council</i>	Reino Unido	Financiación	<i>ESRC data policies and standards</i>	2000 (desde 1994)
<i>Engineering and Physical Sciences Research Council</i>	Reino Unido	Financiación	<i>EPSRC policy framework on research data</i>	2011
<i>European Commission</i>	Unión Europea	Financiación	<i>Communication calling for uniform policies across member nations</i>	No se aplica
<i>European Science Foundation</i>	Unión Europea	Financiación	<i>ESF provee financiación principalmente para crear equipos</i>	No se aplica
<i>Genome Canada</i>	Estados Unidos	Financiación	<i>Genome Canada data release and resource sharing policy</i>	2008 (desde 2005)
<i>Gordon and Betty Moore Foundation</i>	Estados Unidos	Financiación	<i>GBMF data sharing philosophy and plan</i>	2008 (desde 2005)
<i>Medical Research Council</i>	Reino Unido	Financiación	<i>MRC data sharing and preservation policy</i>	2006
Diferentes organizaciones	Global	Diversos	<i>Sharing research data to improve public health: joint statement of purpose</i>	2011
<i>National Institutes of Health</i>	Estados Unidos	Financiación	<i>NIH data sharing policy</i>	2003
<i>National Science Foundation</i>	Estados Unidos	Financiación	<i>NSF data sharing policy and data management plan requirements</i>	2001
<i>Natural Environmental Research Council</i>	Reino Unido	Financiación	<i>NERC data policy</i>	2008 (desde 1996)
<i>Wellcome Trust</i>	Reino Unido	Financiación	<i>WT policy on data management and sharing</i>	2010
<i>Wellcome Trust Sanger Institute</i>	Reino Unido	Investigación	<i>Data sharing policy and guidelines</i>	2009 (desde 1998)

Basado en: *Biosharing* <http://biosharing.org> y *Centro Nacional de Datos Polares* <http://hielo.igme.es>

Tabla 3. Principales políticas de *data sharing* de organismos de investigación

Así, en el *Protocolo de remisión, almacenamiento y difusión de datos antárticos*, redactado por el *Comité Polar Español* se explicita que “los investigadores principales de los proyectos polares deberán enviar los datos brutos y los de las calibraciones instrumentales para su archivo en el *Centro Nacional de Datos Antárticos*, en un plazo máximo de tres meses posteriores a la campaña finalizada [...]. La no remisión de dichos datos hará incompatible al investigador principal para nuevos proyectos antárticos” (*Centro Nacional de Datos Polares*, 2004). Estos datos se almacenan y gestionan en el *Centro Nacional de Datos Polares*¹⁵, creado en 2004 (**Bermúdez; Barragán; Alonso**, 2011).

Por otro lado y al margen de los contextos nacionales, en la esfera internacional quien más ha impulsado el *data sharing* ha sido la *OECD (Organización para la Cooperación y el Desarrollo Económicos)*. Como resultado de una reunión de París en 2004 (*OECD*, 2007; **Arzberger et al.**, 2004) dicho organismo hizo públicas los *OECD Principles and guidelines for access to research data from public funding* (2007), donde se fijan recomendaciones y directrices para gobiernos y agencias públicas de investigación, que sirven como base de

sus políticas de acceso abierto. En el caso de la Unión Europea, la preocupación por la preservación de los datos se manifiesta a partir de 2007, con la comunicación de la *Comisión Europea (CE) On scientific information in the digital age: access, dissemination and preservation*¹⁶. En ella se exponen proyectos específicos dedicados al tema como *Caspar (Cultural, artistic and scientific knowledge for preservation, access and retrieval)*¹⁷, que estudia cómo se acceden y preservan los datos, o *Seadatanet*¹⁸, que permitió la creación de la infraestructura para compartir datos marinos de más de 40 países. Asimismo, la *CE* ha impulsado a través del *7º Programa marco* la cofinanciación de repositorios convencionales y de datos, dando luz verde a proyectos como *ODE (Opportunities for data exchange)*¹⁹, que analiza las opiniones de expertos y percepciones de los investigadores sobre *data sharing*; *Aparsen (Alliance for permanent access)*²⁰, una red de excelencia que centraliza los esfuerzos investigadores y garantiza la preservación de los datos científicos; *Parse. Insight (Permanent access to the records of science in Europe)*²¹, que establece el plan de actuación que deberá seguir la Unión Europea en esta materia; y por último, el *Scidip-es*

(*Science data infrastructure for preservation - Earth science*)²², un proyecto centrado en establecer la infraestructura necesaria para permitir compartir datos de investigación en ciencias de la Tierra y garantizar su preservación basándose en el modelo presentado en el proyecto *Caspar*.

A pesar de estos intentos, en gran parte de las agencias de investigación de la UE todavía no existen políticas bien documentadas (Field *et al.*, 2009). No obstante ya se están marcando los objetivos que la UE persigue en este sentido, con la publicación del informe *Riding the wave: how Europe can gain from the raising tide of scientific data*²³, en el que justifica la necesidad de establecer una política que permita maximizar los beneficios en materia científica mediante la creación de una infraestructura común que permita compartir datos científicos. Del mismo modo, señala los grandes retos que se le presentan y establece el año 2030 como fecha límite para alcanzar dichos objetivos.

Junto a estos organismos existen otros actores relevantes que están fomentando un cambio de actitud entre los investigadores y éstas son las revistas científicas. Actualmente las grandes publicaciones científicas (*Nature*, *Science*, *British medical journal*, etc.) comienzan a incidir con mayor o menor ahínco en sus instrucciones a autores en la necesidad, y en algunos casos en la obligación, de que los autores compartan los datos (tabla 4). Casi siempre la decisión de una revista de apoyar la libre disposición se realiza en aras

de la transparencia, la necesidad de favorecer la replicación de los estudios, facilitar el trabajo de los revisores así como de política de apoyo y aprovechamiento de las infraestructuras referenciadas en los párrafos anteriores. Por ello, uno de los cauces habituales que siguen las revistas es indicar en las instrucciones dónde deben los autores depositar los datos. Lo común es indicar el banco de datos donde deben hacerlo y la necesidad de cumplir los estándares establecidos llegando incluso al extremo de *Plos one*, que señala que la decisión de los autores de no compartir datos conforme a los estándares²⁴ puede dar lugar a que el trabajo sea rechazado²⁵.

Asimismo, uno de los requisitos comunes es que no sólo se realicen depósitos en el banco de datos adecuado, sino que se debe además proporcionar el *accession number* (número de registro) o DOI que identifica unívocamente dichos datos de manera que puedan ser recuperados y asociados al trabajo. En el caso de que no existan repositorios, algunas de las revistas establecen mecanismos que permitan a los autores compartirlos. Así, *Science* ofrece a los autores la posibilidad de que ofrezcan materiales suplementarios a los trabajos científicos que cuelgan junto al artículo de la sede web de la propia revista, garantizando de esta forma el propio editor la accesibilidad a los datos y otros materiales complementarios²⁶. En última instancia, cuando esto no es posible, las revistas científicas, como el caso de *Nature*, instatan al investigador a que muestre una disposición favorable

Revista	Declaración de las revistas sobre su política de <i>data sharing</i> (extraída de las normas para autores)
<i>Science</i> Multidisciplinar	<i>Science</i> apoya los esfuerzos de las bases de datos para compartir datos para el uso de la comunidad científica. Por ello, los sets de datos (incluyendo datos de <i>microarrays</i> –chips con fragmentos de ADN–, proteínas, secuencias de DNA, coordenadas atómicas, mapas microscópicos de electrones para estructuras macromoleculares, o datos meteorológicos) deberán ser depositados en un repositorio reconocido, y deberá indicarse en el artículo el número de registro o la dirección para su localización. Se recomienda el cumplimiento de las directrices <i>Mibbi</i>
<i>Plos one</i> Multidisciplinar	<i>Plos one</i> promueve la investigación abierta y pretende que todos los trabajos que publica puedan servir como punto de partida para futuros científicos. Por ello, requerimos la aceptación de los estándares existentes para el depósito público de datos (por ejemplo, secuencias genéticas o datos de expresiones de <i>microarrays</i> –chips con fragmentos de ADN–). Cualquier estándar para comunidades concretas también deberá ser acatado. No cumplir estos requisitos puede suponer el rechazo del manuscrito.
<i>Nature Publishing Group</i> Multidisciplinar	Una condición para publicar en una revista de <i>NPG</i> es que los materiales, datos y protocolos asociados estén disponibles para ser consultados por quien lo desee. Los sets de datos deberán estar accesibles de manera gratuita desde la fecha de publicación y deberán suministrarse a los editores y revisores en el momento de enviar el manuscrito con el objetivo de evaluar mejor la calidad del mismo. Para los siguientes tipos de datos, es obligatorio enviarlos a repositorios públicos validados por la comunidad. Deberá indicarse en el manuscrito el número de registro.
<i>British medical journal</i> Medicina	Animamos a los autores a que enlacen sus artículos publicados en <i>BMJ</i> con los datos brutos de sus estudios, depositándolos donde ellos consideren (<i>BMJ</i> no ofrece, al menos por el momento, la opción de almacenar sets de datos en <i>bmj.com</i>). Pedimos esto, no sólo por el deseo de maximizar la utilidad y el uso de los datos y fomentar la transparencia, sino también porque muchas agencias financiadoras fomentan o incluso obligan a la distribución pública de datos. Entendemos el deseo de muchos autores de guardar sus datos hasta haber publicado sus investigaciones y somos conscientes de lo difícil que resulta compartir datos.
<i>New England journal of medicine</i> Medicina	Los datos de <i>microarrays</i> obtenidos deberán incluirse en repositorios como <i>Gene expression omnibus</i> o <i>ArrayExpress</i> antes de enviar el manuscrito. Los sets de datos brutos y transformados de cada experimento de <i>microarrays</i> deberán ofrecerse a través de un repositorio, con un número de registro para cada experimento en la sección de métodos. Si los datos están protegidos, en la carta de presentación deberán indicarse el nombre de usuario y la contraseña. Uno de los criterios para la publicación es el acceso a todos los datos relevantes en un repositorio público.
<i>Annals of internal medicine</i> Medicina	En aras al fomento de la transparencia y de la reproducibilidad de las investigaciones, <i>Annals</i> indicará junto a cada artículo de investigación original la voluntad de los autores de compartir los siguientes elementos con el público: protocolo de estudio (original y arreglos), código estadístico usado para generar los resultados y set de datos de donde se extraen los resultados. <i>Annals</i> no obliga a que se compartan estos elementos, pero sí obliga a los autores a que indiquen su voluntad de hacerlo, así como cualquier restricción.

Tabla 4. Ejemplo de políticas editoriales en relación al *data sharing* de algunas de las principales revistas científicas multidisciplinarias y médicas

a compartir sus datos y se anima a los lectores, en el caso de que alguien se los solicite, a denunciar los casos en los que no reciba los datos²⁷, una situación que se produce con más frecuencia de la deseable (**Savage; Vickers**, 2009).

En el caso de las revistas del área de medicina, la adopción de políticas editoriales en torno al *data sharing* también se está produciendo, pero no sin problemas. Al ser un área donde se trabaja con pacientes y con nuevos fármacos los datos son más sensibles, puesto que son confidenciales y tienen un potencial valor económico. Uno de los problemas principales son los datos de los pacientes, ya que deben ser totalmente anonimizados antes de compartirlos y en algunos casos obtener su consentimiento (**Stone; Ling**, 2005), situación que ralentiza y complica el proceso de compartir. Como consecuencia, las referencias a las políticas de datos suelen ser menos exigentes y en ocasiones tan sólo son meras declaraciones de apoyo. En líneas generales las políticas en las revistas de medicina son por tanto, tímidos acercamientos, pero reconocen la importancia de este movimiento y auguran su implantación en las prácticas científicas en el corto, medio plazo. Recomiendan al investigador principal a que declare tener acceso directo a los datos en caso de que fueran requeridos por un tercero. Algunas de las revistas punteras ya se hayan posicionado como por ejemplo el *British medical journal (BMJ)* en 2010 siguiendo la estela de *Annals of internal medicine* o del *American journal of epidemiology*. En el caso concreto de *BMJ*, la respuesta por parte de los autores está siendo lenta (**Groves**, 2010), con casos excepcionales de puesta en práctica de *data sharing*.

Finalmente señalamos las evidencias obtenidas por **Piwovar** (2011) en relación a las revistas sobre descripción de expresiones génicas, donde se observó que suelen ser las revistas con políticas editoriales de *data sharing* las que tienen un mayor factor de impacto. Asimismo son las revistas de acceso abierto y dependientes de sociedades científicas y no de editoriales comerciales las que muestran posturas más definidas (**Piwovar; Chapman**, 2008b).

5. Consideraciones finales: papel de los profesionales de la información

Pese a las ventajas es necesario mencionar una serie de obstáculos que complican el acto de compartir los datos de forma pública. Tal vez las reticencias más severas provengan del propio investigador, al temer que se ponga en cuestión su trabajo y el tratamiento que han recibido los datos recogidos así como los resultados extraídos de los mismos (**Vickers**, 2006). Otro de los impedimentos es la falta de reconocimiento y protección a los investigadores que comparten datos (**Borgman**, 2008; **Gardner**, 2003).

Como profesionales de la información, un aspecto crucial a superar es la dificultad técnica en la implantación y gestión de políticas de *data sharing*, dado el alto coste que supone el desarrollo de la infraestructura y la estandarización de los datos. A ello se suma la dificultad añadida de modificar una serie de prácticas que hasta ahora pertenecían a la dinámica interna de los grupos de investigación (**Tenopir et al.**, 2011). Por ello uno de los principales retos es no sólo asegurar que los datos sean accesibles, sino que también sean reutilizables, haciéndose necesaria la implementación de una serie

de protocolos para poner los datos a disposición de la comunidad científica como *Miame (Minimum information about a microarray experiment)*²⁸ o *Minseq (Minimum information about a high-throughput sequencing experiment)*²⁹. Es imprescindible una cantidad mínima de metadatos detallados, ya que sin ellos es difícil entender el contexto en que se generaron los conjuntos de datos, pudiendo llevar a malas interpretaciones de los mismos (**Gadner**, 2003).

Por los problemas técnicos subrayados y por las necesidades de apoyo que muchos investigadores manifiestan, es necesario resaltar el papel primordial que los profesionales de la información deben jugar en la gestión de datos científicos. Como indican **Benett y Nicholson** (2011), los bibliotecarios pueden promover el almacenamiento y la compartición de datos por parte de sus usuarios, ayudándoles en la localización de sets de datos que puedan ser relevantes para sus investigaciones. Asimismo, indican que otras de sus tareas podrían ser la coordinación de iniciativas para la elaboración de directrices o normativas sobre gestión de datos científicos, o servir de intermediarios entre los investigadores y los repositorios institucionales. No en vano algunos de los repositorios de datos más conocidos cuentan con la biblioteca como valedora del mismo. Éste es el caso de la *Cornell University* con *Datastar*³⁰ o la *British Library* con *Dryad*³¹, lo que demuestra el creciente potencial de las bibliotecas en los servicios de apoyo a la gestión de datos científicos (**Kirlew**, 2011).

También habría que sumar el servicio que se puede ofrecer a los investigadores de la institución en la gestión de sus propios datos (**Torres-Salinas**, 2010b), lo que se empieza a conocer como *campus data management* (**Newton; Miller; Bracke**, 2010). Finalmente, a esto se suma el incipiente desarrollo académico de la especialidad, con el establecimiento de cursos específicos de *data curation* en algunas universidades como la *University of Illinois (Master of science: Specialization in data curation)*³² o la de *North Carolina (DigCCurr)*³³. **Kim** (2011) señala dos líneas académicas diferenciadas: una enfocada a la preservación digital de los datos científicos, y otra más volcada en las ciberinfraestructuras y en la vertiente tecnológica. En este sentido, el proyecto *Data curation profiles*³⁴ de la biblioteca de la *Purdue University* es un valioso recurso de ayuda para los bibliotecarios que desean familiarizarse con la conservación y gestión de datos científicos.

En Gran Bretaña, pese a las reticencias de algunos bibliotecarios en asumir roles de gestores de datos, casi dos tercios de los investigadores piensan que los bibliotecarios académicos deben asumir tareas de conservación y de gestión de las políticas, implementando estándares y metadatos para el intercambio de información y ontologías que ayuden a organizar los datos con vistas a su posterior recuperación (**RIN**, 2007).

En un reciente estudio de repositorios de datos (**Kirlew**, 2011) se menciona que el uso por parte de los bibliotecarios académicos es sensiblemente menor al detectado para los científicos; sin embargo, los *datbanks* suponen un campo especializado de creciente interés para los bibliotecarios de ciencias. **Kirlew** (2011) recuerda también que una vez que los profesionales de la información se familiarizan con los

repositorios disponibles en su campo (lo cual no siempre es fácil dada la complejidad y la variedad de los datos a almacenar), pueden proporcionar servicios de apoyo muy valiosos a los investigadores, y sugiere que el asesoramiento en almacenamiento y conservación de datos científicos será un servicio habitual que los bibliotecarios deberán prestar en un futuro próximo.

Los bibliotecarios expertos en datos científicos optan a un nuevo nicho de mercado de enorme potencial. Encuadrado en las profesiones de la e-ciencia pero también en el perfil del bibliotecario “integrado” (Kim; Addom; Stanton, 2011), que presta apoyo especializado y personalizado a los investigadores, mimetizándose con su equipo de trabajo (Torres-Salinas, 2011). Los profesionales del *data sharing* tienen ante sí el reto de intermediar con investigadores y gestores académicos y mostrarles los beneficios y oportunidades de compartir datos científicos en el marco de la e-ciencia.

6. Notas

1. http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm
2. <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>
3. <http://www.bbsrc.ac.uk/organisation/policies/position/policy/data-sharing-policy.aspx>
4. Cita original: “I have begun to think that no one ought to publish biometric results, without lodging a well arranged and well bound manuscript copy of all his data, in some place where it should be accessible, under reasonable restrictions, to those who desire to verify his work”.
5. <http://www.pdbe.org>
6. <http://www.rcsb.org/pdb/statistics>
7. <http://neurodatabase.org/dataserver/goto.do?page=home>
8. <http://cancergenome.nih.gov>
9. <https://www.dataone.org>
10. <http://www.ncbi.nlm.nih.gov/genbank>
11. http://grants.nih.gov/grants/policy/data_sharing/data_sharing_faqs.htm
12. <http://www.esrc.ac.uk/about-esrc/information/data-policy.aspx>
13. <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>
14. <http://www.ncbi.nlm.nih.gov/guide/data-software>
15. <http://hielo.igme.es>
16. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2007:0056:FIN:EN:PDF>
17. <http://www.casparpreserves.eu>
18. <http://www.seadatanet.org>
19. <http://www.alliancepermanentaccess.org/index.php/current-projects/ode>
20. <http://www.alliancepermanentaccess.org/index.php/current-projects/aparsen>

21. <http://www.parse-insight.eu>
22. <http://www.alliancepermanentaccess.org/index.php/current-projects/scidip-es>
23. <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>
24. http://biosharing.org/standards_view
25. <http://www.plosone.org/static/policies.action#sharing>
26. http://www.sciencemag.org/site/feature/contribinfo/prep/prep_online.xhtml
27. <http://www.nature.com/authors/policies/availability.html>
28. <http://www.mged.org/Workgroups/MIAME/miame.html>
29. <http://www.mged.org/minseq>
30. <http://datastar.mannlib.cornell.edu>
31. <http://www.datadryad.orgsame>
32. http://www.lis.illinois.edu/academics/programs/ms/data_curation
33. <http://www.ils.unc.edu/digccurr/aboutII.html>
34. <http://www4.lib.purdue.edu/dcp>

7. Bibliografía

- Arzberger, Peter; Schroeder, Peter; Beaulieu, Anne; Bowker, Geof; Casey, Kathleen; Laaksonen, Leif; Moorman, David; Uhlir, Paul; Wouters, Paul.** “An international framework to promote access to data”. *Science*, 2004, v. 303, n. 5665, pp. 1777-1778.
<http://www.sciencemag.org/content/303/5665/1777.full?ijkey=sgWI1mlejCudY&keytype=ref&siteid=sci>
<http://dx.doi.org/10.1126/science.1095958>
- Axelson, Ann-Sofie; Schroeder, Ralph.** “Making it open and keeping it safe: enabled *data sharing* in Sweden”. *Acta sociologica*, 2009, v. 52, n. 3, pp. 213-226.
http://oxford.academia.edu/RalphSchroeder/Papers/856843/Making_it_Open_and_Keeping_it_Safe_enabled_Data_Sharing_in_Sweden_and_Related_Issues
- Berman, Helen M.** “The protein data bank: a historical perspective”. *Acta crystallographica. Section A*, 2007, v. 64, n. 1, pp. 88-95.
- Bermúdez, Óscar; Barragán, Antonio; Alonso, Francisco.** “La gestión de los datos polares en España: una aproximación a la contribución de las ciencias de la vida”. *Ecosistemas*, 2011, v. 20, n. 1, pp. 94-103.
[http://rua.ua.es/dspace/bitstream/10045/18229/1/ECO_20\(1\)_11.pdf](http://rua.ua.es/dspace/bitstream/10045/18229/1/ECO_20(1)_11.pdf)
- Borgman, Christine.** “Data, disciplines, and scholarly publishing”. *Learned publishing*, 2008, v. 21, n. 1, pp. 29-38.
<http://www.ingentaconnect.com/content/alpsp/lp/2008/0000021/00000001/art00005>
<http://dx.doi.org/10.1087/095315108X254476>
- Cabezas-Clavijo, Álvaro; Torres-Salinas, Daniel; Delgado-López-Cózar, Emilio.** “Ciencia 2.0: catálogo de herramientas

- e implicaciones para la actividad investigadora". *El profesional de la información*, 2009, v. 18, n. 1, pp. 72-79.
<http://eprints.rclis.org/handle/10760/12811>
<http://dx.doi.org/10.3145/epi.2009.ene.10>
- Centro Nacional de Datos Polares, Instituto Geológico y Minero de España. *Protocolo de remisión, almacenamiento y difusión de los datos antárticos*, 2004.
http://www.micinn.gob.es/stfls/MICINN/Investigacion/FICHEROS/Comite_Polar_definitivo/Protocolo.pdf
- Crawford, Susan Y.; Hurd, Julie M.; Weller, Ann C.** *From print to electronic: the transformation of scientific communication*. Medford: ASIS, 1996.
- Field, Dawn; Sansone, Susanna; Collis, Amanda, et al.** "Omics data sharing". *Science*, 2009, v. 326, n. 5950, pp. 234-236.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2770171/?tool=pubmed>
<http://dx.doi.org/10.1126/science.1180598>
- Gardner, Daniel; Toga, Arthur; Ascoli, Giorgio, et al.** "Towards effective and rewarding data sharing". *Neuroinformatics*, 2003, v. 1, n. 3, pp. 289-95.
<http://dx.doi.org/10.1385/NJ:1:3:289>
- Ginsparg, Paul.** "ArXiv at 20". *Nature*, 2011, v. 476, pp. 145-147.
<http://www.nature.com/nature/journal/v476/n7359/full/476145a.html>
<http://dx.doi.org/10.1038/476145a>
- Godin, Benoît.** "From eugenics to scientometrics: Galton, Cattell, and men of science". *Social studies of science*, 2007, v. 37, n. 5, pp. 691-728.
http://www.csiic.ca/PDF/Godin_32.pdf
- Green, Toby.** "We need publishing standards for datasets and data tables". *OECD Publishing white papers*, 2009.
<http://dx.doi.org/10.1787/787355886123>
- Groves, Trish.** "Managing UK research data for future use". *British medical journal*, 2009, v. 338, b1252.
<http://dx.doi.org/10.1136/bmj.b1252>
- Groves, Trish.** "The wider concept of data sharing: view from the BMJ". *Biostatistics*, 2010, v. 11, n. 3, pp. 391-392.
http://hinxtongroup.files.wordpress.com/2010/10/groves_biostatistics_20101.pdf
<http://dx.doi.org/10.1093/biostatistics/kxq031>
- Hrynaskiewicz, Iain; Altman, Douglas G.** "Towards an agreement on best practice for publishing raw clinical trial data". *Trials*, 2009, v. 10, n. 17.
<http://www.trialsjournal.com/content/10/1/17>
<http://dx.doi.org/10.1186/1745-6215-10-17>
- Kim, Youngseek; Addom, Benjamin K.; Stanton, Jeffrey M.** "Education for eScience professionals: integrating data curation and cyberinfrastructure". *International journal of digital curation*, 2011, v. 6, n. 1, pp. 125-138.
<http://www.ijdc.net/index.php/ijdc/article/view/168>
- Kirlew, Peter W.** "Life science data repositories in the publications of scientists and librarians". *Issues in science and technology librarianship*, 2011, n. 65.
<http://www.istl.org/11-spring/refereed1.html>
- Kirwan John R.** "Making original data from clinical studies available for alternative analysis". *Journal of rheumatology*, 1997, v. 24, n. 5, pp. 822-825.
- Kowalczyk, Stacy; Shankar, Kalpana.** "Data sharing in sciences". *Annual review of information science and technology*, 2011, v. 45, n. 45, pp. 247-294.
- Marshall, Eliot.** "Bermuda rules: community spirit, with teeth". *Science*, 2001, v. 291, n. 5507, p. 1192.
<http://dx.doi.org/10.1126/science.291.5507.1192>
- Martínez-Urbe, Luis; Macdonald, Stuart.** "Un nuevo cometido para los bibliotecarios académicos: data curation". *El profesional de la información*, 2008, v. 17, n. 3, pp. 273-280.
<http://www.elprofesionaldeinformacion.com/contenidos/2008/mayo/03.pdf>
<http://dx.doi.org/10.3145/epi.2008.may.03>
- Nagelkerke, Nico J. D.; Bernsen, Roos M. D.; Rizk Daa E. E.** "Authors should publish their raw data". *International urogynecology journal*, 2007, v. 18, n. 12, pp. 1387-90.
<http://dx.doi.org/10.1007/s00192-007-0464-z>
- Nature. "Data's shameful neglect". *Nature*, 2009, v. 461, n. 7261, p. 145.
<http://www.nature.com/nature/journal/v461/n7261/full/461145a.html>
<http://dx.doi.org/10.1038/461145a>
- Nelson, Bryn.** "Empty archives". *Nature*, 2009, v. 461, n. 7261, pp. 160-163.
<http://www.nature.com/news/2009/090909/full/461160a.html>
<http://dx.doi.org/10.1038/461160a>
- Newton, Mark P.; Miller, C. C.; Bracke, Marianne-Stowel.** "Librarian roles in institutional repository data set collecting: outcomes of a research library task force". *Collection management*, 2010, v. 36, n. 1, pp. 53-67.
<http://www.tandfonline.com/doi/full/10.1080/01462679.2011.530546>
<http://dx.doi.org/10.1080/01462679.2011.530546>
- Nicholson, Sawn W.; Bennett, Terrence B.** "Data sharing: academic libraries and the scholarly enterprise". *Portal. Libraries and the academy*, 2011, v. 11, n. 1, pp. 505-516.
http://muse.jhu.edu/journals/portal_libraries_and_the_academy/toc/pla.11.1.html
- NIH (National Institutes of Health). *Final NIH Statement on sharing data*. National Institutes of Health, 2003.
- OECD (Organisation for Economic Co-operation and Development). *OECD Principles and guidelines for access to research data from public funding*. Paris: OECD Publications, 2007.
<http://www.oecd.org/dataoecd/9/61/38500813.pdf>
- OECD (Organisation for Economic Co-operation and Development). *Biomedicine and health innovation: synthesis report*. Paris: OECD Publications, 2010.
<http://www.oecd.org/dataoecd/42/56/46925602.pdf>
- Pennisi, Elizabeth.** "Will computers crash genomics?" *Science*, 2011, v. 331, n. 6018, pp. 666-668.

<http://dx.doi.org/10.1126/science.331.6018.666>

Perneger, Thomas V. "Sharing raw data: another of Francis Galton's ideas". *British medical journal*, 2011, v. 342, d3035.

<http://dx.doi.org/10.1136/bmj.d3035>

Piwowar, Heather A. "Who shares? Who doesn't? Factors associated with openly archiving raw research data". *Plos one*, 2011, v. 6, n. 7, e18657.

<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0018657>

<http://dx.doi.org/10.1371/journal.pone.0018657>

Piwowar, Heather A.; Chapman, Wendy W. "Identifying data sharing in biomedical literature". *Nature proceedings*, 2008b, pp. 596-600.

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2655927/?tool=pubmed>

Piwowar, Heather A.; Becich, Michael J.; Bilofsky, Howard; Crowley, Rebecca S. "Towards a data sharing culture: recommendations for leadership from academic health centers". *Plos medicine*, 2008a, v. 5, n. 9, pp. 1315-1319.

<http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.0050183>

<http://dx.doi.org/10.1371/journal.pmed.0050183>

Piwowar, Heather A.; Day, Roger S.; Fridsma, Douglas B. "Sharing detailed research data is associated with increased citation rate". *Plos one*, 2007, v. 2, n. 3, p. e308.

<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0000308>

<http://dx.doi.org/10.1371/journal.pone.0000308>

Ramasamy, Adaikalavan; Mondry, Adrian; Holmes, Chris C.; Altman, Douglas G. "Key issues in conducting a meta-analysis of gene expression microarray datasets". *Plos medicine*, 2008, v. 5, n.9, e184.

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2528050/>

<http://dx.doi.org/10.1371/journal.pmed.0050184>

Rennolls, Keith. "Science demands data sharing". *British medical journal*, 1997, v. 315, n. 7106, pp. 486.

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2127344/pdf/9284679.pdf>

RIN (Research Information Network). *Researchers' use of academic libraries and their services*, 2007.

<http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/researchers-use-academic-libraries-and-their-serv>

RIN (Research Information Network). *Data centres: their use, value and impact*, 2011.

<http://www.rin.ac.uk/data-centres>

Savage, Caroline J.; Vickers, Andrew J. "Empirical study of data sharing by authors publishing in *Plos* journals". *Plos one*, 2009, v. 4, n. 9, p. e7078.

<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0007078>

<http://dx.doi.org/10.1371/journal.pone.0007078>

Hanson, Brooks; Sugden, Andrew; Alberts, Bruce. "Making data maximally available". *Science*, 2011, v. 331, n. 6018, p. 649.

<http://dx.doi.org/10.1126/science.1203354>

Schneiderman, Ben. "Science 2.0". *Science*, 2008, v. 319, n. 5868, pp. 1349-1350.

<http://dx.doi.org/10.1126/science.1153539>

Stone, Margaret A.; Ling, Jennifer T. "Sharing patient data: competing demands of privacy, trust and research in primary care". *British journal of general practice*, 2005, v. 55, n. 519, pp. 783-789.

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1562354>

Swan, Alma; Brown, Sheridan. *To share or not to share: publication and quality assurance of research data outputs. A report commissioned by the Research Information Network (Main report)*, 2008a.

<http://www.rin.ac.uk/our-work/data-management-and-curation/share-or-not-share-research-data-outputs>

Tenopir, Carol; Allard, Suzie; Douglass, Kimberly; Aydinoglu, Arsev-Umur; Wu, Lei; Read, Eleanor; Manoff, Maribeth; Frame, Mike. "Data sharing by scientists: practices and perceptions". *Plos one*, 2011, v. 6, n. 6, p. e21101.

<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0021101>

Torres-Salinas, Daniel. "Compartir datos (data sharing) en ciencia: contexto de una oportunidad". *Anuario Thinkepi*, 2010a, pp. 262-265.

<http://www.thinkepi.net/compartir-datos-data-sharing-enciencia-el-contexto-de-una-oportunidad>

Torres-Salinas, Daniel. "Hacia la gestión de datos de investigación en las universidades: la *data asset framework*". *Anuario Thinkepi*, 2010b, pp. 258-261.

<http://www.thinkepi.net/primeros-pasos-hacia-la-gestion-de-datos-de-investigacion-en-las-universidades-la-iniciativa-daf>

Torres-Salinas, Daniel. "Integrados en la investigación: los *embedded librarians*". *Anuario Thinkepi*, 2011, v. 5, pp. 48-51.

<http://www.thinkepi.net/incrustados-e-integrados-en-la-investigacion-los-embedded-librarians>

Tramullas, Jesús; Garrido-Picazo, Piedad. "Software libre para repositorios institucionales: propuestas para un modelo de evaluación de prestaciones". *El profesional de la información*, 2006, v. 15, n. 3, pp. 171-181.

http://eprints.rclis.org/bitstream/10760/9345/1/vol15_3.1.pdf

UK Data Archive. *Managing and sharing data: best practice for researchers*, 2011.

<http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>

Vickers, Andrew J. "Whose data set is it anyway? Sharing raw data from randomized trials". *Trials*, 2006, v. 7, p. 15.

<http://www.trialsjournal.com/content/7/1/15>

<http://dx.doi.org/10.1186/1745-6215-7-15>