



GESTIÓN DE DATOS DE INVESTIGACIÓN: INFRAESTRUCTURAS PARA SU DIFUSIÓN



Luis-Millán González, Tomás Saorín, Antonia Ferrer-Sapena, Rafael Aleixandre-Benavent y Fernanda Peset



Luis-Millán González, doctor por la *Universitat de València (UV)*, es profesor desde 1998 en el *Departamento de Educación Física* de la UV donde imparte las asignaturas de metodología de la investigación y ejercicio físico para la calidad de vida. Sus intereses son: gestión de los datos de investigación, arquitectura de la información, estadística, software documental y ejercicio físico en la patología músculo-esquelética. Ha publicado en las principales revistas de medicina del deporte indexadas en los *Journal Citation Reports*.

<http://orcid.org/0000-0002-6478-4014>

Universitat de València, Facultat de Ciències de l'Activitat Física i l'Esport (FCAFE)
Gascó Oliag, 3, 46010 Valencia, España
luis.m.gonzalez@uv.es



Tomás Saorín es profesor asociado en la *Facultad de Comunicación y Documentación* de la *Universidad de Murcia* y documentalista de la *Comunidad Autónoma de Murcia* donde ha participado en la puesta en marcha de proyectos de gestión de contenidos e información institucional en las áreas de servicios sociales, trabajo, empleo y bibliotecas. Participa en el capítulo español de *Wikimedia* para el conocimiento libre y en acciones de divulgación del movimiento *GLAM-Wiki* para bibliotecas, archivos y museos. Ha investigado sobre estrategia digital y de edición electrónica, así como sistemas de gestión de contenidos y proyectos colaborativos.

<http://orcid.org/0000-0001-9448-0866>

Servicio Regional de Empleo y Formación, Unidad de Documentación
Av. Infante Juan Manuel, 14. 30011 Murcia, España
tsp@um.es



Antonia Ferrer-Sapena es licenciada en geografía e historia contemporánea por la *Universidad de Valencia*, y doctora en técnicas y métodos de información y documentación por la *Universidad Politécnica de Valencia (UPV)*. Es coordinadora de investigación en *Florida Centre de Formació* y miembro de su Comité de Innovación. Entre los proyectos que dirige destaca el portal de economía social *EcSocial.com*. Es profesora de la UPV, y miembro del foro *Innovem Junts*, del *Instituto de la Pequeña y Mediana Industria (Impiva)* de la *Generalitat Valenciana*.

<http://www.ecsocial.com>

<http://orcid.org/0000-0001-6432-917X>

Universidad Politécnica de Valencia
Camino de Vera, s/n. 46022 Valencia, España
anfersa@upv.es



Rafael Aleixandre-Benavent es científico titular del *Consejo Superior de Investigaciones Científicas (CSIC)* y catedrático acreditado de biblioteconomía y documentación. Es doctor en medicina, especialista en documentación médica por la *Universitat de València (UV)* y codirector de la *Unidad de Información e Investigación Social y Sanitaria (Uisys)* de la UV. Sus principales líneas de trabajo son la evaluación de la investigación y de las publicaciones científicas y los estudios sobre el acceso abierto a los datos científicos

<http://orcid.org/0000-0002-6678-8844>

Instituto de Historia de la Medicina y de la Ciencia López Piñero
Plaza Cisneros, 4. 46003 Valencia, España
rafael.aleixandre@uv.es



Fernanda Peset es profesora titular de la *Universidad Politécnica de Valencia, Departamento de Comunicación Audiovisual, Documentación e Historia del Arte* y coordinadora del programa de doctorado. Es doctor por la *Universidad de Murcia* en 2002. Su docencia y publicaciones se orientan a comunicación científica, acceso abierto, implantación del protocolo *OAI-PMH*, normalización de la información, descripción de documentos, sistemas de documentación de museos, y datos abiertos y enlazados. Participa en proyectos como *IraLIS, E-LIS, Grupo CIEPI, ODiSEA* y dirige el proyecto *I+D Datasea*.

<http://orcid.org/0000-0003-3706-6532>

Universitat Politècnica de València
Camino de Vera, s/n. 46022 Valencia, España
mpesetm@upv.es

Artículo recibido el 04-02-2013

Aceptación definitiva: 19-06-2013

Resumen

El intercambio de datos de investigación es un tema candente debido a la expansión de los métodos de trabajo colaborativos con un uso intensivo de tecnología (e-ciencia). El volumen y el intercambio de datos electrónicos están aumentando con rapidez, y si no se tratan correctamente sus beneficios no llegarán a sus potenciales interesados (investigadores, gobiernos y organismos de financiación públicos, directores de investigación y bibliotecas, sector editorial...). Este trabajo revisa su aparición y los movimientos de acceso abierto, describe un modelo de ciclo de vida para la preservación de datos, y revisa las iniciativas que tratan sobre metodologías y tecnologías. De forma paralela se presentan los avances internacionales y las necesidades españolas.

Palabras clave

Datos de investigación, Intercambio de datos, Acceso abierto, E-ciencia, Bancos de datos, Preservación, Política científica.

Title: Managing and disseminating research data

Abstract

The sharing of research data is a hot topic at the international level, especially due to the development of e-science, based on collaborative working methods that use technology in an intensive manner. The amount of digital data and of its sharing is growing fast, and a lack of proper management could limit the potential benefits to stakeholders (researchers, governments and public funding agencies, research managers and libraries, and finally the publishing sector). The aim of this review is first to place the sharing of electronic research data in the context of its appearance and of open environments; secondly, it describes the life cycle model of data preservation; finally, it describes several initiatives dealing with methodologies and technologies. In parallel, international developments and highlights needs that exist in Spain are described.

Keywords

Research data, Data-sharing, Open access, E-science, Databank, Preservation, Scientific policy.

González, Luis-Millán; Saorín, Tomás; Ferrer-Sapena, Antonia; Aleixandre-Benavent, Rafael; Peset, Fernanda (2013). "Gestión de datos de investigación: infraestructuras para su difusión". *El profesional de la información*, septiembre-octubre, v. 22, n. 5, pp. 415-423.

<http://dx.doi.org/10.3145/epi.2013.sep.06>

Introducción¹

La preservación y el intercambio de datos científicos se han convertido en temas de interés a escala internacional para quienes financian, gestionan y generan la investigación, algo que ya recogían desde 2003 las pautas sobre acceso abierto a la información científica de los *National Institutes of Health (NIH)*. **Borrego** (2012), basándose en el informe *Opportunities for data exchange (ODE, 2011)*, cifra en un 60% el número de investigadores que desearían utilizar datos primarios ajenos, mientras que aproximadamente un 40% muestra problemas para compartir los suyos.

La investigación científica genera cada vez un volumen mayor de datos digitales. Algunos se usan en publicaciones pero buena parte permanecen ocultos (*Nesta, 2010*). De la encuesta *Parse.insight* (2009) se desprende que ha aumentado el volumen de datos que corren peligro de perderse cuando ha finalizado la financiación, ya que se almacenan a nivel departamental y si se comparten es por canales no formalizados como correo o webs.

Preservar los datos digitales procedentes de la investigación con una infraestructura adecuada es prioritario para la comunidad científica y para la sociedad. Muestra de ello es la financiación del *National Science Foundation's DataNet Program* de EUA a grandes proyectos de investigación en esta área.

Existen multitud de tipos de datos digitales utilizados en los procesos de investigación: publicaciones, *raw data* (datos

brutos, sin tratar), estadísticas, gráficos, vídeos... El *W3Consortium* aconseja la utilización de formatos estándar (csv, xls, xml) para poder compartirlos con mayor facilidad. Su almacenamiento necesita instrucciones exactas, ya que no son igual de reutilizables datos que ya han sido analizados, que los directamente salidos de los instrumentos, por ejemplo.

El intercambio de resultados publicados y de datos de las investigaciones (*data sharing*) estimula descubrimientos adicionales.

La gestión de los datos de investigación es insuficiente tanto en el ámbito internacional como, especialmente, en el

Datos de investigación

La definición que tiene más consenso, al haber sido adoptada por entidades como los *National Institutes of Health (NIH)* de EUA y la *OECD* (2007), es "todo el material registrado durante el proceso investigador, reconocido por la comunidad científica y que sirve para certificar los resultados de la investigación que se realiza". Deben provenir de una fuente única y ser difíciles o imposibles de obtener de nuevo (**Torres-Salinas; Robinson-García; Cabezas-Clavijo, 2012**).

nacional. El volumen de datos electrónicos y su intercambio por métodos informales aumentan con rapidez y se corre el riesgo de perderlos al acabar la financiación de los proyectos. En España no se ha investigado sistemáticamente de qué situación partimos. La *Ley 14/2011* de 1 de junio de la ciencia, la tecnología y la innovación, que obliga a depositar las publicaciones en abierto, no incluye mención sobre los datos científicos. Este tema apenas está tratado por universidades y organismos públicos de investigación (OPI), que son quienes producen más datos.

Ante esta situación uno de los mayores retos es la investigación de nuevos métodos, tecnologías y modelos de gestión que sean capaces de identificar, clasificar, preservar y distribuir los conjuntos de datos (*datasets*), así como los flujos y ciclos de vida desde que son generados hasta que son reutilizados.

Este trabajo es una revisión de algunas tendencias internacionales que pueden ser de referencia para España. Existen cuestiones que deben ser dirimidas y armonizadas a escala internacional, como modelos, estándares y tendencias, mientras que otras están determinadas por el contexto en que se producen, como mandatos, agentes involucrados, estados de opinión y conductas, o valorización de los datos. Por esta razón se revisa el contexto internacional y se proponen aspectos a indagar en España.

Desde 2003 las pautas sobre acceso abierto de los *National Institutes of Health* contemplan los datos científicos además de la publicación de los resultados

En primer lugar se sitúa el objeto de estudio en el contexto de su aparición, los movimientos de acceso abierto; en segundo lugar se describe un modelo de ciclo de vida para su preservación; y por último se revisan métodos y técnicas aplicables. Asimismo se da cuenta de los avances internacionales y de las necesidades españolas.

Movimientos de acceso abierto

La investigación científica ha evolucionado en los últimos años a causa de varios fenómenos vinculados a la e-ciencia y los movimientos *open*. Se observa que las formas de investigación han cambiado (*The Royal Society*, 2012). El trabajo colaborativo -que ya existía por ejemplo en genómica- avanza hacia la e-ciencia, hacia la investigación facilitada por las infraestructuras electrónicas/TIC, en las que el intercambio de datos es esencial (*Shneiderman*, 2008). Un buen ejemplo es el desarrollo de vacunas contra la gripe porcina (H1N1),

The screenshot shows the NIH website's 'Grants & Funding' section. The main heading is 'NIH Data Sharing Policy'. Below the heading, there is a paragraph stating: 'Data sharing is essential for expedited translation of research results into knowledge, products and procedures to improve human health.' This is followed by a sub-heading: 'The Final NIH Statement on Sharing Research Data was published in the NIH Guide on February 26, 2003. This is an extension of NIH policy on sharing research resources, and reaffirms NIH support for the concept of data sharing. The new policy becomes effective with the October 1, 2003 receipt date for applications or proposals to NIH.' There are two bullet points: 'Data Sharing Regulations/Policy/Guidance Chart for NIH Awards (08/30/2006) - (MS Word - 58 KB) - This chart is designed as a quick guide only for the purpose of identifying various data sharing regulation/policy/guidance documents applicable to NIH funding.' and 'NIH Guide Notice (02/26/2003) - Final NIH Statement on Sharing Research Data.'

http://grants.nih.gov/grants/policy/data_sharing

que se produjo en tan sólo tres meses, frente al año que había necesitado la vacuna contra el *SARS coronavirus* diez años antes.

Existe una brecha entre el deseo de utilizar los datos ajenos (60%) y el de compartir los propios (40%)

La *Agenda digital europea 2003* abogaba ya por una mayor reutilización de los datos gubernamentales, pero también del patrimonio cultural y de investigación, como complemento del acceso abierto a las publicaciones (*Comisión Europea*, 2011b; 2012; *Riding the wave*, 2010).

Preservación de datos

Preservación digital es el "proceso de establecer y mantener a largo plazo repositorios digitales de referencia capaces de salvaguardar los datos actuales y futuros" (*Higgins*, 2008). Otros la definen como "la administración de datos científicos digitales a los que se proporciona un valor añadido dándoles un contexto y una vinculación a su origen de manera que se facilite su reutilización, integración y promoción" (*Rusbridge et al.*, 2005). Para la preservación y la reutilización de datos no sólo hay que resolver aspectos técnicos, sino también de organización y de procedimiento, económicos, financieros y de personal, administración de su propiedad, obligaciones legales, requisitos de auditoría, restricciones sociales de uso, etc. Este conjunto de acciones configura lo que se denomina *data curation*, actividad que recibió un gran impulso con la creación en marzo de 2004 del *Digital Curation Centre (DCC)* en Gran Bretaña, con financiación del *Joint Information Systems Committee (JISC)*. Actualmente es un centro de referencia mundial, siendo sus propósitos crear estándares y ayudar en la generación de planes de gestión (*Borgman*, 2012).

Las bibliotecas pueden asumir una importante responsabilidad en la preservación, pues están especializadas en conocer las fuentes de información digital y los métodos y estándares para su almacenamiento y conservación (*Ogburn*,

Tabla 1. Iniciativas de carácter general

País Institución	Proyecto
Reino Unido JISC	-DCC (<i>Data Curation Centre</i>), 2004. -Ukrds (<i>United Kingdom Research Data Service</i>), 2007-2010. Estableció un modelo sostenible de costes. -DAF (<i>Data Asset Framework</i>), 2009. Creó una metodología para asesorar a quienes han de gestionar este tipo de información.
Unión Europea (Reino Unido, Holanda, Alemania, Francia, además de Suiza). FP7	<i>Parse.Insight (Permanent access to the records of science in Europe)</i> , 2008-2010. Se ocupó de la preservación de la ciencia (datos primarios y publicaciones). Llevó a cabo estudios en profundidad para diseñar una hoja de ruta que orientase la estrategia de la <i>Comisión Europea</i> sobre infraestructura técnica. Resultados de la encuesta: -Los investigadores consideran importante un reanálisis de los datos ajenos. -Los gestores ven necesario disponer en abierto los datos obtenidos con financiación pública -Los editores creen que poner los datos en abierto hará posible el avance científico
Unión Europea. FP7	Los informes de ODE (<i>Kotarski et al., 2012; Dallmeier-Tiessen et al., 2012</i>) recogen buenas prácticas y opiniones de legisladores, financiadores, operadores de infraestructuras, centros de datos, proveedores y usuarios de datos, bibliotecas y editoriales.
Unión Europea. FP7	<i>OpenAirePlus (Open access infrastructure for research Europe)</i> . Toma como base las publicaciones enriquecidas: <i>Enhanced publications</i>
Unión Europea. FP7	<i>DataCite</i> es una organización británica sin ánimo de lucro que desde 2009 asesora para que se reconozcan como contribuciones científicas los datasets compartidos. Ha implementado un esquema de metadatos (<i>DataCite</i> , 2009) basado en DOI.
Estados Unidos. <i>National Science Foundation</i>	Las agencias de financiación tienen un papel relevante en la preservación y sostenibilidad de los datos (<i>Keefer</i> , 2011). Ello repercute en las bibliotecas académicas estadounidenses, que ayudan a elaborar los planes que exige la NSF en su apartado <i>Sharing of findings, data, and other research products</i> y formularios como el <i>Data conservancy</i> .

2010). Fuera de España algunas bibliotecas de investigación ya están desarrollando estos servicios de apoyo (**Choudhury**, 2009; **Gold**, 2010).

En los últimos años la atención se ha desplazado hacia la gestión del material según su ciclo de vida: creación, intercambio y preservación. Algunos objetos tienen una historia larga y complicada, pero tener en cuenta la “continuidad del ciclo de vida” mejora su longevidad y sus perspectivas de reutilización (**Higgins**, 2008). Para ello se registra la información relevante a lo largo del tiempo y los departamentos o secciones responsables de ese objeto en particular. Por ejemplo, en una institución de educación superior los profesores crean objetos digitales, pero es posible que ciertos detalles sobre los autores residan en otra unidad de gestión, que los usuarios del objeto estén registrados previamente en otra diferente o que se envíe una copia del objeto a la biblioteca, archivo o *digital curation centre* para su preservación.

El ciclo de vida de *digital curation* sigue estos pasos:

- conceptualización del plan para la fase de creación de datos;
- creación de objetos digitales con los metadatos adecuados;
- garantía de fácil acceso por los usuarios;
- selección de objetos a conservar, respetando sus dere-

chos y asegurando las copias;

- plan de expurgo y destrucción segura del resto de objetos;
- ingesta de contenidos en el repositorio;
- preservación;
- reevaluación de los objetos que no cumplen los procedimientos;
- almacenamiento: mantenimiento de los datos de manera segura;
- garantía del acceso y reutilización de los datos públicos;
- transformación en nuevos objetos digitales a partir de los originales.

Estas actividades han estado tradicionalmente vinculadas a las bibliotecas —recolección, producción, gestión y conservación— por lo que sus profesionales podrían evolucionar hacia los roles de *data manager* o *data curator*. Sin embargo el camino no es fácil, pues sus conocimientos han de completarse con aspectos muy específicos sobre la creación científica y la transformación de datos.

En cualquier caso el elemento clave para facilitar el acceso y movimiento de datos son los metadatos, en particular los de preservación, pues informan, describen y registran este tipo de actividades. Actualmente existen varias posibilidades para realizar estas funciones (**Martínez-Urbe**; **Macdonald**,

Tabla 2. Proyectos por disciplinas

Tema	Proyecto
Científico y cultural	<i>Caspar (Cultural, artistic and scientific knowledge preservation, for access and retrieval)</i> . Destaca la dificultad de encontrar un marco común para todo el ciclo de vida de una información que es muy heterogénea: científica, artística y cultural; sin embargo, el modelo de preservación <i>Oais</i> lo hizo posible.
Ciencias físicas, biomédicas, sociales y humanidades	<i>KE (Knowledge exchange)</i> recoge las buenas prácticas de los gobiernos británico, alemán, holandés y danés. Se focaliza en la necesidad de fomentar el intercambio de datos, facilitar programas de formación y estudiar la infraestructura necesaria.

2008; Educause, 2013): DDI 3.0 u *Open archival information system (OAIS) (ISO 14721:2003)*. La opción más utilizada es *OAIS*, un marco de referencia que define las responsabilidades obligatorias de un archivo digital y expone las actividades y relaciones básicas que deben llevarse a cabo para que el sistema funcione.

Para la puesta en valor de los datos se necesitan sin embargo otro tipo de ontologías destinadas a describir el ámbito disciplinar, cómo se han tomado los datos o el formato en que se encuentran.

Infraestructuras para difusión de datos

En relación con las cuestiones revisadas en los apartados precedentes se pueden identificar diversos tipos de personas, iniciativas e infraestructuras. Hay tres grupos de personas:

- científicos, que son productores y consumidores de los datos de investigación;
- gobiernos y agencias públicas, que financian el entramado de la ciencia, si bien en este grupo se incluyen en numerosas ocasiones las bibliotecas y servicios de investigación de los OPIs;
- sector editorial científico que hace pública la ciencia.

En las iniciativas se puede distinguir:

- portales orientados a estudiar métodos y a asesorar;
- tecnología para compartir datos y directorios que recogen los bancos de datos.

Metodologías

En este campo las iniciativas son numerosas y generalmente previas a establecer las infraestructuras técnicas. Publican metodologías y protocolos para un campo científico o se constituyen en plataformas internacionales con servicios de asesoramiento (tabla 1).

Por otra parte, existen proyectos que se limitan a disciplinas (tabla 2).

También las sociedades científicas adquieren protagonismo en tanto que productoras de datos. Por ejemplo la *International Spinal Cord Society (ISCoS)* recomienda el modelo *Common data elements* del *National Institute of Neurological Disorders and Stroke (Ninds)*, de los *NIH*, su agencia natural de financiación. Por otra parte, la *American Psychological Association (APA)* enlaza multitud de políticas y repositorios autorizados para distribuir los datos.

En cuanto a las revistas, el análisis de las políticas de las editoriales sobre datos muestra una correlación entre índices de impacto altos y el hecho de que las revistas acepten y hagan disponibles *datasets*. Se ha estudiado para el área de salud (Piwowar; Chapman, 2008; Piwowar, 2011) y para todas las disciplinas, escogiendo las revistas más citadas (Alsheikh-Ali et al., 2011; García-García et al., 2012a). Qui-

<http://www.dcc.ac.uk>

zá sería necesario realizar un estudio sistemático de estas políticas, teniendo en cuenta que el papel de las editoriales se considera crucial en el intercambio de datos (Dallmeier-Tiessen et al., 2012).

Como vemos, existen multitud de actores para un escenario que está en plena ebullición. En España el grupo de trabajo sobre datos de *Recolecta* ha publicado un informe preliminar (*Grupo de trabajo*, 2012) que ofrece un panorama de la gestión de datos a escala internacional y española. Simultáneamente el *Plan Nacional de I+D+i* ha financiado en su convocatoria de 2012 el proyecto *Opendatascience, centro de recursos para la preservación y gestión de datos abiertos de investigación* (CSO2012-39632). También las fundaciones privadas están interesadas en investigar en este campo y, por ejemplo, la *Fundación Mapfre* subvenciona el proyecto *Estrategias para la gestión y la promoción del uso compartido de datos científicos sobre la salud*. Ambos proyectos están siendo desarrollados por las tres instituciones del *VLC/ Campus-Valencia Campus de Excelencia Internacional: Universidad de Valencia, Universidad Politécnica de Valencia y CSIC*.

Las políticas sobre los datos científicos no tienen únicamente un carácter técnico

Tecnologías

Existen múltiples formas de compartir datos, más o menos formalizadas, pero las predominantes son las no registradas o dispersas, como se deduce de los escasos y heterogéneos casos de éxito que recogen **Torres-Salinas, Robinson-García y Cabezas-Clavijo** (2012): *ProteinDataBank* y *GenBank*.

En cuanto a la difusión formalizada, pueden adoptarse varias vías que casi coinciden con las conocidas ruta verde y dorada para fomentar el libre acceso a las publicaciones, aunque no usan el protocolo *OAI-PMH* para cosechar los

Tabla 3. Iniciativas relacionadas con el depósito de los datos

Clasificación	Nombre	Descripción
Servicios nacionales	<i>ANDS (Australian National Data Service)</i>	Objetivo: publicar, descubrir, acceder y reutilizar datos de investigación. <i>Research Data Australia</i> es un inventario y <i>My Data</i> un servicio de citación basado en DOI.
	<i>SND (The Swedish National Data Service)</i>	Coordina las bases de datos y guía a los científicos durante su investigación y en el acceso a datos externos
Infraestructuras temáticas	<i>Seadatanet (2007). Unión Europea</i>	Sistema interoperable para la gestión de datos e información marina
	<i>IDIP-ES (Science data infrastructure for preservation - Earth science), 2011. Unión Europea</i>	Infraestructura de datos sobre ciencias de la tierra
Servicios generalistas centrados en los contenidos	<i>Figshare o Dryad</i>	Permiten publicar los datos de investigación en abierto, facilitando su intercambio, cita y recuperación
	<i>Dataverse. Harvard University (EUA)</i>	Aplicación de código abierto para publicar, compartir, citar, extraer y analizar datos de investigación. Garantiza el uso de protocolos estandarizados y el reconocimiento de la autoría de los datos. Se organiza por colecciones creadas por cada institución que usa la plataforma.

metadatos. **Borrego (2012)**, siguiendo *ODE*, menciona repositorios y editoriales como las vías de almacenamiento preferidas por los investigadores. La ruta verde sería el depósito de *datasets* en repositorios o bancos específicos por disciplinas; y la dorada, almacenarlos en las plataformas editoriales junto a la publicación.

El concepto de continuidad del ciclo de vida mejora la longevidad y la reutilización de los datos

Torres-Salinas, Robinson-García y Cabezas-Clavijo (2012) siguiendo a **Kowalczyk (2011)** clasifican los repositorios y bancos de datos en centralizados por institución, descentralizados de varias instituciones, federados con una plataforma común de consulta y ciberestructuras. Nuestra propia indagación nos conduce a una clasificación más práctica:

- servicios nacionales de datos, con el caso australiano a la cabeza;
- plataformas por disciplinas; y
- plataformas de interés general, orientadas al investigador, preocupado especialmente por no perder el reconocimiento de su trabajo (**Starr; Gastl, 2011**).

Se prevén aún más líneas de actuación de cara al lanzamiento de *Horizon 2020* en la UE, el nuevo programa marco de investigación e innovación que integrará todas las líneas de investigación de la *Comisión Europea*, incluido el programa marco vigente todavía.

La tabla 3 muestra algunos ejemplos de lo que se puede llamar vía verde.

En cuanto a la vía que hemos comparado con la ruta dorada -el almacenamiento de datos en las editoriales- obviamos describir la tecnología utilizada, ya que se trata de las mismas plataformas de las publicaciones. Suele ser tecnología

propietaria en el caso de editoriales convencionales como *Sciverse* y *Wiley-Blackwell (Navarro-Molina, 2012)* o sistemas *OAI-PMH* compatibles como *Open Journal System (OJS)* o *Biomed Central* para el caso de las revistas *open access*.

Por último *Thomson Reuters* ha entrado en escena pisando fuerte con el lanzamiento del *Data citation index* en la *Web of knowledge*. Gracias a este producto, los resultados de las búsquedas mostrarán los *associated records* si la publicación tiene archivados datos en uno de los repositorios indexados (69 en enero de 2013).

Las infraestructuras que recopilan *datasets* pueden ser muy variadas en muchos sentidos -modelo, disciplina, tipos de datos, grado de reutilización, tecnología-. En la tabla 4 se listan algunos directorios.

Se buscaron repositorios de datos españoles en *DataBib* y *ODiSEA (International registry on research data)*. En el primero no aparece ninguno, y en el segundo dos (entre los 183 registrados a noviembre de 2012): *Ceacs*, de la *Fundación Juan March*, y *Digital.CSIC*; la *Universitat Pompeu Fabra* también han dado sus primeros pasos.

Según *OpenDoar*, sólo 82 repositorios ofrecen *datasets* en todo el mundo (3,6%). De los 97 repositorios españoles, que suponen el 4% del total, sólo uno acepta *datasets (Digital.CSIC)* (1%).

Por otra parte, suponiendo que los investigadores españoles estén depositando sus datos en bancos internacionales, se ejecutaron búsquedas en *ODiSEA* por los términos *Spain* y *Spanish*³: unos 20 bancos (11%) ofrecían conjuntos de datos españoles.

Conclusiones

Debido a la irrupción de los movimientos de acceso abierto en todos los ámbitos de la sociedad, el sector de la gestión de la información está en un momento de cambio, al que se suma las normativas que fomentan la reutilización de los datos para darles nuevo valor social y económico. El acceso

Tabla 4. Directorios de conjuntos de datos

Directorio	Detalles
<i>Research Data Australia</i> http://researchdata.andis.org.au	Carácter nacional
<i>CKAN/DataHUB</i> http://datahub.io	Múltiples recursos también relacionados con <i>opendata</i> y <i>linkedopendata</i>
<i>Dataverse</i> http://thedata.org	Plataforma de colecciones
<i>ODiSEA</i> http://odisea.ciepi.org	Incluye desde 2011 únicamente las bases de datos sobre datos de investigación (García-García <i>et al.</i> , 2012b y 2012c). Ante la falta de un inventario mundial de datos, similar a los de repositorios de publicaciones DOAR o ROAR, se seleccionaron aproximadamente 200 bancos.
<i>DataBib</i> http://databib.org	Inventario mundial colaborativo que incluye casi 600 items, entre los cuales aparecen también webs

a datos digitales, con los beneficios futuros que se deriven de ello, depende de las estrategias y procedimientos de preservación digital que se adopten ahora y del respaldo que reciban de las autoridades.

No obstante, entendemos que los organismos públicos de investigación y la universidad tienen características y funciones muy diferentes a las del resto de la administración pública, pues los datos que generan son la materia prima para su trabajo.

Es necesario conocer aspectos como:

- quiénes son los agentes involucrados en el flujo de los datos de investigación;
- cuáles son sus prácticas, necesidades y límites a la hora de compartir sus datos;
- qué datos se están generando, registrándolos y calificando especialmente su grado de apertura;

El trabajo de promoción y asesoría, al modo de bibliotecas como las del MIT, *Purdue University* o *California University*, es una oportunidad para nuestro sector. Unirlo a las estrategias que se están llevando a cabo para los repositorios institucionales valorizará esta información, hasta ahora escondida en los discos de los investigadores.

La complejidad de un escenario con tantos actores obliga a incluir a “infomediadores” que se encarguen de valorizar la información de cara al usuario final. El trabajo de promoción o *advocacy*, tan necesario, ha de ser liderado desde varios frentes. Minoritariamente incluirá al sector bibliotecario, junto a otros agentes relacionados con la investigación, como las oficinas de transferencia de resultados de la investigación (OTRI). Esos frentes habrán de constituirse en una red de influencia con el objetivo de contribuir a modificar los hábitos de la producción, tratamiento y difusión de los datos de investigación.

Notas

1. Este trabajo recibe financiación del Plan Nacional I+D+i del *Ministerio de Economía y Competitividad (Mineco)*: *Opendatascience*, Centro de recursos para la preservación y gestión de datos abiertos de investigación (CSO2012-39632), y de la *Fundación Mapfre*, convocatoria 2012 de ayudas a la investigación

http://www.fundacionmapfre.org/fundacion/es_es/default.jsp

2. Para escribir este artículo se ha consultado *Sciencea.com*, que ofrece servicios de valor añadido sobre los datos de *Cordis*.
<http://www.sciencea.com>
http://cordis.europa.eu/home_es.html

3. Los autores agradecen la ayuda de **Florencia Dieci**, colaboradora de *ODiSEA*.

Bibliografía

Alsheikh-Ali, Alawi A.; Qureshi, Waqas; Al-Mallah, Mouaz H.; Ioannidis, John P. A. (2011). “Public availability of published research data in high-impact journals”. *PLoS one*, v. 6, n. 9, pp. e24357.
<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0024357>

ANDS. *Australian National Data Service*.
<http://www.andis.org.au>

APA. *Links to data sets and repositories*.
<http://www.apa.org/research/responsible/data-links.aspx>

Arano, Silvia; Martínez, Gemma; Losada, Marina; Villegas, Marta; Casaldàliga, Anna; Bel, Núria (2011). “La comunidad ‘Recursos y datos primarios’ de la *Universitat Pompeu Fabra*: los repositorios institucionales como infraestructuras científicas. Estudio de caso”. *Revista española de documentación científica*, v. 34, n. 3, pp. 385-407.
<http://goo.gl/sfpnN>
<http://dx.doi.org/10.3989/redc.2011.3.834>

Borgman, Christine L. (2012). “The conundrum of sharing research data”. *Journal of the American Society for Information Science and Technology*, v. 63, n. 3, pp. 1059-1078.
<http://dx.doi.org/10.1002/asi.22634>

Borrego, Àngel (2012). “Los retos de la gestión de datos de investigación”. *Blok de bid*, 6 nov.
<http://goo.gl/10h9S>

Caspar. *Cultural artistic and scientific knowledge preservation for access and retrieval*.
<http://www.casparpreserves.eu>

Choudhury, Sayeed (2009). *Rethinking scholarly communi-*

cation: building data curation infrastructure.

<http://www.it.utah.edu/leadership/research/ciday/2009/notes/choudhury.pdf>

Comisión Europea (2003). *Agenda digital europea.*

http://europa.eu/rapid/press-release_IP-11-1524_es.htm

Comisión Europea (2010). *Riding the wave: How Europe can gain from the rising tide of scientific data.* Final report of the High Level Expert Group on Scientific Data.

<http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>

Comisión Europea (2011a). *Datos abiertos: Un motor para la innovación, el crecimiento y la gobernanza transparente.*

<http://goo.gl/4n5Wpw>

Comisión Europea (2011b). *Propuesta de directiva del Parlamento Europeo y del Consejo por la que se modifica la Directiva 2003/98CE relativa a la reutilización de la información del sector público.*

<http://goo.gl/9iVHTH>

Comisión Europea (2012). "Dictamen del Comité de las Regiones — Revisión de la Directiva relativa a la reutilización de la información del sector público y los datos abiertos (2012/C 391/12)". *Diario oficial de la Unión Europea*, 18.12.2012.

<http://goo.gl/ef3nE>

Comisión Europea (2012). *Online survey on scientific information in the digital age.* Luxembourg: Publications Office of the European Union.

<http://goo.gl/jVX0k>

<http://dx.doi.org/10.2777/7549>

Dallmeier-Tiessen, Sunje; Darby, Robert; Gitmans, Kathrin; Lambert, Simon; Suhonen, Jari; Wilson, Michael (2012). *Compilation of results on drivers and barriers and new opportunities.*

<http://goo.gl/ChTK6>

Data asset framework (2009). *Implementation guide.* Londres: JISC.

<http://www.data-audit.eu>

Data citation index.

http://wokinfo.com/products_tools/multidisciplinary/dci

DataCite (2009). *Metadata schema for the publication and citation of research data. Version 2.2.*

<http://goo.gl/AD4sgs>

<http://dx.doi.org/10.5438/0005>

DataConservancy. *Leading the movement to build data management tools and services across institutions and disciplines.*

<http://dataconservancy.org>

Dataverse network project.

<http://thedata.org>

DCC digital curation lifecycle.

<http://www.dcc.ac.uk/digital-curation/what-digital-curation>

DDI 3.0

<http://www.ddialliance.org>

Educause (2013). *Developing an institutional research data management plan service.*

<http://net.educause.edu/ir/library/pdf/ACT11301.pdf>

Figshare.

<http://figshare.com>

García-García, Alicia; García-Massó, Xavier; Ferrer-Sapena, Antonia; González, Luis-Millán; Peset, Fernanda; Aleixandre-Benavent, Rafael (2012a). "Mejores prácticas en reuso de conjuntos de datos publicados online como material adicional a los artículos". En: *2ª Conferencia sobre calidad de revistas de ciencias sociales y humanidades (CRECS 2012)*, Valencia.

<http://www.thinkepi.net/crecs2012>

García-García, Alicia; García-Massó, Xavier; Ferrer-Sapena, Antonia; González, Luis-Millán; Peset, Fernanda; Rodríguez-Gairín, Josep-Manuel; Saorín, Tomás (2012b).

"ODiSEA: International registry on research data". En: *5ª Jornadas OS-repositorios "La motricidad de los repositorios de acceso abierto"*, 23-25 mayo 2012.

<http://e-spacio.uned.es/ocs/index.php/osrepositoriosBilbao/motricidad/paper/view/15>

García-García, Alicia; Rodríguez-Gairín, Josep-Manuel; Saorín, Tomás; González, Luis-Millán; García-Massó, Xavier; Ferrer-Sapena, Antonia; Peset, Fernanda (2012c) "ODiSEA: International registry on research data". *BiD: textos universitaris de biblioteconomia i documentació*, n. 29.

<http://www.ub.edu/bid/29/garcia2.htm>

Gold, Anna (2010). *Data curation and libraries: short-term developments, long-term prospects.* April, 4.

http://digitalcommons.calpoly.edu/lib_dean/27

Grupo de Trabajo *Depósito y Gestión de datos en Acceso Abierto* del proyecto *Recolecta* (2012). *Informe preliminar para la conservación y reutilización de los datos científicos en España*. Madrid: Fundación Española para la Ciencia y la Tecnología (Fecyt).

<http://www.fecyt.es>

Higgins, Sarah (2008). "The DCC curation lifecycle model". *The international journal of digital curation*, v. 3, n. 1, pp 134-140.

<http://goo.gl/WTlgt>

ISCOs. International SCI data sets.

<http://www.iscos.org.uk/international-sci-data-sets>

Keefer, Alice (2011). "La preservación de los datos de investigación y las agencias de financiación de la I+D". *Blok de bid*, 11 de octubre.

<http://www.ub.edu/blokdebid/es/node/130>

Knowledge Exchange-KE

<http://www.knowledge-exchange.info>

Kotarski, Rachael; Reilly, Susan; Schrimpf, Sabine; Smit, Eefke; Walshe, Karen (2012). *Report on best practices for citability of data and on evolving roles in scholarly communication. Executive summary.*

<http://goo.gl/6qktq>

Kowalczyk, Stacy; Shankar, Kalpana (2011). "Data sharing in sciences". *Annual review of information science and technology*, n. 45, pp. 247-294.

- Martínez-Urbe, Luis; Macdonald, Stuart** (2008). "Un nuevo cometido para los bibliotecarios académicos: data curation". *El profesional de la información*, v. 17, n. 3, pp. 273-280.
<http://eprints.rclis.org/19304>
<http://dx.doi.org/10.3145/epi.2008.may.03>
- Martínez-Urbe, Luis; Macdonald, Stuart** (2009). "User engagement in research data curation". En: *ECDL'09 Proceedings of the 13th European conference on research and advanced technology for digital libraries*.
<http://www.era.lib.ed.ac.uk/handle/1842/3206>
- Navarro-Molina, Carolina** (2012). *Propuesta de criterios de calidad formal para revistas científicas electrónicas basadas en aspectos de interacción persona-ordenador*. Tesis doctoral. Universidad de Valencia.
<http://roderic.uv.es/handle/10550/23755>
- Nesta (National Endowment for Science Technology and the Arts) (2009 - 2010). *Open science case studies*.
<http://www.rin.ac.uk/our-work/data-management-and-curation/open-science-case-studies>
- NIH (2003a). *Data sharing faqs*.
http://grants.nih.gov/grants/policy/data_sharing/data_sharing_faqs.htm
- NIH (2003b). *Final NIH statement on sharing research data*.
<http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>
- NIH. Office of Extramural Research. *NIH Data sharing policy*.
http://grants1.nih.gov/grants/policy/data_sharing
- Ninds *Common data elements CDE*.
<http://www.commondataelements.ninds.nih.gov>
- Ninds (2012). *Maximizing what we learn from clinical trials: Ninds common data elements*.
http://www.ninds.nih.gov/about_ninds/message/message-CDE.htm
- NSF (2009). *Grant general conditions (GC-1)*.
<http://www.nsf.gov/pubs/gc1/jan09.pdf>
- NSF. Directorate for Social, Behavioral and Economic Sciences. *Data archiving policy*.
<http://www.nsf.gov/sbe/ses/common/archive.jsp>
- ODE (2010). *Opportunities for data exchange*.
<http://www.alliancepermanentaccess.org/index.php/community/current-projects/ode>
- OECD (2007). *Principles and guidelines for access to research data from public funding*.
<http://www.oecd.org/science/scienceandtechnology/policy/38500813.pdf>
- Ogburn, Joyce L.** (2010). "The imperative for data curation". *Portal-libraries and the academy*, v. 10, n. 2, pp. 241-246.
<http://dx.doi.org/10.1353/pla.0.0100>
- Parse.insight (2009). *Insight into digital preservation of research output in Europe. Survey report*.
http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf
- Piwowar, Heather A.** (2011). "Who shares? Who doesn't? Factors associated with openly archiving raw research data". *PLoS one*, v. 6, n. 7, p. e18657.
<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0018657>
- Piwowar, Heather A.; Chapman, Wendy W.** (2008). "A review of journal policies for sharing research data". *AMIA Annual symposium proceedings archive*.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2655927/?tool=pubmed>
- Rusbridge, Chris; Burnhill, Peter; Ross, Seamus et al.** (2005). "The digital curation centre: a vision for digital curation". En: *Proceedings from local to global: data interoperability - challenges and technologies*, 2005. Forte Village Resort, Sardinia, Italy, pp. 1-11.
<http://eprints.erpanet.org/82>
- Scidip-es. *Science data infrastructure for preservation - Earth science*.
<http://www.scidip-es.eu>
- The Royal Society (2012). *Science as an open enterprise*. The Royal Society Science Policy Centre report. ISBN: 978 0 85403 962 3
<http://royalsociety.org/policy/projects/science-public-enterprise/report>
- Seadatanet.
<http://www.seadatanet.org>
- Shneiderman, Ben** (2008). "Science 2.0". *Science*, v. 319, n. 5868, pp. 1349-1350.
<http://dx.doi.org/10.1126/science.1153539>
- Starr, Joan; Gastl, Angela** (2011). "isCitedBy: a metadata scheme for DataCite". *D-lib magazine*, v. 17, n. 1-2
<http://dlib.org/dlib/january11/starr/01starr.html>
<http://dx.doi.org/10.1045/january2011-starr>
- SND. Swedish National Data Service.
<http://snd.gu.se/en>
- Torres-Salinas, Daniel** (2010a). "Compartir datos (data sharing) en ciencia: el contexto de una oportunidad". *Anuario ThinkEPI*, v. 4, pp. 258-261.
- Torres-Salinas, Daniel** (2010b). "Primeros pasos hacia la gestión de datos de investigación en las universidades: la iniciativa DAF". *Anuario ThinkEPI*, v. 4, pp. 262-265.
- Torres-Salinas, Daniel; Robinson-García, Nicolás; Cabezas-Clavijo, Álvaro** (2012). "Compartir los datos de investigación: introducción al data sharing". *El profesional de la información*, v. 21, n. 2, pp. 173-184.
<http://ec3.ugr.es/publicaciones/datasharing.pdf>
<http://dx.doi.org/10.3145/epi.2012.mar.08>