

DESARROLLO DE UNA BASE DE DATOS PARA EL SERVICIO DE PRODUCCIÓN CIENTÍFICA EN LOS CENTROS DE INVESTIGACIÓN ESPECIALIZADOS

Gómez Sánchez, Alicia Fátima. Responsable del Servicio de Biblioteca y Documentación. Fundación Centro Nacional de Investigaciones Cardiovasculares, Melchor Fernández Almagro, 3 - 28029 Madrid, afgomez@cnic.es; **López Álvarez, Victoria.** Responsable del Servicio de Biblioteca y Documentación. Fundación Centro Nacional de Investigaciones Oncológicas, Melchor Fernández Almagro, 3- 28029 Madrid, vlopez@cnio.es.

Resumen: Los servicios que prestan las bibliotecas de los centros de investigación en la actualidad van más allá de los tradicionales como la contratación de recursos, el préstamo interbibliotecario o el servicio de referencia, y a éstos se suman otros, entre los que creemos que merecen una mención especial los relacionados con el seguimiento y la evaluación de la producción científica, y que principalmente tienen como objetivo final la realización de estudios bibliométricos y memorias.

La experiencia nos ha mostrado que para cubrir esta necesidad de ofrecer un servicio de apoyo a la gestión y a la investigación que resulte eficiente es imprescindible el desarrollo de una base de datos de producción científica propia y actualizada que incluya las referencias de todas las publicaciones del personal investigador y de los demás servicios científicos o unidades técnicas. A cada una de estas referencias bibliográficas se añaden además otros datos bibliométricos como el FI o los deciles y cuartiles por áreas temáticas, así como datos de afiliación y colaboraciones. Asimismo es interesante también disponer de cada uno de los pdfs, si bien su uso a menudo puede quedar restringido al ámbito interno por temas de derechos. Aunque pensamos que lo ideal es el desarrollo de una base de datos *ad hoc*, cabe también la posibilidad de utilizar otros recursos, como por ejemplo algunos gestores bibliográficos ya existentes en el mercado.

En definitiva, el hecho de que el servicio de producción científica cuente con una base de datos de estas características es fundamental para poder responder con la calidad y rapidez adecuada al realizar estudios bibliométricos sobre la producción científica del centro. A su vez permite también prestar un apoyo documental y facilitar la actualización de los *curricula* y demás informes de los propios investigadores del centro.

Palabras clave: repositorios, bibliometría, producción científica.

Keywords: repository, bibliometry, scientific production.

1. Introducción: Los repositorios y su evolución

En los últimos años ha aumentado de manera muy notable el interés de las instituciones por la creación de repositorios con el objetivo de archivar, gestionar y difundir la propia producción científica. Según los datos de OpenDOAR, el número de repositorios en el mundo ha pasado de ser 800 en 2007 a una cifra que en la actualidad supera los 1400. En el caso de España en 2007 eran una decena y hasta la fecha de hoy se han registrado en este directorio 46¹.

Podemos considerar los repositorios como bases de datos que recogen documentos seleccionados y organizados en función de diversos criterios. Así, podemos hablar de dos grupos principales de repositorios, los temáticos y los institucionales. Los repositorios temáticos recogen documentos de un mismo tema y en ocasiones de una misma disciplina, como por ejemplo los siguientes: *PsyDok*, que contiene textos en alemán sobre psicología; el *Social Science Open Access Repository (SSOAR)*, que contiene textos completos sobre ciencias sociales; el repositorio de *HighWire Press*, que es una división de las Bibliotecas de la Universidad de Stanford y provee acceso abierto al texto completo de artículos del área de las ciencias de la vida; *OpenMED@NIC*, que es un archivo también para ciencias médicas en los que los propios autores pueden auto-archivar sus documentos, al igual que el conocido *ELIS (E-prints in library information science)*, dedicado a la publicación en acceso abierto de todo tipo de documentos sobre bibliotecología. Por otra parte están los repositorios institucionales, cuyo objetivo es recoger y almacenar todas las publicaciones de un organismo o centro. La creación de repositorios institucionales tiene lugar principalmente en el ámbito académico, y de manera especial en el círculo de las universidades. De hecho, la mayoría de los repositorios e incluso las definiciones de este concepto parten de dicho contexto. Precisamente el primer repositorio de nuestro país, *Tesis Doctorales en Xarxa (TDX)*, creado en el año 2001, nació también del ámbito universitario.

Clifford Linch (2003, p. 2) define un repositorio –en su caso universitario– como “una serie de servicios que la universidad ofrece a su comunidad para la gestión y difusión de los materiales digitales creados por dicha institución y los miembros de su comunidad. Fundamentalmente se trata de un compromiso organizativo para la administración de dichos materiales digitales, que incluye su preservación a largo plazo en un lugar adecuado, así como su gestión y su acceso o distribución”. Según Crow (2002, p.16) los repositorios han de “pertenecer a una institución; ser de ámbito académico; ser acumulativos y perpetuos; ser abiertos e interactivos”.

En definitiva, se trata de bases de datos que:

- reúnen y almacenan los documentos que recogen la producción científica de dicha institución.
- son propias de dicha institución.
- permanecen en el tiempo.

Hay además otras características comunes a la mayoría de los repositorios, como pueden ser su carácter de recurso abierto y gratuito o que provienen del ámbito académico. Sin embargo esto no tiene que ser una premisa, ya que podemos encontrar repositorios cerrados

o de pago, así como repositorios de empresas, como el repositorio de *Novartis*. En cualquier caso, la idea principal es que los repositorios institucionales sirvan de herramienta para:

- Llevar a cabo un registro unificado de todas las publicaciones del personal de cada centro, de modo que se asegure y se preserve cada uno de los documentos que componen la producción científica de la institución, independientemente de que el autor o incluso el departamento deje de formar parte del centro.
- Permitir la consulta centralizada de la producción científica del centro, que por su naturaleza o por el tipo de publicación de otra manera estaría dispersa en numerosas revistas, actas de congresos, bases de datos, etc. De esta forma se favorece una mayor visibilidad y con ello el prestigio de la institución.
- Asimismo, el hecho de que quede constancia y copia de cada uno de los trabajos publicados dentro del marco de la institución sirve para que se protejan los derechos de copyright de los autores.

Por último, uno de los aspectos que ejerce una influencia adicional en las instituciones para la creación de un repositorio es el propio marco legal. Como señala Barton (2004, p.10), el *House of Commons Science and Technology Committee* de Gran Bretaña ya en el año 2004 hacía la recomendación de que “todas las instituciones de educación superior usen Repositorios Institucionales donde se pueda almacenar lo que se publique y donde se pueda leer de forma gratuita en línea”. El informe también sugiere a los organismos económicos gubernamentales “que obliguen a sus investigadores beneficiarios a depositar una copia de sus artículos de esta forma”.

En España el interés por el archivo de la producción científica aún tiene trabajo por delante, si bien es cierto que en el Capítulo III del Anteproyecto de la nueva Ley de la Ciencia y la Tecnología que trata sobre la regulación de la difusión de resultados y la cultura científica se incorpora un artículo referido a la publicación en acceso abierto, y en él se dispone que “todos los investigadores cuya actividad haya sido financiada con los presupuestos generales del Estado, estarán obligados a publicar en acceso abierto una versión electrónica de los contenidos aceptados para publicación en publicaciones de investigación” y además “se encomienda a los distintos agentes del Sistema el desarrollo de repositorios institucionales de acceso abierto”. En este sentido estamos de acuerdo con Melero (2008) quien señalaba que “las bibliotecas son, en muchos casos, las promotoras, creadoras y gestoras de los repositorios, sin embargo su buen trabajo puede quedar frustrado si no existe la participación de los autores y de las autoridades que deben dictar las políticas de acceso a la documentación que albergan sus depósitos digitales”.

2. ¿Por qué hablamos de base de datos para el servicio de producción científica?

Los centros de investigación especializados tienen características propias que difieren de las de las instituciones universitarias, de las cuales destacamos dos que a nuestro entender son fundamentales:

Especificidad de la temática: En las universidades los repositorios son de temática multidisciplinar mientras que en este tipo de centros la temática es muy concreta; en el caso que abordamos aquí son la cardiología y la oncología respectivamente. Este aspecto tiene implicaciones a la hora de clasificar los contenidos y de asignar descriptores. Es cierto también que de este modo, a la vez que creamos un repositorio institucional estamos creando al mismo tiempo una base de datos con un valor temático especializado añadido.

La tipología de los documentos: Las universidades son instituciones académicas y uno de sus objetivos esenciales es la formación, de modo que, aunque la mayoría comenzaron creando repositorios que recogían las tesis leídas en los distintos departamentos y a éstas se unían contribuciones de su personal docente e investigador en libros, revistas, congresos, etc., es cierto que cualquier tipo de material docente o de apoyo para la docencia puede resultar interesante como candidato a ser incluido en dicho repositorio². Sin embargo, el tipo de los documentos que se reúnen y almacenan en un centro de investigación especializado es mucho más restringido y se compone principalmente de artículos publicados en revistas científicas, pudiendo completar estos fondos con tesis y capítulos en libros o textos presentados en congresos (ponencias, comunicaciones o pósters).

A estas dos características hemos de añadir que la causa y la finalidad principales de la creación de una base de datos de publicaciones en nuestras instituciones no fue únicamente la de tener un repositorio que almacenara la producción científica del centro, sino también la de crear una base de datos que sirviera de punto de partida para proporcionar una herramienta eficiente a la hora de realizar informes bibliométricos y de evaluación de la producción científica del centro, así como para el desempeño de otras tareas como la publicación de la memoria del centro, la generación de informes para los planes estratégicos, o incluso para dar apoyo a los investigadores en la preparación de documentación para presentar solicitudes de ayudas para proyectos, contratos, becas, etc. o simplemente para mantener actualizados sus currículos.

3. Creación de un repositorio para necesidades específicas

3.1 Inicio del proceso y elección del software

En un primer momento comenzamos a realizar el registro de las publicaciones de nuestro centro utilizando el gestor bibliográfico EndNote, sin embargo, al poner en práctica los propósitos de nuestra base de datos nos dimos cuenta de que teníamos que pensar en un programa adecuado que nos diera la posibilidad de añadir otros datos no bibliométricos y que nos permitiera distintos tipos de consultas y de extracción de información.

La creación de un repositorio es sin embargo algo que no se limita a la característica de ser simplemente una base de datos con documentos, se trata de algo más complejo. Como señala Linch (2003, p. 2) requiere de la colaboración de distintos grupos de profesionales que, por supuesto, incluye a los bibliotecarios y los documentalistas:

“While operational responsibility for these services may reasonably be situated in different organizational units at different universities, an effective institutional repository of necessity represents a collaboration among librarians, information technologists, archives and records managers, faculty, and university administrators and policymakers (...). An institutional repository is not simply a fixed set of software and hardware”.

Así pues, una vez que habían quedado claros el objetivo y el empleo que queríamos darle a nuestra base de datos de producción científica se puso en marcha el proyecto en el que se establecía la colaboración entre el Servicio de Documentación y Biblioteca, el Departamento de Gestión Científica y el Departamento de Informática.

La mayoría de los repositorios desarrollados en universidades y en centros de investigación, tanto en España como en otros muchos países, utilizan principalmente software libre. Nosotros también pensamos en partir de un software de este tipo pero atendiendo a la estructura interna de nuestros centros y contando con la disponibilidad de un departamento de informática finalmente pensamos que lo más rápido y lo ideal en nuestro caso era el desarrollo de una base de datos *ad hoc*.

Tras meses de trabajo, en la actualidad contamos con una base de datos que incluye las referencias de las publicaciones del personal investigador y de los demás servicios científicos o unidades técnicas. Como especificaremos con más detalle en el siguiente punto, a cada una de estas referencias bibliográficas se añaden además otros datos de tipo bibliométrico, así como datos de afiliación y colaboraciones.

3.2 Estructura de la base de datos y descripción de sus campos

Antes de centrarnos en los campos de nuestro repositorio nos parece interesante comentar el proceso de localización de artículos publicados por nuestros investigadores: Principalmente trabajamos a través de búsquedas individualizadas y alertas en el motor de búsqueda *PubMed*, que permite el acceso a la base de datos bibliográfica Medline; estas alertas se revisan periódicamente en función de los cambios en el directorio de personal. Asimismo, en la Intranet de cada centro tenemos un formulario a disposición de los investigadores para que puedan avisar al Servicio de Biblioteca y Documentación cada vez que publiquen. Dicho esto, pasamos a explicar en detalle los distintos campos que forman cada uno de los registros de nuestra base de datos, agrupados en función de su tipología.

Datos referentes a los artículos:

El primer bloque de datos incluye los estos campos que contienen información puramente bibliográfica. Estos datos se pueden importar al repositorio de producción científica directamente desde el motor de búsqueda *PubMed* o se pueden introducir de manera manual. Los campos son:

- **Pumed Id:** código numérico que corresponde a la referencia bibliográfica del artículo que se encuentra recogido en la base de datos de *PubMed*. Gracias a la introducción

de este campo, los listados que se obtienen de realizar la consulta en la herramienta contarán bajo la referencia el link del artículo de tal forma que, si se pincha en el enlace, éste nos llevará al abstract del trabajo que aparece en *PubMed*.

- Título del artículo: contendrá el título del trabajo de forma tal y como aparece en la publicación periódica.
- Revista: el registro de los datos relativos a la revista o a la publicación periódica en la que aparece el artículo se realiza incluyendo el ISSN de la publicación y de forma automática nos aparecerán el título completo y el título abreviado de la revista o serie. Para ello tenemos un documento asociado en el que junto al ISSN aparecen estos datos, lo cual minimiza la posibilidad de errores ortotipográficos y asegura una normalización que evita problemas de ruido o silencio en las búsquedas, además de que nos permite conocer información cualitativa de la revista de manera muy detallada.
- Año: en este campo introduciremos el año en el que se publica el trabajo, si bien este dato puede variar en tanto en cuanto la publicación pasa de estado *e-pub ahead of print* a estar en prensa.
- Estado: en este indicaremos si se encuentra *press* o *in press*, en función del estado del artículo.
- Volumen, número y páginas inicial y final: también indicaremos el número de volumen y número, si lo hubiere, así como las páginas con las que cuenta el trabajo de investigación. No hay que olvidar que este campo debe permitir la introducción de caracteres alfanuméricos.

Por último, aunque no es un dato bibliográfico, sino la referencia en sí misma, se incluye también el enlace a la versión del artículo en pdf.

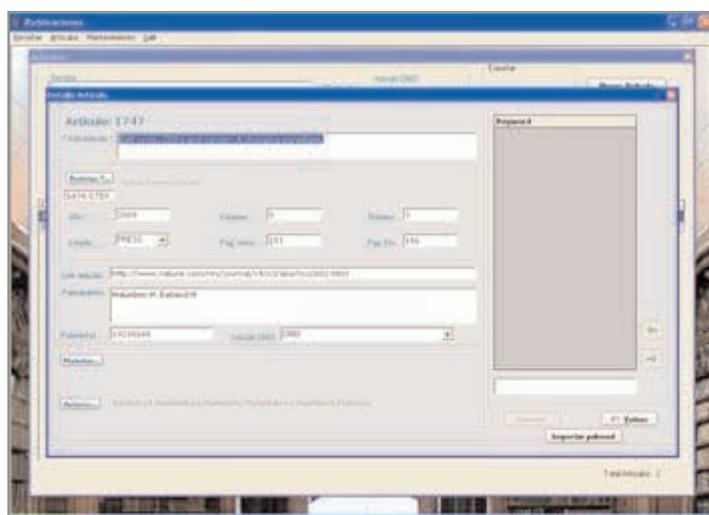


Figura 1: Campos necesarios para catalogar el artículo.

Datos referentes a la autoría y la afiliación

Dentro de este grupo gestionamos los campos referidos a los datos que van ligados a la autoría del artículo:

- **Participantes:** en este campo se indican los nombres de todos los autores del artículo siguiendo el formato de *PubMed*: Apellido Inicial, Apellido Inicial, etc. Aquí se incluyen todos los autores del artículo, con independencia de que pertenezcan o no a la institución.
- **Autores:** a diferencia del campo anterior, este campo se reserva a los nombres de aquellas personas que pertenecen a alguno de los organismos (CNIO o CNIC) señalando además a la unidad y programa al que pertenecen e indicando además su perfil como investigador, esto es, el cargo que pueda tener en el grupo o unidad y el lugar en el que aparece en el artículo. Al igual que ocurre con los datos relacionados con las revistas, los nombres de los autores van ligados al directorio de personal de cada centro, de manera que los datos sobre departamentos, laboratorios, unidades, etc. se cargan de forma automática, evitando también posteriores problemas en la recuperación de la información.
- **Afiliación y autoría:** en este apartado informaremos si el artículo puede ser considerado como fruto del trabajo de investigadores de la propia institución o si se trata de una colaboración extranjera o nacional. Por otra parte, queda señalado también todo autor que aparezca como ‘primer autor’, ‘último autor’ o ‘autor para correspondencia’ (para obtener esta información nos fijaremos en el *reprint adress* del artículo).

Datos bibliométricos:

Como ya hemos comentado anteriormente, la inclusión de datos de tipo bibliométrico es una de las características esenciales para poder cumplir los objetivos marcados para nuestra base de datos. Estos campos incluyen el factor de impacto de la revista, así como los cuartiles y deciles en los que se clasifica dicha revista. La ventaja de trabajar con este tipo de datos es que son fijos, a diferencia de lo que ocurre con las citas, que están en constante evolución.

Para obtener la información bibliométrica requerida de cada revista utilizamos el documento asociado que contiene el listado de las revistas en las que el ISSN, además de estar relacionado con el título completo y el título abreviado de la publicación, también relaciona el factor de impacto, así como el decil y el cuartil (siempre que corresponda). Este documento se genera a partir de los datos del *Journal Citation Report* de la base de datos *Web of Knowledge* y, al igual que esta base de datos, también se actualiza anualmente.

4. Explotación y uso de la base de datos

4.1 Tipos de consultas y obtención de informes:

Los tipos de consultas que se pueden realizar son muy variados y flexibles y dependen de la estructuración de los campos que componen cada uno de los registros. El resultado informativo tendrá carácter conciso o general según sea la necesidad informativa.

Por ejemplo, podemos consultar teniendo en cuenta el tipo de autoría y así obtener informes, con todas aquellas publicaciones de una institución determinada para realizar comparativas con otros centros. La consulta puede ser más concisa en tanto en cuanto establezcamos un rango de tiempo y especifiquemos un título de publicación en concreto. Sin embargo, si lo que necesitamos es hacer un estudio bibliométrico interno podemos por ejemplo realizar una consultas que incluyan los propios departamentos. Por otra parte, en cada consulta y en los informes derivados, en función del tipo de necesidad que se tenga y el nivel de profundidad que se necesite, se podrán seleccionar los campos que interesen en cada caso.

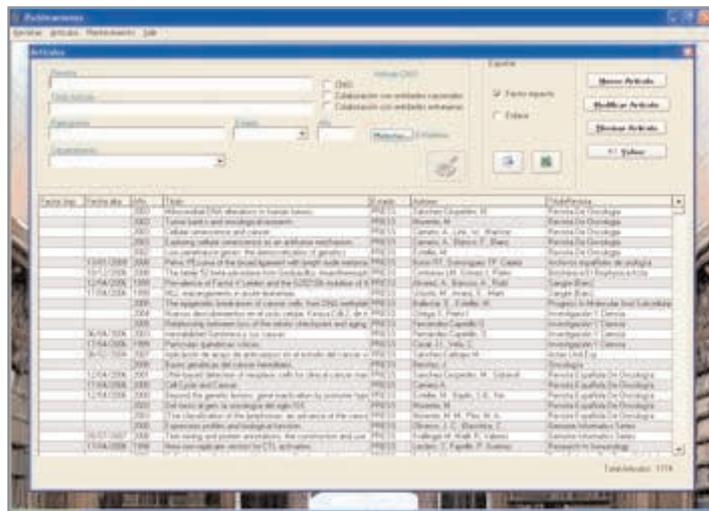


Figura 2. Formulario y visualización de la pantalla de consultas.

Asimismo podemos obtener documentos en formato texto o en forma de hoja de cálculo. La necesidad de poder ofrecer estas dos posibilidades viene determinada por el destino que tenga el informe, es decir, si lo que vamos a necesitar son diagramas o estadísticas será conveniente exportar nuestros resultado en formato de hoja de cálculo, mientras que si queremos los resultados para la confección de una memoria científica será más útil disponer de esta información en formato texto.

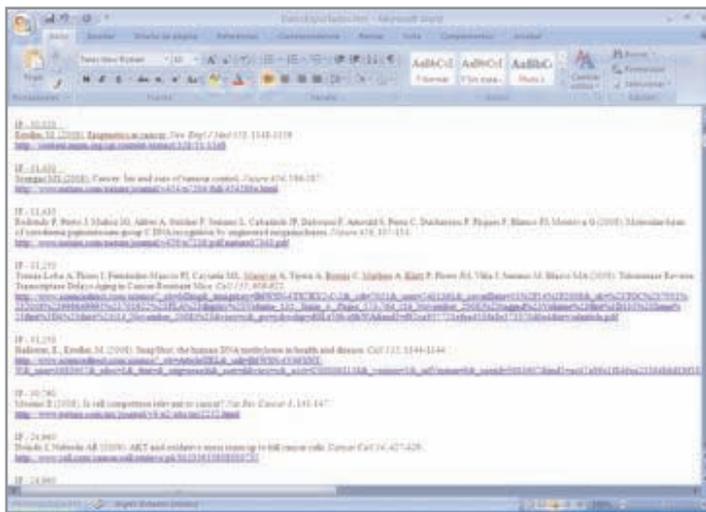


Figura 3. Ejemplo de informe obtenido en formato texto.

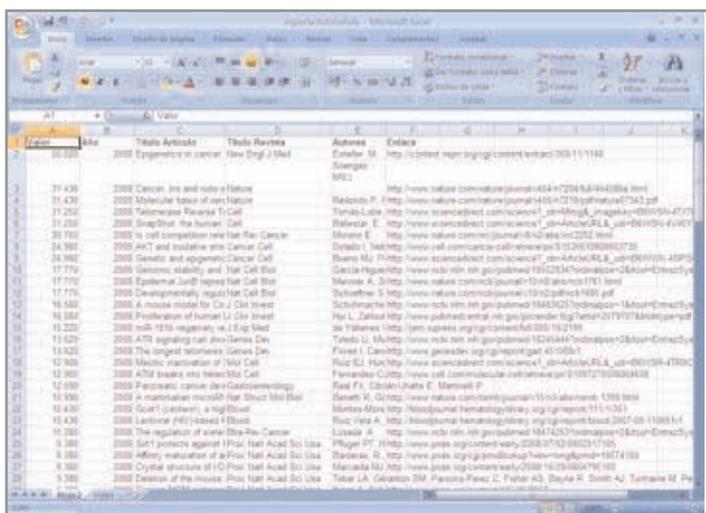


Figura 4. Ejemplo de informe obtenido en formato de hoja de cálculo.

4.2 Perfiles de usuarios

Los usuarios finales de la base de datos se pueden dividir en distintos grupos. Por una parte está el Servicio de Documentación y Biblioteca que es el usuario gestor-administrador y como tal es el encargado de introducir los datos bibliográficos y de mantener actualizada la información que haya en el repositorio. Por otra parte están los usuarios invitados, que serían los que tienen acceso a la consulta de datos del repositorio, así como a la obtención de algunos informes. Estos últimos tendrán limitado su uso, de tal forma que sólo podrán

realizar las consultas bibliográficas y obtener tipos concretos de informes sin tener la opción a realizar cualquier modificación, mientras que el usuario gestor-administrador será el único con derechos para dar de alta, actualizar o dar de baja registros, además de poder realizar y crear consultas e informes en función de sus necesidades.

5. Conclusiones

Sin duda, el objetivo primero y fundamental de la creación de repositorios institucionales es el incremento de la visibilidad y el prestigio de la propia institución, sin embargo, pensamos que la explotación de estas bases de datos puede ser mucho mayor si conseguimos que no se limiten a ser simplemente un recurso de almacenaje sino que sean herramientas a las que se dote de funcionalidades bibliométricas adicionales. En nuestra opinión, el valor añadido de este tipo de bases de datos de publicaciones que recogen la producción científica de nuestros centros estriba sobre todo en las posibilidades de realizar consultas y de generar informes tanto de tipo bibliográfico como bibliométrico.

Para finalizar hemos de señalar que aunque hasta el momento nuestros repositorios centran su contenido únicamente en artículos o colaboraciones pertenecientes a publicaciones periódicas, entra en nuestros planes incluir en un futuro esperamos que no muy lejano otros tipos de documentos académicos como tesis, capítulos de libros, contribuciones a congresos, pósters y patentes.

Referencias bibliográficas

Barton, Mary R. *Creating an Institutional Repository: LEADIRS Workbook*. [en línea] Cambridge, MA: MIT, 2004. <<http://dspace.org/implement/leadirs.pdf>>. [Consulta: 4/09/2009].

Crow, Raym. *The Case for Institutional Repositories: A SPARC Position Paper. Discussion Paper*. [en línea] Washington, D.C.: Scholarly Publication and Academic Resources Coalition, 2006. <http://scholarship.utm.edu/20/1/SPARC_102.pdf>. [Consulta: 4/09/2009].

Lynch, Clifford A. “Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age” [en línea] *ARL: A Bimonthly Report*, 2003, febrero, n. 226, p. 1-7. <<http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>>. [Consulta: 3/09/2009].

Melero, Remedios. “El paisaje de los repositorios institucionales *open access* en España” [en línea] *BiD: textos universitaris de biblioteconomia i documentació*, 2008, junio, n. 20. <http://www2.ub.edu/bid/consulta_articulos.php?fichero=20meler4.htm> [Consulta: 4/09/2009].

Ministerio de Ciencia e Innovación. *Borrador del Anteproyecto de la nueva Ley de la Ciencia y la Tecnología*. [en línea] 2009. <<http://lcyt.fecyt.es>>. [Consulta: 4/09/2009].

¹ The Directory of Open Access Repositories – OpenDOAR. [en línea]
<http://www.opendoar.org/index.html> [Consulta 14.09.09].

² Como ejemplo de repositorio universitario podemos echar un vistazo a *Phaidra*, el repositorio de la Universidad de Viena [en línea] <<https://phaidra.univie.ac.at/>>, en el que se pueden consultar todo tipo de materiales desde libros, artículos de revistas o actas de congresos hasta mapas o incluso folletos. [Consulta 10.09.09].