

Depósitos digitales 2.0: la experiencia de la UAB

Cristina Azorín (*)
Gestora de Serveis d'Informació
Universitat Autònoma de Barcelona
Teléfono: 93 581 16 34 Fax: 93 581 32 19
E-mail: Cristina.Azorin@uab.es

Ferran Jorba
Cap de Projectes Informàtics
Universitat Autònoma de Barcelona
Teléfono: 93 581 42 40 Fax: 93 581 20 94
E-mail: Ferran.Jorba@uab.es

Beatriu Piera
Gestora de Projectes
Universitat Autònoma de Barcelona
Teléfono: 93 581 48 15 Fax: 93 581 32 19
E-mail: Beatriz.Piera@uab.es

En estos momentos las universidades españolas se encuentran en pleno desarrollo de la gestión de la documentación digital. En unos casos se trata de recoger la documentación institucional, en otros de gestionar compras o donativos, y en los más afortunados se dispone de presupuestos especiales para la digitalización. Las bibliotecas han reaccionado con la creación de depósitos institucionales y de sistemas de gestión de biblioteca digital.

Inmerso en el proceso de adaptación al nuevo Espacio Europeo de Educación Superior, basado en el autoaprendizaje, la formación continua y la adquisición de competencias en gestión de la información, el Servicio de Bibliotecas de la Universidad Autónoma de Barcelona ha realizado importantes inversiones en su biblioteca digital y ha apostado por la creación de un Depósito Digital de Documentos (DDD) propio.

El programa CDS Invenio (elaborado por el CERN) ha permitido integrar colecciones de muy diferente tipología, temática, origen e incluso formato. Los registros en MARC21 permiten una mayor flexibilidad en la descripción de los recursos y a la vez unas mayores capacidades de búsqueda.

Describiremos los procedimientos que se han seguido (identificación de problemas, selección del programa, coordinación, gestión del mantenimiento, calendario...) con el objetivo de llegar a conclusiones que puedan ser de utilidad para iniciativas similares.

Palabras clave: depósito digital ; repositorio ; CDS Invenio ; CDSware ; Universidad Autónoma de Barcelona

1. ¿Quiénes?

El origen del DDD (Depósito Digital de Documentos, <http://ddd.uab.es>) se produce de un modo inverso al de buena parte de depósitos digitales abiertos. En nuestro caso, no podemos desvincular la creación del DDD de la necesidad urgente de contar con una herramienta de gestión de la colección digital local, y ello ha marcado sin duda su consiguiente definición y el planteamiento de los objetivos.

La literatura creciente sobre el tema, el software de continua aparición, y los enfoques que las diferentes instituciones están haciendo de sus depósitos digitales hace que resulte muy difícil (cuando no imposible) hablar de un modelo único o de una tipología clara. Lo que sí existe en las bibliotecas universitarias es una tendencia unánime de intentar proporcionar un sistema de preservación y distribución de la investigación y los materiales docentes que se producen en formato digital.

Las categorías pueden variar mucho según el contenido (tesis, revistas, pre-publicaciones...), su propósito (mejora del acceso, gestión del fondo digital, preservación...), o el enfoque al usuario (aumento de la visibilidad, introducción directa a Internet, copia de protección...).

El nacimiento de DDD se produce como consecuencia de importantes cambios en la tecnología aplicada a nuestros servicios de información, en una andadura que comienza con la participación de la UAB en el proyecto europeo Decomate, Delivery of Copyright Material to End-Users (marzo de 1995). El objetivo de este proyecto era el desarrollo de un sistema informático que permitiese la gestión local de documentos proporcionados directamente por los productores comerciales en formato electrónico, básicamente revistas de Elsevier Science y Kluwer Academic Publishers, además de las propias de la UAB. Los documentos estaban sujetos a unas condiciones de acceso negociadas para garantizar la consulta del texto completo a los usuarios de la propia universidad. Este proyecto nos permitió tener experiencia en el desarrollo y mantenimiento de servicios de gestión local de revistas digitales.

En el año 2005 se toma la decisión de cerrar el servidor que daba acceso a las revistas digitales que habían participado en el proyecto Decomate, la mayoría se encontraban ya disponibles a través de otras plataformas comerciales. En el momento de cerrar el servicio alojaba 362.073 artículos científicos, en texto completo, de 436 revistas científicas de 3 editores, ocupando los PDFs un total de 177 GB.

2. ¿Qué y cuándo?

En nuestro caso no es exacto afirmar que hemos creado un depósito *institucional* universitario ya que no nos limitamos exclusivamente a recoger los materiales digitales creados por la institución y/o por los miembros de nuestra comunidad universitaria.

The screenshot shows the homepage of the 'Depósito Digital de Documentos de la UAB - Servicio de Bibliotecas'. At the top left is the logo with the text 'Dipòsit Digital de Documents UAB'. A navigation bar includes links for 'visitante', 'sesión', 'avisos', 'cestas', and 'identificación'. Below the logo, it states 'Página principal' and 'Depósito Digital de Documentos de la UAB - Servicio de Bibliotecas'. A search bar indicates 'Buscar en 6,902 registros por:' with a dropdown menu set to 'cualquier campo' and buttons for 'Buscar' and 'Explorar'. Below the search bar are links for 'Consejos para la Búsqueda' and 'Búsqueda Avanzada'. A section titled 'Búsqueda por colección:' lists several categories with checkboxes and counts: 'UABPubD: Publicacions digitals de la UAB (3,561)', 'Revistes científiques (3,517)', 'Butlletins corporatius (13)', 'Llibres i col·leccions (6)', 'Eines d'Innovació Docent en Educació Superior (3)', 'Butlletins d'estudiants (1)', 'Otras publicaciones (22)', 'Lengua y literatura catalanas (1,921)', 'Història i crítica de la narrativa catalana contemporània (1,295)', 'Traducció i recepció de les literatures estrangeres a la literatura catalana, 1992-1939 (626)', 'Materiales de curso (1,387)', 'Programas de curso (349)', 'Exámenes (1,008)', 'Documentación empresarial (63)', and 'Informes anuales (1)', 'Declaraciones medioambientales (62)', 'Informes de sostenibilidad (0)'. On the right side, there is a 'NOTICIAS Y NOVEDADES' section with a short paragraph. At the bottom, there is a footer with contact information, a language selection menu (Català, Español, English), and a copyright notice: 'Última actualización 22 Aug 2006 11:12:19 CEST'.

Fig. 1: página principal de DDD

El objetivo principal de DDD es facilitar el acceso y mejorar la visibilidad del material digital que se encuentra en la institución o que puede ofrecerse a través del Servicio de Bibliotecas por acuerdos con los editores.

Como objetivo secundario, más amplio pero no menos importante, se trataría de tener en funcionamiento un servicio que promueva el acceso a la producción docente del profesorado dentro del proceso de transformación que ha introducido el Espacio Europeo de Educación Superior en las universidades (EEES).

DDD sería un modelo híbrido entre un depósito institucional y un gestor de biblioteca digital. Es fruto de la colaboración con otros servicios y departamentos de la Universidad Autónoma de Barcelona, y recibe un claro apoyo por parte de la Vicegerencia de Asuntos Académicos, lo cual puede dar un fuerte empuje a nuestro deseo de difundir en el mayor grado posible la producción científica institucional.

La indefinición en este caso es por tanto explícita: DDD no se concibe como un depósito de acceso abierto que recoge un tipo específico de material, sino como un nuevo servicio que se ofrece a la comunidad universitaria para describir y difundir los materiales digitales. La preocupación por el tema de los depósitos institucionales por parte del Servicio de Bibliotecas de la Universidad Autónoma de Barcelona se muestra claramente con la participación, desde sus inicios, en los proyectos que el Consorcio de Bibliotecas Universitarias de Cataluña (CBUC) pone en marcha en colaboración con el Centro de Supercomputación de Cataluña (CESCA).

Según la propia definición de Ricard de la Vega (CESCA) son tres "depósitos electrónicos colectivos institucionales de e-información" que facilitan la adopción de procedimientos comunes, reducen costos y tienen ventajas a nivel de consolidación y prestigio (políticas de preservación y mayor visibilidad).



Fig. 2: imagen de cada uno de los depósitos

En febrero de 2001 nació el primer depósito del CBUC i el CESCA con el nombre de TDC@t, actualmente TDR, Tesis Doctorales en Red (<http://www.tesisenred.net/>), que tiene como finalidad recoger las tesis doctorales leídas en las universidades de Cataluña.

En las mismas fechas se empieza a desarrollar el segundo depósito consorciado: RACO, Revistas Catalanas en Acceso Abierto (<http://www.raco.cat/>). RACO permite la consulta en acceso abierto y en texto completo de los artículos de las revistas científicas, culturales y eruditas catalanas.

RECERCAT, Depósito de la Investigación de Cataluña (<http://www.recercat.net/>) sería nuestro tercer depósito, en funcionamiento desde septiembre de 2005.

La mayor parte del material recogido en DDD es de carácter textual, pero los registros bibliográficos (en MARC21) permiten tanta tipología de documentos como cualquier catálogo de biblioteca. Es fundamental el tratamiento que damos a la información, la descripción e indización de los recursos por parte de personal bibliotecario.

Respecto a su contenido, es muy variado: desde material docente como los programas o exámenes, a artículos científicos de revista, pasando por alguna base de datos bibliográfica o cartas manuscritas. Hay un deseo expreso de flexibilidad en el tipo de material que debe incluirse.

Las colecciones y subcolecciones (en forma de árbol) son el criterio principal de agrupación de los registros en CDS Invenio. Una colección es un conjunto de registros que cumplen una condición de búsqueda determinada, y las hay de dos tipos: las colecciones regulares y las colecciones virtuales.

Al introducir un registro nuevo siempre debemos adscribirlo a una colección regular. Las colecciones regulares se definen por un valor (y sólo uno) en el subcampo 980\$a, por el que se determina la inclusión exclusiva, es decir, que el registro que pertenece a una colección no pertenece a ninguna otra. La idea es que la suma de todos los registros de las colecciones regulares corresponda exactamente a la suma de todos los registros de la base de datos.

Paralelamente, además, podemos definir colecciones virtuales, agrupando un conjunto de registros que tienen una característica transversal (por ejemplo, un mismo autor, lengua, o temática). A partir de esas asociaciones un registro puede pertenecer sin problemas a varias colecciones virtuales.

Artículos científicos de acceso abierto

The screenshot displays the search interface of the DDD system. At the top, it shows a search bar with the text "Buscar en 3,228 registros por:" and a dropdown menu set to "cualquier campo". There are "Buscar" and "Explorar" buttons. Below the search bar, there is a link for "Consejos para la Búsqueda :: Búsqueda Avanzada".

The main content area is divided into two columns. The left column, titled "Últimas adquisiciones:", lists three recent acquisitions with their dates, times, authors, titles, and links to the full records. The right column, titled "Enfocado a:", lists various virtual collections with their respective counts.

Últimas adquisiciones:	
2006-07-10 16:38	Goffman, E. Sacks, H. Cicourel, A. Poller, M. y Díaz, F. Sociologías de la Situación / Vánegas, Guillermo E. [...] Athenea digital : revista de pensamiento e investigación social, N. 0 (2001) http://ddd.uab.es/pub/athdig/15788946n0a8.htm Registro completo - Registros similares
2006-07-10 15:39	Castien Maestro, Juan Ignacio. Las Astucias del Pensamiento. Creatividad ideológica y adaptación social entre los inmigrantes marroquíes en la Comunidad de Madrid / Astudillo Moreno, Sandra Juan Ignacio CASTIEN MAESTRO, Las Astucias del Pensamiento. [...] Athenea digital : revista de pensamiento e investigación social, N. 7 (2005) http://ddd.uab.es/pub/athdig/15788946n719.htm Registro completo - Registros similares
2006-07-07 16:35	Sáez, Javier. Teoría Queer y psicoanálisis / Córdoba García, David Javier SÁEZ, Teoría Queer y psicoanálisis. [...] Athenea digital : revista de pensamiento e investigación social, N. 7 (2005) http://ddd.uab.es/pub/athdig/15788946n728.htm Registro completo - Registros similares

Enfocado a:
Anàlisi (216)
Athenea digital (300)
Ciències (47)
Catalan Journal of linguistics (47)
Catalan working papers in linguistics (38)
Documents d'anàlisi geogràfica (213)
Enrahonar (122)
Enseñanza de las ciencias (299)
Educar (181)
Faventia (224)
HMiC (53)
Links & Letters (62)
Locus amoenus (133)
Manuscripts (112)
Orsis (79)
Papers (339)
Publicacions matemàtiques (219)
Quaderns (221)

Fig. 3: colección regular a la izquierda y colecciones virtuales a la derecha

La Web 2.0, según Tim O'Reilly, es una evolución de la Web tradicional donde, entre otras características, se fomenta la participación del usuario final para enriquecer el contenido de las páginas. El Web 2.0 es una actitud y no precisamente una tecnología. En este ámbito

pueden enmarcarse algunas opciones que ofrece actualmente DDD o que están en estudio para ser implementadas en próximas versiones:

1. Registros similares. Permite la consulta de registros que tengan similitudes con el que estamos consultando en ese momento: el mismo autor, la misma materia...
2. Otros registros consultados, nos muestra qué otras páginas vieron los usuarios que también consultaron el registro que nos interesa. Se establecen redes sociales dentro del depósito.
3. A partir de la versión 0.90, los usuarios pueden añadir reseñas y comentarios a los registros bibliográficos, con lo que se enriquece la información catalográfica clásica con estas notas.

3. ¿Cómo?

A nivel mundial la mayoría de depósitos digitales de acceso abierto se generan a partir de software libre, entre sus ventajas destacaremos tres: en muchos casos no hay un gasto económico (aunque no siempre libre quiere decir gratuito), son en general programas muy flexibles que se adaptan a muchas necesidades, y por último, pero no menos importante, tienen detrás amplias comunidades de usuarios que participan en su continuo desarrollo.

Por parte del Consorcio de Bibliotecas Universitarias de **Cataluña** (CBUC) se creó en 2004 un grupo de informáticos y bibliotecarios para estudiar la selección de un programa que permitiese la gestión de los depósitos digitales institucionales de libre acceso. Las dos condiciones básicas es que se tratase de software libre y que se pudiese implementar de manera inmediata. La elección final fue el programa DSpace (<http://www.dspace.org/>), desarrollado conjuntamente por las bibliotecas del MIT (Massachusetts Institute of Technology) y la empresa Hewlett-Packard. La descripción de los objetos digitales se hace a través de metadatos en formato Dublin Core. Existe ya una federación de depósitos de DSpace denominada FeDCOR (Federación DSpace utilizando CORDRA) que se encuentra en una etapa de pruebas.

El Servicio de Bibliotecas de la UAB, en cambio, se decantó por el programa CDS Invenio, entre otras razones porque DSpace no permitía un tratamiento y visualización ágil y flexible de las publicaciones periódicas, que era uno de los objetivos más urgentes del proyecto.

CDS Invenio (<http://cdsware.cern.ch/invenio/>) es el resultado de la unificación de diferentes aplicaciones del CERN (European Organization for Nuclear Research) para facilitar la gestión de su colección digital de documentos. Esta aplicación, como su hermana CDS Indico –que gestiona conferencias y congresos–, está escrita en el language de programación Python. Aunque en su momento también había módulos escritos en PHP, en estos momentos se están eliminando y reescribiendo en Python. La base de datos es MySQL, pero utilizando unos índices propios, ya que, según las pruebas que hicieron en el CERN, los índices nativos de MySQL no garantizaban un buen rendimiento con más de 100.000 registros.

En el caso de CDS Invenio, como en la mayoría de las aplicaciones de software libre, es muy importante la comunidad de usuarios con desarrolladores del programa fuera de lo que es el núcleo inicial de creación del mismo, que añaden necesidades diferentes, sugieren soluciones y aportan mejoras. CDS Invenio es un sistema seguro (en el sentido de compatibilidad futura, ya que sigue los estándares actuales), económico (inversión en servidores y recursos

humanos) y rápido. Por nuestra parte, la implementación completa ha durado aproximadamente un año.

La cooperación con otras bibliotecas a nivel de gestión interna es importante. Como hemos comentado a lo largo de este texto, la UAB participa en los depósitos institucionales que actualmente tiene el CBUC en colaboración con el CESCO. En este sentido resulta imprescindible hacer referencia a las actividades en pro de la digitalización de contenidos que la Unión Europea promueve dentro del programa **plurianual** (2004-2008) E-Content plus (http://europa.eu.int/information_society/activities/econtentplus/programme/index_en.htm). Este programa **centra** su interés en todas aquellas actividades y herramientas que faciliten el acceso a contenidos digitales, su uso y explotación, así como en la calidad de los contenidos a través del uso de metadatos y estándares bien definidos.

El hecho de que el programa genere registros en MARC21 hace que trabajemos necesariamente con un mínimo de rigor profesional. No descartamos que en algún momento el autor original introduzca los documentos directamente, pero procurando que, de forma invisible, se establezca una conversión interna entre los campos del formulario (donde éste introduce los datos) y los registros en MARC21 que pasan al sistema. Nuestro mayor desafío en este sentido es conseguir alcanzar unos niveles de estandarización que hagan posibles nuestras metas de cooperación sin alterar por ello las posibilidades que nos ofrece el autoarchivo.

4. ¿Por qué?

Los puntos fuertes que nos llevaron a seleccionar este programa frente a otros que se encuentran disponibles para la misma función fueron:

- MARC21 como formato bibliográfico interno. Las posibilidades de gestión van más allá del Dublin Core, aunque, por otra parte, también aumenta la complejidad de la configuración.
- Interfaz multilingüe y multialfabeto.
- Grandes posibilidades de búsqueda: por índice, palabra clave, expresiones regulares, registros parecidos...

Los campos de búsqueda se pueden definir específicamente por cada colección, de manera que lo que a nivel de MARC21 es una etiqueta 100 en todos los casos, en unas colecciones se puede definir como autor y en otras como profesor.

Exámenes

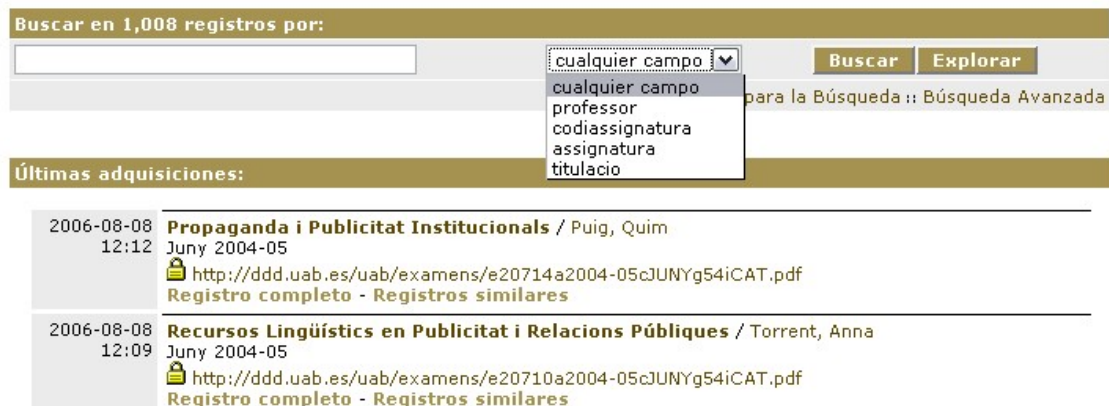


Fig. 4: desplegable de búsqueda de la colección Exámenes

- Gran velocidad de respuesta independientemente de la cantidad de registros que contenga, en concreto la base de datos del propio CERN (<http://cdsweb.cern.ch/>) consta ya de más de 900.000.
- Admite registros de tipología muy variada: *pre-prints*, conferencias, congresos, colecciones históricas de cartas, fotografías, vídeos...
- Permite limitar el acceso al texto completo. El hecho de que muchos recursos sólo puedan consultarse desde la red propia de la UAB no es una característica negativa del sistema. La garantía de un acceso restringido favorece que los autores o editores estén dispuestos a llegar a acuerdos para la difusión de material.
- Personalización por parte del usuario.

Como limitaciones decir, por ejemplo, que no es compatible con Z39.50.

4. Conclusiones

Frente a la dispersión en cuanto a tipología y a software el Servicio de Bibliotecas de la UAB ha optado por crear un solo depósito, con un único software que integre todos los materiales y formatos que se necesiten.

Viendo el ritmo acelerado con el que avanzan las tecnologías, el factor principal a la hora de crear un depósito digital es utilizar estándares abiertos que garanticen un traspaso fácil de los datos en caso de cambio en el futuro y una interoperatividad con otras plataformas para rentabilizar nuestro esfuerzo.

Para que los usuarios consideren que el servicio es útil se debe disponer de una masa crítica de documentos; no recomendamos iniciar una campaña de difusión de un nuevo depósito digital si no se tiene un cierto volumen de entradas que justifiquen el esfuerzo de su consulta por parte del usuario final.

Uno de los objetivos principales era conseguir un sistema de búsqueda global para todas las colecciones. Era muy importante para nosotros que se pudiesen hacer búsquedas tipo Google,

por palabra clave en cualquier campo; en caso de no encontrar la palabra exacta aparece un índice de términos aproximados.

Las universidades han aprovechado las actuales condiciones que facilitan la gestión de recursos digitales: ancho de banda, software libre, creación digital en origen, servidores institucionales... El prestigio que da la visibilidad de las investigaciones en Internet es básico y es un compromiso social publicar contenidos científicos de acceso gratuito. Ese es nuestro futuro, debemos generar contenidos de calidad que fomenten las investigaciones científicas.

Además tenemos previsto implementar en breve el protocolo OAI-PMH para poder intercambiar nuestros metadatos (en formato Dublin Core) con otras plataformas y sobre todo con los depósitos de acceso abierto, ya creados o de futura creación, del CBUC.

5. Bibliografía consultada

BORRÀS, A. ; CARREÑO, J. ; JORBA, F. ; PRATS, J. ; ROS, R.. Programari de codi lliure per gestionar dipòsits digitals: el procés de tria dut a terme al CBUC. *BiD : textos universitaris de biblioteconomia i documentació*, 2006, n. 16. [en línea]. [Consulta: 10 de septiembre de 2006]. Disponible en:
http://www2.ub.edu/bid/consulta_articulos.php?fichero=16borras.htm

CERN. *Workshop on Innovations in Scholarly Communication: implementing the benefits of Oai*. Geneva, 12- 14th February. Geneva: CERN, 2004 [en línea]. [Consulta: 16 de noviembre de 2006]. Disponible en: <http://oai3.web.cern.ch/oai3/>

LYNCH, C. A. ; LIPPINCOTT, J. K. Institutional repository deployment in the United States as of early 2005. *D-Lib magazine*, 2005, vol. 11, n. 9. [en línea]. [Consulta: 22 de junio de 2006]. Disponible en: <http://www.dlib.org/dlib/september05/lynch/09lynch.html>

DE LA VEGA, R., [et al.]. "Dipòsits col·lectius d'e-informació". En: *V Jornades de Programari Lliure, 5-8 de julio de 2006*. Barcelona: Escola Tècnica Superior d'Enginyeria Industrial, 2006 [en línea]. [Consulta: 15 de noviembre de 2006]. Disponible en: <http://portal.jornadespl.org/ponencies/diposits-colb7lectius-d-e-informacio-programari-lliure-i-acces-obert-un-binomi-guanyador/>

JEREZ, H.; MANEPALLI, G.; NELSON, M. L. FeDCOR: un registro institucional CORDRA. *D-Lib Magazine*, 2006, vol. 12, n. 2. [en línea]. [Consulta: 18 de noviembre de 2006]. Disponible en: http://www.cnri.reston.va.us/papers/FeDCOR_spanish.pdf

HIXSON, C. Un cop creats, què? El futur dels dipòsits institucionals. *BiD : textos universitaris de biblioteconomia i documentació*, 2005, n.15. [en línea]. [Consulta: 22 de junio de 2006]. Disponible en:
http://www2.ub.edu/bid/consulta_articulos.php?fichero=15hixso1.htm

NISO Framework Advisory Group. *A framework of guidance for building good digital collections*. 2nd ed. Bethesda, MD: National Information Standards Organization, 2004. [en línea]. [Consulta: 22 de junio de 2006] Disponible en:
<http://www.niso.org/framework/framework2.pdf>

O'REILLY, T. *What Is Web 2.0 : Design Patterns and Business Models for the Next Generation of Software* [en línea]. [Consulta: 15 de junio de 2006] Disponible en: <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>

SITAS, A. CDSware (CERN Document Server Software). *Library hi tech*, 2006, v. 24, n. 3, p. 420-429 [en línea]. [Consulta: 23 de octubre de 2006]. Disponible en: <http://www.emeraldinsight.com/Insight/ViewContentServlet?Filename=Published/EmeraldFullTextArticle/Articles/2380240309.html>

STEENBAKKERS, J. *Setting up a Deposit System for Electronic Publications. The NEDLIB Guidelines*. The Hague: Titia van der Werf, 2002. (NEDLIB Report Series, 5).

VESELY, M. *CERN Document Server Software: document management system for grey literature in networked environment, presentación a GL5*. Amsterdam, 2003 [en línea]. [Consulta: 16 de noviembre de 2006] Disponible en: <http://cdsware.cern.ch/publications/gl5.pdf>

VIVES I GRÀCIA, J. Aspectos de propiedad intelectual en la creación y gestión de repositorios institucionales. *El Profesional de la Información*, 2005, v. 15, n. 4, p. 267-278. [en línea]. [Consulta: 18 de noviembre de 2006] Disponible en: <http://eprints.rclis.org/archive/00004230/01/267.pdf>