Consorci de
Biblioteques Universitàries
de Catalunya

Biblioteca Digital de Cataluña : Continente y contenido

5ª Jornada sobre la Biblioteca Digital Universitaria, Los Polvorines (Argentina), 8 y 9 de noviembre de 2007

Lluís M. Anglada i de Ferrer, Director del CBUC, langlada@cbuc.es
Joan Cambras, Jefe del Departamento de Sistemas del CESCA, jcambras@cesca.es
Sandra Reoyo Tudó, Gestora de proyectos del CBUC, sreoyo@cbuc.es
Ricard de la Vega, Técnico líder del área de repositorios y portales del CESCA, rdelavega@cesca.es

1. Introducción y resumen

Desde 1999 el Consorcio de Bibliotecas Universitarias de Cataluña (CBUC) ha creado una nueva línea de trabajo, junto con el Centro de Supercomputación de Cataluña (CESCA), para promocionar la investigación que se lleva a cabo en Cataluña y al mismo tiempo contribuir al movimiento mundial de depositar la producción académica y de investigación en la red de forma gratuita. Este movimiento mundial, que recibe el nombre de *Open Access*, ha sido puesto en marcha por las instituciones que financian la investigación con la finalidad de crear alternativas al paradigma de pagar por tener acceso a la información que se ha elaborado en la propia institución. Esta nueva línea de trabajo son los repositorios institucionales.

En la tercera edición de estas mismas Jornadas presentamos una comunicación sobre los repositorios institucionales que por entonces teníamos implementados[1]. Ahora después de dos años presentamos su estado actual y reflexionamos sobre la estrategia para lograr la inclusión de documentos (políticas institucionales, comités científicos, etc.), su contenido (estándares usados, derechos de autor, preservación, etc.), su continente (programas y tecnología utilizada, protocolos, etc.) y comentamos los resultados obtenidos en los seis años de experiencia desde que se puso en marcha el primero.

[1] Anglada i de Ferrer, Lluís M. [et al.]. "Los repositorios como componentes esenciales de las bibliotecas". Presentado en: *3ª Jornada sobre la Biblioteca Digital Universitaria, Córdoba (Argentina), 27-28 de octubre de 2005*. <http://hdl.handle.net/2072/1167>. [Consulta: 03/09/07].

2. El Consorcio de Bibliotecas Universitarias de Cataluña: del catálogo colectivo a los repositorios electrónicos

Hace un par de días el Consorcio de Bibliotecas Universitarias de Cataluña (CBUC) celebró su onceavo aniversario (ya que la fecha oficial de nacimiento fue el 6 de noviembre de 1996). A lo largo de este período el CBUC ha iniciado diversos programas de servicios. El Consorcio se creó con el objetivo de constituir un catálogo colectivo que fuera un instrumento de información bibliográfica y de catalogación por copia.

La actividades del CBUC se agrupan bajo tres líneas estratégicas: la primera está destinada a mejorar el acceso a las colecciones impresas, la segunda gira alrededor de la información digital y la tercera está orientada a la innovación. Dentro del primer grupo de actividades se incluye el catálogo colectivo, el préstamo interbibliotecario y la creación de un almacén de preservación de libros y revistas impresas. El CCUC[2] (Catálogo Colectivo de las Universidades de Cataluña) se empezó a confeccionar en 1994 y fue el 4 de marzo de 1996 cuando se inició la catalogación en línea. Actualmente el CCUC da noticia de más 3 millones de documentos diferentes. En el mismo 1996 se realizaron los trabajos preparatorios del programa de Préstamo Interbibliotecario, servicio que entró en funcionamiento en abril de 1997 y que hoy tramita casi 35.000 suministros anuales y que incluye también acuerdos de acceso mutuo a las bibliotecas miembros del Consorcio y de préstamo 'in situ'. En 2003 se inició el proyecto de crear un almacén cooperativo para preservar los documentos de bajo o nulo uso de las bibliotecas del CBUC.

Las actividades de fomento de la innovación del CBUC son diversas. Por ejemplo, realizamos cursos de formación desde 1999 y también desde 2001 traducciones al catalán de textos de la literatura profesional que contribuyen a la formación de los bibliotecarios. Así mismo el Consorcio ha realizado tareas de apoyo a las dos evaluaciones institucionales de las bibliotecas universitarias hechas en 1998-2002 y 2006-2007. También se ha encargado de coordinar el cambio del sistema automatizado de las bibliotecas consorciadas y de la organización de les reuniones internacionales de SELL y de ICOLC en Barcelona.

Entre 1998 y 1999 nace la Biblioteca Digital de Cataluña (BDC). Bajo el nombre de Biblioteca Digital se incluye la información electrónica contratada conjuntamente por los miembros del CBUC y también los repositorios de información electrónica propia. La información electrónica contratada actualmente a editores comerciales incluye unas 8.700 revistas electrónicas, unos 8.000 libros electrónicos y 37 bases de datos. La importancia de las contrataciones de recursos electrónicos hechas bajo el nombre de BDC no nos tiene que hacer olvidar que casi al mismo tiempo se pusieron los fundamentos del primer repositorio electrónico cooperativo de los cinco que mantienen hoy las universidades de Cataluña y la Biblioteca de Cataluña a través del CBUC –consorcio de bibliotecas- y del CESCA –consorcio de supercomputación-. El primer repositorio colectivo de documentos electrónicos fue el de tesis –TDR- que se inició en 1999 aunque no entró en funcionamiento hasta el 2001. Posteriormente se han creado otros repositorios en acceso abierto para revistas -RACO-, para documentos de investigación –RECERCAT-, para imágenes –MDC- y para material docente –MDX-.

[2] http://ccuc.cbuc.es/vtls24/vtls/spanish/. [Consulta: 05/09/07].

3. Definiendo una estrategia para los repositorios

El CBUC y el CESCA empezaron a construir un repositorio para documentos electrónicos bastante antes que estos recibieran el nombre de repositorios institucionales (con el que han sido reconocidos en la literatura profesional posteriormente[3]) y que se acuñara el término del Acceso abierto (OA, *Open Access*). La situación actual es de cinco repositorios activos que utilizan tres software distintos. ¿Esta era la situación que podíamos esperar alcanzar cuando empezamos con el primer repositorio ahora hará unos siete años?
El primer repositorio surgió de una oportunidad. El Gobierno de Cataluña elaboró un plan estratégico para la sociedad de la información en el seno del cual se pudieron formular propuestas innovadoras que favorecieran el uso de Internet en las universidades. Al mismo tiempo de una iniciativa de administración electrónica, de un sistema de compra de libros de las ediciones universitarias a través de la Red, o de creación de material docente digital, el CBUC y el CESCA propusieron la creación de un repositorio electrónico de tesis doctorales a texto completo.

La primera opción estratégica se tomó entonces y se hizo de forma consciente. Entre la opción de crear un software propio para la gestión del repositorio o la de usar uno de ya existente, se optó por esta última. Influyó aquí la experiencia de las bibliotecas de las universidades de Cataluña de los años 80 de elaborar programas 'in house' para la gestión de sus bibliotecas[4]. Los diez años de experiencia elaborando software mostraron que se invierten muchos recursos (entre los cuales el tiempo no es el menos importante) en el desarrollo de programas y que los resultados no siempre se ajustan a los recursos invertidos. Un informe encargado por el CBUC sobre la situación internacional de la digitalización de tesis doctorales nos permitió conocer la existencia de la Networked Digital Library of Theses and Dissertations (NDLTD)[5], que proporcionaba entonces un programa de código libre: ETD[6]. Fue necesario adaptar el programa a los objetivos propios de hacer un repositorio colectivo, pero se sacaron aprendizajes positivos del hecho de utilizar herramientas construidas por otros para propósitos similares a los nuestros.

El segundo gran aprendizaje fue que la principal dificultad en la construcción de un repositorio no era la puesta a punto del programa informático sino la de establecer un marco legal que permitiera iniciar el proyecto y, sobre todo, convencer a la comunidad universitaria de que poner documentos (en este caso las tesis) en acceso abierto y en un repositorio no solo es fácil sino que tiene además muchas ventajas. Más adelante en esta ponencia desarrollaremos los mecanismos creados para el repositorio de tesis y para los demás para facilitar la introducción de documentos y para disminuir las resistencias de editar electrónicamente y en abierto.

---

[3] Crow, Raym. "The Case for Institutional Repositories: a SPARC Position Paper. Release 1.0". Washington: Scholarly Publishing & Academic Resources Coalition, 2002. <http://www.arl.org/sparc/bm~doc/ir_final_release_102.pdf>. [Consulta: 31/07/07].
[4] Anglada i de Ferrer, Lluís M. "Veinticinco años de automatización de bibliotecas en Cataluña". *BiD: textos universitaris de biblioteconomia i documentació*, núm. 16 (juny 2006). <http://www.raco.cat/index.php/BiD/article/view/40499/41021>. [Consulta: 03/09/07].
[5] Dijkstra, J. M. M. "Towards Storage and Access of Electronic Theses [manuscrito no publicado]: A Proposal of Its Organisation for the Consortium of Catalan Libraries". Barcelona, 1999.
[6] http://etd.vt.edu. [Consulta: 04/09/07].

Cuando se puso en funcionamiento el repositorio de tesis se barajó también la posibilidad de proponer crear otro para revistas. Por cuestiones de oportunidad la idea quedó aplazada. Pero el CBUC mantenía desde 1998 una base de datos de sumarios electrónicos que con modificaciones leves permitió añadir a los sumarios de revistas el texto completo de artículos. Así fue como en 2001 pudimos poner la primera revista en Internet a texto completo. El proyecto recibió más tarde el nombre de RACO pero a pesar del acierto del objetivo y de los esfuerzos que se le dedicaron quedó claro que usar un programa no adecuado limitaba el proyecto hasta el punto de hacerlo inviable.

Entre 2001 y 2003 el CBUC llevo a cabo un seguimiento de la literatura profesional referente a programas de automatización de bibliotecas ya que se hacía necesario la sustitución del que entonces estaba en uso. Esto permitió observar que había nuevas realidades emergentes en los servicios bibliotecarios: los repositorios institucionales (IR, *Institutional Repositories*) y que estaban surgiendo nuevos programas (muchos de ellos de código libre) para la gestión de la 'biblioteca digital' o más específicamente para la gestión de objetos digitales. En este período se empezó a gestar un cambio en lo que había sido un sólido paradigma a lo largo le los últimos 20 años: los sistemas automatizados de gestión de las bibliotecas que habían encontrado su éxito en la integración de funciones empezaron, gracias a los estándares de interoperabilidad, a ser concebidos 'desintegrados'. Esto hacía que el panorama fuera distinto y que se considerara conveniente crear un sistema de gestión de objetos digitales (aún no lo llamábamos repositorio) al margen del sistema integrado de gestión bibliotecaria.

Desde 1999 se había estado trabajando también, aunque a nivel teórico, con los esquemas entonces emergentes de metadatos y concretamente con la aplicación de Dublin Core para la 'catalogación' de recursos bibliotecarios. Pronto se llegó a la conclusión de que sin un repositorio en funcionamiento en base real nunca se llegaría a poder concretar un conjunto de metadatos que facilitara el descubrimiento y la usabilidad de los documentos digitales. Entre finales de 2003 y principios de 2004 nos propusimos crear un repositorio de objetos digitales no con la finalidad de que este fuera un servicio en si mismo, sino como un soporte a las actividades 'digitales' de la biblioteca universitaria (el propio OPAC, los sistemas de aprendizaje electrónico o la edición digital) y como un modo de mejorar la preservación de los documentos digitales entonces ya frecuentes en las universidades. Se acordó seleccionar conjuntamente un mismo programa con los objetivos de experimentar con los procesos de captación y catalogación de objetos digitales teniendo en cuenta que el objetivo de las bibliotecas es facilitar la consulta (actual y futura) de los objetos digitales.

Se pretendía entonces usar este repositorio de forma inmediata para revistas electrónicas o para versiones digitales de revistas elaboradas originalmente en papel y para ampliar más adelante el dominio del repositorio a *working papers*, materiales docentes, ficheros audio y vídeo, etc. No se tenía claro si el repositorio sería único o no, pero se creía firmemente que un solo programa informático permitiría gestionar la variedad de objetos digitales mencionados.

El grupo de trabajo para la selección del software, gracias a que se pudo basar en una guía de la Soros Foundation[7], concluyó su tarea en poco tiempo y seleccionó como programa más adecuado para nuestras necesidades y objetivos: DSpace[8]. Dentro de

[7] Crow, Raym. "A Guide to Institutional Repository Software". 3rd. ed. New York: Open Society Institute, August 2004. <http://www.soros.org/openaccess/software>. [Consulta: 29/09/05].
[8] http://www.dspace.org/. [Consulta: 04/09/07].

2004 empezaron los trabajos de adaptación de DSpace a las necesidades propias y a ponerlo en funcionamiento, pero pronto se vio que el programa se adaptaba perfectamente a los *working papers* y a la literatura gris generada en el proceso de la investigación universitaria pero no tanto a las revistas[9] o a otros objetos digitales. El repositorio se concretó en un repositorio de literatura gris de investigación (excepto tesis) con el que las bibliotecas de las universidades de Cataluña empezaron a determinar políticas y procesos de trabajo comunes en el marco de DSpace (por ejemplo, definir comunidades, subcomunidades y colecciones), tomar decisiones respecto a metadatos y a experimentar con los circuitos y procedimientos para hacer edición electrónica en el marco de los repositorios.

Los trabajos continuaron y a lo largo del curso académico 2004-2005 se eligió el programa OJS[10] para el desarrollo del repositorio de revistas en OA y, un poco más tarde, el programa comercial CONTENTdm[11] de OCLC para la gestión del repositorio Memoria Digital de Cataluña (para imágenes y reproducciones de objetos). Más recientemente, de hecho hace solo algunos meses, se ha decidido iniciar un repositorio de material docente (MDX) y usar para ello también DSpace.

Después de lo contado podría parecer que no hay estrategia en la creación de repositorios, pero no es así. El repositorio de tesis ya había permitido fijar dos aprendizajes: preferir los programas ya desarrollados (aunque se tuvieran que adaptar) a los perfectos pero aún no existentes y asumir que las mayores dificultades estaban en convencer a la comunidad universitaria que 'publicar' en electrónico y en abierto era bueno y además no era difícil. El proceso (fallido) de búsqueda de un software ideal para propósitos múltiples nos convenció que las características de algunos tipos de documentos requerirían aún durante algún tiempo que tuviéramos que usar programas distintos. Se fijó entonces una política de no cuestionar (al menos de momento) los contenedores ya seleccionados para centrarse en llenarlos de contenido.

Se tiene la visión (que se retomará en el apartado de conclusiones) de que algunos desarrollos notables en el ámbito de los repositorios están por venir y que el impacto de las búsquedas federadas puede ser determinante para decidir hasta que punto es mejor el uso de un solo programa de propósito general para todo tipo de objetos o el de programas específicos para cada tipo de documento.

[9] No obstante, hay diversas experiencias de revistas electrónicas sostenidas por DSpace, por ejemplo el repositorio "Revistes i congresos UPC" (https://upcommons.upc.edu/revistes/) de la Universidad Politécnica de Cataluña, universidad miembro del CBUC.
[10] http://pkp.sfu.ca/?q=ojs. [Consulta: 04/09/07].
[11] http://www.contentdm.com/. [Consulta: 04/09/07].

4. Continente de los repositorios

Los repositorios de la BDC tienen diversas características comunes: son de acceso abierto; cumplen con el protocolo de interoperabilidad de la Open Archives Initiative[12]; usan metadatos Dublin Core; están construidos y alimentados de manera cooperativa; y cuatro de los cinco repositorios están desarrollados con software libre: TDR, RECERCAT, RACO y MDX. MDC, como se ha comentado anteriormente, se ha instalado utilizando un software propietario, CONTENTdm con licencia para un número ilimitado de objetos.

Estos repositorios están instalados en un clúster Linux de alta disponibilidad con las características de balanceo de carga de las consultas que reciben, y de tolerancia a fallos en caso de desastre en alguno de los nodos que componen la plataforma.

A continuación se especifican los software libres escogidos para cada repositorio, las principales modificaciones que se han realizado y la interoperatividad de estos con otros repositorios nacionales e internacionales a través del protocolo OAI-PMH.

4.1 Software libre

La tendencia mundial es implementar los repositorios institucionales con software libre. Son muchas las ventajas que este tipo de software ha aportado a los repositorios pero, sin lugar a dudas, la principal ha sido la adaptabilidad, ya que ha permitido realizar mejoras que han acercado los programas a los requisitos establecidos para cada uno de ellos. A parte de la adaptabilidad y de los costes, los repositorios también se han beneficiado de otras ventajas del software libre, como el escrutinio público, donde el código está sometido a un proceso de revisión que aporta dinamismo a la corrección de errores, o la independencia del proveedor, hecho que garantiza la continuidad teórica del software.

Las comunidades internacionales que respaldan los programas en que se basan los repositorios (comunidad DSpace, OJS, ETD-main, etc.) son importantes (por la cantidad de instituciones implicadas, por el volumen de usuarios, etc.), lo cual ya de por sí es una garantía de continuidad.

En TDR, a nivel de software, pueden diferenciarse dos capas, la de la gestión de contenidos de la colección de tesis y la de la indización y publicación de los datos. Para la primera se adaptó el Electronic Theses and Dissertations (ETD) de la Virginia Tech University (VT) como un proyecto de colaboración entre la Graduate School de VT, la Digital Library and Archives (DLA) y la Networked Digital Library of Theses and Dissertations (NDLTD), de la que TDR forma parte. Para la segunda capa se usan Glimse y WebGlimse.

RECERCAT está implementado a partir del programa de código abierto DSpace (actualmente, en su versión 1.4). Este software fue creado por el Massachussets Institute of Technology (MIT) y Hewlett Packard (HP). DSpace es un repositorio digital innovador que captura, almacena, indiza, preserva y redistribuye los resultados de la investigación de una organización. Se han realizado diferentes adaptaciones para la búsqueda con diacríticos, la clasificación por materias de los documentos según la Clasificación Decimal Universal (CDU), el control de autoridades y las búsquedas por tipo de documento. También se ha implementado un módulo de estadísticas personalizado ya que los diferentes *pluggins* que se ofrecen para DSpace no cumplían con los requisitos que se

[12] http://www.openarchives.org/. [Consulta: 04/09/07].

deseaban para el proyecto. Dichas estadísticas se pondrán en producción en otoño de 2007, y se podrán consultar de manera gráfica: los documentos incorporados y sus accesos por institución, los accesos por países y también clasificados por las materias de los documentos.

RACO se basa en el programa de código abierto Open Journal Systems (OJS), un software desarrollado por el Public Knowledge Project (PKP) con el objetivo de promover el acceso a la investigación, facilitando la administración y publicación de revistas científicas. Se han desarrollado dos nuevos módulos: uno para implementar una clasificación temática de revistas y otro para simplificar el proceso de publicación.
La condición de software libre nos ha permitido también, adaptar cada repositorio a distintos idiomas y después contribuir con estas traducciones a la comunidad internacional. Todos los repositorios son accesibles en catalán, castellano e inglés y además, TDR también lo es en gallego.

Finalmente en cuanto a costes, cabe remarcar que el de adquisición del software libre es inexistente pero no así el de desarrollo e implantación. Para adaptar los software seleccionados a las necesidades establecidas de cada repositorio es necesario invertir tiempo y recursos tanto materiales como de personal.

4.2 Open Archives Initiative

Los repositorios utilizan un protocolo de interoperabilidad creado por la Open Archives Initiative (OAI) que aumenta la visibilidad de la e-información compartiendo el metadato del repositorio con otros repositorios internacionales, como OAIster[13]. El protocolo de OAI para recolección de metadatos (OAI-PMH) es un mecanismo de bajo nivel para la interoperabilidad entre repositorios. Hay dos actores posibles, por un lado, los proveedores de datos (*data providers*) que son repositorios que exponen metadatos estructurados vía el protocolo OAI-PMH. Por otro lado, los proveedores de servicio (*service providers*) que efectúan peticiones OAI-PMH a los proveedores de datos para recolectar sus metadatos.

TDR es el repositorio que hace más tiempo que es proveedor de datos de este protocolo (desde 2003). En parte, esto puede explicar el gran aumento en las consultas que recibe. Se pasó de 278.708 consultas anuales en 2003 a 1.184.750 en 2004, 2.211.960 en 2005 y 3.543.640 en 2006. En el primer cuatrimestre de 2007 se han recibido 1.369.112 consultas.

Los cuatro repositorios de la BDC que están en funcionamiento actualmente son proveedores de datos del protocolo, obteniendo así una mayor visibilidad de los contenidos. Además, dos de ellos también son proveedores de servicios: RECERCAT y TDR.

En septiembre de 2006 se adaptó DSpace con un conector que permite que RECERCAT sea también recolector de metadatos. De esta manera, el repositorio se convirtió en híbrido al permitir que alguna de sus instituciones participara aportando una recolección de metadatos de su propio repositorio institucional.

Después de potenciar la visibilidad de las tesis doctorales que contiene TDR en los últimos años, desde finales de 2006 se ha trabajado para incrementar el valor del

[13] http://www.oaister.org/. [Consulta: 05/09/07].

repositorio. Actualmente se pueden consultar desde TDR no sólo las tesis de las universidades participantes, sino también todas aquellas tesis de otros repositorios españoles que las hacen accesibles con el protocolo OAI-PMH (actualmente: Dialnet y la Universidad Carlos III de Madrid). Esta búsqueda recibe el nombre, en la interfaz de TDR, de "Búsqueda global". Para poner en marcha esta nueva funcionalidad, se ha utilizado el módulo MetaIndex y las librerías X-Server del software MetaLib de ExLibris, que permite importar los metadatos de los diferentes repositorios que operan el protocolo OAI-PMH. De esta manera, las tesis continúan estando almacenadas en el repositorio original, pero TDR permite consultarlas a través del metadato. Con esta nueva funcionalidad, TDR se ha convertido en el primer repositorio del Estado que, además de dar acceso a sus propias tesis, permite también buscar entre las tesis doctorales de otros repositorios del Estado que ofrecen metadatos OAI-PMH.

5. Contenido

La Comisión Técnica del CBUC, uno de los órganos de gobierno de nuestro consorcio formado por los directores de las bibliotecas de las instituciones miembro, estableció a finales de 2005 como principal objetivo para los repositorios y para el periodo de 2006-2008, el de recaptar el máximo número de documentos e introducirlos en los repositorios colectivos. Desde el Consorcio se ha trabajado para facilitar esta tarea (recomendaciones, edición de materiales de difusión, etc.), pero el trabajo principal lo llevan a cabo desde la propias instituciones participantes en cada repositorio.

Si hacemos una mirada retrospectiva al pasado, vemos que la puesta en marcha del primer repositorio (TDR) no fue rápida en cuanto a contenido: más de 16 meses para alcanzar las primeras 500 tesis, mientras que actualmente sólo se tarda alrededor de 6 meses en incorporar 500 nuevas tesis. La proporción de tesis introducidas ha mejorado año tras año, no sólo a causa de los nuevos participantes, sino también porque las universidades consiguen reunir e incorporar más tesis en TDR, alguna de ellas gracias a que lo han convertido en un requisito para poder obtener el título de doctor. En este momento, por término medio, una de cada dos tesis leídas en las universidades catalanas es introducida en el repositorio.

En este apartado comentaremos los principales aspectos implicados antes, durante y después de la introducción de los documentos en nuestros repositorios colectivos. Cabe destacar que siempre que es posible los cuatro repositorios utilizan procedimientos comunes.

A continuación ofrecemos un breve estado de la cuestión en cuanto a materiales que incluyen (tipo y cantidad) e instituciones que participan actualmente.

Repositorio de Tesis Doctorales en Red[14]

Contiene el texto completo de las tesis doctorales leídas en las universidades de Cataluña y de otras comunidades autónomas españolas. Participan actualmente 17 universidades con un total de más de 4.700 tesis y ha recibido más de 3,5 millones de búsquedas en 2006.

[14] http://www.tesisenred.net/. [Consulta: 05/09/07].

Repositorio de la Investigación de Cataluña[15]

Contiene literatura de investigación de universidades e instituciones de investigación de Cataluña a texto completo. Participan actualmente 14 instituciones con un total de más de 3.600 documentos.

Revistas Catalanas con Acceso Abierto[16]

Contiene artículos a texto completo de revistas científicas, culturales y eruditas catalanas. Participan actualmente 30 instituciones con 139 revistas y 36.000 artículos.

Memoria Digital de Cataluña[17]

Este es el último repositorio que se ha puesto en marcha (noviembre de 2006) y por ello actualmente el menos consolidado aunque está teniendo una muy buena aceptación. Es un repositorio de fotografías, carteles, manuscritos y tiene como principal finalidad difundir y preservar la memoria de Cataluña. Participan actualmente 7 instituciones con aproximadamente unos 65.000 documentos.

Materiales Docentes en Red

Es un repositorio para materiales docentes. Se ha empezado a definir este mismo año y aun no se encuentra en producción. Su creación ha partido de un grupo de trabajo que ha definido sus características y procesos de trabajo. Actualmente este grupo ha presentado su propuesta a la comisión que toma las decisiones y se ha aprobado su creación.

5.1 Aspectos legales

Hoy en día, uno de los principales obstáculos para introducir documentos en los repositorios de acceso abierto es el miedo al plagio por parte de los autores. Desde que pusimos en marcha el primer repositorio, TDR, hemos trabajado este tema desde diferentes vertientes: ofreciendo asesoramiento legal, estableciendo procesos en caso de detección de plagio, recopilando y publicando en el web las respuestas a las preguntas más frecuentes[18], etc.

Para cada uno de nuestros repositorios hemos encargado, a un especialista en derecho de la información, un dictamen jurídico que tiene como objetivo estudiar la situación de los derechos de autor y ofrecer las herramientas que permitan a las instituciones participantes solicitar el permiso de los autores para publicar sus trabajos a través de dichos repositorios. Cuando una institución se integra en un repositorio firma un convenio de colaboración con las instituciones coordinadoras (CBUC y CESCA) y se le facilita dicha documentación jurídica.

En el caso de TDR, para cada una de las tesis se debe firmar un contrato de edición entre la universidad donde se ha leído y el autor, con el cual este autoriza a la universidad a publicar la tesis en el entorno de Internet y se reserva el derecho de publicarla en otros

[15] http://www.recercat.net/. [Consulta: 05/09/07].
[16] http://www.raco.cat/. [Consulta: 05/09/07].
[17] http://www.cbuc.cat/mdc. [Consulta: 05/09/07].
[18] Por ejemplo las de TDR: http://www.tesisenred.net/es/pmf.html. [Consulta: 05/09/07].

soportes y/o editoriales. Existe un modelo base de contrato que cada universidad ha personalizado.

En RECERCAT, todos los documentos incluidos están sujetos a la licencia *Creative Commons* (CC) de Reconocimiento-NoComercial-SinObraDerivada (actualmente, en su versión 2.5). Esta licencia establece que el trabajo puede ser copiado, distribuido y comunicado públicamente, siempre que los autores y sus instituciones sean citados y que no se haga ningún uso con finalidades comerciales ni obra derivada. Por obra derivada se entiende ese documento que ha sido editado, traducido, combinado con materiales de terceros, cambiado de formato o modificado de cualquiera otra forma. Si no es el propio autor el que introduce el documento (*self-archiving*) y por lo tanto el que acepta las condiciones de esta licencia CC se le debe solicitar consentimiento escrito para introducirlo en el repositorio con estas condiciones. Los programas informáticos de código libre se incluyen bajo la licencia *Creative Commons GNU GPL*.

Para RACO el procedimiento es similar, cuando se incorpora una nueva institución se le facilita la documentación jurídica con los diferentes modelos de documentos para solicitar permiso a los autores de los artículos. Así mismo se ofrece un modelo de cláusulas para incorporar en los contratos de edición o en las condiciones de publicación de las revistas.

Para los dos últimos repositorios establecidos, MDC y MDX, los aspectos legales están en fase de estudio y definición.

Desde que empezamos a trabajar con los repositorios, y hasta el momento, han sido detectados solo dos casos de usos indebidos de tesis incluidas en TDR. Cuando se conocieron estos usos, un especialista estudio el problema y se estableció el procedimiento para actuar. Fueron resueltos rápidamente y sin complicaciones.

5.2 Introducción de los documentos

Actualmente la mayoría de las universidades que forman parte de nuestro consorcio y que participan en los repositorios colectivos también disponen de repositorios locales o están en vías de implementarlos. La introducción de los documentos se realiza en ambos repositorios ya que cada uno tiene su finalidad principal y además es otra forma de contribuir a su preservación (al estilo del programa *LOCKSS – Lots of Copies Keep Stuff Safe*).

Los repositorios colectivos proporcionan más visibilidad y difusión de los documentos (más masa crítica, más usuarios, etc.) y los repositorios locales contribuyen a incrementar y consolidar la visibilidad institucional. Como ya se ha comentado en el apartado especifico de software, para facilitar la participación en ambos repositorios se han modificado los programas que gestionan los repositorios colectivos para que permitan tanto la introducción directa de metadatos y datos (los ficheros con el texto completo) como la introducción indirecta de metadatos.

La introducción directa de metadatos y datos la usan básicamente aquellas instituciones que no tienen repositorio local o bien que quieren tener más de una copia de los documentos para facilitar, como hemos comentado, la preservación. Esta introducción se lleva a cabo desde las propias instituciones participantes y son ellas mismas las que deciden como. Normalmente el servicio de bibliotecas tiene siempre un papel activo pero no tiene porque ser el responsable. Por ejemplo en el caso de TDR en cada universidad

funciona de manera diferente: coordinado por el servicio de publicaciones, por el de gestión de tercer ciclo o bien por la biblioteca. En el caso de RECERCAT normalmente la introducción se realiza desde la biblioteca o bien desde los propios departamentos pero es el servicio de bibliotecas el que acaba de validar el documento.

La introducción indirecta consiste en introducir los metadatos y datos en los repositorios locales y recopilar los metadatos a través del protocolo OAI-PMH para facilitar su consulta. La consulta de los datos (texto completo) se hace finalmente en el repositorio local.

Los documentos que se introducen en los repositorios, tanto en los colectivos como en los locales, se describen con metadatos Dublin Core cualificados y se ofrecen especificaciones de cómo dar la información de los diferentes elementos. Por ejemplo, los puntos de acceso de los autores, materias, etc. se establecen tomando como referencia el Catálogo Colectivo de las Universidades de Cataluña.

Anualmente se llevan a cabo diferentes actuaciones de calidad como por ejemplo: la revisión de los índices de autores para evitar las duplicidades, tesis que no tienen resumen, etc.

5.3 Retrodigitalización y preservación

Con el objetivo de experimentar con digitalización retrospectiva e incrementar los documentos contenidos en estos repositorios, en el año 2005 llevamos a cabo una prueba piloto de digitalización que gracias a su buen resultado nos ha permitido continuar trabajando en ello.

Disponemos de diversas empresas que trabajan con nuestros estándares[19] y criterios y que una vez realizada la digitalización introducen el fichero de consulta (PDF) directamente en los repositorios.

Para RACO en estos tres años (2005, 2006 y 2007) se han digitalizado 50 revistas que representan más de 195.000 páginas y para TDR se ha empezado por primera vez este año en el que se digitalizarán unas 400 tesis (160.000 páginas). Estas digitalizaciones han sido posible gracias a una subvención del gobierno catalán.

Uno de los objetivos de los repositorios institucionales es garantizar el acceso permanente a los documentos, es decir preservarlos a través del tiempo independientemente de su formato, tamaño, etc. Actualmente se llevan a cabo acciones básicas de preservación como hacer copias de seguridad, comprobación de la integridad de los documentos (*checksums*), URL permanentes, etc.

[19] "Estándares de digitalización. Elementos mínimos. Propuesta del CBUC (actualizado en julio de 2005)". http://docs.cbuc.cat/fitxers/5digital/Public0507Minims_Spa.pdf. [Consulta: 05/09/07].

Para ampliar estas acciones y mejorar y asegurar este aspecto hemos encargado un estudio[20] en el que se detectan las principales acciones que debemos llevar a cabo. Las principales conclusiones de este informe hacen referencia tanto a aspectos tecnológicos como organizativos. El objetivo para este año es empezar a aplicar las principales recomendaciones de este informe en los diversos repositorios.

6. Algunas reflexiones finales a modo de conclusión

En esos últimos años hemos visto como aquellas primeras tentativas, tímidas y dubitativas, de establecer 'almacenes' de documentos electrónicos (por ejemplo, de tesis) han evolucionado de forma conceptual y pragmática hasta establecerse como realidades consolidadas en el panorama bibliotecario mundial bajo el nombre de repositorios institucionales[21]. Al mismo tiempo, y partiendo del concepto del autoarchivo de los artículos de revistas científicas y de ponencias a congresos se ha consolidado el movimiento del Acceso Abierto a la información científica. Una de las principales (aunque no la única) declaraciones a favor del Acceso Abierto, la de Berlín, ha encontrado en España el apoyo de quince universidades.

El repositorio de documentos de física y matemáticas arXiv.org ha sido citado a menudo como una muestra de lo que podría conseguirse en el resto de disciplinas del saber humano, pero arXiv.org ha mostrado también que la existencia de un repositorio disciplinar potente no ha invalidado la existencia de revistas que continúan añadiendo valor a los *preprints* almacenados de forma conjunta y digital en un silo. Las dos vías que ha encontrado el Acceso Abierto para expresarse han sido los repositorios y las revistas[22] y, probablemente, las dos, más que vías alternativas (como algunos debates parecen dar a suponer), se mostraran como complementarias. Ni las revistas OA ni los repositorios parecen tener el camino totalmente despejado para su desarrollo futuro. Centrándonos en estos últimos, creemos que la comunidad internacional tienen tres importantes tareas pendientes: incrementar el autoarchivo en IR, integrar estos en los instrumentos bibliográficos ya existentes y encontrar un equilibrio entre el carácter institucional de los repositorios y su voluntad científica.

En nuestra ponencia de hace dos años ya indicábamos que en nuestro caso, lo que parecía un obstáculo difícil de superar (la tecnología asociada al establecimiento de un repositorio) había sido la parte fácil de la tarea mientras que la difícil había sido la de restablecer mecanismos e inercias de alimentación de los repositorios. Diversos artículos recientes [23] [24] [25] muestran como las dificultades que están experimentando las bibliotecas

[20] Gallart Marsillas, Núria; Keefer, Alice. "Preservació de dipòsits digitals". Barcelona: Consorci de Biblioteques Universitàries de Catalunya, 2007. <http://hdl.handle.net/2072/4078>. [Consulta: 31/07/2007].

[21] Ver la situación a nivel mundial en los artículos de Clifford A. Lynch y Joan K. Lippincott, "Institutional repository deployment in the United States as of early 2005". *D-Lib magazine*, vol. 11, no. 9 (2005), <http://www.dlib.org/dlib/september05/lynch/09lynch.html>, y de Gerard van Westrienen y Clifford A. Lynch, "Academic institutional repositories: deployment status in 13 nations as of Mid 2005". *D-Lib magazine*, vol. 11, no. 9 (2005), <http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>. [Consulta: 04/09/07]

[22] Guédon, Jean-Claude. "The "Green" and "Gold" Roads to Open Access: the case for mixing and matching". *Serials Review*, vol. 30, no. 4, (2004), p. 315-328.

[23] Davis, Philip M.; Connolly, Matthew J. L. "Evaluating the Reasons for Non-use of Cornell University's Installation of DSpace". *D-Lib Magazine*, vol. 13, no. 3-4 (2007). <http://www.dlib.org/dlib/march07/davis/03davis.html>. [Consulta: 31/07/07].

[24] Gierveld, Heleen. "Considering a Marketing and communications approach for an institutional repository". *Ariadne*, issue 49 (October 2006). <www.ariadne.ac.uk/issue49/gierveld>. [Consulta: 02/01/07].

universitarias de todo el mundo es la de establecer un entorno que ayude a la captación de contenidos para los repositorios institucionales. La solución más obvia sería que las universidades hicieran obligatorio el autoarchivo de los documentos de investigación en los repositorios institucionales, pero a pesar de las muchas adhesiones a declaraciones como la de Berlín, la verdad es que las políticas de obligación son excepción. Cierto es también que la tarea de instituciones relevantes en la financiación de la investigación se están posicionando claramente a favor del OA y recomiendan encarecidamente de introducir en repositorios de acceso abierto la investigación producida con fondos públicos [26].

Durante décadas, los investigadores se han acostumbrado a 'descubrir' los documentos que les podían ser de ayuda en su trabajo en dos instrumentos: bibliografías o bases de datos de resúmenes e índices y en los catálogos de las bibliotecas. Pero las posibilidades de descubrir materiales relevantes para una investigación se han incrementado enormemente en los últimos años: los buscadores en Internet (sean Google o Scopus), los 'paquetes' de revistas contratados conjuntamente de forma consorciada, las búsquedas federadas a múltiples instrumentos... En este nuevo contexto, ¿dónde situar a los IR? ¿tenemos que incorporar los metadatos de los documentos de los repositorios en los catálogos actuales o hacer búsquedas federadas entre estos y los repositorios? ¿la diferencia de granularidad, no conducirá en este caso a la desorientación del usuario? [27]

Finalmente, con los repositorios hemos podido constatar la contradicción que se produce entre los legítimos deseos de una universidad o institución de investigación de hacer aparecer los resultados de la investigación producida en su seno bajo su 'marca' y las igualmente legitimas necesidades de la comunidad investigadora de prescindir de búsquedas múltiples y de las fronteras institucionales para acceder a la globalidad de la producción de investigación clasificada en cualquier caso solo por criterios de calidad. Los repositorios nacidos como repositorios locales buscan, para ser eficaces, ser consultados conjuntamente, como parecen indicar las iniciativas italiana (PLEIADI[28]), holandesa (DAREnet[29]), australiana (ARROW[30]) o el proyecto europeo DEEP[31] que tiene como objetivo recoger los metadatos de tesis doctorales europeas.

En el contexto internacional y vistos estos últimos años a cierta distancia, el elemento diferencial de la iniciativa de las universidades de Cataluña respecto a repositorios ha sido la elección consciente de hacer el camino de forma conjunta y cooperativa (en la selección de software, en el establecimiento de un servidor en la fijación del conjunto de metadatos y de los procesos de captación, etc.) ¿Qué ventajas nos ha supuesto hacer conjuntamente lo que mayoritariamente se ha hecho de forma individual? A nuestro

---

[25] Henty, Margaret. "Ten major issues in providing a repository service in Australian universities". *D-Lib Magazine*, vol. 13, no. 5-6 (May-June 2007). <http://www.dlib.org/dlib/may07/henty/05henty.html>. [Consulta: 05/07/07].

[26] Ver, por ejemplo: "European Research Advisory Board. Final report. Scientific publication: policy on Open Access". EURAB, 2006. <http://ec.europa.eu/research/eurab/pdf/eurab_scipub_report_recomm_dec06_en.pdf>. [Consulta: 31/07/07].

[27] Dempsey, Lorcan. "The library catalogue in the new discovery environment: some thoughts". *Ariadne*, issue 48 (July 2006). <http://www.ariadne.ac.uk/issue48/dempsey>. [Consulta 14/07/07].

[28] http://www.openarchives.it/pleiadi/

[29] http://www.darenet.nl

[30] http://search.arrow.edu.au

[31] http://elib-a.ucl.ac.uk/

entender los beneficios específicos han sido de haber conseguido más resultados con más facilidad, más visibilidad y más valor añadido.

El primer beneficio para el sistema universitario catalán del trabajo colectivo en repositorios ha sido la facilidad en conseguir resultados, o dicho de otra forma, la maximización de resultados por el esfuerzo invertido. El trabajo conjunto es más lento que el individual en producir resultados pero es más efectivo. Los acuerdos tomados por más de una universidad se aplican con más facilidad que los específicos y la adopción de procedimientos comunes y permite a cada institución concentrarse en la captación de contenidos sin tener que distraer su atención en el desarrollo del repositorio.

En segundo lugar, los repositorios cooperativos proporcionan más visibilidad a los documentos que recogen. Esto es igualmente cierto para las búsquedas 'humanas' (¿quién elegirá repetir la misma búsqueda en diferentes lugares pudiendo hacer una de sola en un mismo lugar?) como para las realizadas a través de motores de búsqueda. Esta visibilidad mayor redunda en beneficio de los autores, de las instituciones específicas en el marco de las cuales realizan específicamente la investigación, y, finalmente, del sistema universitario de conjunto.

Finalmente, el tercer gran beneficio es que conjuntamente se pueden conseguir mejores resultados de valor añadido. En algunos casos esto se deriva sencillamente de la sinergia del trabajo conjunto. Por ejemplo, algunas soluciones a problemas de desarrollo de distintas aplicaciones (el de la indización de palabras con signos gráficos, por ejemplo) han sido encontradas 'a nivel local' y más tarde depuradas para incorporarse a los repositorios y para todas las instituciones. En otros casos es por economía de escala. Por ejemplo, la preservación a largo plazo de los documentos digitales de los repositorios es una tarea que reviste aún a nivel internacional numerosos interrogantes, pero parece claro que no va a conseguirse solo con el repositorio de un documento en un repositorio. Van a ser necesarias operaciones tediosas, almacenamientos redundantes y operaciones sofisticadas. Todas estas pueden sin duda hacerse a nivel local, pero con menor probabilidad de éxito que si son hechas a escala masiva.

7. Más bibliografía

Beagrie, Neil. “The continuing access and digital preservation strategy for the UK Joint Information Systemes Committee (JISC)”. *D-Lib Magazine*, vol. 10, no. 7/8 (July/August 2004). <http://www.dlib.org/dlib/july04/beagrie/07beagrie.html>. [Consulta: 31/07/07].

“Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities”. Presentado en: *Conference on Open Access to Knowledge in the Sciences and Humanities, Berlin, 20-22 october, 2003*. <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>. [Consulta: 31/07/07].

Borràs, Antoni [et al.]. “Programari de codi lliure per gestionar dipòsits digitals: el procés de tria dut a terme al CBUC”. *BiD: textos universitaris de biblioteconomia i documentació*, núm. 16 (juny 2006). <http://www2.ub.edu/bid/consulta_articulos.php?fichero=16borras.htm>. [Consulta: 31/07/07].

Chapman, Stephen. “Counting the costs of digital preservation: is repository storage affordable?”. *Journal of digital information*, vol. 4, no. 2, article no. 178. <http://jodi.tamu.edu/Articles/v04/i02/Chapman/>. [Consulta: 31/07/07].

Lynch, Clifford A. “Institutional repositories: essential infrastructure for scholarship in the digital age”. *ARL Bimonthly Report*, no. 226 (February 2003). <http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>. [Consulta: 31/07/07].

Reoyo, Sandra [et al.]. “RECERCAT: el dipòsit de la recerca de Cataluña». Presentado en: *10es Jornades Catalanes de Documentació: un espai de reunió, de diàleg, de participació*. Barcelona: Col·legi Oficial de Bibliotecaris-Documentalistes de Catalunya, 2006.
p. 331-347. También disponible en línea en: <http://hdl.handle.net/2072/2195>. [Consulta: 31/07/07].

Rosenthal, David S. H. [et al.]. “Requirements for digital preservation systems”. *D-Lib magazine*, vol.11, no.11 (November 2005). <http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>. [Consulta: 31/07/07].