

Los repositorios electrónicos cooperativos de la Biblioteca Digital de Cataluña

Lluís M. Anglada i de Ferrer, Director del CBUC, langlada@cbuc.es

Sandra Reoyo Tudó, Gestora de proyectos del CBUC, sreoyo@cbuc.es

Joan Cambras, Jefe del Departamento de Sistemas del CESCO, jcambras@cesca.es

Ricard de la Vega, Técnico líder del área de repositorios y portales del CESCO, rdelavega@cesca.es

La versión en castellano de este artículo ha sido realizada por Noelia Martínez Delgado (Consortio de Bibliotecas Universitarias de Cataluña).
--

Resumen

Entre 1998 y 1999, el Consorcio de Bibliotecas Universitarias de Cataluña (CBUC) creó una nueva línea de trabajo: la Biblioteca Digital de Cataluña. Ésta nació con la finalidad de contratar información electrónica interdisciplinar para la comunidad universitaria e investigadora de las bibliotecas miembro del Consorcio. Casi al mismo tiempo, esta línea de trabajo se amplió con la vertiente de la información propia generada por esta comunidad y se empezó a trabajar con la definición e implementación de repositorios electrónicos cooperativos. Estos últimos serán el objeto de este estudio. Se analiza su evolución, su estado actual, la estrategia para conseguir la inclusión de documentos (políticas institucionales, comités científicos, etc.), su contenedor (programas, tecnología usada, protocolos, etc.) y su contenido (estándares utilizados, derechos de autor, preservación, etc.). Por último se reflexiona sobre las ventajas, extraídas de la propia experiencia, de estos repositorios electrónicos cooperativos.

1. La Biblioteca Digital de Cataluña: de las contrataciones a los repositorios electrónicos

El Consorcio de Bibliotecas Universitarias de Cataluña (CBUC) se creó con el objetivo de constituir un catálogo colectivo que fuera un instrumento de información bibliográfica y de catalogación por copia, pero rápidamente amplió sus actividades. Éstas se agrupan bajo tres líneas estratégicas: la primera está destinada a mejorar el acceso a las colecciones impresas, la segunda se desarrolla alrededor de la información digital y la tercera está orientada a la innovación.

Entre 1998 y 1999 nació la Biblioteca Digital de Cataluña (BDC), la cual hoy en día incluye tanto información electrónica contratada conjuntamente por los miembros del CBUC como también los repositorios electrónicos con información propia. La información electrónica contratada actualmente a editores comerciales incluye unas 10.600 revistas electrónicas, unos 8.000 libros electrónicos y 37 bases de datos. La importancia de las contrataciones de recursos electrónicos realizadas bajo el nombre de BDC no nos debe hacer olvidar que casi a la vez se pusieron los cimientos del primer repositorio electrónico cooperativo de los cinco que mantienen hoy las universidades de Cataluña y la Biblioteca de Cataluña a través del CBUC y del Centro de Supercomputación de Cataluña (CESCA).

La construcción de un repositorio para documentos electrónicos comenzó bastante antes que éstos recibieran el nombre de repositorios institucionales (con el que han sido reconocidos en la literatura profesional posteriormente¹) y que se utilizara el término de acceso abierto (OA, *Open Access*). La situación actual es de cinco repositorios activos que utilizan cuatro programas diferentes. ¿Era ésta la situación que podíamos esperar cuando empezamos a hacer el primer repositorio hace 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. Éstas fueron: una iniciativa de administración electrónica, un sistema de compra de libros de las editoriales universitarias a través de la red, la creación de material docente digital y la que hicieron el CBUC y el CESCA de crear un repositorio electrónico de tesis doctorales a texto completo.

Para implementar este repositorio se tomó la opción estratégica de no hacerlo desarrollando un programa informático propio para su gestión sino hacerlo usando uno de ya existente. Aquí influyó la experiencia de las bibliotecas de las universidades de Cataluña de los años 80 de crear y usar software de gestión de bibliotecas². 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)³, que proporcionaba entonces un programa de código libre: el ETD. Fue necesario adaptar el programa al objetivo propio de hacer un repositorio cooperativo, pero se obtuvieron aprendizajes positivos del hecho de utilizar herramientas construidas por otros para propósitos similares a los nuestros.

El principal aprendizaje de aquella experiencia inicial –TDR- fue que la mayor 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, sobretodo, convencer a la comunidad universitaria que poner documentos (en este caso las tesis) en acceso abierto no sólo es fácil sino que tiene además muchas ventajas. Más adelante en este artículo detallaremos los mecanismos creados para el repositorio de tesis, y para los otros, para facilitar la introducción de documentos y disminuir la resistencia de editar electrónicamente y en abierto.

Cuando se puso en funcionamiento el repositorio de tesis se contempló también la posibilidad de crear otro para revistas. Por cuestiones de oportunidad la idea quedó aplazada. No obstante, desde 1998 el CBUC tenía una base de datos de sumarios electrónicos que, con modificaciones leves, permitió añadir a los sumarios de revistas el texto completo de los 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 (Revistas Catalanas con acceso abierto) pero a pesar del acierto en el objetivo y de los esfuerzos

¹ 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: 26/02/08].

² 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: 13/02/08].

³ Dijkstra, J. M. M. "Towards Storage and Access of Electronic Theses [manuscrit no publicat]: A Proposal of Its Organisation for the Consortium of Catalan Libraries". Barcelona, 1999.

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 llevó a cabo un seguimiento de la literatura profesional referente a programas de automatización de bibliotecas ya que se hacía necesaria la sustitución del que entonces estaba en uso (VTLS). Esto permitió observar que había nuevas realidades emergentes en el uso de servicios bibliotecarios: los repositorios institucionales (IR, *Institutional Repositories*) y que estaban surgiendo nuevos software (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 comenzó a gestar un cambio en el que había sido un sólido paradigma a lo largo de los últimos 20 años: los sistemas automatizados de gestión de bibliotecas que habían encontrado su éxito en la integración de funciones comenzaron, 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 (todavía 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 a la “catalogación” de recursos bibliotecarios. En seguida se llegó a la conclusión que sin un repositorio en funcionamiento real nunca se llegaría a poder concretar un conjunto de metadatos que facilitaran 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 estos fueran un servicio en sí mismos, 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 una manera de mejorar la preservación de los documentos digitales entonces ya frecuentes en las universidades. Se acordó seleccionar conjuntamente un mismo programa con el objetivo 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 de audio y vídeo, etc. No se tenía claro si el repositorio sería único o no, pero se creía firmemente que un único programa informático permitiría gestionar la variedad de objetos digitales mencionados.

Se creó un grupo de trabajo del Consorcio con la finalidad de seleccionar un software. La tarea se vio facilitada por una guía de la Soros Foundation⁴ de manera que este grupo concluyó su tarea en poco tiempo y seleccionó como programa más adecuado para nuestras necesidades y objetivos, DSpace. En 2004 comenzaron 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 artículos, *working papers*, informes de investigación, etc. pero no tanto a las revistas⁵ o a otros objetos digitales. El repositorio se concretó en un repositorio de literatura gris de investigación (excepto tesis)

⁴ Crow, Raym. “A Guide to Institutional Repository Software”. 3rd. ed. New York: Open Society Institute, August 2004. <<http://www.soros.org/openaccess/software/>>. [Consulta: 26/02/08].

⁵ No obstante, hay diversas experiencias de revistas electrónicas sostenidas por DSpace, por ejemplo el repositorio “Revistes i congressos UPC” (<https://upcommons.upc.edu/revistes/?locale=es>) de la Universidad Politècnica de Catalunya, universidad miembro del CBUC.

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 escogió el programa OJS para el desarrollo del repositorio de revistas en OA y, un poco más tarde, el programa comercial CONTENTdm de OCLC para la gestión del repositorio Memoria Digital de Cataluña (para imágenes y reproducciones de objetos). Más recientemente, en 2007, se decidió iniciar un repositorio para material docente (MDX) y utilizar también DSpace.

Después de lo que hemos explicado 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 (a pesar de tener que adaptarlos) a los perfectos pero todavía 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 todavía durante algún tiempo la utilización de programas diferentes. Se fijó entonces una política de no cuestionar (al menos de momento) los contenedores ya seleccionados para centrarse en llenarlos de contenido.

Actualmente en el seno del CBUC se tiene la visión de que algunos desarrollos notables en el ámbito de los repositorios están por llegar y que el impacto de las búsquedas federadas puede ser determinante para decidir hasta qué punto es mejor el uso de un único programa de propósito general para todo tipo de objetos o el de programas específicos para cada tipo de documento.

2. Los repositorios electrónicos cooperativos

Actualmente la BDC incluye cinco repositorios electrónicos cooperativos: TDR (tesis), RACO (revistas), RECERCAT (documentos de investigación), MDC (imágenes) y MDX (material docente) que entrará en funcionamiento este año. Los objetivos comunes de estos repositorios son:

1. Facilitar la introducción de documentos en la red contribuyendo así al movimiento mundial del acceso abierto
2. Proporcionar un instrumento de visualización y búsqueda conjunta de documentos que fomente el uso de los mismos
3. Crear servicios de valor añadido (como por ejemplo: mecanismos de preservación, estadísticas, URL permanentes, etc.)

Los repositorios de la BDC tienen diversas características comunes: son de acceso abierto; cumplen con el protocolo de interoperabilidad de la Open Archives Initiative; usan metadatos Dublin Core; están contruidos y alimentados de manera cooperativa; y cuatro de los cinco repositorios están desarrollados con software libre. 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 escogidos para cada repositorio, las principales modificaciones que se han realizado y la

interoperabilidad de éstos con otros repositorios nacionales e internacionales a través del protocolo OAI-PMH.

2.1 Software

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 duda, 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. Aparte de la adaptabilidad y de los costes, los repositorios también se han beneficiado de otras ventajas del software libre, como el escrutinio público, ya que 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 apoyan los programas en los que se basan los repositorios (comunidades DSpace, OJS, ETD-main, etc.) son importantes (por la cantidad de instituciones implicadas, por el volumen de usuarios, etc.), la cual cosa es ya en sí misma una garantía de continuidad.

En TDR, a nivel de software, se pueden diferenciar tres capas, la de la gestión de contenidos de la colección de tesis, la de la indización y publicación de los datos y la de la recolección de tesis de otros repositorios españoles que usan el protocolo OAI-PMH. Para la primera se adaptó el Electronic Theses and Dissertations (ETD)⁶ de la Virginia Tech University (VT) como un proyecto de colaboración entre la Graduate School de la VT, la Digital Library and Archives (DLA) y la Networked Digital Library of Theses and Dissertations (NDLTD), de la cual TDR forma parte. Para la segunda capa se usan Glimse y WebGlimse y para la tercera se ha utilizado el módulo MetaIndex y las librerías X-Server del software MetaLib de ExLibris.

RECERCAT está implementado a partir del programa de código abierto DSpace⁷ (actualmente, en su versión 1.4). Este software fue creado por el Massachusetts 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 *plugins* que se ofrecen para DSpace no cumplían con los requisitos que se deseaban para el proyecto. MDX, que se pondrá en funcionamiento este año 2008, también se implementará con DSpace. Se utilizará la última versión estable (1.4.2) y se añadirán las librerías Manakin. Éstas permiten una mayor personalización de las comunidades y colecciones. En función de los resultados obtenidos se prevé usar también Manakin en RECERCAT.

RACO se basa en el programa de código abierto Open Journal Systems (OJS)⁸, 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 módulos nuevos: uno para implementar una clasificación temática de revistas y otro para simplificar el proceso de publicación.

⁶ <http://etd.vt.edu>. [Consulta: 26/02/08].

⁷ <http://www.dspace.org/>. [Consulta: 26/02/08].

⁸ <http://pkp.sfu.ca/?q=ojs>. [Consulta: 26/02/08].

La condición de software libre nos ha permitido también, adaptar cada repositorio a diferentes 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 coste de desarrollo e implementación. Para adaptar el software seleccionado a las necesidades establecidas de cada repositorio es necesario invertir tiempo y recursos tanto materiales como de personal.

Anteriormente hemos comentado que se han utilizado cuatro software diferentes para los cinco repositorios activos en la actualidad. Tres son software libre, que son los que se han comentado hasta ahora y el cuarto es un software comercial. Éste es CONTENTdm⁹ de OCLC para el repositorio de imágenes MDC. Este software se ha contratado con licencia para objetos ilimitados.

2.2 Open Archives Initiative

Los repositorios utilizan un protocolo de interoperabilidad creado por la Open Archives Initiative (OAI)¹⁰ que aumenta la visibilidad de la información electrónica compartiendo el metadato del repositorio con otros repositorios internacionales, como OAIster¹¹. 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; y 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 más de 3 millones y medio durante 2006 y 2007.

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 estado trabajando para incrementar el valor del 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 se recolectan: las de Dialnet, las de la Universidad Carlos III, las de la Complutense de Madrid y las de Alicante. En total, un fondo de casi 10.000 tesis doctorales a texto completo. Esta búsqueda recibe el nombre, en la interfaz de TDR, de “Búsqueda global”¹².

⁹ <http://www.contentdm.com/>. [Consulta: 26/02/08].

¹⁰ <http://www.openarchives.org/>. [Consulta: 26/02/08].

¹¹ <http://www.oaister.org/>. [Consulta: 26/02/08].

¹² <http://global.tesisenred.net/>. [Consulta: 28/02/08].

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 español que, además de dar acceso a sus propias tesis, permite también buscar entre las tesis doctorales de otros repositorios españoles que ofrecen metadatos OAI-PMH.

2.3 Contenido

La Comisión Técnica del CBUC, formada por los directores de las bibliotecas de las instituciones miembro del Consorcio, establecieron a finales de 2005 como principal objetivo para los repositorios y para el período 2006-2008, el de recolectar el máximo número de documentos e introducirlos en los repositorios cooperativos. Desde el Consorcio se ha trabajado para facilitar esta tarea (coordinación, recomendaciones, edición de materiales de difusión, etc.), pero el trabajo principal lo llevan a cabo las instituciones participantes en cada repositorio.

Si miramos hacia el pasado, vemos que la puesta en marcha del primer repositorio (TDR) no fue rápida en cuanto a contenido: se tardó más de 16 meses en llegar a las primeras 500 tesis, mientras que actualmente cada año se incorporan unas 1.000. 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, de 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 los repositorios de la BDC. Cabe destacar que siempre que es posible los repositorios utilizan procedimientos comunes. A continuación ofrecemos un breve estado de los repositorios en cuanto a materiales que incluyen (tipo y cantidad) e instituciones que participan actualmente.

Tesis Doctorales en Red¹³



Contiene el texto completo de las tesis doctorales leídas en las universidades de Cataluña y otras comunidades autónomas españolas. Participan actualmente 20 universidades con más de 5.300 tesis y ha recibido más de 3,5 millones de consultas en 2007.

Repositorio de la Investigación de Cataluña¹⁴



Contiene literatura de investigación a texto completo de universidades e instituciones de investigación de Cataluña. Participan actualmente 17 instituciones con más de 4.300 documentos y ha recibido casi 400.000 consultas en 2007.

¹³ <http://www.tesisenred.net/>. [Consulta: 26/02/08].

¹⁴ <http://www.recercat.net/>. [Consulta: 26/02/08].



Revistas Catalanas con Acceso Abierto¹⁵

Contiene artículos a texto completo de revistas científicas, culturales y eruditas catalanas. Participan actualmente 39 instituciones con 177 revistas y 38.000 artículos.



Memoria Digital de Cataluña¹⁶

Es el último repositorio que se ha puesto en marcha (noviembre de 2006) y por este motivo, actualmente el menos consolidado aunque está teniendo una muy buena aceptación. Es un repositorio de imágenes (fotografías, carteles, manuscritos, etc.) que tiene como principal finalidad difundir y preservar las colecciones especiales de las bibliotecas de Cataluña. Participan actualmente 7 instituciones con aproximadamente unos 100.000 documentos.



Material Docente en Red

Es un repositorio para materiales docentes. Se ha empezado a definir en 2007 y se pondrá en funcionamiento este año. Su creación se ha realizado a partir de la labor de un grupo de trabajo que ha definido las características y procesos de trabajo del repositorio.

Está previsto que durante este año se incluyan estos cinco repositorios cooperativos en el MetaBUscador¹⁷ (usa el programa comercial MetaLib de ExLibris) de la BDC y también en los metabuscadores locales. De esta manera será posible consultar conjuntamente todos los recursos (tanto información contratada como información propia) de la Biblioteca Digital de Cataluña con una única búsqueda.

2.4 Aspectos legales

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 la web las respuestas a las preguntas más frecuentes¹⁸, etc. Para cada repositorio de la BDC se ha encargado 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 los repositorios. Cuando una institución se integra en un repositorio firma un convenio de colaboración con las instituciones coordinadoras (CBUC y CESA) y se le facilita la documentación jurídica.

En el caso de TDR, para cada tesis se debe firmar un contrato de edición entre la universidad donde se ha leído y el autor, con el cual éste autoriza a la universidad a publicar la tesis en el entorno de Internet y se reserva el derecho de publicarla en otros soportes y/o editoriales. Existe un modelo base de contrato que cada universidad ha personalizado.

¹⁵ <http://www.raco.cat/>. [Consulta: 26/02/08].

¹⁶ <http://www.cbuc.cat/mdc>. [Consulta: 26/02/08].

¹⁷ <http://www.cbuc.cat/metilib>. [Consulta: 28/02/08].

¹⁸ Por ejemplo las de TDR: <http://www.tesisenred.net/es/pmf.html> [Consulta: 28/02/08].

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 el 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 sólo dos casos de usos indebidos de tesis incluidas en TDR. Cuando se conocieron estos usos, un especialista estudió el problema y se estableció el procedimiento para actuar. Fueron resueltos rápidamente y sin complicaciones.

2.5 Introducción de los documentos

Actualmente la mayoría de las universidades que forman parte del Consorcio y que participan en los repositorios cooperativos 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 manera de contribuir a su preservación (al estilo del programa *LOCKSS – Lots of Copies Keep Stuff Safe*).

Los repositorios cooperativos, en tanto que 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 específico de software, para facilitar la participación en ambos repositorios se han modificado los programas que gestionan los repositorios colectivos de manera 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, que es la forma mayoritaria, se realiza en los repositorios cooperativos desde las propias instituciones participantes y son ellas mismas las que deciden cómo. Normalmente el servicio de bibliotecas tiene siempre un papel activo pero no tiene porqué ser el responsable. Por ejemplo, en el caso de TDR 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 quien 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 a través del repositorio colectivo. La consulta de los datos (texto completo) se realiza 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 Cataluña¹⁹. Anualmente se llevan a cabo diferentes actuaciones de calidad como por ejemplo: la revisión de los índices de autores para evitar duplicados, tesis que no tienen resumen, etc.

Con el objetivo de experimentar con la digitalización retrospectiva e incrementar los documentos introducidos en estos repositorios, en 2005 se llevó a cabo una prueba piloto de digitalización que gracias a su buen resultado nos ha permitido continuar trabajando en ello. La tarea se externaliza a empresas que trabajan con los estándares²⁰ y criterios consorciados 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 un total de 52 revistas que representan más de 200.000 páginas y para TDR se ha comenzado a hacer por primera vez en 2007 y se han digitalizado 400 tesis que representan más de 150.000 páginas. Este año 2008 también se han creado dos nuevas convocatorias para facilitar la digitalización de documentos a incluir en estos dos repositorios. Estos programas de retrodigitalización han sido posibles gracias a ayudas de la Generalidad de Cataluña²¹.

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. Para ampliar estas acciones y mejorar y asegurar este aspecto, el CBUC encargó un estudio²² en el que se detectan las principales acciones que se deben llevar a cabo. Les principales conclusiones de este informe se refieren tanto a aspectos tecnológicos como organizativos.

¹⁹ <http://ccuc.cbuc.es/vtls24/vtls/spanish/>. [Consulta: 28/02/08].

²⁰ "Estándares de digitalización. Elementos mínimos. Propuesta del CBUC (actualizado en julio de 2005)". http://docs.cbuc.cat/fitxers/5digital/Public0507Minims_Spa.pdf. [Consulta: 28/02/08].

²¹ De la Secretaría de Telecomunicaciones y de la Sociedad de la Información hasta 2007 y para 2008 de la Dirección General de Cooperación Cultural del Departamento de Cultura y Medios de Comunicación.

²² Gallart Marsillas, Núria; Keefer, Alice. "Preservació de dipòsits digitals". Barcelona: Consorci de Biblioteques Universitàries de Catalunya, 2007. <<http://hdl.handle.net/2072/4078>>. [Consulta: 28/02/2008].

3. Reflexiones finales

En estos últimos años hemos visto cómo 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 institucionales²³. A la vez, 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.org²⁴ ha sido citado a menudo como una muestra de lo que se podría conseguir 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 revistas²⁵ y, probablemente, las dos, más que vías alternativas (como algunos debates parecen dar a entender), se mostrarán como complementarias. Ni las revistas OA ni los repositorios parecen tener el camino totalmente libre para su desarrollo futuro. Centrándonos en estos últimos, creemos que la comunidad internacional tiene tres importantes tareas pendientes: incrementar el autoarchivo en los repositorios, integrar éstos en los instrumentos bibliográficos ya existentes y encontrar un equilibrio entre el carácter institucional de los repositorios y su voluntad científica.

Decíamos al inicio de este artículo 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 establecer mecanismos e inercias de alimentación de los repositorios. Diversos artículos recientes^{26 27 28 29} muestran cómo la dificultad que están experimentando las bibliotecas 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

²³ 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]

²⁴ <http://arxiv.org/>. [Consulta: 28-02-08].

²⁵ 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. Disponible la versión en catalán en: <http://hdl.handle.net/2072/3726>.

²⁶ 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].

²⁷ 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].

²⁸ 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].

²⁹ Miguel Ferreira, Ana Alice Baptista, Eloy Rodrigues y Ricardo Saraiva. “Carrots and sticks: some ideas on how to create a successful institutional repository”. *D-Lib Magazine*, vo. 14, no. ½ (2008). <http://www.dlib.org/dlib/january08/ferreira/01ferreira.html>. [consulta: 22.01.08].

Berlín, la verdad es que las políticas de obligación son una excepción. Ciertamente es también que instituciones relevantes en la financiación de la investigación se están posicionando claramente a favor del OA y recomiendan encarecidamente introducir en repositorios de acceso abierto la investigación producida con fondos públicos³⁰. El movimiento parece que está adquiriendo cierta fuerza en Europa en los últimos meses³¹.

Durante décadas, los investigadores se acostumbraron a “descubrir” los documentos que les podían ser de ayuda en su trabajo en dos instrumentos: en 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 en múltiples instrumentos... En este nuevo contexto, ¿dónde situar los IR? ¿debemos 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?³²

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 bajo su “marca” y las igualmente legítimas 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 sólo por criterios de calidad. Los repositorios nacidos como repositorios locales buscan, para ser eficaces, ser consultados conjuntamente, como parecen indicar las iniciativas italiana (PLEIADI³³), holandesa (DAREnet³⁴), australiana (ARROW³⁵) o el proyecto europeo DART-Europe³⁶ que tiene como objetivo recoger los metadatos de tesis doctorales europeas.

En el contexto internacional y vistos estos últimos años con cierta distancia, el elemento diferencial de la iniciativa de las bibliotecas del Consorcio respecto a los 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 comportado hacer conjuntamente lo que mayoritariamente se ha hecho de forma individual? A nuestro entender los beneficios específicos han sido los 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 para el esfuerzo invertido. El trabajo conjunto es más lento

³⁰ 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_sci pub_report_recomm_dec06_en.pdf>. [Consulta: 28/02/08].

<http://ec.europa.eu/research/eurab/pdf/eurab_sci pub_report_recomm_dec06_en.pdf>. [Consulta: 31/07/07].

³¹ “L’Open Access a Europa”. En el blog: Bdig (biblioteques digitals i cooperació), viernes 15.02.08. <<http://bdig.blogspot.com/2008/02/lopen-access-europa.html>>. [Consulta: 26/02/08]

³² 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].

³³ <http://www.openarchives.it/pleiadi/>

³⁴ <http://www.darenet.nl>

³⁵ <http://search.arrow.edu.au>

³⁶ <http://elib-a.ucl.ac.uk/>

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 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 escogerá repetir la misma búsqueda en diferentes lugares pudiendo hacer una sola en un mismo sitio?) como para las realizadas a través de motores de búsqueda. Esta mayor visibilidad repercute 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) se han encontrado “a nivel local” y más tarde se han depurado 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 tiene todavía a nivel internacional diversos interrogantes, pero parece claro que no se conseguirá sólo con el depósito de un documento en un repositorio. Serán necesarias operaciones tediosas, almacenamientos redundantes y operaciones sofisticadas. Todas éstas pueden sin duda hacerse a nivel local, pero con menor probabilidad de éxito que si se hacen a escala masiva.