

ETD 2012 : Simposio Internacional de tesis y disertaciones electrónicas = ETD 2012 : International Symposium Electronic Theses and Dissertations (15º : Lima, Perú : 12-14 sep 2012).

Título

Repositorios sostenibles, reflexiones a partir de la experiencia española

Autores

Ciro Llueca
Universidad de Barcelona, España
cllueca@ub.edu

Sandra Reoyo
Consorcio de Bibliotecas Universitarias de Cataluña, España
sroyo@cbuc.cat

Profesores de *Repositorios digitales* en el Máster Online en Documentación Digital, de la Universidad Pompeu Fabra, España.

Resumen

La creación y posterior gestión de un repositorio institucional no tiene sentido si no cuenta con un número significativo de documentos y un crecimiento permanente de la colección, a un coste razonable.

A partir del estudio de repositorios institucionales y de colecciones patrimoniales de bibliotecas, archivos y museos de España, los autores comparten sus reflexiones bajo la premisa de producir repositorios sostenibles, promoviendo la autosuficiencia en el incremento de sus fondos, la garantía de la financiación permanente por parte de la institución que los integra y, especialmente, el uso de los documentos depositados por parte de la comunidad a la que sirve la institución.

Tras un breve repaso a los procesos de adaptación de la filosofía del acceso abierto a los repositorios existentes, se establece una hoja de ruta para el diseño e implementación de un nuevo repositorio, teniendo en cuenta la cobertura estratégica y legal del proyecto, las opciones de hardware y software más populares, así como la planificación de los procesos de trabajo y la adopción de metadatos de descripción e interoperabilidad. Se presentan estrategias de difusión y evaluación de los repositorios. Finalmente, se aportan recomendaciones básicas de preservación digital, a la espera de una solución global.

Texto Completo

1. Introducción

Es evidente que el uso de Internet significa una oportunidad de desarrollo para las organizaciones y que la sociedad, las empresas e instituciones están cambiando y adaptando sus funciones a la nueva cultura de la información y de la comunicación. Los repositorios son producto de esta nueva cultura en el sector de bibliotecas, archivos y museos, y no es aventurado afirmar que los repositorios son ya un complemento imprescindible de los catálogos.

Como señaló la UNESCO en sus *Directrices para la preservación del patrimonio digital* (2003), los repositorios son también la respuesta local al reto global. La amenaza es clara: lo que no exista en las redes, no existirá como materia de información o de investigación.

Pero la creación de un repositorio no tiene sentido si no cuenta con un número significativo de documentos y un crecimiento permanente de la colección, a un coste razonable.

A partir del estudio de repositorios institucionales y de colecciones patrimoniales de bibliotecas, archivos y museos de España, los autores comparten sus reflexiones bajo la premisa de producir repositorios sostenibles, promoviendo la autosuficiencia en el incremento de sus fondos, la garantía de la financiación permanente por la institución que los integra y el uso por parte de la comunidad a la que sirve la institución.

1.1. Open Access

En las últimas décadas del siglo XX hubo tres fenómenos que dieron lugar al movimiento *Open Access*. Por una parte la necesidad de diseminar la literatura gris, aquella investigación que solo se difundía a través de los contactos del propio autor y que no dejaba de estar oculta; por otra, la llamada “*serial crisis*”, el aumento constante del coste de suscripción a las revistas por parte de las bibliotecas; y por último los avances en la edición digital y en la comunicación a través de la red. Estos tres fenómenos dieron lugar a un nuevo modelo de comunicación científica: el acceso abierto u *Open Access*, que se define como el movimiento mundial puesto en marcha por las instituciones que financian la investigación, con la finalidad de depositar en la red y de forma gratuita la producción académica y de investigación de todas las disciplinas. La idea es que los resultados de la investigación financiada con dinero público deben ser de acceso público.

Desde que surgió este movimiento han sido muchas las declaraciones que se han firmado para apoyarlo y las políticas institucionales, nacionales e internacionales elaboradas para seguirlo. Son conocidas las 3 B (Budapest, Bethesda y Berlín) y en el ámbito europeo han sido especialmente relevantes las recomendaciones realizadas por la Comisión Europea -- que han dado lugar a la creación del proyecto *OpenAIRE* (<http://www.openaire.eu/>)-- en la que se destaca la importancia de establecer políticas que favorezcan en los países miembros el acceso abierto a su producción científica. Complementariamente, los países del

sur de Europa publicaron una declaración específica a favor del acceso abierto, la *Declaración de La Alhambra* (2010).

Como ha señalado Dominique Babini (2011), en Latinoamérica el avance del movimiento internacional de acceso abierto, con mandatos que exigen en autoarchivo de la producción de los investigadores en repositorios, anticipa que en forma creciente gran parte de la producción científica publicada por América Latina y el Caribe en revistas internacionales estará también en acceso abierto. Los resultados actuales muestran liderazgo de Brasil y una fuerte presencia de universidades de México, Colombia, Argentina, Chile y Venezuela.

1.2. Producción científica invisible

En entornos no anglosajones, por otra parte, el uso de repositorios (y las políticas de uso y promoción de los mismos) no son únicamente recomendables, son imprescindibles para la supervivencia: el escenario de la producción científica latinoamericana refleja que gran parte de la producción es desconocida, y su área de influencia local o limitada. De hecho, las revistas iberoamericanas difícilmente ingresan en las bases de datos de mayor prestigio. En los estudios de Eduardo Aguado et al. (2006, 2009) se analizan algunas cifras determinantes: en 2004 64 países concentraron el 98% de los artículos altamente citados, mientras 162 países solo contribuyeron al 2% del total. Por otra parte, de los 1.222 científicos más citados, el 66% trabajan en instituciones de USA; el 20,5% en el Reino Unido; y el 5,1% en Alemania.

Por contra, como se menciona en el portal de divulgación del acceso abierto de la *Universidade de São Paulo* (<http://www.acessoaberto.usp.br/>), los artículos científicos publicados en acceso abierto (en revistas, o en repositorios) reciben, de media, un 8% más de citas que los de acceso restringido. Y en países en desarrollo, como Brasil, Argentina, o Rusia, ese porcentaje llega al 25% de citas adicionales.

En cifras totales, y según el directorio mundial de repositorios académicos de acceso abierto, *OpenDOAR*, en agosto de 2012 existen más de 2.190 depósitos digitales, de los cuales únicamente 179 (el 8,1%) pertenecen a instituciones de América Latina o el Caribe.

1.3. La experiencia española, brevemente

En España, desde la creación en 2001 del repositorio en acceso abierto para tesis doctorales TDR (*Tesis Doctorales en Red*, <http://www.tesisenred.net/>) se han creado 94 repositorios institucionales, un 4,3% de la producción mundial.

Además de la adaptación de los mandatos internacionales o europeos ya mencionados, cabe destacar la *Ley 14/2011 de la Ciencia, la Tecnología y la Innovación*, que en cierta manera obliga a los investigadores a publicar en acceso abierto la actividad financiada mayoritariamente con fondos de los presupuestos estatales; así como el *Real Decreto 99/2011*, que obliga a las universidades a publicar en abierto sus tesis doctorales.

En España, como en el resto de la comunidad mundial de repositorios institucionales de universidades y centros de investigación, las tendencias apuntan a:

- Asegurar el compromiso institucional por medio de la publicación de mandatos u otros textos institucionales.
- Unificar y normalizar procesos de depósito de documentos a partir de los currículums normalizados del personal docente e investigador, incluyendo para ello procesos de importación y exportación del repositorio al catálogo, y viceversa.
- Analizar la integración de sistemas de *Linked Open Data* para abrir y enlazar datos que puedan ser utilizados en la web semántica.
- Ampliar las colecciones con contenido heterogéneo: científico, docente, administrativo y patrimonial; procedente del autoarchivo, del depósito promovido por unidades bibliotecarias de gestión, o de la digitalización retrospectiva.
- Impulsar el uso y reutilización de los documentos depositados, por medio de la promoción de los servicios de repositorio mediante Internet, redes sociales, aplicaciones para móviles, etc.
- Avanzar tímidamente en servicios y procesos que garanticen la preservación digital de los documentos depositados, especialmente de ámbito cooperativo.

2. Repositorios sostenibles, reflexiones a partir de la experiencia española

Sin extendernos en definiciones que tienen su precedente genérico en Clifford Lynch (2003), entenderemos un repositorio "como un conjunto de servicios prestados por una institución a su comunidad para recopilar, gestionar, difundir y preservar su producción digital a través de una colección organizada, de acceso abierto e interoperable" (Barrueco et al., 2010). Dicho de otra manera, un repositorio sirve para publicar en Internet, normalmente con acceso libre, una base de datos que incluye los documentos digitales o digitalizados que forman los fondos y colecciones de una unidad de información.

Los repositorios representan el espacio natural de crecimiento de los servicios de información: los procesos de catalogación dieron lugar a los catálogos automatizados, éstos a los catálogos en línea, y éstos han sido complementados por los repositorios.

Por repasar sintéticamente los beneficios narrados por Alma Swan (2011) y muchos otros autores, un repositorio que albergue la producción de su institución consigue más visibilidad, difusión y prestigio de sus autores; de su producción documental y científica; y de sus centros de investigación y divulgación. La visibilidad clásica de la producción investigadora se abre al entorno internacional, con lo que se garantiza más audiencia y mejor impacto científico. Además, contribuye al movimiento mundial de acceso abierto gracias a que deposita y difunde su producción académica de forma gratuita. Se posibilita una mejor preservación de sus fondos, puesto que se promueve la digitalización retrospectiva de sus colecciones y por ello el acceso a la documentación original puede disminuir sustancialmente. El trabajo en escala permite reducir costes, y se perfeccionan los sistemas de evaluación de impacto, pues se facilita la obtención de datos reales de uso.

2.1. Pensar antes que actuar: cobertura estratégica y legal

Queda claro que el objetivo de una política institucional de acceso abierto se orientará a facilitar, promover e incentivar que los investigadores depositen sus publicaciones en repositorios y/o se publiquen en revistas de acceso abierto.

Establecer estas políticas no es obligatorio y dependerá de la institución en la que se implemente el repositorio y la relación y proximidad con los autores. No obstante, se recomienda promover estas políticas y mandatos para afianzar el respaldo permanente de la dirección de la institución. Para elaborar una política institucional según el modelo que más se aproxime a las características y necesidades de la institución se puede consultar ROARMAP (*Registry of Open Access Repository Material Archiving Policies*, <http://roarmap.eprints.org/>) o, para instituciones de habla hispana, MELIBEA (<http://www.accesoabierto.net/politicas/>).

Pero el primer paso para la implementación de un sistema de repositorio pasa necesariamente por realizar un diagnóstico de situación que deberá tener en cuenta la planificación estratégica de la institución en la que se quiere crear el repositorio, para sumar esfuerzos entre las áreas de la organización, y dar cumplimiento a la estrategia genérica existente. Más allá de la coherencia estratégica, se trata de preparar el proyecto para que sea financiable y tenga el interés de los directivos.

Lógicamente el diagnóstico debe responder a qué necesidad real responde la voluntad de crear el repositorio. Tras identificar los objetivos marco a los que puede dar cumplimiento la creación de un repositorio, el análisis debe continuar con la descripción de los recursos necesarios el proyecto: de personal, de infraestructura y, especialmente, de las colecciones susceptibles de ser depositadas, sean de documentos nacidos digitales o de documentos analógicos. Complementariamente, es necesario prever las dimensiones que alcanzará el repositorio, que formatos aceptará, y qué usos tendrá.

El proceso de diagnóstico debe incorporar la búsqueda de fuentes alternativas de financiación a las de la propia institución, bien por medio de ayudas públicas, bien a partir de patrocinios privados.

Un ejemplo de ese trabajo precordinado es el portal español ARCA (*Archivo de Revistas Catalanas Antiguas*, <http://www.bnc.cat/digital/arca/castella/index.html>), que contiene publicaciones periódicas que ya no se publican y que han sido representativas de la cultura y de la sociedad catalana. Impulsado por la *Biblioteca de Catalunya* y el CBUC, *Consortio de Biblioteques Universitaries de Catalunya*, el éxito de la iniciativa se basa en cuatro ejes:

- Se publica en acceso abierto las colecciones de revistas.
- Se contribuye a la preservación de las revistas originales, que al ser publicadas en Internet no necesitan ser siempre consultadas manualmente.
- Gracias al proceso de OCR se ofrece búsqueda a texto completo de las revistas, imposible antes de la digitalización.
- Y se fomenta la cooperación entre instituciones bibliotecarias diversas, así como de archivos y museos, puesto que se completan en la red colecciones físicas que no tiene en exclusiva ningún centro.

Insistimos en la necesidad de seleccionar cuidadosamente los materiales que queremos incluir en el repositorio. La definición dependerá del marco de actuación de cada organización, pero lógicamente debemos planificar qué prioridades tendrán un mejor impacto, puesto que la financiación de proyectos de repositorio debe convencer a las estructuras directivas en el momento inicial, y a medio y largo plazo.

Un buen ejemplo de esa planificación corresponde al *Archivo Municipal de Girona* (100.000 habitantes), que además de una fabulosa colección de 4 millones de fotografías digitalizadas y de acceso en línea, incluye en sus repositorios una sección de archivo administrativo y otra de archivo histórico (a partir del s. XIV). El proceso de digitalización se inició en 1990, y contrariamente al inicial deseo de los historiadores locales, se centró justamente en los documentos producidos más recientemente por el ayuntamiento: las actas de los plenarios municipales. Los responsables políticos dieron soporte unánime a un proyecto que les ayudaba en su tarea cotidiana, facilitando así la financiación permanente del resto de sus fondos. La estrategia fue fruto de un cambio de filosofía, al pasar la prioridad de la custodia y preservación de la documentación analógica, al acceso por medio de las tecnologías a la documentación administrativamente activa.

Sin abundar en lo descrito, durante la planificación deberemos decidir también si para nuestra institución, usuarios y características de los documentos que queremos incluir en el repositorio es más conveniente implementar un repositorio institucional propio o bien participar en un repositorio cooperativo ya existente. Si se dispone de recursos suficientes para implementar un repositorio local es aconsejable, ya que contribuirá a incrementar y consolidar la visibilidad institucional. No obstante, se debe considerar igualmente participar en proyectos más globales tanto a nivel territorial, como de tipología documental y temática, pues darán más visibilidad y difusión a los documentos, a sus autores y a la institución.

Finalmente, tras la identificación de las políticas de la organización en las que podemos vehicular el proyecto de repositorio, y tras seleccionar qué colecciones documentales formarán parte de éste, es necesario tener en cuenta los condicionantes legales que afectan a esas colecciones: la cobertura legal.

Antonia Ferrer et al. (2005) han listado los materiales no sujetos a derechos de explotación de la legislación española que más habitualmente forman parte de bibliotecas digitales, que no reproduciremos por no ser universalmente aplicables. Pero en todo caso, como ha indicado Josep Vives (2005), disponemos de buenos y suficientes argumentos para convencer a nuestros futuros depositantes de la bondad de los repositorios.

En el caso de los repositorios institucionales, es habitual que se autoarchiven versiones pre o post-print de artículos publicados en revistas académicas y profesionales, que normalmente permiten esta práctica, así como presentaciones de las sesiones presenciales docentes, o trabajos académicos realizados por estudiantes. En ese caso es necesario que el autor, que realiza el autoarchivo, exprese por escrito su consentimiento a incluir su documento en el depósito y su posterior comunicación pública, según el nivel de privacidad de acceso que se determine. Como *SHERPA/RoMEO* (<http://www.sherpa.ac.uk/romeo/>),

Dulcinea (<http://www.accesoabierto.net/dulcinea/>) es un proyecto para conocer las políticas editoriales de las revistas españolas en relación al autoarchivo en repositorios.

En los documentos nacidos digitales que se incluyen en los repositorios en acceso abierto se están usando ampliamente las licencias *Creative Commons* (CC)

Sin ánimo de contrariar lo expuesto hasta el momento, no han sido pocos los repositorios iniciados por pequeñas unidades de bibliotecas o departamentos universitarios, sin más amparo que la motivación de sus miembros y sin más recursos que los puntualmente obtenidos para proyectos tecnológicos. El *Depósito Digital de Documentos* (<http://ddd.uab.cat/?ln=es>), gestionado por la *Universidad Autónoma de Barcelona* (43.000 estudiantes) ocupa actualmente la primera posición española del ranking web de repositorios. Con 82.000 registros, comprende artículos de la comunidad universitaria, materiales docentes, e incluso fondos personales que han sido legados a la universidad. Creado en 2006 con el software CDS Invenio, se alimenta por autoarchivo; copia de registros del catálogo de la biblioteca universitario --a los cuales se añade el documento digitalizado--; compilación de metadatos procedentes de repositorios consorciados; y cargas de bases de datos de digitalizaciones retrospectivas. Su responsable, Cristina Azorín, ha señalado (2010) que el crecimiento del repositorio responde principalmente a la labor constante del personal de las bibliotecas, más que a las políticas de autoarchivo de docentes e investigadores. De hecho, el mandato institucional de la universidad no fue aprobado hasta abril de 2012 (<http://ddd.uab.cat/record/89641?ln=es>).

Tras contar con cobertura estratégica y legal para publicar las colecciones susceptibles de formar parte del repositorio, es el momento de identificar los recursos que necesitaremos para implementarlo, así como el sistema de metadatos que utilizaremos en el repositorio.

2.2. Herramientas hardware & software

Por lo que afecta al tipo de servidor que soportará el repositorio, la elección entre un sistema de arquitectura informática local o externa depende de diversos factores, básicamente relacionados con las finanzas de la institución. Clobridge (2010, p. 44) ha analizado a fondo las ventajas e inconvenientes según el tipo de servidor para un repositorio económico. Pero es evidente que no hay una única respuesta. Si la institución no tiene un equipo informático estable y personal técnico, será más rápido y barato empezar ubicando el repositorio en un sistema externo. Pero a largo plazo será más caro mantener esa estructura externa.

Otros factores que determinarán el grado de complejidad de la arquitectura son el volumen de documentos a depositar (especialmente en el caso de documentos muy "pesados", como las imágenes), y el nivel de consultas de los usuarios potenciales. A más expectación, más elaborada deberá ser la arquitectura técnica.

Por citar un ejemplo real, un sistema básico de infraestructura técnica local para un repositorio comprendería un nodo de la serie Proliant DL360 (HP, o su equivalente en Sun o Dell), con 2 procesadores de 6 núcleos más 24 GB de RAM y un disco duro interno de 300

GB, con una inversión inicial de unos 9.000 USD, dependiendo del proveedor y del momento de la compra.

Además del sistema básico de servidor será necesario contar con estaciones de trabajo (equipos informáticos de sobremesa, escáneres o cámaras digitales) con la estructura mínima para el trabajo de edición de los documentos a depositar. Complementariamente, un sistema que garantice alimentación ininterrumpida y un dispositivo de *backup*.

Imaginamos que un servicio de información que desee implementar un sistema de repositorio ya ha tenido en cuenta estos elementos de mínimos. Cuando la medida de las organizaciones lo permite, lo lógico es establecer alianzas con los departamentos informáticos de la propia institución para no duplicar tareas ni infraestructuras.

Por lo que afecta al software, la tendencia mundial es implementar los repositorios 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 teóricamente la continuidad del software. Las comunidades internacionales que apoyan los programas en los que se basan los repositorios (DSpace, OJS, EPrints, etc.) son importantes por la cantidad de instituciones implicadas y por el volumen de usuarios, lo cual es una garantía de continuidad.

La elección del software se produce normalmente en base al tipo de documentación que deba almacenar, y no existe un único sistema con soluciones ideales para todos los casos. En el caso de depositar imágenes, el sistema debe permitir editarlas, descargarlas, revisar múltiples imágenes para comparar detalles, poder hacer uso del zoom, y comprobar los metadatos visibles. Los manuscritos deben poder, además, transcribir la información o el texto. Los archivos de imagen y vídeo requieren más espacio que el resto. Los trabajos académicos y, en general, los documentos con texto, deben tener funciones de búsqueda en texto libre. A título de ejemplo, OJS parece el más indicado para gestionar la publicación de revistas digitales, CONTENTdm podría ser una buena solución para gestionar imágenes y DSpace para documentos de investigación, materiales docentes, tesis doctorales, etc.

A nivel internacional, entre los repositorios de acceso abierto, existen unos 70 programas en uso. Los dos sistemas más comunes según *OpenDOAR* son DSpace (41%) y EPrints (14%). Estos son fáciles de instalar y mantener, y gestionan todo tipo de formatos. Fedora, según la mayoría de análisis, es el tercero en nombre de instalaciones, pero más complejo: apto para todo tipo de documentos digitales, requiere instalación y desarrollo técnico experto. Greenstone, de código abierto, es muy usado por instituciones de dimensión mediana y pequeña porque es sencillo de instalar y mantener. CONTENTdm, ofrecido por OCLC y otros consorcios a sus socios para colecciones de tamaño limitado, permite poca personalización. Los dos casos son de fácil instalación y gestión, y tienen un coste bajo.

Como han señalado Abadal y Anglada (2010) el programa más utilizado en España para los repositorios de acceso abierto es DSpace.

Justamente ese popular software libre, DSpace, fue el elegido para el repositorio *Digital CSIC* (<http://digital.csic.es/>) por el *Centro Superior de Investigaciones Científicas* (CSIC), la mayor agencia española de investigación científica, la tercera en la Unión Europea, con 15.000 personas en plantilla, de las que 3.000 son investigadores. El CSIC genera el 20% de la producción científica en España. Su repositorio, con 51.000 registros, realiza un tratamiento centralizado del depósito de documentos científicos: el 40% una oficina técnica, el 40% las diferentes bibliotecas sectoriales, y el 20% restante procede del autoarchivo. Isabel Bernal, su responsable, aspira a convertir el repositorio en la memoria digital de la ciencia española.

CONTENTdm, por su parte, es el principal software propietario utilizado en España. Pese a que permite gestionar todo tipo de documentos, son las colecciones de imágenes las que predominan en su cartera de clientes.

La Fundación Joan Miró, dedicada al artista surrealista del mismo nombre, es una fundación privada que en 1975 abrió sus puertas en Barcelona para acoger una colección de 14.000 piezas del artista, así como de otros contemporáneos (Duchamp, Clader, Chillida). Su repositorio (<http://www.fundacionmiro-bcn.org/colecciojoanmiro.php?idioma=6>), gestionado por CONTENTdm, contiene unos mil registros de bocetos, dibujos y pinturas del artista.

Si el objetivo del repositorio es para uso personal, doméstico, o de un centro con recursos mínimos, puede ser buena idea utilizar algunos de los sistemas gratuitos que se asemejan a los sistemas repositorios, como Omeka, el Picasa de Google, o el popular Flickr.

Más allá del programario que gestione la colección, es una recomendación generalizada que los repositorios de dimensiones medias y altas se construyan sobre un sistema de bases de datos, como Oracle, Microsoft SQL, o MySQL, para garantizar la flexibilidad a largo plazo y las posibilidades técnicas que surgen al trabajar en estructuras más sólidas. En caso de querer trasladar un repositorio de un sistema a otro se debe tener la seguridad de que los datos estarán en formatos estándares, como XML.

Existen diversas comparativas que calculan las diferencias en materia de coste-beneficio en las novedades que se producen en cada programa. Un ejemplo reciente es la propuesta de Singh y Witt y Salo (2010) presentada en la *International Conference on Open Repositories*.

Finalmente, si se desea externalizar el proyecto, empresas como *Open Repository*, impulsada por *BioMed Central*, ofrecen servicios integrales para repositorios que pueden incluir instalación, desarrollo, *hosting* y mantenimiento de un repositorio con DSpace desde 11.600 USD por licencia, y mantenimiento anual de 13.400 USD. En la publicidad de sus productos se comprometen a poner en funcionamiento un repositorio en tres meses.

2.3. Metadatos y protocolos para la interoperabilidad

Los documentos que se introduzcan en el repositorio deberán estar descritos utilizando un conjunto de metadatos basados en estándares internacionales y mínimamente normalizados. El estándar de metadatos más ampliamente utilizado es el Dublin Core Cualificado. Es importante establecer unas especificaciones sobre cómo dar la información en los diferentes elementos para asegurar su homogeneidad y normalización: por ejemplo, los puntos de acceso de los autores y materias se pueden dar tomando como referencia el catálogo de la biblioteca.

Si los documentos se incluyen a través del autoarchivo, es decir que es el propio autor del documento quien introduce en el repositorio el registro descriptivo mínimo (por ejemplo: autor, título y fecha) y el fichero, se debe establecer un proceso de trabajo que permita la revisión al personal bibliotecario para completar y validar la información.

Por otra parte, la mayoría de los repositorios cumplen con el protocolo de interoperabilidad creado por la Open Archives Initiative: OAI-PMH (*Open Archives Initiative - Protocol for Metadata Harvesting*). El cumplimiento de este protocolo es totalmente recomendable (si no imprescindible) ya que asegura que los recursos se puedan recuperar en otros repositorios o agregadores, e incrementar de este modo la visibilidad de los documentos contenidos. Si queremos participar en determinados portales, como por ejemplo DRIVER o --en el caso de los países europeos-- *Europeana*, deberemos cumplir con sus concreciones de metadatos.

Un ejemplo de cooperación internacional por medio de la interoperabilidad de los registros es el portal de *Movimientos Migratorios Iberoamericanos* (<http://pares.mcu.es/MovimientosMigratorios>). Coordinado por el ministerio español de Cultura, cuenta con la colaboración del *Archivo General de la Nación de México*, el *Archivo General de la Nación de la República Dominicana*, el *Archivo General de la Nación de Uruguay*, y la *Fundación Complejo Cultural Parque de España*, de Argentina, y está prevista la incorporación del *Archivo Nacional de la República de Cuba*. Tiene el objetivo de facilitar el acceso a la documentación administrativa relativa a la emigración española a América Latina. Procesado por un software hecho a medida, contiene 34.000 registros.

En la medida de lo posible es recomendable también implementar el uso de metadatos PREMIS (*PREservation Metadata: Implementation Strategies*), estándar administrado por el Comité Editorial PREMIS (<http://www.loc.gov/standards/premis/>) en colaboración con la *Library of Congress* estadounidense. El uso de estos metadatos está destinado a apoyar la preservación a largo plazo de los objetos digitales en los repositorios. Y también de METS (*Metadata Encoding and Transmission Standard*), otro estándar usado para codificar los metadatos descriptivos, administrativos y estructurales que está basado en el esquema XML. Está administrado por la *Network Development and MARC Standards Office* de la misma biblioteca nacional. Su finalidad, a través del lenguaje XML y de sus siete secciones, es describir la estructura jerárquica de los objetos digitales y los nombres y ubicación de los ficheros que los componen.

2.4. Promoción para la sostenibilidad

Es esencial programar, desde el inicio del proceso de implementación y carga de objetos del repositorio, las estrategias de comunicación que se llevarán a cabo, teniendo en cuenta la diversidad del tipo de usuarios del repositorio.

Como ha señalado Isidro Aguillo, responsable del *Cybermetrics Lab* (<http://internetlab.cindoc.csic.es/>), el secreto de un buen repositorio está en las 3C: Contenido, Contenido, y Contenido. En su decálogo de buenas prácticas para el posicionamiento: la asignación de URL como marca única; la creación de nuevos y conversión de antiguos contenidos; la interconexión con la comunidad científica; la inclusión de ficheros ricos y multimedia, etc.

En esa misma línea Dorner & Revell (2012) afirmaban, a partir de una serie de entrevistas a bibliotecarios referencistas neozelandeses --sin duda con una perspectiva algo naif-- que un repositorio exitoso es fruto de la inclusión de contenidos y el acceso (y uso) a esos contenidos. De hecho, por centrar el estudio en el Simposio de nos ocupa, lo que más han valorado los bibliotecarios especializados de las universidades estudiadas es el acceso a las tesis doctorales (y el resto de los trabajos finales de estudios).

Por citar un par de buenas prácticas: en RiuNet (<http://riunet.upv.es/>), el repositorio institucional de la *Universidad Politécnica de Valencia* (36.000 estudiantes), el 73% de su contenido (11.444 registros, en agosto de 2012) son de aprendizaje o de docencia. Ello significa que el personal de la biblioteca ha llevado a cabo una ardua tarea de promoción del autoarchivo entre profesores para conseguir no solo incorporar al repositorio copias de las publicaciones en revistas, sino también materiales inéditos. La tarea del autoarchivo va seguida de un trabajo de edición de metadatos, para garantizar coherencia en la descripción bibliográfica de la colección. En la *Universidad de Salamanca* (30.000 estudiantes), el repositorio *Gredos* (<http://gredos.usal.es/>) ha apostado también por un contenido heterogéneo: en sus 93.000 ítems se incluyen colecciones de contenido administrativo y docente, promovido sin duda por los esfuerzos del personal bibliotecario.

Establecer alianzas estratégicas, dentro y fuera de la organización, es un paso ineludible. Como han señalado Jenny Fry et al. (2011), los investigadores no asocian acceso abierto con autoarchivo, y tienden a ser conservadores: un repositorio no reemplaza a las herramientas de difusión que ya existen.

En ese mismo sentido se expresaron Davis & Connolly (2007), con los motivos esgrimidos por los investigadores para no usar el repositorio de su institución: confusión de los autores ante la variedad de repositorios; falta de tiempo para aprender el sistema de funcionamiento; confusión sobre los derechos de autor; temor al plagio; temor de revelar el estado de la cuestión sus proyectos; percepción psicológica de baja calidad de los materiales depositados; duda sobre si el material depositado se considera publicado o no.

Un ejemplo destacable de promoción del autoarchivo se halla en el repositorio *Acceda* (<http://acceda.ulpgc.es/>), de la *Universidad de Las Palmas de Gran Canaria* (24.000 estudiantes). Creado en 2009, contiene 5.700 registros gestionados con DSpace. Además de diversos vídeos de ayuda para el autoarchivo, se reclama en portada la colaboración de

la comunidad docente e investigadora. Un blog creado al efecto permite estar al día de las novedades del repositorio y la cultura *Open Access* (<http://bibwp.ulpgc.es/accesoabierto/>).

Máxime en el caso de repositorios institucionales, dar a conocer y formar en el uso a los clientes potenciales de la propia institución o empresa es el paso previo a dar publicidad exterior del repositorio, mediante el registro en directorios y portales, como *OAI Data providers*, *OpenDOAR*, *BuscaRepositorios*, y --en el caso español-- *Hispana*, o en sistemas de ranking web, como el de *Repositorios del mundo* (<http://repositories.webometrics.info/>) que mantiene el *Laboratorio de Cibermetría* del CSIC, el *Centro Superior de Investigaciones Científicas* español.

La presencia en congresos y publicaciones científicas y profesionales de bibliotecas, archivos y museos es una buena práctica, puesto que da visibilidad de la actividad profesional en la organización responsable del servicio de información.

El repositorio *Memoria Digital Vasca* (DSpace, 3.000 registros, accesible en: <http://www.memoriadigitalvasca.es/>) fue creado por la *Fundación Sancho El Sabio* con sede en Vitoria, donde se halla la entidad bancaria que la impulsó en 1955, *Caja Vital Kutxa*. Comprende manuscritos, mapas, material gráfico, monografías y revistas, con un sistema propio de clasificación. Desde su creación, en 2009, sus responsables han publicitado el repositorio en ámbitos académicos y culturales, de la misma manera que desde su *apertura* al público, en 2011, el repositorio *Clamor: Colección Digital de Música Española* (<http://digital.march.es/clamor/>) ha tenido presencia en foros profesionales. *Clamor* ha sido creado por la *Fundación Juan March*, institución privada que desarrolla sus actividades en el campo de la cultura humanística y científica desde 1955. Contiene grabaciones en vivo, programas de mano, información sobre los intérpretes, fotografías, etc. Su software es *Islandora*, una adaptación de *Fedora*.

Por último, utilizar de manera sistemática y programada las herramientas 2.0 y las aplicaciones para móviles es recomendable para garantizar un posicionamiento del repositorio en el sector en el que se encuentre la institución o empresa responsable de haberlo impulsado. Ello incluye actualmente el uso permanente de redes sociales genéricas como Facebook, Twitter, LinkedIn, y conseguir la presencia con impactos positivos en banners o enlaces de otras páginas web de su ámbito, para garantizar una posición sólida en Internet. En un reciente estudio de la *Bibliothèque Nationale de France* se afirma que los investigadores realizan el seguimiento de su disciplina científica a través de las redes sociales (Chevalier e Illien, 2011). En uno de los casos descritos el investigador entrevistado afirma desconocer a menudo si está investigando o jugando, debido al tiempo que empleaba en la lectura de tweets y actualizaciones de estado de Facebook, como la manera más útil de saber cuáles eran los avances de sus compañeros.

2.5. Evaluación para la excelencia

La *Fundación Española para la Ciencia y la Tecnología* (FECYT) y la *Red de Bibliotecas Universitarias* (REBIUN) editaron la ejemplar *Guía para la evaluación de repositorios institucionales de investigación*, coordinada por José Manuel Barrueco en 2010, que

sistematiza un proceso de evaluación de repositorios institucionales, y entendemos es ampliable a todo tipo de depósitos digitales.

Recomendamos utilizar esta guía antes de la implementación de cualquier repositorio, y con posterioridad a su creación, a fin de mejorarlo bajo criterios objetivos. Entre los aspectos a evaluar, la visibilidad (mención del repositorio en la página principal de la institución, presencia en directorios nacionales e internacionales, nombre inequívoco y normalizado, etc.); políticas (si existe una declaración sobre la misión y objetivos del repositorio, etc.); aspectos legales (cumplimiento de los derechos de propiedad intelectual, etc.); metadatos (aplicación de sistemas normalizados, etc.); interoperabilidad (uso del protocolo OAI-PMH, etc.); logs y estadísticas; seguridad, autenticidad e integridad de los datos (uso de identificadores persistentes, etc.).

Las interfaces de consulta y visualización tienen sus sistemas específicos de evaluación, como el propuesto por Estivill, Gascón y Sulé (2011).

2.6. Estrategia y táctica en preservación digital

Damos por supuesto el consenso sobre el hecho que digitalizar y depositar no es sinónimo de preservación. Es quizás el primer paso, pero no el único.

La preservación digital requiere un esfuerzo permanente, que incluye continuos cambios y actualizaciones. Alice Keefer y Núria Gallart (2007) han señalado acertadamente que a diferencia de la preservación tradicional, orientada esencialmente a mantener el documento en su estado original, los documentos digitales deben ser modificados y transformados continuamente para que continúen siendo accesibles e inteligibles: se trata de preservación por alteración.

Porque como ha sintetizado recientemente Tránsito Ferreras (2010), las amenazas son numerosas: los objetos digitales son inestables, los equipos y programas tienen un breve ciclo vital, las responsabilidades están poco definidas, existen dificultades legales, etc. Pero probablemente el mayor reto al que se enfrentan los profesionales de bibliotecas, archivos y museos es la planificación de la preservación digital: de los documentos digitales y de los metadatos que les acompañan. Es inevitable tener en cuenta tres objetivos básicos en materia de preservación digital, que deben ser garantizados permanentemente (Wheatley, 2004), según el necesario equilibrio entre coste y beneficio:

- Debe impedirse que los datos depositados se pierdan, dañen o alteren. Por ejemplo mediante una estructura sólida de hardware de seguridad consistente en sistemas de alimentación ininterrumpida y un dispositivo riguroso de *back up*, complementada con programas que revisen periódicamente los contenidos del repositorio.
- Debe promoverse el acceso permanente a los datos. Por ejemplo mediante el uso de identificadores persistentes.
- Debe actualizarse de manera permanente la integridad y acceso a los documentos, superando retos actuales y de futuro, puesto que los cambios en formatos y soportes son impredecibles a largo plazo. Por ejemplo, con las técnicas más habituales en

materia de preservación: migración (de formatos: de una versión antigua a una nueva de un software determinado, o de un programa propietario a uno libre), *refreshing* (de soportes: de un DVD a un disco duro externo, por ejemplo), y emulación (uso de hardware y software originales o que simulan la originalidad de los equipos y versiones).

Además de la adopción de medidas propias, como el contar con un sistema de caja fuerte o archivo oscuro, de acceso restringido, una opción es sumarse a un proyecto de preservación ya existente. Es el caso del repositorio español para tesis doctorales TDR, que desde 2011 forma parte de la cooperativa *MetaArchive*, con el objetivo de asegurar la preservación de las tesis españolas, y de participar en una de las iniciativas pioneras en el mundo en este ámbito. Esta cooperativa, formada por instituciones importantes de todo el mundo, dispone de una red de intercambio de información que asegura que otras instituciones mantienen una copia del contenido digital a preservar. La preservación se lleva a cabo a través del software LOCKSS que constantemente compara las diferentes copias por, si se da el caso, avisar y restaurar de manera automática cualquier fallo. Un repositorio será más seguro si cuenta con un sistema de réplica en otras instituciones.

Tras numerosos informes, estudios y proyectos, no existe en la actualidad un sistema informático que cumpla todos los requerimientos de preservación digital. Por ello, para garantizar el acceso a los fondos digitales, se combinan estrategias a medio plazo y actividades mínimamente experimentadas.

3. Referencias

Abadal, E.; Anglada, L., eds. (2010). *Open Access in Southern European countries*. Madrid: FECYT. http://oaseminar.fecyt.es/Resources/Documentos/OASouthern_Vweb.pdf

Aguado-López, E.; Rogel-Salazar, R. (2006). "Redalyc: Red de Revistas Científicas de América Latina, el Caribe, España y Portugal : un balance de tres años". D. Babini, J. Fraga. *Edición electrónica, bibliotecas virtuales y portales para ciencias sociales en América Latina y el Caribe*. Buenos Aires: Consejo Latinoamericano de Ciencias Sociales (CLACSO), p. 209-233.

Aguado-López, E., et al. (2009). "Presencia de universidades en la red: la brecha digital entre Estados Unidos y el resto del mundo". *Revista de Universidad y Sociedad del Conocimiento (RUSC)*. V. 6, núm. 1. <http://rusc.uoc.edu/ojs/index.php/rusc/article/viewFile/18/72>

Aguillo, I. F., et al. (2010). "Indicators for a webometric Ranking of Open Access Repositories". *Scientometrics*, 82, n. 3, p. 477-486. <https://digital.csic.es/bitstream/10261/32190/1/Ranking%20of%20Repositories.pdf>

Anderson, C. G.; Maxwell, D. C. (2004). *Starting a digitization center*. Oxford: Chandos.

Andrews, J.; Derek, L., eds. (2004). *Digital libraries: policy, planning and practice*. Hants: Ashgate.

Anglada, L.; Reoyo, S.; Cambras, J.; De la Vega, R. (2008). "Los repositorios electrónicos cooperativos de la Biblioteca Digital de Cataluña". *BiD: textos universitaris de biblioteconomia i documentació*, núm. 20 (jun). <http://www.ub.edu/bid/20angla2.htm>

Azorín, C.; Jorba, F.; Piera, B. (2006). "Repositorio digital de documentos: ¿diferente? ¿definido? ¿desafío?". *VI Workshop REBIUN sobre proyectos digitales*. (2006 : Madrid). <http://eprints.rclis.org/handle/10760/8995#.UDS8qaDQuuM>

Azorín, C., et al. (2010). "Els repositoris institucionals: entre la gestió i l'avaluació". *12es Jornades Catalanes d'Informació i Documentació*. (2010 : Barcelona). http://www.cobdc.org/jornades/12JCD/materials/comunicacions/AZORIN_repositoris_institucionals.pdf

Babini, D. (2011). "Acceso abierto a la producción científica de América Latina y el Caribe : identificación de principales instituciones para estrategias de integración regional". *Revista CTS*, vol. 6, núm. 17 (abr). <http://revistacts.net/files/Volumen%206%20-%20N%C3%BAmero%2017/Babini.pdf>

Barrueco, J. M., coord. (2010). *Guía para la evaluación de repositorios institucionales de investigación: v. 1.0 diciembre 2010*. Madrid: FECYT, RECOLECTA; CRUE. <http://www.recolecta.net/buscador/documentos/GuiaEvaluacionRecolectav1.0-1.pdf>

"Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities" (2003). *Conference on Open Access to Knowledge in the Sciences and Humanities* (Berlin : 20-22 oct). <http://www.zim.mpg.de/openaccessberlin/berlindeclaration.html>

Bernal, I.; Pemau, J. "Estadísticas para repositorios: sistema métrico de datos en Digital.CSIC". *El Profesional de la información*. vol. 19, núm. 5 (sep-oct). <http://elprofesionaldelainformacion.metapress.com/media/cmw75594tjdryl80uw1h/contributios/t/1/h/3/t1h3750537433248.pdf>

Bor Ng, K.; Kucsma, J., eds. (2010). *Digitization in the real world: lessons learned from small and medium-sized digitization projects*. New York: Metropolitan New York Library Council.

Chevalier, P.; Illien, G. (2011). *Les archives de l'Internet: une étude prospective sur les représentations et les attentes des utilisateurs potentiels*. Paris: Bibliothèque nationale de France. http://www.bnf.fr/documents/enquete_archives_web.pdf

Clobridge, A. (2010). *Building a digital repository program with limited resources*. Oxford: Chandos.

Crow, R. (2002). *The Case for Institutional Repositories: a SPARC Position Paper: release 1.0*. Washington: Scholarly Publishing & Academic Resources Coalition.
http://www.arl.org/sparc/bm~doc/ir_final_release_102.pdf

Davis, P. M.; Connolly, M. J. L. (2007). "Institutional repositories: evaluating the reasons for non-use of Cornell University's installatin of DSpace". *D-Lib Magazine*, vol. 13, n. 3/4.
<http://www.dlib.org/dlib/march07/davis/03davis.html>

Dorner, D. G.; Revell, J. (2012) "Subject librarians' perceptions of institutional repositories as an information resource", *Online Information Review*, Vol. 36, Iss. 2, p. 261-277.
<http://dx.doi.org/10.1108/14684521211229066>

España (2010). *Declaración de la Alhambra sobre Open Access*. Granada: Ministerio de Ciencia e Innovación.
http://oaseminar.fecyt.es/Resources/Documentos/ADeclaration/AD_OA_Spanish.pdf

España (2011). "Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación". *Boletín Oficial del Estado*, 131 (2 jun 2011), p. 54.387 y ss.
<http://www.boe.es/boe/dias/2011/06/02/pdfs/BOE-A-2011-9617.pdf>

España (2011). "Real Decreto 99/2011, de 28 de enero, por el que se regulan las enseñanzas oficiales de doctorado". *Boletín Oficial del Estado*, 35 (10 feb 2011), p. 13.909 y ss.
<http://www.boe.es/boe/dias/2011/02/10/pdfs/BOE-A-2011-2541.pdf>

Estivill, A.; Gascón, J.; Sulé, A. (2011). "Interfaces de consulta en las colecciones digitales patrimoniales españolas". *Anales de documentación*, vol. 14, núm. 2.
<http://revistas.um.es/analesdoc/article/download/113931/124971>

European Commission. Digital Libraries Initiative. Comité des Sages (2011). *The new renaissance: [report on Bringing Europe's Cultural Heritage Online]*.
http://ec.europa.eu/information_society/activities/digital_libraries/doc/executivesummery/final_renaissace_es.pdf

Ferreira, M., et al. (2008). "Carrots and sticks: some ideas on how to create a successful institutional repository". *D-Lib Magazine*, vol. 14.
<http://www.dlib.org/dlib/january08/ferreira/01ferreira.html>

Ferrer, A., et al. (2005). *Guía metodológica par la implantación de una biblioteca digital universitaria*. Gijón: Trea.

Ferreras, T. (2010). *Preservación digital en repositorios institucionales: Gredos [Tesis de Máster]*. Salamanca: Universidad de Salamanca.
<http://eprints.rclis.org/handle/10760/16356#.UDS9YaDQuuM>

Fry, J., et al. (2011). *PEER Behavioural research: authors and users vis-à-vis journals and repositories. final report.* (aug).
http://www.peerproject.eu/fileadmin/media/reports/PEER_D4_final_report_29SEPT11.pdf

Jacobs, N., ed. (2006). *Open access: key strategic, technical and economic aspects*. Oxford: Chandos.

Jones, R.; Andrew, T.; MacColl, J. (2006). *The Institutional repository*. Oxford: Chandos.

Keefer, A.; Gallart, N. (2007). *La preservación de recursos digitales: el reto para las bibliotecas del siglo XXI*. Barcelona: Universitat Oberta de Catalunya (UOC).

Kresh, D., ed. (2007). *The Whole digital library handbook*. Chicago: American Library Association.

Llueca, C.; Reoyo, S. (2012). "Creación de un repositorio digital con recursos limitados". Rovira, C.; Marcos, M. C.; Codina, L. (dirs.). *Máster en Documentación Digital*. Barcelona: Universidad Pompeu Fabra. <http://www.documentaciondigital.org>

Lynch, C. A. (2003). "Institutional repositories: essential infrastructure for scholarship in the digital age" *ARL*, no. 226 (feb), p. 1-7. <http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>

Swan, A. (2011). "Institutional repositories: now and next". Dale, P.; Beard, J.; Holland, M., eds. *University libraries and digital learning environments*. <http://eprints.ecs.soton.ac.uk/21471/>

Singh, S. K.; Witt, M.; Salo, D. (2010). "A comparative analysis of institutional repository software". *International Conference on Open Repositories* (2010 : Madrid). <http://or2010.fecyt.es/Resources/documentos/GSabstracts/AComparativeAnalysisInstitutionalRepositorySoftware.pdf>

Térmens, M. (2009). "Investigación y desarrollo en preservación digital: un balance internacional". *El Profesional de la información*, vol. 18, núm. 6 (nov-dic). <http://elprofesionaldelainformacion.metapress.com/media/6n99d5ttlj4uuuj2g9m7j/contributions/f/8/h/1/f8h1x20t283mn152.pdf>

UNESCO (2003). *Directrices para la preservación del patrimonio digital*. Canberra: National Library of Australia. <http://unesdoc.unesco.org/images/0013/001300/130071s.pdf>

Vives, J. (2005). "Aspectos de propiedad intelectual en la creación y gestión de repositorios institucionales". *El profesional de la información*, vol. 14, núm. 4 (jul-ago), p. 267-278. <http://eprints.rclis.org/handle/10760/6505>

Weathley, P. (2004). *Technology watch report: institutional repositories on the context of digital preservation*. Leeds: University of Leeds; Digital Preservation Coalition. http://www.dpconline.org/component/docman/doc_download/90-institutional-repositories-in-the-context-of-digital-preservation