

Las colecciones digitales en España: situación actual y perspectivas de futuro

Por María-Antonia Carrato-Mena

Resumen: La Dirección General del Libro, Archivos y Bibliotecas, Subdirección General de Coordinación Bibliotecaria ha puesto en marcha una serie de iniciativas para difundir y preservar el patrimonio bibliográfico de nuestro país. La digitalización de fondos y su distribución por internet (siguiendo las normas más implantadas DC, OAI-PMH, OAIS, Premis, SRU/SW) constituye el más reciente de sus proyectos. Además de todos los beneficios que ello supone, cabe valorar especialmente el esfuerzo para que sean proyectos de consenso. Por un lado ha captado la cooperación de las comunidades autónomas; y por otro todas sus acciones son, por la tecnología y procedimientos que usa, acordes con el marco internacional, como la acción Digital Libraries, del VII Programa marco de la Unión Europea.



María-Antonia Carrato-Mena es subdirectora general de Coordinación Bibliotecaria del Ministerio de Cultura desde comienzos de 2005. Su Subdirección es responsable del desarrollo y la coordinación del Sistema Bibliotecario Español, del Catálogo Colectivo del Patrimonio Bibliográfico y de la promoción de relaciones internacionales en materia de bibliotecas. Previamente se ocupó de la creación de infraestructuras y servicios bibliotecarios integrados en la Comunidad de Madrid, y también realizó diferentes trabajos en la Biblioteca Nacional de España. Es miembro del Comité ejecutivo de Naple (National Authorities on Public Libraries in Europe), preside el Foro Iberoamericano de Responsables Nacionales de Bibliotecas Públicas y forma parte del proyecto ED-Local de la Biblioteca Digital Europea.

Palabras clave: Colecciones digitales, España, Digitalización, Patrimonio documental, DC, OAI-PMH, OAIS, Premis, SRU/SW, Dirección General del Libro, Archivos y Bibliotecas, Subdirección General de Coordinación Bibliotecaria.

Title: Digital collections in Spain: Current situation and future perspectives

Abstract: Spain's national agency for library coordination (Dirección General del Libro, Archivos y Bibliotecas, Subdirección General de Coordinación Bibliotecaria) has launched a series of initiatives to disseminate and preserve the national bibliographic heritage. The digitization of collections and their internet distribution (using well-known standards such as DC, OAI-PMH, OAIS, Premis, SRU/SW) is their most recent project. In addition to all the benefits this brings, their efforts to build consensus for the project is especially appreciated. Of particular significance, they have secured the cooperation of Spain's autonomous communities (i. e., regional governments), all of their initiatives incorporate technology and procedures compliant with the action Digital Libraries of the EU's VII Framework Programme.

Keywords: Digital collections, Spain, Digitalization, DC, OAI-PMH, OAIS, Premis, SRU/SW, Collaboration, Library coordination, Archives, Bibliographic heritage.

Carrato-Mena, María-Antonia. "Las colecciones digitales en España: situación actual y perspectivas de futuro". En: *El profesional de la información*, 2008, julio-agosto, v. 17, n. 4, pp. 418-421.

DOI: 10.3145/epi.2008.jul.09

EL MINISTERIO DE CULTURA a través de la Dirección General del Libro, Archivos y Bibliotecas, ha puesto en marcha una serie de iniciativas con el objetivo principal de asegurar la preservación del patrimonio bibliográfico mediante políticas de digitalización coordinadas entre el propio Ministerio y las comunidades autónomas. Además, persigue garantizar el acceso permanente de todos los ciudadanos a dicho patrimonio, así como ampliar al máximo su visibilidad en internet.

Estos proyectos en colaboración con las comunidades autónomas tienen muy presente las buenas prácticas y objetivos establecidos a nivel internacional. En su planteamiento, desarrollo y perspectivas de futuro representan la plasmación concreta de la normativa internacional, de la IFLA y de la normativa europea.

Catálogo colectivo del patrimonio bibliográfico

Ya desde 1988 el Ministerio de Cultura a través de la Subdirección general de Coordinación Bibliote-

caria (SGCB), y las comunidades autónomas han venido colaborando en un proyecto para facilitar el acceso a la información del patrimonio bibliográfico español: el *Catálogo colectivo del patrimonio bibliográfico*.

En la actualidad permite consultar en red más de 831.767 ediciones y 2.259.270 ejemplares de fondos patrimoniales repartidos entre 729 bibliotecas públicas y privadas.

<http://www.mcu.es/bibliotecas/MC/CCPB/index.html>



Catálogo colectivo del patrimonio bibliográfico

Biblioteca virtual de prensa histórica

De todos los conjuntos que conforman el patrimonio bibliográfico español, el que se halla en mayor peligro de desaparecer es la prensa y por ello ha sido el primero en rescatarse.

Incluye casi cuatro millones de páginas de más de 1.900 títulos, volumen comparable al de los repositorios más importantes de la *Unión Europea*. En 2008 se digitalizarán 400.000 páginas más. Se trata de un proyecto en marcha, resultado de un proceso de digitalización cooperativa del *Ministerio de Cultura*,

las comunidades autónomas y otras instituciones de la memoria. Se han ido incorporando bibliotecas públicas del Estado y también se han firmado convenios con otras instituciones que poseen este tipo de material: ateneos, bibliotecas universitarias, bibliotecas municipales, editores de cabeceras históricas de prensa que se siguen publicando, fundaciones, archivos, etc.

Se pretende simultáneamente preservar y hacer accesibles unos materiales bibliográficos que se caracterizan en la mayor parte de los casos por ser ejemplares únicos, y por lo tanto de difícil acceso.

A principios de 2008 se inició el reconocimiento óptico de caracteres de los pdfs de este repositorio y cuando escribimos este texto (finales de junio) se dispone de las primeras 550.000 páginas. Los ficheros se obtienen en el formato *METS/ALTO* (*Metadata encoding and transmission standard / Analyzed layout and text object*), que es el empleado en la mayoría de los proyectos de digitalización para establecer no ya la estructura de los ficheros sino también las coordenadas de cada uno de los caracteres. Una vez analizados los resultados se proseguirá de forma más afinada con el resto de la prensa y se empezará a aplicar a la *Biblioteca virtual del patrimonio bibliográfico*.

<http://prensahistorica.mcu.es>

Biblioteca virtual del patrimonio bibliográfico

El *Ministerio* ha avanzado en la preservación y difusión de los fondos incluidos en el *Catálogo colectivo del patrimonio bibliográfico*. La puesta en marcha en 2007 de la *Biblioteca virtual del patrimonio bibliográfico* supone la digitalización de cuantiosas colecciones de manuscritos e impresos antiguos que forman parte del patrimonio histórico. Cuenta ya con 850.000 páginas pertenecientes a colecciones de manuscritos, incunables e impresos antiguos, y este año se prevé digitalizar 300.000 páginas más.

En 2008, además, con motivo del segundo centenario de la Guerra de la Independencia, se ha iniciado una nueva colección que incluye ya más de 250 títulos. Por otra parte, se están incorporando fondos bibliográficos históricos de institutos de bachillerato que formarán parte de una nueva colección: *Bibliotecas escolares históricas*. Ambas colecciones dispondrán de los respectivos micrositos que permitirán el acceso directo a las mismas.

<http://bvpb.mcu.es>



Biblioteca virtual de prensa histórica



Biblioteca virtual del patrimonio bibliográfico

Directorio y recolector de recursos digitales

Puesto que uno de los objetivos de las colecciones digitales que se crean es contribuir a la *Biblioteca digital europea*, cabe recordar que toda la normativa aplicada se ajusta a la que utiliza la misma. En particular se aplica el protocolo OAI (siglas en inglés de Iniciativa de Archivos Abiertos) tanto para la recolección (*web harvesting*) como para la creación de repositorios de metadatos conforme a la norma Dublin core / *ISO 15386*. La aplicación de este protocolo logra cumplir con dos objetivos simultáneamente:

1. Disponer de un instrumento de control que favorece la cooperación a la hora de iniciar nuevos proyectos de digitalización.

2. Acceso a más de 230.000 recursos correspondientes a treinta y tres repositorios que soportan el protocolo OAI-PMH.

<http://roai.mcu.es>

Por una parte, el *Directorio y recolector de recursos digitales*

(que tiene 196 colecciones, junio de 2008) recoge los proyectos de digitalización existentes en España para su difusión entre los profesionales de las bibliotecas y los ciudadanos en general y actúa como herramienta para evitar la repetición de reproducciones. Y por otra, crea un mecanismo de recolección de los registros bibliográficos correspondientes a estos recursos. Constituye un sistema de búsqueda y recuperación web de las descripciones de

cada uno de los documentos reproducidos.

Ha sido el primer servicio de este tipo desarrollado en España y uno de los 29 proveedores de servicios registrados de OAI en *Open access registered service providers*.

<http://www.openarchives.org/service/listproviders.html>

El *Directorio* permite detectar las obras que ya están digitalizadas con el fin de evitar duplicar esfuerzos. Por ello, una buena práctica obligada en la selección del futuro material a digitalizar deberá incluir la consulta previa del mismo.

Todas estas iniciativas están en línea con la *Recomendación de la Comisión Europea sobre digitalización y accesibilidad en línea del material cultural y la conservación digital*, de 24 de agosto de 2006, cuyo propósito es contribuir a la acción coordinada de los estados miembros en estos ámbitos para reunir información sobre la digitalización que se encuentra en curso o está prevista.

Directorio y recolector pueden consultarse en todas las lenguas oficiales tal y como promueve *MichaelPlus (Multilingual inventory of digital cultural heritage in Europe)*, acción de la que está previsto que forme parte.



Directorio y recolector de colecciones y recursos digitales

La última versión del *Directorio*, disponible desde el primer trimestre de 2008, incorpora el esquema de datos *MARC XML* para recuperar registros bibliográficos, de manera que se convierte en un catálogo colectivo. Al incluir estructuras *XML* para los *e-prints* será posible recuperar recursos electrónicos de naturaleza académica. Por último también se pretende recolectar los metadatos de los archivos mediante la aplicación de la *DTD* de *EAD* (*encoding archival description*), pero para que esto sea posible debe aumentar la implantación de repositorios *OAI* y, en general, del conjunto de normativa inherente a él, muy en particular de *Dublin core*.

Ayudas para la creación de repositorios

Para fomentar la creación y transformación de repositorios *OAI*, la *SGCB* puso en marcha un programa de ayudas (*BOE* de 25 de mayo de 2007 y *BOE* de 31 de marzo de 2008). Ha financiado proyectos de digitalización por un total de 1.950.000 euros, de los que 1.600.000 se destinan a las distintas administraciones y el resto, 350.000, a fundaciones y entidades sin ánimo de lucro.

Gracias a esta acción, que tiene continuidad en 2008 con la misma inversión, se prevé publicar en internet anualmente más de cinco millones de páginas.

Uso de estándares

En la apuesta por las nuevas tecnologías, la *SGCB* promueve los nuevos estándares. Se está experimentando el más avanzado sistema de recolección de la información basado en los protocolos *SRU* (*Search/Retrieval via URL*) / *SRW* (*Search/Retrieval Web Service*), recomendados por la *Biblioteca digital europea*.

También fomenta la aplicación del formato de preservación de colecciones digitales *Premis* (*Preservation metadata implementation strategies*), *ISO 14721*.

Se dispone de las últimas versiones de las *DTDs* de *Marc-XML*, *MADS* (*Metadata authority description schema*) para el intercambio de información de autoridades y puntos de acceso y *MODS* (*Metadata object description schema*) para el intercambio de información de objetos digitales. Todo ello encapsulado en las estructuras *METS* (*Metadata encoding transmission standard*) para el intercambio de metadatos y teniendo en cuenta la arquitectura de la información que prevé el modelo *OAIS* (*Open archival information system*), norma *ISO 14721*.

La *Biblioteca virtual de patrimonio bibliográfico*, la *Biblioteca virtual de prensa histórica* y algunos proyectos de las comunidades autónomas ya incluyen la totalidad de esta normativa, a la que también se ajustarán a lo largo de 2008 los 32 proyectos subvencionados en 2007 por la *SGCB*. Asimismo, los proyectos se adaptan a un aspecto requerido en Europa y que se adecua con la realidad multilingüe española: la citada normativa *Michael* y *MichaelPlus* abogan por el protocolo *SRU/SRW* utilizando puntos de acceso multilingües. Por ello se están aplicando modelos de datos que incluyen el acceso multilingüe según *Viaf* (*Virtual international authority file*) y conforme al formato *MARC 21*, a la *DTD MADS* y a *Marc-XML* para autoridades.

A medio plazo la *SGCB* tiene previsto aplicar las directrices de *Driver* (*Digital repository infrastructure vision for European research*) a los repositorios de sus bibliotecas virtuales. Una de sus ventajas es poder diferenciar los

registros que se encuentran a texto completo-*OCR* (con texto buscable) de los que están en pdf-*imagen*, así como el tipo de material, análisis de contenido o fecha de las distintas colecciones digitales. En la convocatoria de ayudas de 2008 para proyectos de digitalización se valorará la aplicación de la metodología de *Driver*.

Al reunir los metadatos de los repositorios sobre patrimonio bibliográfico marcadas con las directrices de *Driver*, el *Directorio* y *recolector de recursos digitales* contendrá un primer esbozo de los datos y metadatos de los objetos digitales de dichas bibliotecas. Estos registros y sus URLs se incluirán posteriormente en la *Biblioteca virtual del patrimonio bibliográfico* y en *Biblioteca virtual de prensa histórica*, aunque su integración definitiva se llevará a cabo vía *METS*.

Conclusión

El reto más importante de la *Dirección General del Libro, Archivos y Bibliotecas*, a través de la *Subdirección General de Coordinación Bibliotecaria*, es que las bibliotecas españolas participen en los proyectos europeos que buscan precisamente financiar el desarrollo de colecciones digitales que faciliten el acceso, la difusión y la preservación del patrimonio bibliográfico español. Ello se conseguirá con las líneas de trabajo descritas anteriormente, que se ajustan estrictamente a las directrices del programa *Digital libraries* de la *Unión Europea*.

María-Antonia Carrato-Mena, Subdirectora general de Coordinación Bibliotecaria
Plaza del Rey, 1. 28004 Madrid.
Tel.: +34-917 017 265;
fax: 917 017 339
<http://www.mcu.es/bibliotecas/>